Alexis Quesada-Arencibia
José Carlos Rodríguez
Gabriele Salvatore de Blasio
Carmelo Rubén García
Roberto Moreno-Díaz (Eds.)

EUROCAST 2024

Computer Aided Systems Theory

EXTENDED ABSTRACTS

19th International Conference on Computer Aided Systems Theory
Las Palmas de Gran Canaria, Spain, February 2024

Nineteenth International Conference on
COMPUTER AIDED SYSTEMS THEORY

EUROCAST 2024

Edited by

Alexis Quesada-Arencibia
José Carlos Rodríguez-Rodríguez
Gabriele Salvatore de Blasio
Carmelo Rubén García
Roberto Moreno-Díaz

ISBN: 978-84-09-58721-6
REGISTRO: 2024009406

eurocast 2024

Nineteenth International Conference on Computer Aided Systems Theory

February 25 – March 1, 2024

Museo Elder de la Ciencia y la Tecnología, Las Palmas de Gran Canaria
Canary Islands, Spain

Supporter Institutions:

Universidad de Las Palmas de Gran Canaria
Las Palmas de Gran Canaria, Spain

Johannes Kepler Universität Linz
Linz, Austria

University of Applied Sciences
Upper Austria

Museo Elder de la Ciencia y la Tecnología
Las Palmas de Gran Canaria, Spain

Fundación Universitaria de Las Palmas
Las Palmas de Gran Canaria, Spain

With the collaboration of

City Council of Gáldar
Convention Bureau

Conference Chairs

Founder and Honorary: Franz Pichler (Linz)

General: Roberto Moreno-Díaz (Las Palmas)

Program: Michael Affenzeller (Hagenberg)

Logistics: Alexis Quesada-Arencibia (Las Palmas)

Technical Secretariat

Instituto Universitario de Ciencias y Tecnologías Cibernéticas
Universidad de Las Palmas de Gran Canaria
Campus de Tafira
35017 Las Palmas de Gran Canaria, Spain

Phone: +34-928-457108
Fax: +34-928-457099
E-mail: eurocast@iuctc.ulpgc.es

Preface

The Eurocast Conferences are particularly unique among the European Scientific-Technical Congresses because it is one of the few periodic meetings that is promoted and organized exclusively by university and socio-cultural institutions, without the tutelage, direction or funding of associations, professionals or companies. It is currently the oldest of those. It is celebrated every two years. Initially, alternating Las Palmas de G.C. and a university in continental Europe, and since 2001, always in Las Palmas de G.C.

The idea of the first Eurocast was developed in 1988 by Prof. Franz Pichler, of the University of Linz and Prof. Roberto Moreno, at a meeting in Vienna promoted by the past Honorary President, the late Dr. Werner Schimanovich. The first meeting, Eurocast 1989, took place in February of that year, in Las Palmas School of Industrial Engineers, promoted by the Faculty of Informatics of Las Palmas and the Institute of Systems of the University of Linz. The Opening Session took place in the town of Gáldar, February 26th, 1989.

Science, and especially Technology, have moved in an almost vertiginous way, driven by the need and the promotion of consumerism, associated with the change of values that has been printed in the new generations. And Eurocast, within what we understand as a certain freedom, and with prudence, has been adapting the profile of its organization from a meeting of very specific specialists, to a practically multidisciplinary, flexible and changing conference, which in each event try to attract the experts and especially young researchers, facilitating the interaction between them, which is a generator of creativity.

The key to the success of Eurocast for 35 years has been in the quality of the contributions of its participants. This has to be recognized in the first place. They have made possible, with the help of the Springer Verlag publications in Computer Science, the worldwide distribution of the most important effect of Eurocast: that of joining together for many years, scientists and engineers of ages, training, interests and from very different European and non-European institutions. And that they could share their experiences in the design and analysis of systems using the most advanced mathematical methods to make efficient models and algorithms in computers. And this from the socio-economic, biological, medical technologies and sciences and information and communication engineering topics. All in a multidisciplinary atmosphere, which has facilitated the appearance and discussion of new and creative ideas and developments.

In this open multidisciplinary spirit, the 2024 edition consists of 14 major thematic blocks, which sweep a broad spectrum of cutting-edge research in computer and systems sciences and technologies, including systems theory, applications, pioneers and landmarks, theory and applications of

metaheuristic algorithms, mechatronic product development, model-based system design, verification and simulation, applications of signal processing technology, applied data science and engineering for intelligent transportation systems and smart mobility, computer and systems based methods and electronic tools in clinical and academic medicine, systems in industrial robotics, automation and IoT, systems thinking: applications in technology, science and management, data science in medical and bio-informatics, modeling, simulation and optimization in production and logistics, "Green AI" and SW-Tools for sustainable energy and materials consumption, stochastic models, statistical methods and applied systems simulations and systems cybersecurity technologies and quantum approaches potentials.

Among the about 150 submissions of papers to participate in the Conference, around 128 have been selected and will be presented and defended by their authors during the four days of scientific sessions. Among them, we will proceed to a final selection, after the Conference, which will be published by Springer Verlag and distributed to the scientific community.

But, again, the kernel of the success of Eurocast lays in the right proposals of subjects for Workshops, their resonance and impact, their diffusion and their strict selection of the many intended contributions, all by the Workshops Chairpersons. They are Eurocast.

Three invited Plenary Conferences are presented: the first, by Prof. Maynar from the University of Las Palmas de Gran Canaria, known expert in Computer and Systems based Methods and Electronic Tools in Clinical and Academic Medicine. The second, by Prof. Klempous from Wroclaw University of Technology, known expert in Algorithms, Computing in Mathematics, Natural Science, Engineering and Medicine and Human-computer Interaction. And the third, by Prof. Jacob from Kempten Technological University and expert in Manufacturing Engineering, Industrial Engineering and Mechanical Engineering and Prof. Quesada-Arencibia from the University of Las Palmas de Gran Canaria, expert in Image Processing, Intelligent Transport Systems and Indoor Positioning Systems.

Special thanks to our hosts, partners and most of all, friends, colleagues of Elder Museum.

The Eurocast 2024 Conference, in line with its 35 years of history, once again offers an opportunity for Canary, Spanish, European and global science and technology for openness and international relations, which will benefit our society. Welcome to **Eurocast 2024**.

Las Palmas de Gran Canaria, February 2024.
The Editors.

Table of Contents

Systems Theory, Applications, Pioneers, and Landmarks

Theory and Applications of Metaheuristic Algorithms

Mechatronic Product Development

Model-Based System Design, Verification and Simulation

Applications of Signal Processing Technology

Applied Data Science and Engineering for Intelligent Transportation Systems and Smart Mobility

Computer and Systems Based Methods and Electronic Tools in Clinical and Academic Medicine

Systems in Industrial Robotics, Automation and IoT

Systems Thinking: Applications in Technology, Science, and Management

Data Science in Medical and Bio-Informatics

Modeling, Simulation, and Optimization in Production and Logistics

"Green AI" and SW-Tools for Sustainable Energy and Materials Consumption

Stochastic Models, Statistical Methods, and Applied Systems Simulations

Systems Cybersecurity Technologies and Quantum Approaches Potentials

Application of Named Entity Recognition

Christian Nitzl[1], Achim Cyran[1], Sascha Krstanovic[2], and Uwe M. Borghoff[1,3]

[1] Center for Intelligence and Security Studies (CISS), Neubiberg, Germany
[2] Aleph Alpha, Heidelberg, Germany
[3] Institute for Software Technology, University of the Bundeswehr Munich, Neubiberg, Germany
ciss@unibw.de

Extended Abstract

Named Entity Recognition (NER) can be used to identify, among other things, times, locations, organizations, and persons in texts. In this study, the potential added value of NER in the military analysis process is tested by means of an experiment. The military analysis process is a highly sensitive area of work that requires a high degree of maturity of NER. As part of the experimental testing, a demonstrator deepCOM was developed. The results of the experiment show, on the one hand, that there is a clear added value by artificial intelligence in military information processing. On the other hand, it is shown in the specific for NER that the added value can be further increased by more accurate labelling.

Digitization offers a wide range of possibilities and opportunities. This particularly applies to the vulnerable area of national and alliance defence. Military intelligence is playing a pioneering role in digitization and intelligence studies [7]. Developments in artificial intelligence (AI) and the associated automated processing of information are having a tremendous impact on the area of information gathering and analysis. The task of military intelligence is initially to ensure analysis and reporting of militarily relevant developments abroad. Due to the increased and still increasing volume of information, analysis software with integrated AI components is required, which must first and foremost ensure the analysis capabilities of an operator. Mature analysis software provides a real-time picture of the current situation and is the prerequisite for a functioning crisis early warning system [2].

We explain the Processing phase based on the NATO Allied Joint Doctrine for Intelligence Procedures [1]. It serves to pinpoint a possible added value of NER in the military analysis process. The goal of the Processing phase is to meet the information needs of command personnel as well as to recommend possible courses of action. An overview of the sub-steps of the phase Processing of the NATO Allied Joint Doctrine is shown in Fig. 1.

deepCOM provides a military analyst with interactive access to a database consisting of texts and images. The core functionality of the system is a semantic search. This allows the user to ask direct questions which are answered by the system, indicating the sources used [3]. When present in the text, tags are derived

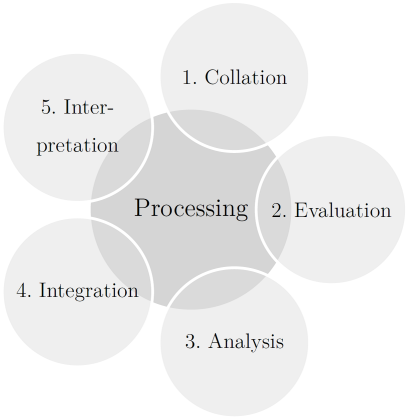


Fig. 1. Phase of Processing.

from mentions of times, locations, organizations, and persons, which are color-coded to the user both while identifying relevant sources and while reading [4, 8]. An implementation of Contrastive Language-Image Pretraining (CLIP) allows the same prompt to also search for images that occur within sources by having the operator describe in their own words what is seen in the graphic or photograph [6].

The NER used in the experiment was not trained for any specific application, but for German texts in general. However, it could be customized for the entities that occur exclusively in military source texts (e.g., weapon designations, names of military leaders). Another possibility for further improvement would be automated domain-specific dictionary creation [5].

References

1. AJP-2.1: Allied joint doctrine for intelligence procedures – NATO Standardization Office (2016)
2. Albrecht, K., Nitzl, C., Borghoff, U.M.: Transdisciplinary software development for early crisis detection. In: EUROCAST. pp. 3–10. LNCS 13789, Springer (2022)
3. Deb, M., et al.: AtMan: Understanding Transformer Predictions Through Memory Efficient Attention Manipulation. arXiv preprint arXiv:2301.08110 (2023)
4. Devlin, J., et al.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
5. Häffner, S., et al.: Introducing an interpretable deep learning approach to domain-specific dictionary creation: A use case for conflict prediction. Political Analysis pp. 1–19 (2023)
6. Radford, A., et al.: Learning transferable visual models from natural language supervision. In: ICML. pp. 8748–8763 (2021)
7. Scheffler, A.C., Jeraj, B., Borghoff, U.M.: The rise of intelligence studies: A model for Germany? Connections: The Quaterly Journal **15**(1), 79–106 (2016)
8. Yadav, V., Bethard, S.: A survey on recent advances in named entity recognition from deep learning models. arXiv preprint arXiv:1910.11470 (2019)

A System Theoretical Multi-Agent Approach to Human-Computer Interaction (Extended Abstract)

Uwe M. Borghoff[1⋆], Paolo Bottoni[2], and Remo Pareschi[3]

[1] Computer Science, University of the Bundeswehr Munich, Neubiberg, Germany
[2] Computer Science, Sapienza University of Rome, Rome, Italy
[3] Università degli Studi del Molise, Campobasso, Italy

We explore a novel perspective on Human-Computer Interaction (HCI), conceptualizing it as a dynamic interplay among a diverse array of agents, both human and computational, within a networked system. The approach draws inspiration from models of living systems and systems theory, emphasizing the autonomous yet interdependent nature of these agents [1]. The essence of our model lies in the understanding that HCI is not merely about the interface between humans and computers but is fundamentally about coordination and communication within a network of heterogeneous agents. These agents, each with distinct capabilities, roles and goals, engage in a continuous process of interaction, governed by well-defined protocols and procedures, to achieve collective goals, from which possibly unforeseen behaviours may emerge.

Central to our discussion is the distinction between human and computational agents. Human agents, localized within nodes of the network, are pivotal in determining the outcomes of interactions and their subsequent application in relevant activities. In contrast, computational agents, with the ability to be distributed across the network, engage in activities that are either reactive or proactive, adhering strictly to predefined protocols. This dichotomy underscores the multifaceted nature of HCI, where human intuition and judgment are complemented by the computational efficiency and consistency of software agents.

Our model posits that interactive computational systems can be viewed as populations of these heterogeneous agents, whose evolution and interactions are steered by human users, acting as a selective factor. This perspective aligns with the notion of systems as open entities, where multiple concurrent processes interact and constrain each other, maintaining the system's identity until a terminal point, akin to the concept of 'death' in living systems. In this context, system identity correlates with its ability to maintain a state within a viable range, a state that is the cumulative result of the activities of various agent populations that may either cooperate or compete, following different 'metabolic pathways'.

At any moment during the evolution of the system, each agent is interacting with a limited number of its peers, depending on space, time, and channel constraints. Hence, we can assume that each communication act has effects on the system's next state depending on the *group* of entities which can communicate with each other under the current conditions (i.e., their *communication space*).

⋆ Corresponding author: uwe.borghoff@unibw.de

```
Establish(agent) {
/* Determine communication spaces for active modality */
       foreach typeOfMessage the agent can issue {
             let rec[typeOfMessage] be the set of agents that can
             receive typeOfMessage }
/* Determine communication spaces for passive modality */
       foreach typeOfMessage the agent can receive {
             let snd[typeOfMessage] be the set of agents that can
             issue typeOfMessage }
/* Establish active and passive links */
       foreach group in rec do snd { FormGroup(group,agent) }   }
```

Fig. 1. Establishing instantaneous agent communication spaces.

Since agents are mobile, the group composition may change over time. We model this situation by associating each communication space with a dedicated computational agent whose state includes a description of the group composition and an indication of its activity state. Fig. 1 offers a depiction of the model for coordinating the activities of agents based on group communication, see also [2].

The practical application of this model is particularly relevant in the realm of system design and evaluation. By employing this agent-based approach, we can model both existing and future computational systems, assessing their potential for success based on their capacity to form a stable and enduring identity over time. The approach has also practical implications in predicting the success of emerging technologies. For instance, the widespread adoption and resilience of technologies like browsers and smartphones can be attributed to their ability to form systems with stable identities. Applying this model to nascent technologies, such as the metaverse and generative AI platforms, provides a framework to evaluate their potential in creating similarly resilient systems.

Furthermore, the incorporation of quantitative criteria for measurement and methodologies could lead to a systematic approach to optimizing agent interactions. Methodologies like genetic algorithms and network optimization algorithms suggest themselves as suitable to the nature of the system, which is both reminiscent of biological ecosystem and apt to be modeled in terms of complex networks. This not only could enhance the efficiency and effectiveness of HCI systems but also ensure their adaptability and sustainability over time. In conclusion, our model offers a comprehensive framework for understanding and designing HCI systems, emphasizing the synergistic and evolving nature of human-computer interactions within a network of diverse agents.

References

1. Borghoff, U.M., Bottoni, P., Mussio, P., Pareschi, R.: Multi-agent coordination in living systems. In: BIES '95, Tokyo, Japan. pp. 93–100 (1995)
2. Meng, X., Tan, Y.: Learning group-level information integration in multi-agent communication. In: AAMAS '23. pp. 2601–2603. ACM (2023)

The Yetman Transmitting Machine

Berthold Kerschbaumer[1]

[1] School of Informatics, Communications and Media - Campus Hagenberg
University of Applied Sciences Upper Austria
Softwarepark 11, 4232 Hagenberg, Austria
Berthold.kerschbaumer@fh-hagenberg.at

Keywords: telegraph typewriter, transmitting typewriter.

Extended Abstract

The "Yetman Transmitting typewriter" was an innovative invention that attempted to create a multifunctional typewriter at the turn of the 19th and 20th century to establish a multifunctional typewriter on the telegraph market. Although there were numerous attempts to equip typewriters with electromechanical add-ons [1] to create "character transmission machines", the function as a "pure" typewriter was lost. There was hardly any need for a machine that went beyond the Morse telegraph to transmit character transmission. The low acquisition costs of the telegraph equipment and the specific skills of the telegraphs stood in the way of the widespread use of character transmission machines with keyboards. Although some telegraphs with keyboards were already on the market, their acceptance and distribution were rather low.

Charles Elmer Yetman (*1863, Darlington, Pennsylvania +1949 New York), the inventor of the machine, received his education at Oberlin College in Ohio, where he was also trained as a telegraph operator. He gained practical experience as a telegraph operator in the following years at various railroad companies in Ohio.[2]

In the 1890s and 1900s, Yetman was granted three comprehensive patents in the USA for an "automatic telegraphing machine" (US534025, 1895), "a combind typewriter and telegraphic transmitter" (US637059A, 1899) and "a type-writing machine" (US760719, 1904), among others. One of which was granted both in Canada and the United Kingdom and another one only in the UK (CA57689, 1897; GB17271A, 1901; GB20115A, 1906). In addition, the patenting of a spring-loaded motor (to drive the signal coding shaft, US807529, 1905), ball-bearing type levers (US884347, 1908) or a unique ink ribbon transport (US760676, 1904) impressively demonstrate Charles Yetman's inventive spirit.

From 1895, Yetman was in close contact with the Remington Standard Typewriter Co. and the first prototypes of the "Remington telegraphing typewriter" were built in Illion,PA and field-tested in 1900.[3] From 1902/03, the Yetman Transmitting typewriter with visible writing was launched on the US market.[4] An advertising campaign

accompanied the market launch of this machine, which was praised as "a new mechanical wonder in telegraphy"[5]. Ultimately, financial problems, caused by speculation on the part of a director, brought the Yetman Typewriter Company into serious turbulence already in 1905 and led to bankruptcy in 1906. Despite its restart in 1907 by Charles Yetman himself, it went broke again a year later, after machines had already been delivered at a small scale.[6]

The upcoming article is intended to highlight Charles Yetman's pioneering achievement, which, although not a commercial success, paved the way for subsequent machines.

References

1. Messenger, R.: From the McLaughlin to the Murray. ETCetera magazine of the Early Typewriter Collectors Association, No. 110, 18–20 (2015)
2. The Telegraph Age, Aug. 1.1903, p.385, New York (1903)
3. Illion History, p.247-251, Illion, (n.d.)
4. The New York times, 26.07.1902, New York (1902)
5. Association of Railway Telegraph Superintendents. Proceedings of the Annual Meeting, Milwaukee (1903)
6. Best, E. M., Business Equipment Topics. Vol. XII, No.2, June 1909, The Typewriter Publishing Co., New York, (1909)

Data Transformation: Limitless Strategies in Energy Consumption

Margaret Miró-Julià[1 [0000−0002−9300−8698]], Monica J. Ruiz-Miró[1,2], Ricardo Alberich[1 [0000−0002−0707−8607]], and Francisco Cordero Piñero[2]

[1] Departament de Ciències Matemàtiques i Informàtica, Universitat de les Illes Balears, 07122 Palma, Spain
margaret.miro@uib.es, monica.ruiz@uib.es, r.alberich@uib.es
http://www.uib.es
[2] BALANTIA, Energy Solutions & Technologies, 07012 Palma, Spain
mruiz@balantia.com, fcordero@balantia.com
http://www.balantia.com

Extended abstract

Artificial Intelligence and Data Mining techniques can be properly combined to analyze complexity in specific domain problems and provide understanding of the phenomena. In real-world applications of Artificial Intelligence, considerable human effort is used in designing the representation of any domain-specific problem in such a way that generalised algorithms can be used to deliver decision-support solutions to human decision makers [1]. On the other hand, Data Mining emerges in response to technological advances and considers the treatment of large amounts of data [2]. Data Mining is the discovery of interesting, unexpected or valuable structures in large data sets.

Energy is a crucial feature in modern life, and advancements in technology have led to the development of smart meters that generate a significant amount of data on power consumption (electricity, gas, water, ...). The knowledge of energy consumers' profile has been an important tool when making decisions in the energy sectors. The aim of this paper is to explain and evaluate the use of different Data Mining techniques and learn which strategies are more adequate to optimize energy efficiency and thus reduce costs. Nowadays, efficient management of energy resources is an important aspect that must be thoughtfully considered. Responsible, efficient and environmentally aware energy consumption behavior is a necessity in present-day life.

With the advancements in sensor technology, by using smart meters, raw data can be easily collected and stored in a database. Extensive volumes of data with high velocity and veracity attributes is generated. Such data have a time-series notion typically consisting of customers' energy usage measurements of over a time interval [3] together with a detailed level of consumer profile data.

Patterns corresponding to customers' energy usage can be mined and extracted to facilitate effective energy management. However, the energy usage database grows incrementally over time. The knowledge discovered from the original database may no longer be useful at the present time. Obviously, re-mining

the updated energy database from scratch each time is inefficient. Therefore, it is necessary to design algorithms capable of maintaining the mining results incrementally. Furthermore, in order to deal with big data, it is convenient to prepare and process the original database for further analysis. Energy temporal data is a high time-resolution data; energy consumption can be presented using different intervals and granularity (per minutes or hourly basis). This transformation helps reduce the volume of data points. Different time resolution scenarios must be studied to decide the best source data for the next step of the Data Mining process. That is, what subsequence of the daily complete sequence, or what days and times offer the more interesting patterns in the energy data.

Energy usage is different for each and every consumer resulting in data that will also depend on specific geographical area, weather situations, different weekdays, consumer behavior, customer characteristics, ... It is necessary to build models that approximate energy usage for all kinds of criteria (certain subsets of customers, certain subsets of days). Data Mining provides different techniques to the problem of extracting meaningful knowledge from the noisy and large energy database [4]. To analyze energy consumption, exploratory data analysis, data visualization techniques, frequent patterns mining and associations, classification, clustering and anomaly detection should be used. Each technique requires different source data.

Data pre-processing is the foundation for consistent and meaningful Data Mining results. Incorrect data integration leads to unreliable patterns. Data preparation and improvement of data quality is an important part of the Knowledge Discovery Process. In this paper we reflect on these ideas and provide some answers. We propose that by pre-processing the data using various time periods such as 15 min, 30 min, 1 h, 2 h, 6 h, or 12 h and establishing a data management system with a new time-interval parameter will lead to quantitatively better patterns. The model transforms data in such a fashion that a dynamic time-interval can be accommodated without having to restart the process. By altering the time-interval, it is possible to reduce the volume of data considered for mining. Choosing the most adequate data source and transformation, will lead to an improvement in model construction.

References

1. Russell, Stuart and Norvig, Peter: Artificial Intelligence: A Modern Approach, Forth edition, Pearson, 2020.
2. Cios, K. J. and Pedrycz, W. and Swiniarski, R. W. and Kurgan, L. A.: Data Mining. A Knowledge Discovery Approach, Springer, New York, 2007.
3. Chen, Y.C.; Hung, H.C.; Chiang, B.Y.; Peng, S.Y.; Chen, P.J.: Incrementally Mining Usage Correlations among Appliances in Smart Homes. In Proceedings of the 2015 18th International Conference on Network-Based Information Systems (NBiS), Taipei, Taiwan, 2015; pp. 273–279.
4. Pitt, Barnaby: Applications of Data Mining Techniques to Electric Load Profiling, Ph.D. Thesis, University of Manchester, United Kingdom, 2000.

VFLBench: A Practical Benchmark for Vertical Federated Learning in Smart Manufacturing

Nguyen Duy Du[1][0009−0000−3422−1918],
Ramin Nikzad-Langerodi[1][0000−0003−3495−8949], and
Michael Affenzeller[2][0000−0001−5692−5940]

[1] Software Competence Center Hagenberg, Austria
{du.nguyen.duy, ramin.nikzad-langerodi}@scch.at
[2] University of Applied Sciences Upper Austria, Austria
michael.affenzeller@fh-ooe.at

Modern manufacturing value chains often include multiple stages operated by different companies. Due to the complex interactions among stages, the KPIs for a particular stage might be affected by preceding stages controlled by other enterprises. Thus, efficient quality control necessitates a holistic view of the entire value chain, achieved by transcending corporate boundaries and accessing data spanning along the entire chain. However, as concerns over data security and privacy are rising, sharing data between companies has become challenging, if not impossible.

Federated Learning (FL) has recently surfaced as a viable approach for building models across data providers without sharing sensitive data. Vertical Federated Learning (VFL) refers explicitly to scenarios where private datasets share the same sample space but differ in the feature space, commonly found in value chains. Despite its significant potential in industrial applications, VFL has received limited exploration and attention [1]. One of the primary obstacles is the need for a unified and robust benchmark suitable for experimental evaluation. Existing works such as UniFed [2], LEAF [3], or PPSR [4] artificially generate data partitions from a global dataset and allocate them to hypothetical parties. However, such a way of segmentation is often criticized for being unscientific and impractical [5].

To this end, we propose VFLBench, a practical benchmark for VFL focused on manufacturing applications. As shown in Fig. 1, VFLBench includes:

- **Open-source federated datasets** that cover different tasks across various industrial sectors and crucially are partitioned using natural splits. Besides real-world datasets, we introduce a pipeline that enables the generation of synthetic federated datasets with customized data modalities. An overview of the available datasets is provided in Table 1.
- **Open-source benchmarking** provides reference implementations of the state-of-the-art VFL algorithms and a fair performance comparison using unified evaluation metrics. Currently, the list of benchmarked methods includes FedOnce, SecureBoost, and SplitNN.

Moving forward, with the support of the broader research community, we plan to expand the datasets and equip VFLBench with additional techniques

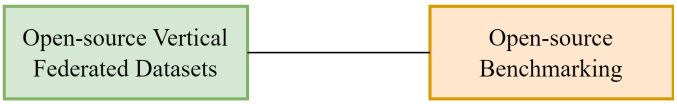


Fig. 1. VFLBench modules.

Table 1. Overview of the available datasets in VFLBench.

Federated dataset	Industrial sector	Task	No. samples	Feature split
Fed-SECOM	Electronics	Anomaly detection	562	C1: 21 features C2: 17 features, 1 target
Fed-CHEM	Chemistry	Regression	176	C1: 12 features C2: 45 features, 1 target
Fed-OIL	Chemistry	Regression	43581	C1: 4 features C2: 8 features C3: 4 features, 1 target
Fed-BOSCH	Metals & Plastics	Classification	18000	C1: 22 features C2: 22 features, 1 target
Fed-SYN	Undefined (synthetic data)	Regression	1000	C1: 10 features C2: 20 features C3: 20 features, 7 targets

to cover a broader range of domains. Minimal requirements and straightforward guidelines will also be provided to make contributing to the benchmark easier. Our ambition is to make VFLBench a living benchmark for VFL to accelerate the development of VFL both theoretically and practically.

References

1. Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., Liu, X. & He, B. A survey on federated learning systems: Vision, hype and reality for data privacy and protection. *IEEE Transactions On Knowledge And Data Engineering*. (2021)
2. Liu, X., Shi, T., Xie, C., Li, Q., Hu, K., Kim, H., Xu, X., Li, B. & Song, D. Unifed: A benchmark for federated learning frameworks. *ArXiv Preprint ArXiv:2207.10308*. (2022)
3. Caldas, S., Duddu, S., Wu, P., Li, T., Konečnỳ, J., McMahan, H., Smith, V. & Talwalkar, A. Leaf: A benchmark for federated settings. *ArXiv Preprint ArXiv:1812.01097*. (2018)
4. Nguyen Duy, D., Affenzeller, M. & Nikzad-Langerodi, R. Towards Vertical Privacy-Preserving Symbolic Regression via Secure Multiparty Computation. *Proceedings Of The Companion Conference On Genetic And Evolutionary Computation*. pp. 2420-2428 (2023)
5. Liu, Y., Kang, Y., Zou, T., Pu, Y., He, Y., Ye, X., Ouyang, Y., Zhang, Y. & Yang, Q. Vertical federated learning. *ArXiv Preprint ArXiv:2211.12814*. (2022)

A Hybrid Cooperative Approach for Symbolic Regression*

Bahareh Etaati[1,2], Stefan Wagner[1], and Michael Affenzeller[1,2]

[1] Heuristic and Evolutionary Algorithm Laboratory
University of Applied Sciences Upper Austria, Softwarepark 11, 4232 Hagenberg
[2] Institute for Computational Perception
Johannes Kepler University, 4040 Linz, Austria
bahareh.etaati@fh-hagenberg.at
stefan.wagner@fh-ooe.at
michael.affenzeller@fh-ooe.at

Symbolic regression is a machine learning technique to discover the underlying complex relationships between variables as mathematical formulas. Unlike traditional regression methods, symbolic regression requires no a-priori knowledge about model structure, as the algorithm searches automatically for model structure and its coefficients. Another advantage of symbolic regression is that mathematical expressions are meaningful, interpretable, and understandable [1]. Nevertheless, the complexity of symbolic regression models can reduce their interpretability due to the bloat phenomenon and the appearance of introns. Additionally, overly complex formulas, especially those that capture noise or outliers in the training data, usually lead to overfitting and thus may have poorer prediction performance on unseen data. Hence, simpler formulas with similar training accuracy tend to be preferred to complex ones [2]. With this in mind, many researchers have considered model accuracy and complexity simultaneously when solving symbolic regression problems. Some studies formulized the symbolic regression problem as a multi-objective problem, with tree length or complexity as another objective besides model accuracy [2, 3].

The most widely used learning algorithm for symbolic regression problems is tree-based genetic programming, which continuously evolves a population of expression trees as candidate solutions by applying genetic operators such as selection, crossover, mutation, and replacement. In recent years, various studies have been conducted to prevent genetic programming from producing overly complex expression trees.

Non-dominated sorting genetic algorithm (NSGA-II) [4] is one of the most popular elitist algorithms for multi-objective optimization. Recently, NSGA-II has been successfully used to solve bi-objective symbolic regression problems [2, 3]. The algorithm ensures elitism by merging the parent population with the generated offspring and integrating the non dominated sorting and crowding distance into the selection process to obtain the new population and a more uniformly distributed Pareto front.

* The project is funded by the FWF DFH 23 doc.funds.connect Human Centered Artificial Intelligence

One widely used method besides Pareto optimality is the weighted sum method, which incorporates various objectives into a single formula [5]. In this approach, the objective functions tend to be continuously normalized as they differ in scale. Also, a weight is assigned to each objective in order to guide the search in different directions within the objective space. Therefore, the algorithm is able to cover various regions in the Pareto front by changing the weights. However, it is difficult for the weighted sum method to find some parts of the Pareto front when dealing with a non-convex problem.

In our work, we propose a hybrid cooperative genetic programming approach for the symbolic regression problem. The proposed algorithm is a hybridization of a weighted sum genetic algorithm and NSGA-II. The objective is that the two algorithms support each other cooperatively by communicating the most promising solutions to create a smoother and well-distributed Pareto front on the NSGA-II side. In the migration phase, the individuals from weighted sum GA are added to the NSGA-II population to cover the less-crowded areas of the Pareto front. On the other hand, the candidate solutions from NSGA-II can contribute to the weighted sum GA to guide the search in different locations within the objective space. For this purpose, having a synchronization strategy to create an appropriate balance between exploration and exploitation is critical. An appropriate synchronization strategy can lead to a better convergence speed and diversity on both sides. With this in mind, the main research contributions are as follows:

1. Proposing a hybrid multi-objective algorithm for a multi-objective symbolic regression problem, optimizing the model accuracy and expression tree length at the same time
2. Making a proper balance between the exploration and exploitation by using the powerful characteristics of both NSGA-II and weighted sum GA

References

1. Michael Affenzeller, Stephan M Winkler, Gabriel Kronberger, Michael Kommenda, Bogdan Burlacu, and Stefan Wagner. Gaining deeper insights in symbolic regression. *Genetic Programming Theory and Practice XI*, pages 175–190, 2014.
2. Michael Kommenda, Gabriel Kronberger, Michael Affenzeller, Stephan M Winkler, and Bogdan Burlacu. Evolving simple symbolic regression models by multi-objective genetic programming. *Genetic Programming Theory and Practice XIII*, pages 1–19, 2016.
3. Christian Haider, Fabricio Olivetti de França, Bogdan Burlacu, and Gabriel Kronberger. Using shape constraints for improving symbolic regression models. *arXiv preprint arXiv:2107.09458*, 2021.
4. Kalyanmoy Deb, Samir Agrawal, Amrit Pratap, and Tanaka Meyarivan. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii. In *Parallel Problem Solving from Nature PPSN VI: 6th International Conference Paris, France, September 18–20, 2000 Proceedings 6*, pages 849–858. Springer, 2000.
5. Wilfried Jakob and Christian Blume. Pareto optimization or cascaded weighted sum: A comparison of concepts. *Algorithms*, 7(1):166–185, 2014.

Comparing Constraint Evaluation Methods for Shape-constrained Regression

Christian Haider[1], Fabricio Olivetti de França[2], and Florian Bachinger[1]

[1] Heuristic and Evolutionary Algorithms Laboratory (HEAL)
University of Applied Sciences Upper Austria Softwarepark 11, 4232 Hagenberg, Austria

[2] Center for Mathematics, Computation and Cognition (CMCC)
Federal University of ABC Santo Andre, Brazil

christian.haider@fh-hagenberg.at

Problem and Motivation Shape-constrained regression is an important desideratum of data-based modeling when you want to enforce your model to possess an expected behavior despite the intrinsic noise of the collected data. Conventional data-based modeling approaches solely reliant on empirical data, often fail to incorporate the essential physical constraints inherent to the underlying systems. This leads to a lack of trust in the model and a reduction of explanatory power[2]. Recognizing these limitations of traditional data-based methodologies, the integration of shape constraints derived from domain expertise or fundamental physical principles emerges as a promising extension to enhance the quality of models. In this paper, we distinguish between single-objective and multi-objective approaches, as well as soft and hard constraint methods to evaluate the constraint violations of models. In the context of this research, the primary focus is on single-objective approaches that incorporate an expanded scope of features and dynamics. Instead of strictly enforcing hard constraints and discarding solutions solely based on constraint violation, the paper presents a dynamic method for calculating constraint violation and prediction error. This approach involves an incremental increase in the weighting of constraint violation during the modeling process, achieving an initial, wide-ranging exploration of the solution space followed by a refined and comprehensive investigation of potential models through an in-depth search. Furthermore, we want to explore the critical issue of premature stagnation, a phenomenon wherein the search for a conformant model, plateaus before reaching an acceptable local optima solution. The paper tries to address this challenge through different strategies to maintain a certain diversity of solutions. The motivation for this research is to advance shape-constrained data-based modeling by delving into various constraint evaluation methods. Specifically, we aim to propose a combined evaluation approach that integrates a dynamic constraint violation measurement with prediction error. By looking into the detailed mechanics of this combined evaluation method aspire to get a deeper understanding of the interplay between shape constraints and modeling accuracy.

Method Our proposed constraint evaluation method represents an advancement in the domain of shape-constrained regression modeling. Therefore, we

employ a single-objective function designed to yield soft constraints. Our approach combines a dynamic shape constraint evaluation with a measurement of prediction error, e.g., using a normalized mean squared error (NMSE). The key of our presented method lies in the dynamic constraint weight adaption throughout the runtime of the algorithm. This strategic approach involves a variation in constraint weights, enabling an extensive exploration of the space in a first step. Subsequently, the algorithm transitions into a deeper and more focused investigation of potential solution candidates. Therefore, we test different methods for incrementing the weights during the evaluation process. These methods include linear increase, exponential increase, and a type of simulated annealing process, each offering a unique perspective on the interplay between shape constraints and prediction accuracy. In our constraint evaluation framework, we explore different methods for calculating constraint violations. The first approach uses interval arithmetic (IA), a method that allows us to handle uncertainties and provide rigorous bounds on constraint violations. As a second approach, we investigate a sampling approach, aiming for a more granular examination of the solution space.

Experimental Setup/Preliminary Results To assess the efficiency of our proposed constraint evaluation approach, we conducted initial experiments using the Feynman datasets[1]. This dataset consists of syntactic problems derived from physics textbooks, which serve as a suitable playground for testing. To construct our training and test dataset we applied different levels of noise and extrapolation scenarios to mimic real-world problems. Moreover, we derived the underlying physical shape constraints by sampling the given equations.

Table 1 presents our preliminary results, showcasing the performance of the proposed evaluation approach. The first results indicate that using our method yields models with comparable and in some cases slightly superior prediction accuracy. These impressions hint that this approach can help us find solutions that reflect the underlying behavior of the system as well as maintain competitive predictive performance. We plan to extend our experimental setup by including a more extensive set of instances and conducting deeper statistical analysis.

Table 1. Median NMSE values for test set.

	GP	GPHard	GPSoft	NSGA-II
I.6.20	0.31	2.25	0.12	**0.10**
I.41.16	**0.25**	5.61	3.42	3.78
I.48.20	**0.00**	**0.00**	**0.00**	**0.00**

References

1. Udrescu, S.M., Tegmark, M.: AI Feynman: A physics-inspired method for symbolic regression. Science Advances **6**(16) (2020)
2. Willard, J., Jia, X., Xu, S., Steinbach, M., Kumar, V.: Integrating physics-based modeling with machine learning: A survey. arXiv preprint arXiv:2003.04919 **1**(1), 1–34 (2020)

Re-evaluation in Dynamic Tree-Search with Backtracking from Known Solutions

Philipp Fleck[1,2], Philipp Neuhauser[1,2], Sebastian Leitner[1,2], and Stefan Wagner[1,2]

[1] Heuristic and Evolutionary Algorithms Laboratory, University of Applied Sciences Upper Austria, philipp.fleck@fh-hagenberg.at
[2] Josef Ressel Center for Adaptive Optimization in Dynamic Environments, University of Applied Sciences Upper Austria

Dynamic optimization problems [2] in industrial applications present numerous challenges. Solution candidates may become outdated due to changes in the environment, different circumstances require different optimization strategies and the formal specification of the optimization problem is often incomplete due to a lack of domain knowledge. In this paper, we use a lot-assignment problem from the steel industry as an example [1], where we aim to efficiently group steel slabs into lots, based on specific requirements and constraints.

To effectively solve the lot-assignment problem, we use different *tree-search algorithms*, such as depth limited search or beam search [3]. Each node in the search space represents an assignment of a steel slab to a transport-lot, and each child adds an additional lot-assignment until all steel slabs are assigned. Since the tree search assumes that the order of children generally reflects their quality, this allows for guidance via *heuristics* and *constraints* given by domain experts. For instance, existing domain experts' strategies for lot-assignments can be incorporated via various heuristics. Overall, the tree-search algorithms embody distinct strategies for traversing the search space, with the specific ordering of child nodes determined by the heuristics serving as a primary guide towards good solutions. Fig. 1 on the left displays a sample search space with the qualities given in the nodes. Note that the ordering of the children is typically indicative of better quality yet does not give a guaranteed optimum order.

Since industrial optimization problems can be extremely complex, with goals and constraints evolving and changing over months or even years, knowledge of good *heuristics* and necessary *constraints* is distributed over multiple experts, sometimes even with opposing views. Therefore, it is impractical to establish optimization parameters that are set in stone, with full knowledge of all necessary information beforehand. Therefore, algorithm developers must anticipate potential changes in the optimization parameters and allow the user to quickly adapt their strategies. To provide the necessary flexibility, user can change

- the *constraints* of solutions (i.e. which child nodes are returned) and
- the *heuristics* guiding the search (i.e. the order of child nodes).

After such changes in search parameters, the current solution must be re-evaluated to check its validity and the algorithm should be re-run in light of the changed heuristic to yield better solutions.

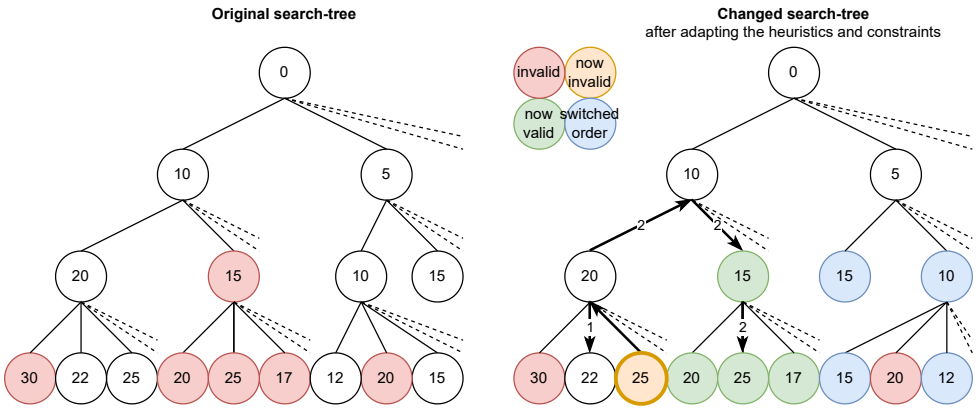


Fig. 1. The original search tree on the left with invalid solutions (red). The changed search tree on the right with solutions now invalid (orange) or now valid (green) due to changed constraints, and altered order of children (blue) due to changed heuristics.

Re-running the entire algorithm without considering the current solution, would discard any valuable findings obtained from it. This paper aims to investigate alternative approaches to utilizing the current solution as an anchor in the tree-search and performing intelligent re-evaluation by backtracking to areas where partial re-runs of the algorithm appear most promising. For instance, considering the changed search tree on the right hand side of Fig. 1, where the previously best solution with quality 25 (thicker borders) is now invalid, requiring an adjustment of this outdated solution. A viable strategy could be to backtrack to the parent-node 20 and descend to the 25's sibling with quality 22 (via line 1). Alternatively, one could also backtrack to the grandparent-node 10, observe that child 15 is now valid, from which one could progress the search and potentially discover the node with a quality of 25 (via line 2). Both backtracking strategies would be more efficient re-evaluation method compared to a complete re-run, particularly in real-world cases where the search-tree is substantial and fast backtracking may prove even more advantageous. The primary objective of this paper is to investigate backtracking strategies and assess their potential.

References

1. Beham, A., Raggl, S., Hauder, V.A., Karder, J., Wagner, S., Affenzeller, M.: Performance, quality, and control in steel logistics 4.0. Procedia Manufacturing **42**, 429–433 (2020)
2. Cruz, C., González, J.R., Pelta, D.A.: Optimization in dynamic environments: a survey on problems, methods and measures. Soft Computing **15**, 1427–1448 (2011)
3. Lemons, S., López, C.L., Holte, R.C., Ruml, W.: Beam search: Faster and monotonic. In: Proceedings of the International Conference on Automated Planning and Scheduling. vol. 32, pp. 222–230 (2022)

Application of Adapt-CMSA to the Electric Vehicle Routing Problem with Simultaneous Pickup and Deliveries

Mehmet Anıl Akbay[1 [0000−0001−7376−7008]], Christian Blum[1 [0000−0002−1736−3559]], and Can Berk Kalayci[2 [0000−0003−2355−7015]]

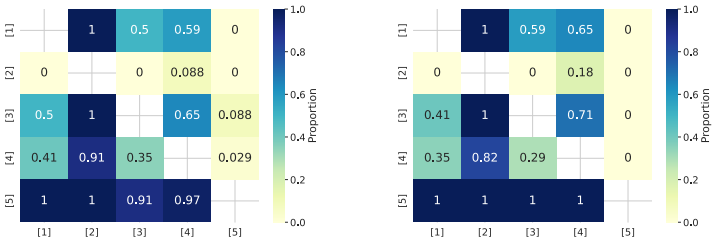
[1] Artificial Intelligence Research Institute (IIIA-CSIC), Campus UAB, Bellaterra, Spain
makbay@iiia.csic.es, christian.blum@csic.es
[2] Department of Industrial Engineering, Pamukkale University, Denizli, Turkey
cbkalayci@pau.edu.tr

The Vehicle Routing Problem (VRP) has been one of the most studied combinatorial optimization problems, aiming to determine the most efficient routes for a fleet of vehicles to deliver goods to a group of customers. As the global emphasis on sustainability and environmental conservation has grown, the logistics sector has begun to adapt and transform its operations to align with this necessity. Introducing Electric Vehicles (EVs) into the logistics landscape has been a significant step in this direction, offering an environmentally friendly alternative to traditional combustion engine vehicles.

However, the integration of EVs into logistics does not come without challenges. The Electric Vehicle Routing Problem with Simultaneous Pickup and Delivery (EVRP-SPD) emerges as a complex variant of the traditional VRP, accounting for the unique constraints of EVs, such as limited battery capacity and en-route recharging necessity. Furthermore, the simultaneous pickup and delivery aspect requires that each customer's demand for both delivery and pickup be satisfied simultaneously, adding another layer of complexity to the problem.

First of all, we formulated the addressed problem as a mixed integer linear programming (MILP) model, which can be solved using general-purpose solvers like CPLEX or Gurobi. However, experiments have shown that the problem becomes increasingly complex in the presence of SPD constraints, in conjunction with the limited driving range of electric vehicles. This complexity asks for the design of more effective algorithms for deriving good-quality solutions. Accordingly, a recent variation of the hybrid metaheuristic "Construct Merge Solve & Adapt" (CMSA) [2]—called Adapt-CMSA—has been designed to solve particularly large-scale problem instances. This self-adaptive variant of the CMSA was developed to handle the parameter sensitivity of the original algorithm [1]. In general, CMSA algorithms rely on the iterative application of an exact solver to specific sub-instances of the original problem instance. This process aims to decrease the search space before employing the exact solver. In order to accomplish search space reduction, a bottom-up strategy is employed that consists of two fundamental stages. Initially, the CMSA algorithm probabilistically generates solutions to the addressed problem. Subsequently, these solutions are merged to form a sub-instance that can then be solved using the exact solver.

The algorithm's performance was evaluated on various problem instances compared with other algorithms, namely a probabilistic C&W Savings algorithm (pC&W), a probabilistic sequential insertion algorithm (pIns), and the VNS variants from [3]. The latter



(a) Best performance (b) Average performance

Fig. 1: Pairwise algorithm performance comparison heatmaps. Algorithms are labeled as: [1]: Probabilistic C&W savings; [2]: Probabilistic Insertion Algorithm; [3]: Reduced VNS-Cyclic; [4]: General VNS-Cyclic; [5]: Adapt-CMSA.

study presents 12 different variable neighborhood search (VNS) algorithm variants designed for the EVRP-SPD. In the context of small problem instances, the performance of the proposed algorithm is also compared with the standalone application of CPLEX for solving our two-index MILP formulation and the three-index formulation from [3].

CPLEX optimally solved every small problem instance (with 5, 10 and 15 customers), regardless of the specific MILP formulation. However, our model enabled this performance with considerably less CPU time than the one from [3]. Concerning the heuristic approaches, VNS and Adapt-CMSA were able to find all the optimal solutions provided by CPLEX. However, Adapt-CMSA presented this performance requiring less CPU time than VNS. Regarding the large-sized EVRP-SDP instances, Adapt-CMSA significantly outperforms the best VNS variants and two construction heuristics regarding best and average results. Note that compared VNS variants are reported as the top two variants among all VNS variants tested in the respective publication.

Figure 1 visually supports this observation, presenting two heatmaps that compare the performance of the algorithms across multiple instances. Each cell in these heatmaps indicates the proportion of instances for which the corresponding row's algorithm outperforms the column's algorithm. Darker shades in the heatmap emphasize a higher proportion, signifying the row algorithm's frequent dominance over the column algorithm, while lighter shades indicate the contrary.

Acknowledgements. Support was received by grants TED2021-129319B-I00 and PID-2022-136787NB-I00 funded by MCIN/AEI/10.13039/501100011033.

References

1. Akbay, M.A., López Serrano, A., Blum, C.: A self-adaptive variant of CMSA: Application to the minimum positive influence dominating set problem. International Journal of Computational Intelligence Systems **15**(1), 1–13 (2022)
2. Blum, C., Pinacho Davidson, P., López-Ibáñez, M., Lozano, J.A.: Construct, merge, solve & adapt: A new general algorithm for combinatorial optimization. Computers & Operations Research **68**, 75–88 (2016)
3. Yilmaz, Y., Kalayci, C.B.: Variable neighborhood search algorithms to solve the electric vehicle routing problem with simultaneous pickup and delivery. Mathematics **10**(17), 3108 (2022)

Optimizing Public Transport Through Integration of Microscopic and Macroscopic Simulations

Kate Han[1], Lee A. Christie[2], Alexandru-Ciprian Zăvoianu[2], and John A.W. McCall[2]

[1] Salford Business School, University of Salford, Manchester, UK
[2] National Subsea Centre, RGU, Aberdeen, Scotland, UK

Workshop: Theory and Applications of Metaheuristic Algorithms

Keywords: multi-modal public transport · micro-level mobility simulations · macro-level mobility simulations · reachability isochrones.

Extended Abstract

The European Commission's net zero goal puts forth a vision of achieving climate-neutral transport by 2050[5]. The UK has enacted legislation aiming to reach this emissions target encompassing the entire economy, including the transport system[6]. Numerous studies have been carried out to assess the potential impacts of connected and autonomous vehicles (CAVs) on emissions[3] also exploring the feasibility of enhancing existing public transport (PT) systems through the use of zero-emission CAVs[4].

Consequently, CAVs could deliver significant social benefits to local communities, particularly in situations where deploying traditional transport options would be economically impractical due to expected low passenger volumes[6].

Recent research suggests macro-level mobility simulations, which rely on open data such as mapping and timetables, can be integrated with optimization strategies. This integration proves beneficial in addressing practical scenarios involving the design of CAV routes[1].

The imperative to transition the current transport system toward achieving net zero is underscored by initial findings and subsequent feedback from transport officials. However, navigating this transformation poses significant challenges on the journey to achieving net zero carbon emissions. It is crucial to identify the most effective changes and elements that will support this transition, necessitating an integrated perspective gained through simulations at various levels of the transport system. This approach is relevant before introducing new routes for CAVs.

In this context, we elaborate on the development of a simulation that integrates both micro-level SUMO[2] simulate traffic data and macro-level mobility open source OSM and GTFS map data. This integration facilitates the utilization of state-of-the-art evolutionary algorithms to uncover realistic optimal transport plans which seamlessly integrate new CAV routes into an existing multi-modal

PT systems. To illustrate this process, we present a case study based on a real-world scenario in the Leeds area.

The findings from our approach provide valuable insights into potential enhancements achievable through the incorporation of micro-level simulation in the designated study area. Specifically, we identified a route with a 3% improvement in mobility using 8 CAVs, increasing to 12.5% with a larger investment of 14 CAVs during the 8 am to 8:30 am peak time on a workday. Additionally, we compared integrated macro-level and micro-level simulation to simple macro-level data simulation. Integrated simulation and macro level identified common stops for constructing CAV routes which we believed would be useful insight to construct CAV routes in real-world scenarios.

Our approach represents a stride towards automating the generation of realistic solutions, specifically addressing the intricate challenges associated with both micro and macro aspects of public transport planning.

Acknowledgements

The authors would like to acknowledge the support of West Yorkshire Combined Authority transport policy officers, DLR Transport experts, and ART-Forum Interreg project supported by the North Sea Programme of the European Regional Development Fund of the European Union.

References

1. Han, K., Christie, L.A., Zăvoianu, A.C., McCall, J.: On discovering optimal trade-offs when introducing new routes in existing multi-modal public transport systems. In: International Conference on Computer Aided Systems Theory. pp. 104–111. Springer (2022)
2. Krajzewicz, D., Hertkorn, G., Rössel, C., Wagner, P.: An example of microscopic car models validation using the open source traffic simulation sumo. In: Proceedings of Simulation in Industry, 14th European Simulation Symposium. pp. 318–322 (2002)
3. Massar, M., Reza, I., Rahman, S.M., Abdullah, S.M.H., Jamal, A., Al-Ismail, F.S.: Impacts of autonomous vehicles on greenhouse gas emissions—positive or negative? International Journal of Environmental Research and Public Health **18**(11), 5567 (2021)
4. Pigeon, C., Alauzet, A., Paire-Ficout, L.: Factors of acceptability, acceptance and usage for non-rail autonomous public transport vehicles: A systematic literature review. Transportation research part F: traffic psychology and behaviour **81**, 251–270 (2021)
5. Plötz, P., Wachsmuth, J., Gnann, T., Neuner, F., Speth, D., Link, S.: Net-zero-carbon transport in europe until 2050—targets, technologies and policies for a long-term eu strategy. Karlsruhe: Fraunhofer Institute for Systems and Innovation Research ISI. Available online: https://www. isi. fraunhofer. de/en. html (accessed on 25 July 2021) (2021)
6. UK, E., Strategy, I.: 'energy white paper: Powering our net zero future. Dept. Bus., Energy Ind. Strategy, UK, Tech. Rep. CCS0220144090, CP **337** (2020)

Automated Inference of Domain Knowledge in Scientific Machine Learning

Florian Bachinger[1], Christian Haider[1], Jan Zenisek[1], and Michael Affenzeller[1]

Heuristic and Evolutionary Algorithms Laboratory
University of Applied Sciences Upper Austria, Hagenberg, Austria

In applications of machine learning (ML) we see the ever-increasing need for interpretable, and domain-informed prediction models that increase trust in model predictions. One research direction tries to explain trained opaque-box models after the fact [2], while another tries to introduce domain knowledge during ML training to ensure that the resulting models adhere to expected behavior. In this work, we try to improve the latter category, by automatically inferring the very knowledge required to guide ML training from a given data.

Background: One category of ML algorithms, capable of domain knowledge integration, is shape-constrained regression (SCR) [1]. Therein, interval bounds are introduced onto the image of the function or any order of its partial derivatives to enforce the shape of the prediction function.

Possible examples for shape-constraints are: a function is known to be solely positive valued $f(X) \geq 0$ (e.g., the power production of a photovoltaic (PV) system), or a function is monotonic increasing over a certain input $\frac{\partial}{\partial x_i} f(X) \geq 0$ (e.g., increasing solar radiation, in turn, increases PV power production). Shape-constraints are especially useful when training data is available in limited amount, or when certain extrapolation behavior of prediction models should be enforced. In short, shape-constraints allow us to define and enforce behavior of models and the desired interactions between inputs with regard to model output.

Motivation: Currently, we rely on domain experts, first principles, and the domain understanding of the data scientists, to gain the prior knowledge for domain-informed ML. However, domain experts might be unaware of the existence of such interactions in their respective fields, or, in the case of material sciences, want to find out about these very interactions, but require empirical research and subsequent analysis of recorded data. *Therefore, we desire an approach that is able to automatically infer knowledge from data of a given domain.*

Solution Concept: We envision our approach as an additional step or tool in the data preparation phase of ML. Data scientists apply our approach on new data to retrieve suitable constraints that can be subsequently integrated into model training. Thereby, our approach will contribute towards guiding the ML algorithm's search and improving resulting prediction performance.

Solution Approach: If certain interactions, such as positivity or monotonicity, are true for the whole data domain, they are also true for any given subset of the data. Thus, in step (1) we split the available data into a training and validation set, but do not apply a random split, instead, we spatially split the data by continuously moving the convex hull of the input space into the

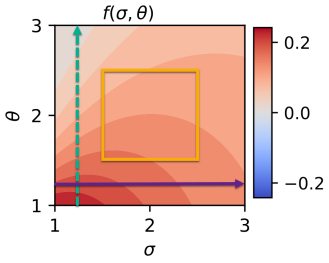


Fig. 1. The contour plot of the AIFeynman expression I.6.2, is monotonic decreasing over θ (dashed turquoise arrow) but exhibits no monotonic behavior over σ. The smaller training subset (yellow square) exhibits the same behavior. Models trained with the right constraints, on data in the yellow square, will exhibit better prediction performance on the validation outside of the yellow square.

training set until the desired split ratio is achieved. See Figure 1 for a visual representation.

In step (2) we train simple ML models (polynomials of degree 3) on the spatially central training data and evaluate their respective extrapolation prediction performance on the validation set. Here, we apply a brute force method and investigate every possible combination of monotonicity constraints for each equation during model training. The assumption being, that correct constraints, and the correct combination thereof, will exhibit the best extrapolation performance as the model is being guided in the correct manner. We rank the resulting models by their validation accuracy to reveal the constraint configurations with the best extrapolation performance.

Experimental Setup: To show the feasibility of our approach, we selected the AI Feynman [3] benchmark set which contains a number of equations as closed form expressions from Richard Feynman's lectures, stemming from the fields of physics and engineering. We can symbolically derive these equations, and thus, determine the actual known constraints beforehand (i.e., analytically determine the actual physical interactions of inputs with regard to the target) that we subsequently want to find ourselves.

From these expressions, we generate data by sampling a uniform random input space and evaluating the equation for this data to calculate the target values. We add normally distributed noise to simulate real world data. The datasets include 120 equations with up to one to nine independent inputs, respectively.

Outlook: In our promising preliminary results, we were able to determine the exact constraints of the original base equation in 28 of the 120 instances. In these instances, the model with the exactly correct constraint combinations ranked first place by order of validation error. We plan to further improve our results and the method of evaluation and reporting thereof for our full contribution.

References

1. Curmei, M., Hall, G.: Shape-constrained regression using sum of squares polynomials. Operations Research (2023)
2. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence **1**(5), 206–215 (2019)
3. Udrescu, S.M., Tegmark, M.: Ai feynman: A physics-inspired method for symbolic regression. Science Advances **6**(16), eaay2631 (2020)

On-demand Automated Guided Vehicles in Yard Logistics *

Ulrike Ritzinger[1], Bin Hu[1], and Martin Reinthaler[1]

AIT Austrian Institute of Technology, Center for Energy
Giefinggasse 4, 1210 Vienna, Austria
[firstName].[lastName]@ait.ac.at

1 Introduction

Connected and automated vehicles have a significant impact on the global automotive industry, particularly the logistics and freight transport sector. The rise of autonomous driving technology is reshaping mobility paradigms, prompting established players and startups alike to adapt. While advancements in AI algorithms show promise in managing the complexities of autonomous driving, the challenges such as high development costs, safety requirements, and legal questions remain significant. Compared to public roads, operation in controlled environments such as yards lowers the risk factors. Still, yard logistics addresses issues such as a shortage of qualified drivers, high accident rates, and underutilized freight capacity.

In the project AWARD (All Weather Autonomous Real logistics operations and Demonstrations) [1], a framework has been developed to support the adoption of autonomous technology in freight transport in real-life logistics scenarios. The goal is to create a system for automated guided vehicles (AGV) that compile with all safety standards, equipped with multiple sensors and a teleoperation system for 24/7 availability. These vehicles will undergo validation in diverse use cases, including warehouse operations, open-road shuttle services, airport baggage dispatching, and port container transfers. A central component is the fleet management system (FMS) for optimizing logistics processes integrating various information sources (e.g., vehicles, logistics systems, and roadside infrastructure). The system is designed to dispatch the most suitable vehicle for a logistics order based on the current conditions.

2 Methodology

This paper focuses on the fleet navigation algorithm of the FMS, which assigns the orders and routes to the AGVs in the system. This is modelled as a dynamic pickup and delivery problem [2] on a road network, which is a directed graph, where nodes represent junctions and arcs represent road segments. Each order

* The project AWARD has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 101006817

has a pickup and a delivery location, a given load, and a given soft time window for each location. The problem is dynamic since orders are added to the operational day on the fly and not known beforehand. In addition, so called *situations*, which impact the speed of the AGVs on defined street segments for a given time period, arise. Situations may arise where certain network areas are temporarily closed or weather conditions affect the roads, consequently the speed of the AGVs. These events also occur dynamically. The solution is a sequence of edges to traverse for each vehicle, so that the fleet can fulfill the maximum number of orders without violating their load and time constraints, and that vehicles do not conflict each other on their route.

Our main contribution is the optimization algorithm tailored to address the multifaceted challenges of real-world freight transport problems involving AGVs. The formulated problem is a dynamic time-dependent pickup and delivery problem with soft time windows, accommodating the variable speeds resulting from dynamically changing situations. The algorithm employs both time-dependent forward and backward Dijkstra algorithms based on [3], considering the departure time, respectively arrival time, to determine the shortest path. Furthermore, to evaluate the feasibility of the time windows, we extended the segment-based evaluation method [4] to account for the impact of time-dependent networks.

Implementing the concept of large neighborhood search [5], the algorithm only accepts conflict-free schedules [6] as new candidate solutions, which is aligned with the move evaluation process. The approach integrates these components to tackle the dynamic nature of the fleet navigation problem, promising practical applicability in optimizing AGV-based freight transport systems.

References

1. AWARD Project Homepage, https://award-h2020.eu/. Last accessed 15 Nov 2023
2. Berbeglia, G., Cordeau, J.-F., Laporte, G.: Dynamic pickup and delivery problems. European Journal of Operational Research **202**(1), 8–15 (2010).
3. Gmira, M., Gendreau, M., Lodi, A., Potvin, J.Y.: Tabu search for the time-dependent vehicle routing problem with time windows on a road network. European Journal of Operational Research **288**(1), 129–140 (2021).
4. Vidal, T., Crainic, T.G., Gendreau, M., Prins, C.: Time-window relaxations in vehicle routing heuristics. Journal of Heuristics **21**, 329–358 (2015).
5. Pisinger, D., Ropke, S.: An Adaptive Large Neighborhood Search Heuristic for the Pickup and Delivery Problem with Time Windows. Transportation Science **40**(4), 455–472 (2006).
6. Hu, B., Seragiotto, C., Prandtstetter, M., Binder, B., Bader, M.: Guiding Autonomous Vehicles past Obstacles - Theory and Practice. In: Computer Aided Systems Theory – EUROCAST 2019. Springer 1–8 (2020).

Diversity Management in Evolutionary Dynamic Optimization*

Bernhard Werth[1,2][0000−0002−7830−3138], Johannes Karder[1,2][0000−0002−7006−8663], Stefan Wagner[1][0000−0002−0323−9274], and Michael Affenzeller[1,2][0000−0001−5692−5940]

[1] Josef Ressel Center for Adaptive Optimization in Dynamic Environments
Heuristic and Evolutionary Algorithms Laboratory
University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria
[2] Institute for Symbolic Artificial Intelligence
Johannes Kepler University, 4040 Linz, Austria

bernhard.werth@fh-hagenberg.at

Dynamic optimization problems are usually defined by a time-dependent objective function, meaning that the decision variables that are considered optimal at one point in time might no longer be suitable later. An underlying assumption of dynamic optimization is that information from previous problem *epochs* can still be utilized to facilitate the search, as opposed to starting the optimizer from scratch every time a change is detected. Although dynamic optimization problems arise in many application areas like production planning, routing, scheduling, data analysis and forecasting, the majority of literature for black-box optimization procedures focuses on static problems. Of those studies that deal with dynamic optimization problems, a prominent concept is the usage of evolutionary algorithms that provide implicit memory in form of a population of solutions [1].

A highly relevant aspect of any dynamic evolutionary algorithm is the careful curation of the genetic diversity in the algorithm's population. Diversity in this context describes the amount and uniqueness of the genetic information present in the population. In static optimization, diversity is usually slowly decreased by selection mechanisms until it finally converges; at this point most members of the population resemble each other both in terms of decision variables and achieved objective value. Conversely, in dynamic optimization, a certain measure of diversity must be retained or periodically reintroduced. An already converged population confronted with a strong change may struggle to re-create essential building blocks.

In this paper we demonstrate and compare different mechanisms of managing diversity in genetic algorithms that are operating on dynamic traveling salesman (TSP) problem instances:

* This research was funded in part by the Austrian Science Fund (FWF) [I 5315-N]. The financial support by the Austrian Federal Ministry for Digital and Economic Affairs and the National Foundation for Research, Technology and Development and the Christian Doppler Research Association is gratefully acknowledged.

- No specific management: In principle, mutation and crossover operations can introduce new diversity, but may severely impact any time performance by destroying important building blocks as it is being constructed.
- Problem-triggered restarts: Recent research [1] reiterates that for certain scenarios restarting a genetic algorithm whenever a change is detected is preferable to complex schemes of memory or diversity management.
- Population-triggered restarts: A simple way to govern algorithm restarts is measuring the likelihood for a child to outperform its parents. Variations of this approach that use dynamic population sizes were already applied to several dynamic scenarios [3].
- Immigration schemes: These algorithms [5] continuously integrate either new solution candidates or memorized ones from previous epochs via immigration into the population, thereby trading the need for restarts for a constant influx of genetic material.
- Layered population structures: The age-layered population structure [2] respects the fact that newly generated solutions often have little chance of outcompeting older members of the population while still potentially containing important building blocks. The population is therefore separated into layers of different age as to limit competition. The youngest layer is periodically reseeded and successful genetic information migrates slowly through the layers, negating the need for discrete restarts.

Performing this analysis on the dynamic TSP [4] provides the advantages of a well-known and well-defined problem, an intuitive encoding where build blocks are represented as sub-tours and the possibility of obtaining the current best solution via exact solvers with negligible computational effort. It allows for an analysis of the age of different optimal building blocks and measuring the correlation of genetic diversity with momentary and overall algorithmic performance. By better understanding the mechanics of diversity retention and loss, we expect to lay the groundwork for more fine-grained and adaptive diversity control mechanisms.

References

1. Alza, J., Bartlett, M., Ceberio, J., McCall, J.: On the elusivity of dynamic optimisation problems. Swarm and evolutionary computation **78**, 101289 (2023)
2. Hornby, G.S.: Alps: the age-layered population structure for reducing the problem of premature convergence. In: Proceedings of the 8th annual conference on Genetic and evolutionary computation. pp. 815–822 (2006)
3. Karder, J., Werth, B., Beham, A., Wagner, S., Affenzeller, M.: Analysis and handling of dynamic problem changes in open-ended optimization. In: International Conference on Computer Aided Systems Theory. pp. 61–68. Springer (2022)
4. Tinós, R., Whitley, D., Howe, A.: Use of explicit memory in the dynamic traveling salesman problem. In: Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation. pp. 999–1006 (2014)
5. Tinós, R., Yang, S.: A self-organizing random immigrants genetic algorithm for dynamic optimization problems. Genetic Programming and Evolvable Machines **8**, 255–286 (2007)

Tackling the α-Domination Problem Heuristically

Enrico Iurlano, Johannes Varga, and Günther R. Raidl

Algorithms and Complexity Group, TU Wien,
Favoritenstraße 9/1921, Vienna, 1040, Austria

{eiurlano|jvarga|raidl}@ac.tuwien.ac.at

The problem of finding a minimum dominating set of vertices of a graph $G = (V, E)$ is one of the best known and intensely studied optimization problems on graphs. Dunbar et al. [4] introduced a variant of this problem called α-*domination problem*, in which for a fixed proportionality parameter $0 < \alpha \leq 1$ the goal is to find a minimum cardinality set $S \subseteq V$ of distinguished vertices such that each vertex $v \in V$ is already contained in S or at least a fraction α of its neighbors is comprised by S; in both cases we say that v is α-dominated. This problem is strongly related to the so-called *Minimum Positive Influence Dominating Set* (MPIDS) problem introduced in [7], in which the aim is to maintain sufficient influence in a social network by utilizing the minimum number of influencing actors (here, even actors need to receive influence from their equals). In the age of constantly growing data on social networks and the increasing relevance of disinformation campaigns [2] and viral marketing [7], the α-dominating set problem seems to us a partially more suitable model to determine how influence—even before consolidating it—can be achieved by using a few simultaneously deployed actors of assumed loyal behavior.

The α-domination problem can be seen as an interpolation between the classical domination problem, for sufficiently small $\alpha > 0$, and the *Vertex Cover Problem*, for $\alpha = 1$. It is shown in [4] that for any $0 < \alpha \leq 1$, it is NP hard to find an optimal solution; moreover, upper and lower bounds for optimal solutions on general graphs but also for special graph classes are derived.

Although practical computational methods for the MPIDS problem and other problems with parallels on a conceptual level (such as, e.g., the *Target Set Selection Problem* [5] studying the influence under a propagation aspect) are recently receiving an increased attention (see, e.g., [1,6]), to the best of our knowledge this was not yet the case for the α-domination problem. Therefore, we study in this work an integer linear programming approach as well as a constructive heuristic and a local search, which are further extended towards a GRASP and a configuration checking [3] metaheuristic. These approaches are experimentally evaluated on a diverse set of graphs in terms of solution quality and speed. In this study, we restrict ourselves on the parameter choice $\alpha = 0.5$, which is readily adopted in the aforementioned comparable settings [5,7].

*We acknowledge the support of the Doctoral Program VGSCO by the Austrian Science Fund (FWF) under grant W1260-N35 and of Austria's Agency for Education and Internationalization under grant BA05/2023.

First, we set up a binary integer linear programming formulation and approach it by the commercial solver Gurobi. We observe that in particular denser graphs with only a few hundred vertices require impractically long runtimes or, if the solver ist stopped early, only poor solutions with large remaining optimality gaps are obtained. Thus, as an alternative, we propose an efficient greedy heuristic that favors vertices leading to a largest immediate increase in newly α-dominated vertices during the process of iteratively adding vertices to S. We supplement this construction with a subsequent local search that always removes a vertex $v \in S \subseteq V$ from S and fixes all resulting violated constraints by trying to make distinguished only a small number of further vertices. The choice of v is based on a score that results from a linear combination of three key quantities of a vertex $v \in S$: its *degree*, its *saturation* (number of distinguished neighbors) and its *indispensability* (number of neighbors that would immediately lose their α-domination status if v was removed from S). Sensibly randomizing the construction heuristic and iterating it together with the local search, we obtain a GRASP metaheuristic. Moreover, we experiment with restricting the actually investigated neighborhoods by means of configuration checking. A particular focus lies in the experimental analysis of the conditions under which the introduction of the configuration checking leads to a further improvement and how the level of detail of the stored configurations is reflected in the solution quality.

The instances we use for evaluation are graphs with up to thousands of vertices (ranging from highly structured to random graphs with different sparsities) already used for similar domination-related problems.

References

1. Akbay, M.A., López Serrano, A., Blum, C.: A self-adaptive variant of CMSA: application to the minimum positive influence dominating set problem. International Journal of Computational Intelligence Systems **15**(1), 44 (2022)
2. Bradshaw, S., Bailey, H., Howard, P.N.: Industrialized disinformation: 2020 global inventory of organized social media manipulation. Tech. rep., Computational Propaganda Research Project at the Oxford Internet Institute (2021), `https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/12/2021/01/CyberTroop-Report-2020-v.2.pdf`, last accessed: 2023-11-15
3. Cai, S., Su, K., Sattar, A.: Local search with edge weighting and configuration checking heuristics for minimum vertex cover. Artificial Intelligence **175**(9-10), 1672–1696 (2011)
4. Dunbar, J.E., Hoffman, D.G., Laskar, R.C., Markus, L.R.: α-domination. Discrete Mathematics **211**(1-3), 11–26 (2000)
5. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 137–146 (2003)
6. Lozano-Osorio, I., Oliva-García, A., Sánchez-Oro, J.: Dynamic path relinking for the target set selection problem. Knowledge-Based Systems **278**, 110827 (2023)
7. Wang, F., Camacho, E., Xu, K.: Positive influence dominating set in online social networks. In: Combinatorial Optimization and Applications, COCOA 2009. LNCS, vol. 5573, pp. 313–321. Springer (2009)

A Comparative Study of Symbolic Regression Methods on the Scaled Nikuradse Dataset

Yousef A. Radwan and Gabriel Kronberger

[1] Nile University, Giza, Egypt
[2] yo.radwan@nu.edu.eg
University of Applied Sciences Upper Austria, School for Informatics, Communication, and Media, Softwarepark 11, 4232 Hagenberg, Austria
gabriel.kronberger@fh-hagenberg.at

Introduction The field of symbolic regression has been growing rapidly in the recent years due to its innate value in generating highly-interpretable models that are highly sought after in real-life applications. This paper compares four of the latest, more advanced methodologies and systems used for symbolic regression: a method proposed by Meta that uses a neural network based on the transformer architecture, two methods based on exhaustive search that rely on heavy pruning of all functions in a function space, and a method based on genetic programming. The four methods are compared on the scaled univariate Nikuradse dataset. The four model generators are compared quantitatively using mean absolute error and R2 score, in addition to qualitatively with the use of visualisation.

Methods The most basic generator is the standard genetic programming based generator [1], this generator uses a population-based optimization method to randomly iterate and move through the function space and find the best function structure and constants it can. This was the leading method for symbolic regression for a large section of its beginnings. However, this method has a number of obvious downfalls including: no guarantee of finding the optimum in the space, and long inference time.
Based on this principle, some generators were created that tried to avoid one of the main downfalls of standard GP (genetic programming) methods: the lack of guarantee that the model would reach the best optimum in the given function space. Two of these methods: ESR (exhaustive symbolic regression) [2] and GE (Grammar enumeration) [4] rely on exhaustive search of the function space to find all possible structures and then heavy pruning to reduce the number of possibilities (this is where they mainly differ) and then evaluation on a specified cost function to find the best fit. ESR does evaluation based on the minimum description length principle.
The final method [3], proposed by Meta, is to challenge the two-step procedure of symbolic regression (finding the structure, finding the constants) entirely and develop a transformer-based model to generate the full expression in one step. The model has shown very promising results and even has the added capability of having the constants refined after the initial generation as a sort of refined initialization. In addition, the use of a transformer tremendously improves inference time over the other methods. This will be referred to as "DeepSR".

Table 1. Examplary expressions and results for the tested methods.

Method	Expression	MSE
DeepSR	$\theta_1 - \theta_2 \lvert \theta_3 \ (\theta_4 \, x)^{\theta_5} - \theta_6 \sin\left((\theta_8 \, x) \, (\theta_9 \, x - \theta_{10})\right) + \theta_{11} \rvert$	0.0044
ESR	$\theta_1 \frac{\theta_2}{\frac{\theta_3}{x} + x}$	0.0136
GE	$\theta_1 e^{\theta_2 x_3^{\theta}} \ln\left(\lvert x + \theta_4 \rvert\right) + \theta_5$	0.0017
GP	$\theta_1 x + \left(\theta_2 x + \theta_3 x \sqrt{e^{\theta_4 x}}\right) \left(\theta_5 x + \theta_6 x x\right) + \theta_7$	0.0016

Discussion It can be seen that is very obvious differences between the performance of each of the generators, the top 2 models being grammar enumeration and the standard genetic programming method. This experiment will be further expanded on more datasets and also with further discussion about the behavior and result characteristics of each model generator.

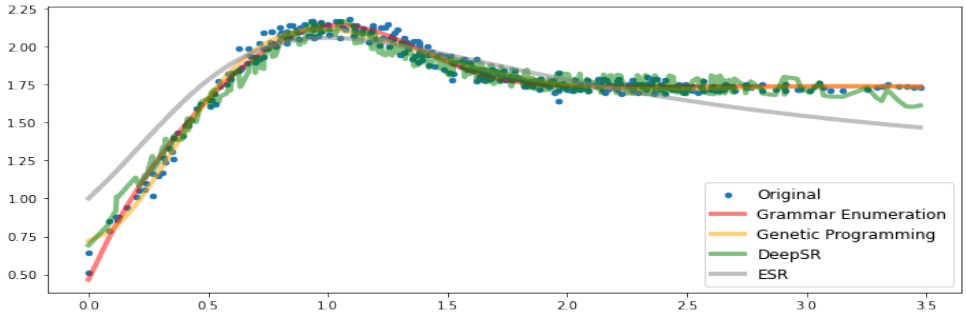


Fig. 1. Comparison of models

References

1. Affenzeller, M., Winkler, S., Wagner, S., Beham, A.: Genetic Algorithms and Genetic Programming: Modern Concepts and Practical Applications (04 2009). https://doi.org/10.1201/9781420011326
2. Bartlett, D.J., Desmond, H., Ferreira, P.G.: Exhaustive symbolic regression. IEEE Transactions on Evolutionary Computation pp. 1–1 (2023). https://doi.org/10.1109/tevc.2023.3280250,
3. Kamienny, P.A., d'Ascoli, S., Lample, G., Charton, F.: End-to-end symbolic regression with transformers (2022)
4. Kammerer, L., Kronberger, G., Burlacu, B., Winkler, S.M., Kommenda, M., Affenzeller, M.: Symbolic regression by exhaustive search: Reducing the search space using syntactical constraints and efficient semantic structure deduplication. In: Genetic Programming Theory and Practice XVII, pp. 79–99. Springer International Publishing (2020). https://doi.org/10.1007/978-3-030-39958-0_5,

Composable Evolutionary Computation

Jan Zenisek[1,2], Florian Bachinger[1], Erik Pitzer[1],
Stefan Wagner[1], and Michael Affenzeller[1,2]

[1] University of Applied Sciences Upper Austria, Softwarepark 11, 4232 Hagenberg

[2] Institute for Symbolic Artificial Intelligence
Johannes Kepler University Linz, Altenberger Straße 69, 4040 Linz

jan.zenisek@fh-hagenberg.at

Background and Motivation: Evolutionary Computation (EC) based studies and applications regularly need long-running times. High dimensional problems, domain-specific target functions (e. g., simulation-based), complex operators (e. g., gradient descent) or large data sets are just some reasons to aim for faster EC algorithms using parallel computing and workload distribution. Moreover, parallel/distributed EC is not only faster, but also enables the development of other algorithmic aspects, e. g., strategies to avoid premature convergence [1].

Software frameworks such as *ECJ* [2], *HeuristicLab* [3], or *DEAP* provide large selections of EC algorithms, parallelization engines to execute them and rich dashboards to analyze their performance. However, without profound programming skills and deeper framework knowledge, it can be quite cumbersome to go beyond the mere use of available algorithms in these tools. In particular, we see a lack of support in the development of custom parallel and distributed EC algorithms, which we want to address in this work.

Method: In this paper, we present an EC framework prototype, which incorporates (a) a library of EC paradigm implementations (b) a description language with corresponding frontend to design new algorithms as operator graph and (c) a kernel, which handles the parallel/distributed execution of algorithms.

(a) Library: We provide implementations for popular EC paradigms such as Genetic Algorithm (GA), Genetic Programming (GP), Evolution Strategy (ES), etc. as basic algorithm building blocks with several standard hyperparameters to tune. The algorithms are made pluggable, so that parallel/distributed concepts, such as Island GA, Cellular EA, or Age Layered Population Structures are easy to design.

(b) Algorithm design: We adopt the operator-graph-algorithm concept of *HeuristicLab* [3] using the graph description language *Sidl*[1] and extend the corresponding Visual Studio Code extension for our purpose. As a first result to this end, we refer to Figure 1, where the design and execution of an Island-GA with an ES, integrated as local search, is demonstrated.

(c) Kernel: The kernel component parses the algorithm description and provisions and orchestrates the necessary environment for execution: Each graph

[1] https://github.com/prescriptiveanalytics/Sidl [2] https://github.com/janzenisek/eurocast2024

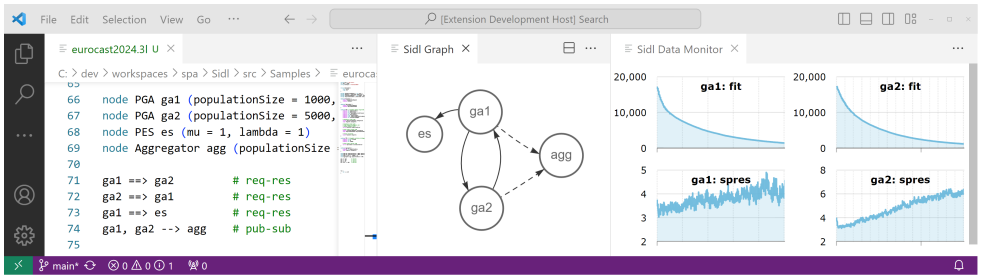


Fig. 1. Design, execution and monitoring of a custom composed EC algorithm: Left-hand side, the language editor with the textual description for a custom algorithm is depicted, while in the middle, a synchronized view of the resulting algorithm graph is shown. A kernel to execute the designed algorithm is accessible via the editor's terminal. On the right-hand side, real-time information regarding best fit, selection pressure, etc. of the currently executed algorithm is displayed.

node is run as an individual, distributable container (cf. docker & kubernetes). Each graph edge, i. e., the communication between nodes, is implemented via sockets, which enable request-response, as well as publish-subscribe messaging. One kernel feature to highlight hereto is the dynamic selection of the optimal communication path: If nodes share memory, they can communicate via custom, high-performance in-memory sockets, otherwise they use network sockets.

In the final paper, we aim to present an implemented and tested concept for design and distribution of EC algorithms in greater detail. However, we do not plan to provide a new, full-fledged EC-framework. We will show serveral algorithm design examples, as well as achieved solution quality and runtime performance on a set of well-known benchmark functions. The current state and future developments can be tracked on our *github* repository[2].

Acknowledgements: The presented work was done within the project "Secure Prescriptive Analytics", funded by the state of Upper Austria as part of the program "#upperVISION2030".

References

1. Harada, T., Alba, E.: Parallel genetic algorithms: a useful survey. ACM Computing Surveys (CSUR) **53**(4), 1–39 (2020), https://doi.org/10.1145/3400031
2. Scott, E.O., Luke, S.: Ecj at 20: toward a general metaheuristics toolkit. In: Proceedings of the genetic and evolutionary computation conference companion. pp. 1391–1398 (2019), https://doi.org/10.1145/3319619.3326865
3. Wagner, S., Kronberger, G., Beham, A., Kommenda, M., Scheibenpflug, A., Pitzer, E., Vonolfen, S., Kofler, M., Winkler, S., Dorfer, V., et al.: Architecture and design of the heuristiclab optimization environment. In: Advanced methods and applications in computational intelligence, pp. 197–261. Springer (2014), https://doi.org/10.1007/978-3-319-01436-4_10

Solving the Manhattan Metric Straddle Carrier Routing Problem with Buffer Areas using a Hybrid Metaheuristic Method

Nina Radojičić[1], Ahmet Cürebal[1], Leonard Heilig[1], and Stefan Voß[1]

Institute of Information Systems,
University of Hamburg,
Von-Melle-Park 5, Hamburg 20146, Germany
{nina.radojici.matic, cahmet.cuerebal, leonard.heilig,
stefan.voss}@uni-hamburg.de

Keywords: Hybrid metaheuristic · VNS · GRASP · port logistics · container terminals · straddle carrier routing problem.

Extended abstract

In recent years, hybrid optimization approaches have been increasingly proposed to solve difficult optimization problems from different domains (see [1]). Discrete optimization problems are prevalent in port logistics, playing a pivotal role in the effective management of intricate operations within seaports. These problems necessitate discrete decision-making to optimize various facets of port activities, including container handling, vessel scheduling, resource allocation, and transportation. Different metaheuristics have been used for solving many of these problems in literature, such as POPMUSIC [5], Variable Neighborhood Search (VNS) [2], etc.

We address an important challenge of optimizing the routing of straddle carriers (SCs) in ports, focusing on their efficient movement of containers between the quay and storage yard while adhering to loading/unloading sequences. Specifically, we tackle the Manhattan Metric Straddle Carrier Routing Problem with Buffer Areas (MSCRB) introduced in [4]. The MSCRB involves optimizing SC routes to minimize vessel turnaround times, considering constraints imposed by loading sequences and limited buffer areas strategically placed within the operational range of quay cranes (QCs). The operational efficiency of such terminals is significantly influenced by the routing of SCs. The routes must consider the order in which containers are loaded and unloaded at quay cranes (QCs), while also accounting for the limited buffer area capacity of each QC. This buffer area temporarily stores containers after being handled by a QC or an SC. Additionally, the solution framework incorporates safety constraints. Efficient SC routing plays a crucial role in minimizing the idle time of QCs, thereby enhancing overall productivity and reducing vessel turnaround time—a key objective of the prob-

lem. The problem is proven to be NP-hard in the strong sense, as demonstrated by [4].

We introduce a hybrid metaheuristic strategy to address this optimization challenge encountered in container terminals where SCs are responsible for transporting containers between stacking areas and the seaside. Variants of Variable Neighborhood Search (VNS) [3] and GRASP [6] methods are implemented to address the MSCRB and the possibilities for their hybridization are explored. These algorithms tackle the problem by integrating four local search operators commonly employed in vehicle routing problems, as well as four proposed shaking operators.

A comparative analysis is performed in which the results of these approaches are evaluated against each other, the exact results obtained and the results from the literature. To compare our results with the exact result, we use the optimization solver GUROBI. The numerical experiments show the superior performance of the proposed algorithms compared to the exact solver, especially for larger instances. Furthermore, a comparison with results from the existing literature shows that the proposed hybrid approaches provide very competitive results. Overall, the proposed methods achieve competitive results that meet the high-performance requirements in real-world scenarios, and the obtained results show the advantages of hybridizing different methods.

Acknowledgments. This study was funded by the German Federal Ministry for Digital and Transport (BMDV) with the funding program for Innovative Port Technologies (IHATEC) within the project TwinSim.

References

1. Blum, C., Raidl, G.R.: Hybrid metaheuristics: powerful tools for optimization. Springer (2016)
2. Cheimanoff, N., Fontane, F., Kitri, M.N., Tchernev, N.: A reduced VNS based approach for the dynamic continuous berth allocation problem in bulk terminals with tidal constraints. Expert Systems with Applications **168**, 114215 (2021)
3. Hansen, P., Mladenović, N.: Variable neighborhood search: Principles and applications. European Journal of Operational Research **130**(3), 449–467 (2001)
4. Kress, D., Meiswinkel, S., Pesch, E.: Straddle carrier routing at seaport container terminals in the presence of short term quay crane buffer areas. European Journal of Operational Research **279**(3), 732–750 (2019)
5. Lalla-Ruiz, E., Voß, S.: POPMUSIC as a matheuristic for the berth allocation problem. Annals of Mathematics and Artificial Intelligence **76**, 173–189 (2016)
6. Resende, M.G., Ribeiro, C.C.: Optimization by GRASP. Springer (2016)

A genetic algorithm-based approach for hotel buildings design

Agnieszka Adamska-Idzikowska[1][0000−0002−5959−9316] and Radosław Idzikowski[2][0000−0001−5232−1998]

[1] Department of Architecture and Visual Arts,
Wrocław University of Science and Technology,
Bolesława Prusa 50/55, 50-317 Wrocław, Poland
agnieszka.adamska-idzikowska@pwr.edu.pl
[2] Department of Control Systems and Mechatronics,
Wrocław University of Science and Technology,
Zygmunta Janiszewskiego 11/17, 50-372 Wrocław, Poland
radoslaw.idzikowski@pwr.edu.pl

Extended Abstracts

Designing hotels is a task that requires the utilization of knowledge from various fields: hotel management, architecture, psychology, marketing, and economics. This work is interdisciplinary in nature, combining expertise in architecture and urban planning with information and communication technologies. Due to the complexity of the subject, the focus has been on aspects directly influencing programming, planning, and designing within the architectural scope.

In each establishment, in addition to complying with building, sanitary, and fire safety regulations, it is necessary to consider requirements regarding equipment and the scope of offered services. The number and complexity of criteria depend on the chosen type of project during the planning stage. Hotels are classified by stars, where one star represents basic accommodation services, and five denote a facility of high standards with a rich range of provided services.

There are numerous publications describing the principles of designing contemporary hotels. In a review paper [4], the focus is on the analysis of multicriteria methods in the context of building design. Six approaches were then compared. Additionally, during the design process, technical aspects of the building can be taken into account, such as energy use and loads, indoor thermal comfort, life-cycle costs, and life cycle CO^2 gas production [6].

In such tasks, solutions based on artificial intelligence, including machine learning, are currently very popular [3]. An alternative are genetic algorithms (GA), which can be applied to various tasks, such as scheduling [5] or designing wireless networks [2]. GA have also been applied in the context of building design, aiming at optimizing building insulation [8] or ecological solutions [7].

The subject of this research is the intelligent support of the hotel design process. The study is divided into two stages. The first stage involves selecting a plot for the hotel project based on individual constraints arising from the Local Spatial Development Plan and technical conditions that buildings and their

placement should meet. These include: a) maximum building footprint, b) minimum biologically active area, c) permissible roof slopes and shapes, d) building height and number of floors, e) applicable and non-exceedable building lines, f) distances from plot boundaries. It is also necessary to introduce additional constraints that do not arise from regulations but have a significant impact on the selection of the functional program and the leitmotiv of the facility. As a result, surfaces are obtained that showcase the potential of each plot (number of rooms, size of functional zones depending on the hotel category).

The second stage begins with the selection of a specific plot. It involves multi-criteria optimization, taking into account criteria such as the number of rooms and attractions, room layout in relation to cardinal directions, and the degree of articulation of the building form. As a result, a Pareto front is obtained, presenting possible hotel form solutions that meet all the specified constraints.

The dynamic development of the hotel industry, coupled with the simultaneous advancement of artificial intelligence, encourages the adoption of intelligent solutions in optimizing the design of contemporary ventures. Furthermore, numerous indicators and forecasts point to ongoing developmental trends in the hotel industry, making it even more worthwhile to delve into the topic associated with successive expansions [1]. The research results will not only present the methodology and operation of algorithms but may also serve as a foundation and inspiration for designing similar solutions in the future. It should be emphasized that, during the literature review, no work was found that covered the scope of the research topic of this paper. This work fills the identified research gap.

References

1. Adamska-Idzikowska, A.: Adaptacje zabytkowych zamków i pałaców województwa dolnośląskiego na współczesne hotele (2023)
2. Altshuler, E., Linden, D.: Wire-antenna designs using genetic algorithms. IEEE Antennas and Propagation Magazine **39**(2), 33–43 (1997)
3. Geyer, P., Singaravel, S.: Component-based machine learning for performance prediction in building design. Applied Energy **228**, 1439–1453 (2018)
4. Mela, K., Tiainen, T., Heinisuo, M.: Comparative study of multiple criteria decision making methods for building design. Advanced Engineering Informatics **26**(4), 716–726 (2012), eG-ICE 2011 + SI: Modern Concurrent Engineering
5. Rudy, J., Zelazny, D.: Memetic algorithm approach for multi-criteria network scheduling. In: Proceedings of the International Conference On ICT Management for Global Competitiveness And Economic Growth In Emerging Economies. pp. 247–261 (2012)
6. Soebarto, V., Williamson, T.: Multi-criteria assessment of building performance: theory and implementation. Building and Environment pp. 681–690 (2001), building and Environmental Performance Simulation:Current State and Future Issues
7. Wang, W., Zmeureanu, R., Rivard, H.: Applying multi-objective genetic algorithms in green building design optimization. Building and Environment **40**(11) (2005)
8. Wright, J.A., Loosemore, H.A., Farmani, R.: Optimization of building thermal design and control by multi-criterion genetic algorithm. Energy and Buildings **34**(9), 959–972 (2002), a View of Energy and Bilding Performance Simulation at the start of the third millennium

Route planning for parcel logistics systems with reusable packaging

Georg Brandstätter, Wolfgang Ponweiser, and Matthias Prandtstetter

Austrian Institute of Technology, Giefinggasse 4, 1210 Wien, Austria
{georg.brandstaetter,wolfgang.ponweiser,matthias.prandtstetter}@ait.ac.at

1 Introduction

In current parcel logistics systems, individual shipments are usually packed in disposable cardboard packages that are discarded by the recipient after delivery. This generates large amounts of waste paper that must be collected and disposed of (or, preferably, recycled).

In total, packaging already accounts for 50% of all paper that is used in Europe [1] – a number that is expected to increase by a further 19% by 2030. In the European Green Deal, the European Commission has set the goal to reduce the amount of packaging waste by 15% (compared to 2018) until 2040 through increased reuse and recycling of, among other things, the packaging used in e-commerce deliveries [2].

If packaging is to be reused, empty packages must be collected from recipients and returned to prospective senders. In the interest of efficiency, this should not be done by dedicated package collection tours, but instead be combined with corresponding delivery tours.

Over the course of a working day, the cargo hold of delivery vehicles becomes progressively emptier as packages are taken out and delivered to their recipients. This excess capacity can then be used to store collected empty packages.

We therefore propose an algorithm for planning combined package delivery and collection tours in a parcel logistics system that uses reusable packaging.

2 Problem Definition

We consider a parcel logistics system covering a designated geographical area in which a given set of package transportation requests must be fulfilled. A request can require either (a) the transportation of full packages (i.e., ones containing goods) from a sender to a recipient, or (b) the transportation of empty packages from a previous recipient to a prospective sender.

Full packages must always be delivered directly from the shipment's designated sender to its recipient, as they represent the transportation of unique goods contained therein. In contrast, empty packages are a *commodity*, i.e., a prospective sender's request for empty packages can be fulfilled from any available source.

Both senders and recipients can be either (a) *local* (i.e., lying within the system's area), or (b) *remote*, (i.e., lying outside the area). For transportation requests between two local parties, the corresponding packages are transported directly from the sender to the recipient (or vice versa). In contrast, requests between a local and a remote party are fulfilled via a designated *hub* (e.g., a parcel distribution center). Instead of delivering directly from (or to) the remote party's address (which lies outside the system's area and might therefore be far away), packages are instead delivered from (or to) the hub instead. Thus, we only consider the last (or first) leg of the package's journey.

Transportation requests are fulfilled by a fleet of vehicles, each with a given package capacity, that start and end their daily tours at their depot.

We consider three distinct operational variants of the aforementioned system. In variant A, recipients return their empty packages to a set of *collection stations* that are spread throughout the system's area, from where they can be picked up by the system's vehicles. In variant B, empty packages are instead collected directly from the recipients' home addresses. In both variants A and B, a single company is responsible for fulfilling all transportation requests.

In contrast, multiple companies operate within the system's area in variant C. Requests for full packages are preassigned to one company, whereas empty packages can be transported by any of them. The assignment of empty package requests to vehicles is done from a global perspective to analyze the system's potential performance in an optimistic case with cooperation between the different companies.

In each variant, the objective is to determine a set of vehicle routes that fulfills all given transportation requests while minimizing the total distance driven.

3 Algorithm and Experiments

We use a metaheuristic algorithm based on variable neighborhood search to compute the aforementioned set of vehicle routes. We then analyze the system's behavior in each of the three previously described variants on a set of benchmark instances generated for different regions of Austria.

References

[1] Patricia Megale Coelho et al. "Sustainability of reusable packaging – Current situation and trends". In: *Resources, Conservation & Recycling: X* 6 (2020), p. 100037. ISSN: 2590-289X. DOI: https://doi.org/10.1016/j.rcrx.2020.100037.

[2] European Commission. *European Green Deal: Putting an end to wasteful packaging, boosting reuse and recycling*. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7155. Accessed: 2023-11-15. 2022.

Selecting User Queries in Interactive Job Scheduling*

Johannes Varga[1], Günther R. Raidl[1], and Tobias Rodemann[2]

[1] Institute of Logic and Computation, TU Wien, Vienna, Austria
{jvarga,raidl}@ac.tuwien.ac.at
[2] Honda Research Institute Europe, Offenbach, Germany
tobias.rodemann@honda-ri.de

Motivation and Problem. We consider a class of job scheduling problems in which human users, e.g., the personnel of a company, need to perform jobs on some shared machines and the availabilities of these users as well as the machines is critical. In such situations it is rarely practical to ask users to fully specify their availability times. Instead we assume users initially only propose a single starting time for each of their jobs, and a feasible and optimized schedule shall then be found within a small number of interaction rounds. In each such interaction round, our scheduling approach may make a small number of b queries to users to find out more about their availabilities. We investigate two different forms of queries. For the first one, the approach proposes alternative time intervals for the users' jobs, which the users either accept or reject, and for the second, users are asked to provide alternative possible starting times for their jobs restricted to certain time intervals. The problem we benchmark our approach on includes multiple identical machines and only one job can be performed in a non-preemptive manner on one machine at any time. As planning horizon we consider several days and time is discretized. Jobs have individual but machine-independent durations and no precedence constraints. Scheduling a job induces time-dependent costs and we allow that jobs remain unscheduled at additional penalty costs.

This interactive job scheduling scenario has originally been considered in [2]. Some approaches to solving a scheduling problem similar to our core problem when neglecting the interaction aspect include mixed integer linear programming [3, 1], a genetic algorithm [3], as well as a greedy heuristic and local search [1].

Solution Approach. The challenge in each round is to find a set of user queries that have a high potential to gain knowledge on further user availabilities for improving the current schedule. For the first form of queries this means that they have to be likely to be accepted and reduce the objective value as much as possible if accepted. To find a good set of queries of the first form, we first collect all meaningful queries, i.e. all time intervals of job lengths fitted into the time horizon, since longer intervals would have a very low chance to be accepted and shorter intervals would not benefit the schedule much. For each of these queries

* J. Varga acknowledges the financial support from Honda Research Institute Europe.

Table 1. Mean %-gaps of objective values after five interaction rounds with different thresholds p^{lim} and for Greedy.

#Machines	#Jobs	#Users	Greedy	0.25	0.4	0.5	0.75	0.85
5	120	30	56.3	46.5	21.4	**9.1**	11.8	25.5
5	240	60	37.3	35.3	27.7	**20.5**	22.3	23.9

we estimate a probability that the user accepts it by approximating the users behavior with a Markov chain that has two states: State 0 represents the user being not available and state 1 represents the user being available. The probabilities p_{01} and p_{10} of the user becoming available or unavailable from one timestep to the next determine the expected time the user is available or unavailable without break and are estimated by assuming a usual workday of eight hours with one break in between. Only queries with an estimated acceptance probability exceeding a certain threshold p^{lim} are considered further and the b most promising, in the sense that they improve the objective the most if accepted, are selected by solving an integer linear program.

To also account for the second form of queries we investigate a stochastic programming approach. The first stage of this stochastic program selects queries from a precomputed set of possible candidate queries, and the second stage estimates the expected solution quality of the schedule given the selected queries, again using the two-state Markov model to approximate user behavior. We obtain solutions using an integer linear program to solve the sample average approximation, which approximates the expected solution quality by taking the mean over a sample of possible user replies.

Results. To test our approach, we generate instances with up to five machines, up to 240 jobs and four jobs per user, by randomly generating user availabilities and using electricity spot market prices for operating costs. Table 1 shows results for the first approach. It gives, for different instance sizes and thresholds p^{lim}, over 30 instances the mean %-gap to the best possible schedule when assuming full knowledge of the users' availabilities. The baseline approach Greedy directly selects queries that can improve the objective the most, without filtering unpromising queries. As can be seen, the probability threshold 0.5 works best.

References

1. Anghinolfi, D., Paolucci, M., Ronco, R.: A bi-objective heuristic approach for green identical parallel machine scheduling. European Journal of Operational Research **289**(2), 416–434 (2021)
2. Varga, J., Raidl, G.R., Rönnberg, E., Rodemann, T.: Interactive job scheduling with partially known personnel availabilities. In: Dorronsoro, et al. (eds.) OLA 2023: Optimization and Learning. Communications in Computer and Information Science, vol. 1824, pp. 236–247. Springer (2023)
3. Wang, S., Wang, X., Yu, J., Ma, S., Liu, M.: Bi-objective identical parallel machine scheduling to minimize total energy consumption and makespan. Journal of Cleaner Production **193**, 424–440 (2018)

Improvements in Large Neighborhood Search for the Electric Autonomous Dial-A-Ride Problem*

Maria Bresich[1]✉, Günther R. Raidl[1], and Steffen Limmer[2]

[1] Institute of Logic and Computation, TU Wien, Vienna, Austria
{mbresich, raidl}@ac.tuwien.ac.at
[2] Honda Research Institute Europe GmbH, 63073 Offenbach/Main, Germany
steffen.limmer@honda-ri.de

Due to increasing mobility and transportation demands, the interest in shared and on-demand transportation services is growing and leads to high practical relevance of this research area. In the dial-a-ride problem (DARP), a fleet of vehicles has to provide service to users with transportation requests consisting of a pickup and a drop-off location and a time window for either the departure or arrival, while allowing different customers to share a vehicle. The goal of the standard DARP is to find minimum cost routes serving all requests and satisfying a set of constraints concerning in particular vehicle capacities and user ride times. Diverse variants of the DARP with different objectives and constraints are studied in the literature.

In this work, we consider specifically the electric autonomous dial-a-ride problem (E-ADARP) as first introduced by Bongiovanni et al. [2]. The E-ADARP is a challenging extension to the DARP, where electric autonomous vehicles are employed and their charging requirements have to be taken into consideration. Route schedules now may include stops at charging stations to (partially) recharge the electric vehicles' batteries and constraints regarding the battery capacity and state of charge must be fulfilled. User inconvenience in terms of user excess ride time also plays an important role and has to be taken into account. User excess ride time is the difference between the direct and the actual travel time from a user's pickup to their drop-off location and is caused by potential detours because of the ride-sharing. The objective of the E-ADARP is to minimize the routes' cost consisting of the weighted sum of total travel times and the total user excess ride times.

Bongiovanni et al. [2] formulated the problem as a mixed integer linear program (MILP), which they successfully applied to instances with up to five vehicles and 50 customers. As such an exact approach has substantial difficulties to scale to larger instances and therefore has only limited practical relevance, Su et al. [5] proposed a deterministic annealing (DA) metaheuristic for solving the E-ADARP. This approach applies an exact route evaluation scheme of linear time complexity based on a forward labeling algorithm. The authors introduce, as a central aspect, the concept of a battery-restricted fragment, which represents a subsequence of pickup and drop-off locations for which the minimum user excess ride time can be independently optimized. Large neighborhood search (LNS) [4] is a metaheuristic

* This project is financially supported by Honda Research Institute Europe GmbH.

that is frequently applied with great success to diverse DARP variants. So did Bongiovanni et al. [1] for the E-ADARP. Moreover, Limmer [3] proposed a bilevel LNS variant in which the outer level employs multiple operators to schedule the charging of the electric vehicles and the inner level greedily inserts customer requests into the routes. This approach is highly scalable, solves instances with up to thousands of requests in reasonable time, and yields on medium-sized and large benchmark instances so far the best results.

We follow this line of research and investigate an LNS for the E-ADARP that combines and significantly extends concepts from the mentioned earlier works. In contrast to previous work, we do not directly deal with charging stops in the destroy and repair operators of the LNS but instead propose a novel route evaluation procedure that inserts them as needed on-the-fly in an optimal way; hereby, also respective charging times are determined. As a baseline approach, we formulate the charging station insertion subproblem as a MILP, but we also propose a computationally more efficient strategy: Considering existing benchmark instances and results reported in the literature, we conclude that optimal solutions rarely contain routes with more than one charging stop. With this in mind, we develop a linear time algorithm that efficiently inserts one charging stop, if needed, in an optimal fashion in each route. The main idea is to find the latest stop in a route before which charging is necessary and then to check for each preceding position in the route whether the insertion of a charging stop can make the route feasible. Over all feasible insertion possibilities, the one inducing the minimum detour length is selected. If a single charging stop is not sufficient to make a route battery feasible, we may either discard the route or apply the MILP. For efficient cost computations, we make use of the concept of battery-restricted fragments but different to [5], we do not enumerate and evaluate all possible fragments of an instance in advance. Instead, we only evaluate occurring fragments on demand and cache results.

Detailed tuning and computational experiments are still work in progress, but the proposed approach already yields excellent results on small to medium-sized instances that often match and sometimes beat those from the literature in terms of quality and runtime.

References

1. Bongiovanni, C., Kaspi, M., Cordeau, J.F., Geroliminis, N.: A machine learning-driven two-phase metaheuristic for autonomous ridesharing operations. Transportation Research Part E: Logistics and Transportation Review **165**, 102835 (2022)
2. Bongiovanni, C., Kaspi, M., Geroliminis, N.: The electric autonomous dial-a-ride problem. Transportation Research Part B: Methodological **122**, 436–456 (2019)
3. Limmer, S.: Bilevel large neighborhood search for the electric autonomous dial-a-ride problem. Transportation Research Interdisciplinary Perspectives **21**, 100876 (2023)
4. Pisinger, D., Ropke, S.: Large neighborhood search. In: Gendreau, M. (ed.) Handbook of Metaheuristics, pp. 399–419. Springer, 2 edn. (2010)
5. Su, Y., Dupin, N., Puchinger, J.: A deterministic annealing local search for the electric autonomous dial-a-ride problem. European Journal of Operational Research **309**(3), 1091–1111 (2023)

A Hybrid Metaheuristic for a Tourist Route Recommender

Cristina González-Navasa[1][0009−0007−6204−4833], José Andrés Moreno Pérez[2][0000−0001−9506−5197], and Julio Brito[3][0000−0003−3857−4707]

[1] I3MA PhD student alu0100457599@ull.edu.es
[2] Universidad de La Laguna, IUDR and DIIS jamoreno@ull.edu.es
[3] Universidad de La Laguna, IUDR and DIIS jbrito@ull.edu.es

Extended Abstract

There is a lot of information on tourist destinations so that tourists can organise their trips in advance. However, all this information is often unmanageable and imprecise, so planning a trip becomes complicated: selecting activities and points of interest (POIs) and organizing the route based on available time and other limitations. It must be taken into account that today, the tourist experience is related to personal satisfaction, so it is increasingly necessary to offer tools that help tourists and propose personalized itineraries based on their interests.

This work is part of a larger project to develop a tourist route recommendation system. For its development, the available data on the activities and POIs in the Canary Islands and the behavior of tourists when visiting the destination are used. The project incorporates techniques that recommend points of interest based on previous experiences. Based on this recommendation of activities and points of interest, the system must propose itineraries that satisfy tourists' preferences, considering tourists' time limitations and the availability of activities and POIs. This work aims to provide a model and appropriate methodologies for planning optimal itineraries that adequately adapt to the available data and to the tourists' preferences and limitations.

At the heart of this research lies the intricate design of tourist itineraries, denoted the Tourist Trip Design Problem (TTDP). These problems have been modelled in the literature using several variants of the Orienteering Problem (OP) [3]. Different versions of OP and Team Orienteering Problems (TOP) are formulated and solved according to their characteristics and limitations, such as the Team Orienteering Problem with Time Windows (TOPTW), which considers the opening window of each POI [1], [5].

The research addresses the complexity inherent in developing personalized itineraries with restrictions on the availability of activities and POIs, specifically the TOPTW. Furthermore, many of these restrictions are established with imprecision and vagueness. Tourists usually have high flexibility, optimizing their time and setting their itineraries.

The application of metaheuristics to address TTDP is widely used [4]. We use hybrid metaheuristics based on constructive procedures and local searches [2].

This research contributes to the growing set of knowledge on the applications of hybrid metaheuristics to TOPTW, shedding light on the potential of hybrid metaheuristics to solve TOPTW. Furthermore, the development of a personalised tourist route recommendation system expands the practical implications of our work, offering a user-centred approach to improve the overall tourist experience. The tool under development selects a personalised set of points of interest, based on tourists' preferences. Future work will expand our exploration of hybrid metaheuristics, incorporating other techniques such as Machine Learning and simulation procedures, and delve into the implementation and evaluation of the route recommendation system.

References

[1] D. Gavalas, C. Konstantopoulos, K. Mastakas, and G. Pantziou, "A survey on algorithmic approaches for solving tourist trip design problems," *Journal of Heuristics*, vol. 20, pp. 291–328, 2014. DOI: 10.1007/s10732-014-9242-5.

[2] J. Ruiz-Meza, J. Brito, and J. R. Montoya-Torres, "A GRASP-VND algorithm to solve the multi-objective fuzzy and sustainable Tourist Trip Design Problem for groups," *Applied Soft Computing*, vol. 131, pp. 109–716, 2022. DOI: 10.1016/j.asoc.2022.109716.

[3] J. Ruiz-Meza and J. R. Montoya-Torres, "A systematic literature review for the tourist trip design problem: Extensions, solution techniques and future research lines," *Operations Research Perspectives*, vol. 9, pp. 100–228, 2022. DOI: 10.1016/j.orp.2022.100228.

[4] M. Tenemaza, S. Lujan-Mora, A. De Antonio, and J. Ramirez, "Improving Itinerary Recommendations for Tourists Through Metaheuristic Algorithms: An Optimization Proposal," *IEEE Access*, vol. 8, pp. 79 003–79 023, 2020. DOI: 10.1109/ACCESS.2020.2990348.

[5] P. Vansteenwegen and A. Gunawan, *Orienteering Problems: Models and Algorithms for Vehicle Routing Problems with Profits*. Springer, 2019. DOI: 10.1007/978-3-030-29746-6.

Learning to Select Promising Initial Solutions for Large Neighborhood Search-Based Multi-Agent Path Finding*

Marc Huber[1], Christian Blum[2], and Günther R. Raidl[1]

[1] Institute of Logic and Computation, TU Wien, Austria
{mhuber,raidl}@ac.tuwien.ac.at
[2] Artificial Intelligence Research Institute (IIIA-CSIC), Campus of UAB, Spain
christian.blum@iiia.csic.es

Multi-Agent Path Finding (MAPF) is the NP-hard problem [5] of computing collision-free paths for a group of agents in a shared space while minimizing their sum of travel times. It has a wide variety of contemporary practical applications, including warehouse logistics, airport operations, and autonomous vehicles [4]. For such applications, MAPF instances can involve several hundred or occasionally thousands of agents. Unfortunately, attempts to solve MAPF instances with already a few hundred of agents using (bounded-sub-)optimal MAPF algorithms typically require too much time or memory. In contrast, fast constructive heuristics usually only find low-quality solutions or no feasible solutions in tightly constrained scenarios at all. An attractive third approach that combines the advantages of both worlds is the family of anytime algorithms. These approaches aim to find feasible solutions quickly and then continuously work on improving them as long as time remains, eventually converging to optimality.

MAPF-LNS [1] is the current state-of-the-art large neighborhood search (LNS)-based approach for anytime, suboptimal MAPF solving. It first produces a feasible initial solution quickly using an efficient existing MAPF algorithm. Then, MAPF-LNS repeatedly applies randomized destroy operators and replans subsets of agents to improve the solution further until the allocated time expires.

Clearly, the time required to generate a feasible initial solution is critical to the success of MAPF-LNS as it determines the remaining time for the LNS. One approach to address this issue is to set a tight time limit for the employed initial solution search algorithm. If it finds a feasible initial solution before the time limit is reached, MAPF-LNS immediately proceeds with the LNS. Otherwise, the state-of-the-art suboptimal algorithm MAPF-LNS2 [2] is invoked to make the solution feasible via LNS before continuing with the LNS of MAPF-LNS.

In this paper, we use prioritized planning (PP) [3] with a randomized priority order of agents and MAPF-LNS2 to find feasible initial solutions for MAPF-LNS.

As we will experimentally analyze, the final solution quality of MAPF-LNS depends significantly on the feasible initial solution, although the feasible initial solution's quality is not correlated much with the final solution quality. In this

* This project is partially funded by the Doctoral Program "Vienna Graduate School on Computational Optimization", Austrian Science Foundation (FWF), grant W1260-N35.

work, we propose to run PP and MAPF-LNS2 many times to create a larger pool of potential feasible initial solutions, from which we then select by means of machine learning (ML) a most promising solution to run the LNS on.

More specifically, we train a LambdaMART model to compute *scores* for the pool of PP and MAPF-LNS2-produced initial solutions and do the selection based on these scores. Thus, the learned scoring function should reflect how promising each initial solution candidate is to ultimately get a final solution of highest quality.

Training data is generated as follows: Given a map M and n agents, we randomly select n start and target vertices to generate a training instance I_M^n. This procedure is repeated several times to obtain a set of training instances $\mathcal{I}_M^n$. For each training instance $I_M^n \in \mathcal{I}_M^n$, PP and MAPF-LNS2 is executed with randomized agent ordering p times to collect p different initial solutions for each training instance (query group). From each of the p initial solutions of I_M^n, we extract 16 hand-crafted agent features such as, e.g., the delay of each agent and the degree of their target vertices. Finally, we compute the minimum, maximum, sum, and average over all agent features with respect to each instance $I_m^n \in \mathcal{I}_M^n$, and normalize them using min-max normalization. The label of a feature vector for the supervised learning is the median of the final solution quality of several LNS runs with respect to the corresponding initial solution and instance.

During testing, we again generate p initial solutions and use the trained ML model to select the most promising solution. The total time limit in the training data generation and final tests is set to 60 seconds.

Empirical experiments on well-established MAPF benchmark instances show that the trained ML model selects the most promising solution from the pool of potentially feasible initial solutions, such that the subsequent LNS leads to a lower-cost final solution than the original combination of MAPF-LNS2 and MAPF-LNS with PP. Moreover, the ML model trained on a map with 50 agents scales well to the same map with hundreds of agents, yielding superior results compared to the original combination of MAPF-LNS2 and MAPF-LNS with PP.

References

1. Li, J., Chen, Z., Harabor, D., Stuckey, P.J., Koenig, S.: Anytime Multi-Agent Path Finding via Large Neighborhood Search. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI). pp. 4127–4135 (2021)
2. Li, J., Chen, Z., Harabor, D., Stuckey, P.J., Koenig, S.: MAPF-LNS2: Fast Repairing for Multi-Agent Path Finding via Large Neighborhood Search. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). pp. 10256–10265 (2022)
3. Silver, D.: Cooperative Pathfinding. In: Proceedings of the First AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment. pp. 117–122 (2005)
4. Veloso, M., Biswas, J., Coltin, B., Rosenthal, S.: CoBots: Robust Symbiotic Autonomous Mobile Service Robots. In: Proceedings of the 24th International Conference on Artificial Intelligence. pp. 4423–4429 (2015)
5. Yu, J., LaValle, S.M.: Structure and Intractability of Optimal Multi-Robot Path Planning on Graphs. In: Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence. pp. 1443–1449 (2013)

A Learning Bilevel Optimization Approach for the Demand Maximizing Battery Swapping Station Location Problem*

Laurenz Tomandl[1], Thomas Jatschka[1], Günther Raidl[1], and Tobias Rodemann[2]

[1] Institute of Logic and Computation, TU Wien, Austria
{ltomandl,jatschka,raidl}@ac.tuwien.ac.at
[2] Honda Research Institute Europe, Offenbach, Germany
tobias.rodemann@honda-ri.de

A problem for the wide-scale adoption of electric vehicles are the usually long battery charging times. To avoid the waiting time for the customer, vehicles with exchangeable batteries and a network of battery swapping stations are a promising solution for smaller-scale vehicles like electric scooters. A customer can drive to a station and exchange their depleted batteries with an already charged battery and thus avoid the waiting that would be necessary otherwise. A good infrastructure is needed to make the implementation of such a system viable. To plan and design this infrastructure we consider the Demand Maximizing Battery Swapping Station Location Problem (DMBSSLP), which is an adaption from the Multi-Period Battery Swapping Station Location Problem (MBSSLP) by Jatschka et al. [2, 3]. We propose a learning bilevel optimization (LBLO) algorithm for solving large-scale instances of the DMBSSLP. This work is an extension of the first author's master thesis [5].

In the DMBSSLP we are given a graph in which vertices represent small geographical areas in which battery stations may be set up and edges represent the connections among these areas with respective travel times. Expected customer trips are described in terms of their origin and destination areas (OD-pairs) and have an amount over a time unit associated. To swap batteries during a trip we assume users are willing to make detours from their fastest paths to reach battery stations, but we expect customer loss if these detours get long. Battery swapping stations can be set up with different capacities. The total number of stations and battery slots is bound by the budget of a problem instance as stations and slots have different associated costs. The goal is to maximize the number of satisfied customers by placing battery swapping stations in convenient areas for the users.

The DMBSSLP is a maximum coverage location problem. Similar problems, in terms of facility management, have been investigated among others by Gazani et al. [1] who consider a superset of constraints considered in the DMBSSLP but only solve small instances with up to 50 areas and 200 OD-pairs using handcrafted heuristics as well as genetic algorithms. Jatschka et al. [2] originally proposed the MBSSLP, which is similar but considers as objective to minimize the budget for achieving a certain coverage instead of maximizing the latter.

* This project was financially supported by Honda Research Institute Europe.

With a large neighborhood search the authors could heuristically solve instances with up to 2000 areas and 8000 OD-pairs reasonably well. In their successive work [3, 4], the scalability of the approach was improved to instances with up to 10000 areas and 100000 OD-pairs by applying multi level optimization (MLO). Critical in multi level optimization is the way how to successively coarsen a problem instance. To this end, Tomandl [5] worked in his master thesis on an improvement in which a neural network is trained on representative instances in order to come up with a better coarsening function.

Building on this result, we now propose the Learning Bilevel Optimization (LBLO) that generates a solution to an MBSSLP instance in just three steps. First the problem size is reduced by clustering and coarsening the instance into a now smaller instance. This clustering is guided by a neural network that predicts the quality of merges for different areas. Then, a solution to this coarsened instance is determined, and finally the obtained solution is projected back to the original problem instance in an advanced way. To address the different subproblems we use mixed integer linear programming (MILP) models that we solve with Gurobi. The simpler bilevel architecture of LBLO but more clever clustering and advanced projection method better utilize the strengths of the MILP solver in particular when dealing with larger problem instances.

We experimentally evaluate LBLO on benchmark instances with up to 10000 areas and 100000 user trips, including comparisons of different methods for performing the clustering. While tuning and detailed tests are still work in progress, initial results already show that the new bilevel optimization approach often yields solutions of significantly higher quality in shorter time in particular for large problem instances than the original MLO approach.

References

1. Hatami Gazani, M., Niaki, S.A.A., Niaki, S.T.A.: The capacitated maximal covering location problem with heterogeneous facilities and vehicles and different setup costs: An effective heuristic approach. International Journal of Industrial Engineering Computations **12**(1), 79–90 (2021)
2. Jatschka, T., Oberweger, F.F., Rodemann, T., Raidl, G.R.: Distributing battery swapping stations for electric scooters in an urban area. In: Olenev, N., et al. (eds.) Optimization and Applications, Proceedings of OPTIMA 2020. LNCS, vol. 12422, pp. 150–165. Springer (2020)
3. Jatschka, T., Rodemann, T., Raidl, G.R.: A multilevel optimization approach for large scale battery exchange station location planning. In: Pérez Cáceres, L., Stützle, T. (eds.) Evolutionary Computation in Combinatorial Optimization. LNCS, vol. 13987, p. 50–65. Springer (2023)
4. Rodemann, T., Kataoka, H., Jatschka, T., Raidl, G.R., Limmer, S., Hiromu, M.: Optimizing the positions of battery swapping stations – pilot studies and layout optimization algorithm. In: Proceedings to the 6th International Electric Vehicle Technology Conference. p. 20231015. JSAE, Pacifico Yokohama, Japan (2023)
5. Tomandl, L.: A learning multilevel optimization approach for a large location allocation problem (2023), MSc thesis, TU Wien, Vienna, Austria

Enhancing Manufacturing Efficiency through Integration of MBSE and Capella in the Digital Thread

Michael Burgstaller[2] [0009-0007-1532-1586], Thomas Schichl[2] [0009-0001-4086-4810] and Mario Jungwirth[1] [0000-0002-1652-8745]

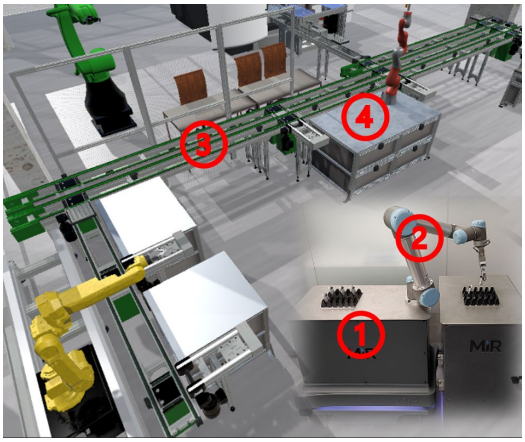
[1] Research Group Smart Mechatronics Engineering
University of Applied Sciences Upper Austria, Wels Campus,
Wels, Austria
mario.jungwirth@fh-wels.at
[2] University of Applied Sciences Upper Austria, Wels Campus,
Wels, Austria

Abstract. The advancement of digitalization in manufacturing has brought about a significant transformation in the industry. The digital thread, a key component in this transformation is defined as the use of digital tools and representations for design, evaluation, and life cycle management. This paper presents a project focused on enhancing the digital thread of a manufacturing line by integrating Model-Based Systems Engineering (MBSE) and Capella. It automates part identification, handling, and transportation in a small-scale manufacturing line, utilizing data from programs like Siemens NX in Capella to generate assembly code for manufacturing. The integration of various software tools like Capella, Siemens NX, and Mechatronics Concept Designer (MCD), along with hardware components like robots and Automated Guided Vehicles (AGVs), is crucial for optimizing manufacturing processes. The physical key components include cameras, sensors, and the UR10 robotic arm, demonstrating a use case of a flexible transport system (FTS) with a gripper and conveyor belt for pick-and-place tasks. The anticipated outcome is improved communication in the digital thread and increased production efficiency through automated processes, reduced manual intervention, and a more agile response to design changes.

Keywords: Digital Thread, MBSE, Siemens NX, Mechatronic Concept Designer, Arcadia method, Capella

1 General setting

This work integrates digital tools to enhance manufacturing processes. It leverages software tools as Capella, Siemens NX and others to automate and optimize manufacturing workflows, highlighting the need for efficient data exchange and system integration [1]. The focus of this project is to define a methodology and implement the requirements in the MBSE environment using various software that interact with each other. Key components like cameras, sensors, grippers, and the UR10 robotic arm are used for demonstrating the use case in a small-scale manufacturing line. This line is part of the Center for Smart Manufacturing (CSM) laboratory at the University of applied sciences upper Austria, Wels Campus.



1 Mobile industrial Robot MiR100

2 Universal robot UR10

3 Conveyor belts

4 Processing station

Fig. 1. A model of the setup in the CSM-laboratory as well as the automated guided vehicle system and robotic arm.

2 Outcome and benefits

Based on MBSE and the Arcadia method, a concept displaying the digital thread should be developed [2], [3]. Interfaces for an efficient bidirectional software data transfer should be engineered to ensure the foundation of an efficient infrastructure.

This also fosters research and development skills, exploring opportunities for optimization and innovation in manufacturing processes [4].

Additionally, it serves as an educational tool by providing hands-on experience in modern manufacturing processes. It offers co-workers and students the opportunity to learn about the intersection of mechanical engineering, computer science, and robotics, thereby enhancing their interdisciplinary understanding. The challenges involved in data and hardware integration help to develop critical thinking and problem-solving skills.

Acknowledgments. This work is part of the project “Requirements-driven Digital Transformation Competences in Mechanical and Plant Engineering (TraceMe)”. FTI Initiative: UpperVISION2030 OÖ Digitale Transformation 2021/22.

References

1. Sinnwell, C.: Methode zur Produktionssystemkonzipierung auf Basis früher Produktinformationen – Ein Beitrag zur Integration von Produktionssystemplanung und Produktentwicklung unter Einsatz des MBSE (Dissertation, ISBN 978-3-95974-127-9). (2019).
2. Shashank P. A.: Evaluating ARCADIA/Capella vs. OOSEM/SysML for System Architecture Development. Technical report Siemens AG. (2019).
3. “Arcadia method” https://mbse-capella.org/arcadia.html. (accessed: 13.11.2023).
4. “System modeling workbench for teamcenter,” https://www.obeo.fr/images/products/misc/Siemens-PLM-System-Modeling-Workbench-for-Teamcenter-fs-5244-A8.pdf. (accessed: 13.11.2023).

Leveraging Learning Factories for Mechatronic Systems Development: A Collaborative Approach

Pavel Kopylov[1] Jule Wegen[1] Thomas Schlechter[2]
Karen Manfredi[3] Letizia Nicoletti[3]
Antonio Padovano[4] Martina Cardamone[4] and Emmanuel Francalanza[5]

[1] EIT Manufacturing CLC North, Gothenburg, Sweden
{pavel.kopylov,jule.wegen}@eitmanufacturing.eu
[2] University of Applied Sciences Upper Austria, 4600 Wels, Austria
thomas.schlechter@ieee.org
[3] CAL-TEK S.r.l., 87036 Rende (CS), Italy
{k.manfredi, l.nicoletti}@cal-tek.eu
[4] Department of Mechanical, Energy and Mechanical Engineering (DIMEG), University of Calabria, 87036 Arcavacata di Rende (CS), Italy
{antonio.padovano,martina.cardamone}@unical.it
[5] Department of Industrial and Manufacturing Engineering, University of Malta, Msida MSD 2080, Malta
emmanuel.francalanza@um.edu.mt

Keywords: Industrial Engineering · Learning Network · Training Collaboration.

This article explores the transformative potential of learning factories in mechatronic systems development. Learning factories offer a dynamic, collaborative environment that bridges the gap between academia and industry, creating a mutually beneficial ecosystem [1]. The case study presented within the context of the LEONARDO project - which is the starting point for the discussion in this paper - underscores the power of international cooperation, the sharing of knowledge, and the collective drive to advance mechatronic education and innovation [2].

In the realm of mechatronics, an intricate fusion of mechanical engineering, electronics, and computer science, the education of future engineers demands an environment that reflects the dynamism of the field. Traditionally, learning factories found their primary home in universities. In the LEONARDO project, we intend to involve not only universities but also commercial companies and industry associations, bridging between theory and practice, offering a vibrant, immersive setting where students learn by doing. The LEONARDO project aims to support the digital transformation plans of higher education institutions (HEIs) by introducing innovative teaching methods and increasing the capacity and readiness of these institutions to manage an effective shift towards digital education [3]. Specifically, LEONARDO focuses on developing a pedagogical approach and curriculum design guidelines that support the teaching of Industry

5.0 content via a combination of digital technologies, such as e-learning, as well as a Learning and Experimenting open-Access Factory (LEAF). The LEAF toolkit and other materials will be developed to support professors and instructors in teaching Industry 5.0 concepts and human-centric factories, effectively "training the trainers". The concept of learning factories is not new, and it has taken various forms over the years. Within the framework of EIT Manufacturing, multiple projects have explored learning factories in the context of manufacturing and Industry 4.0. Initiatives like TFKnownet [4], Lift Europe [5], and Mach4.0 [6] have demonstrated the potential of learning factories in enhancing students' practical knowledge, preparing them for the industry's challenges. However, the competence of learning factories can be shared more effectively. To enhance Europe's competitive position, access to learning factories and resources must be made available to a broader community. EIT Manufacturing is actively constructing a platform for sharing training resources across Europe. Websites like skills.move, an online course offering training in various manufacturing areas, are integral to this effort. The LEONARDO project plays a significant role in this initiative, driven by the belief that the experience gained should be shared with the broader community. Collaboration and knowledge sharing are the keystones to advancing mechatronics. Collaboration fosters a strong, interconnected learning ecosystem that advances mechatronic education. Collective sharing strengthens our knowledge and benefits the entire community. This approach transcends individual institutions and aims to enhance education and innovation on a broader scale. As we journey forward, the LEONARDO project partners are happy to extend an open invitation to institutions and organisations to join us.

References

1. Abele, E., Chryssolouris, G., Sihn, W., Metternich, J., ElMaraghy, H., Seliger, G., Sivard, G., ElMaraghy, W., Hummel, V., Tisch, M., Seifermann, S.: Learning factories for future oriented research and education in manufacturing. CIRP Annals **66**(2), 803–826 (2017)
2. Darun, M., Palm, D., Athinarayanan, R., Hummel, V., von Leipzig, K.: The learning factory – a new stimulus to enhance international collaboration. Procedia Manufacturing **31**, 290–295 (2019), research. Experience. Education. 9th Conference on Learning Factories 2019 (CLF 2019), Braunschweig, Germany
3. Hulla, M., Hammer, M., Karre, H., Ramsauer, C.: A case study based digitalization training for learning factories. Procedia Manufacturing **31**, 169–174 (2019), research. Experience. Education. 9th Conference on Learning Factories 2019 (CLF 2019), Braunschweig, Germany
4. EIT Manufacturing: Teaching Factory Knowledge Sharing Network. https://www.tfknownet.com/
5. EIT Manufacturing: LIFT European Network of Learning Factories. https://www.eitmanufacturing.eu/news-events/activities/lift-european-network-of-learning-factories/
6. EIT Manufacturing: Learning Factory to Implement Industry 4.0 in Machining. https://www.eitmanufacturing.eu/news-events/activities/learning-factory-to-implement-industry-4-0-in-machining/

CNN Based Radar Kick Sensor Gesture Recognition Prototype

Shadman Mahmud[1], Thomas Schlechter[1], and Andreas Loeffler[2]

[1] University of Applied Sciences Upper Austria, 4600 Wels, Austria
shadman.mahmud@students.fh-wels.at, thomas.schlechter@ieee.org
https://www.fh-ooe.at/en/
[2] Continental Business Area Autonomous Mobility, 88131 Lindau, Germany
andreas.3.loeffler@continental-corporation.com
https://www.continental.com/de/

Keywords: Gesture Recognition · FMCW Radar · CNN · Radar Kick Sensor.

The concept of kick sensors is aiding users to open or close vehicle doors applying a simple kick gesture using the foot. These sensors have usually been implemented using ultrasound, capacitive sensing, computer vision, and **24GHz** Continuous Wave (CW) radar. This paper discusses the algorithm development and implementation of a kick sensor on a **60GHz** frequency modulated CW radar platform using deep learning. The goal is to develop a robust yet cost-effective solution for real-time kick gesture recognition. This has been achieved using one transmitting and one receiving antenna, an efficient data compression approach, and a convolutional neural network with a low memory requirement that is capable of achieving **97**% accuracy on test data. The final prototype can detect kicks and send control signals to open or close a vehicle's tailgate at an accuracy level of **88**%. Future improvements are discussed as well.

Capacitive, ultrasonic, and vision-based sensors provide cheap solutions for kick sensing, but their performance is affected by variations in the physical environment around them [1]. Such variations have less of an effect on radar sensors. A Frequency Modulated CW (FMCW) radar sensor can perform kick-sensing operations while providing range and velocity information as well. This makes FMCW kick sensing worth exploring because of the versatility it provides over other technologies. It is a more expensive solution comparatively, but it provides the scope for integrating multiple functions besides the kick-sensing function.

Reference designs posted by Texas Instruments [2, 3] show that capacitive and ultrasonic solutions for kick sensors have been explored. Two vision-based approaches have been presented in [4] and [5] that use the rear fish-eye camera found in most modern vehicles for kick sensing. The former uses computer-vision algorithms like edge detection, whereas the latter uses a deep learning model. The former approach is less adaptable to varying environments, giving the deep learning method an advantage. A 24GHz CW radar is used for kick sensing by [6] and [7]. The classifier model in [7] achieves 99% accuracy while occupying 248kB of memory. More radar-based approaches have been explored, while the fields of application are different, including fall detection [8], human

activity classification [9, 10], and hand gesture recognition [11]. To the best of our knowledge, kick sensing using a 60GHz FMCW radar with deep learning is yet to be explored, which this document deals with.

A highly reliable 'kick' detection can be guaranteed using the approach followed in this research. The used combination of the proposed DSP chain along with a CNN has already exhibited that it can perform well, reaching 97% accuracy on the initial data-set while delivering 88% considering the modified real-world data-set. Future work on this kick sensor can include training the CNN thoroughly while covering multiple scenarios. It is noteworthy that the used Aurix baseboard incorporates a DSP chip for providing hardware acceleration when performing FFT. This has not been used in this study. Using the chip can improve performance such that every frame is processed as soon as it is available without depending on a buffer.

References

1. Kinnaird, C.: Comparing capacitive and ultrasonic kick to open sensing. Online (September 2017), https://e2e.ti.com/blogs_/b/behind_the_wheel/posts/comparing-capacitive-and-ultrasonic-kick-to-open-sensing
2. Texas Instruments: Reference design for capacitive kick sensor. Online (Mar 2023), https://www.ti.com/tool/TIDA-01409
3. Texas Instruments: Reference design for ultrasonic kick sensor. Online (Mar 2023), https://www.ti.com/tool/TIDA-01424
4. Su, C.L., Lu, M.C.: Vision-based automatic trunk opener for automotive electronics. In: 2018 IEEE International Conference of Safety Produce Informatization (IICSPI). IEEE (December 2018).
5. Liu, J., Chen, Y., Yan, F., Zhang, R., Liao, X., Wu, Y., Sun, Y., Hu, D., Chen, N.: Vision-based feet detection power liftgate with deep learning on embedded device. Journal of Physics: Conference Series **2302**(1), 012010 (July 2022).
6. Innosent GmbH: 24ghz cw radar kick sensor. Online (March 2023), https://www.innosent.de/en/automotive/kick-sensor/
7. Shankar, Y., Santra, A.: Valid kick recognition in smart trunks based on hidden markov model using doppler radar. In: 2019 International Radar Conference (RADAR). IEEE (September 2019).
8. Jokanovic, B., Amin, M.: Fall detection using deep learning in range-doppler radars. IEEE Transactions on Aerospace and Electronic Systems **54**(1), 180–189 (February 2018).
9. Kim, Y., Alnujaim, I., Oh, D.: Human activity classification based on point clouds measured by millimeter wave MIMO radar with deep recurrent neural networks. IEEE Sensors Journal **21**(12), 13522–13529 (June 2021).
10. Kim, Y., Moon, T.: Human detection and activity classification based on micro-doppler signatures using deep convolutional neural networks. IEEE Geoscience and Remote Sensing Letters **13**(1), 8–12 (January 2016).
11. Ehrnsperger, M.G., Brenner, T., Hoese, H.L., Siart, U., Eibert, T.F.: Real-time gesture detection based on machine learning classification of continuous wave radar signals. IEEE Sensors Journal **21**(6), 8310–8322 (March 2021).

Comparison of different battery-powered tag positions for lower limb gesture detection

Christina Hofer[1], Julia Mayer[1], Valentin Sturm[1], Klaus Pendl[1], Erwin Schimbäck[1], Christian Kastl[1], and Frederick Runte[2]

[1] Linz Center of Mechatronics GmbH, 4040 Linz, Austria
christina.hofer@lcm.at
https://www.lcm.at/
[2] HANNING ELEKTRO-WERKE GmbH & Co. KG, 33813 Oerlinghausen, Germany
frederick.runte@hanning-hew.com
https://www.hanning-hew.com/

Keywords: Gesture Recognition · Neural Network · Human-Machine-Interaction

Over the last several years, numerous technologies were further developed to simplify human work processes, especially in areas where remote controls and auxiliary services can contribute to time savings and work efficiency. Therefore, gesture recognition can be very useful in many types of processes where touchless control is considered. The definition of gesture recognition is the ability of an electronic device to identify gestures and execute commands based on those gestures. Gesture recognition systems can be applied to other areas such as sign language [1], home automation [2] or computer games [3] and video games[4].
In this work, we propose a gesture recognition system, which functions as a sucessor model of a version presented in [5]. Both versions allow for contactless control of a therapy table which is operated by a pocket-worn tag respectively attached to the body using lower limb gestures. In the first model [5], the tag was placed in the trouser pocket, which already delivered solid results, although there were still a few misbehaviours and a slow reaction time in order to stop the treatment bed's movement. Thus, a new position was considered to counteract the delay.
To evaluate the tag placement, data was recorded simultaneously with three tags placed on the exercising leg. One above the knee, one on the ankle and one on the instep of the foot. A gesture sequence was performed which was determined in advance. The test subject had to perform a sequence of different gestures including short breaks. After subsequent evaluation of the raw data generated by the tags, the position above the ankle was chosen for a more detailed analysis. In addition to the sensor data, videos were recorded to analyse exactly which gesture was performed at what time. Afterwards each frame of the video was manually classified by the corresponding gesture. The resulting labeled dataset was used to train a machine learning model.
A total of nine participants were recorded and numerous data was collected. All gesture sequences were recorded during standing as well as sitting on a chair. Please note that the neural network used is the same architecture as the previous

model, as we wanted to enable an exact and fair comparison of the positions. A convolutional neural network was used for the categorical learning. Gesture classes were categorised from zero to two, where two classes represent the gestures for moving the therapy table upwards respectively downwards, while the remaining class subsumes all other behaviours such as standing still or walking around. Due to the new placement of the tag and the updated data, the average accuracy after training the network was 97.6%. The previous model achieved an average accuracy of 93.3%. The new placement and the new recorded gestures obtained better results and a more robust performance of the system. The wireless control operates significantly smoother and faster, which enables a quick start/stop of the control of the therapy table, which was achieved by the new position. In addition, false detections were reduced as the tag was attached directly to the leg by a velcro fastener, reducing errors due to the orientation filter caused by the tag sliding around.

Acknowledgements This work has been supported by the COMET-K2 Center of the Linz Center of Mechatronics (LCM) funded by the Austrian federal government and the federal state of Upper Austria.

References

1. Tamires Martins Rezende, Silvia Grasiella Moreira Almeida, Frederico Gadelha Guimaraes: Development and validation of a Brazilian sign language database for human gesture recognition. Neural Computing and Applications, 33:10449-10467 (2021)
2. Sehoon Yang, Sangjoon Lee, Yungcheol Byun: Gesture Recognition for Home Automation using Transfer Learning. Artificial Intelligent, Robotics and Human-Computer Interaction, Track 2 (2018)
3. Nhat Vu Le, Majed Qarmout, Cungang Yang, Yu Zhang, Haoren Zhou: Hand Gesture Recognition Systems for Games. 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) (2021)
4. Achim Schade, Jonas Schulz, Vu Nguyen, Christian Scheunert, Sebastian Bodenstedt, Giang T. Nguyen, Stefanie Speidel, Frank H. P. Fitzek: On the Advantages of Hand Gesture Recognition with Data Gloves for Gaming Applications. 2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops): Demos (2023)
5. Christina Tischler, Klaus Pendl, Erwin Schimbäck, Veronika Putz, Christian Kastl, Thomas Schlechter: Lower Limbs Gesture Recognition Approach to Control a Medical Treatment Bed. Eurocast 2022

Comparison of physiological data acquisition for modeling of drivers in autonomous vehicles *

Raul Fernandez-Matellan[1 [0009−0006−0326−8039]], David Puertas-Ramirez[2 [0000−0002−0999−4115]], David Martín Gómez[1 [0000−0003−3764−5083]], and Jesus G.Boticario[2 [0000−0003−4949−9220]]

[1] Intelligent System Lab, Universidad Carlos III, Av. de la Universidad 30, Leganés, Spain
[2] aDeNu, Universidad Nacional de Educación a Distancia, Calle de Bravo Murillo 38, Madrid, Spain

Keywords: Physiological signals · Human-centric symbiotic artificial intelligence · Autonomous driving.

1 Introduction

Humans can undergo rapid emotional shifts, and these changes can significantly impact their ability to perform tasks. For instance, an enraged driver is more likely to exhibit aggressive driving behavior compared to the same driver when they are calm. Consequently, if we are developing human-centric symbiotic artificial intelligence (HCSAI) systems to help autonomous systems and drivers in their interaction, the intelligent system that controls the vehicle must adapt to the user's emotional state. This field of study falls under the umbrella of affective computing. Recognizing that everyone possesses unique characteristics and capabilities, we argue that the next frontier in Artificial Intelligence (AI) lies in creating human-centered models that enhance natural human abilities. HCSAI prioritize collaboration, cooperation, and mutual enhancement between humans and AI agents. They strive to establish an environment where humans and AI agents seamlessly work together, leveraging each other's strengths and compensating for their respective weaknesses. Rather than focusing on replacing human capabilities, AI systems are to be designed to empower them. This symbiotic relationship underscores the importance of collaboration and cooperation between humans and AI agents.

To model the emotional state effectively, the AI system must identify patterns triggered by each person's emotions [2]. These patterns can manifest in facial expressions, gestures, tone of voice, and physiological signals. Physiological signals are involuntary responses of the Autonomous Nervous System (ANS) within the human body. The advantage of using physiological signals is that they cannot be easily controlled. These signals encompass various measurements, such as skin temperature, electrocardiogram, and electrodermal activity (EDA).

* This work was supported by the Spanish Government under Grants TED2021-129485B-C41 and TED2021-129485B-C44.

However, the main challenge in managing these signals is that the most accurate devices tend to be intrusive, impacting user comfort and movement, as they typically require electrodes for recording cardiac and brain activity. Alternatively, some new devices aim to measure physiological signals using wireless technology, but their utility and accuracy have not been sufficiently proven at this point [1]. Hence, we propose achieving a balance between intrusiveness and accuracy using wristbands.

In our research, which is framed within the field of autonomous vehicles, precision and accuracy in the measurement of physiological signals is critical, that is why, we have tested, analyzed, and compared three different wristbands: Fitbit Sense 2, Empatica E4, and Emotibit. Fitbit Sense 2 is designed as a smartwatch for general users to monitor their health status, while Empatica E4 and Emotibit are wristbands primarily intended for research purposes rather than everyday use. Furthermore, there is limited research available on comparing devices for acquiring physiological signals. Building on our previous work, where we emphasized the importance of considering the user's state to enhance the autonomy of autonomous vehicles, this paper presents a comparison of these wristbands' capabilities for measuring physiological signals to predict a driver's state in a Level 3 of autonomy, where vehicles have "environmental detection" capabilities and can make informed decisions for themselves, such as accelerating past a slow-moving vehicle. But they still require human override.

Our analysis begins with an examination of the hardware components of each wristband, highlighting the different sensors, the physical variables they can measure, and their sampling frequencies. We then delve into the software required to operate each wristband. A crucial aspect of our research is having access to raw data, and each device offers different conditions for data access. The sensors in the wristbands are very sensitive to the user's hand and arm movements. We have compared the accuracy of each device both at rest and in motion of the autonomous vehicle user. Finally, we test the wristbands' performance in two distinct driving scenarios: a controlled Level 5 environment at the University Carlos III of Madrid, utilizing autonomous golf carts within the campus(icab), and a real-world driving environment (Levels 2-3) in Scotland, supported by the University of West of Scotland, using a Toyota Prius equipped with Comma OpenPilot technology.

References

1. Gao, Z., Ali, L., Wang, C., Liu, R., Wang, C., Qian, C., Sung, H., Meng, F.: Real-time non-contact millimeter wave radar-based vital sign detection. Sensors 2022, Vol. 22, Page 7560 **22**, 7560 (10 2022). https://doi.org/10.3390/S22197560, https://www.mdpi.com/1424-8220/22/19/7560/htm https://www.mdpi.com/1424-8220/22/19/7560
2. Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X., Yang, X.: A review of emotion recognition using physiological signals. Sensors 2018, Vol. 18, Page 2074 **18**, 2074 (6 2018). https://doi.org/10.3390/S18072074, https://www.mdpi.com/1424-8220/18/7/2074/htm https://www.mdpi.com/1424-8220/18/7/2074

Automatic Control of Robots via 5G Communication from EC

Akira Sasaki[1] and Atsushi Ito[2][0000-0002-9686-4019]

[1] GClue Inc., Aizuwakamatsu, Fukuhima, Japan
[2] Chuo University, Hachioji, Tokyo, Japan

Email: akira@gclue.jp, atc.00s@g.chuo-u.ac.jp

Abstract. This paper describes the results of the development of a POST5G-enabled robot driving and working function with automatic and remote control. This system is unique in that it is mainly configured to use the cloud. All the processes for automatic control of the robot were reallocated to the cloud side, leaving only the minimum configuration on the edge side to build a system that can run automatically.

Keywords: 5G, POST-5G, MEC, ROS.

1 Outline of the problem

Various challenges exist when performing automatic control of robots from MEC via 5G communication. In particular, differences in the communication environment are important. In the WiFi and wired environments, the robot can be controlled freely because the network is local, the delay is within 10 ms, and the math process is performed by a local PC. In contrast, in a 5G environment, the network is a WAN communication, the delay is 30 ms or more, and the computing process is performed in the cloud (MEC), which causes various failures. Examples of failures and countermeasures are described below.

2 Problems

2.1 Initial issues

(1) ROS2 cannot communicate Node search for ROS2 cannot be performed because multicast is not available in the cloud.
(2) Communication environment is not stable Communication is not stable in the 5G NSA in the office.
(3) GUI is required to operate ROS2 Nav2 Screen cannot be displayed on Display in cloud environment.

2.2 Problems that appeared when the robot started actual driving

(1) Self-position estimation is not stable tf2 data sharing does not work well between the robot and the cloud side
(2) ROS (Robot Operating System)2 Nav2 is not stable ROS2 Nav2 periodically crashes due to unknown reasons.

3 Our Solution

It was found that data sharing of "tf" and "tf static" did not go well, and the information that is the base of self-position estimation was not shared well between the edge and the cloud. The "tf" is a coordinate conversion library in ROS that manages the relationship between different coordinate frames of robots and sensors. We reconsidered the node configuration because of two factors: the transfer delay and the data structure of "tf".

4 Result

(1) The robot is built on the edge side with the minimum configuration of Jetson, etc., and the automatic running related parts are placed on the cloud side. The robot was built with the minimum configuration of Jetson, etc. on the edge side, and the automatic driving related parts were placed on the cloud side.
(2) Remote control and remote automatic operation of the robot from the MEC environment were realized using 5G communication. 5G communication was used to realize remote control and remote automatic operation of the robot from the MEC environment.

5 Future Challenges

(1) Currently, Intdash ROS2 Bridge is used for ROS2 data transfer between the edge and the cloud. Establishment of a method to realize the same functionality using OSS.
(2) Currently, the system configuration is based on docomo MEC. Realization of a configuration that works with general clouds such as AWS, Azure, GCP, etc.

Automatic Classification of Grazing Cow Behavior

Atsushi Ito[1], Kota Nagakura[1], Yuko Hiramatsu[1], Yoshikazu Nagao[2], Rika Fukumori[3], Hirosugu Yamamoto[4], Yukitoshi Otani[4], Shiro Suyama[4], Masaki Yasugi[5], Yasutoshi Yoshiura[5]

[1]Chuo University, Faculty of Economics, Tokyo Japan

[2]Utsunomiya University Faculty of Agriculture, Tochigi, Japan

[3]Rakunou Gakuen University, Food and Environment Sciences, Hokkaido, Japan

[4]Utsunomiya University Faculty of Engineering, Tochigi Japan

[5]Fukui Prefectural University, Faculty of Marine Science and Technology, Fukui, Japan

E-mail: [1]atc.00s@g.chuo-u.ac.jp, [1]a22.5j8s@g.chuo-u.ac.jp,[1]susana_y@tamacc.chuo-u.ac.jp, [2] ynagao@cc.utsunomiya-u.ac.jp, [3] fukumori@rakuno.ac.jp, [4] hirotsugu@yamamotolab.science, [4] otani@cc.utsunomiya-u.ac.jp, [4] s_suyama@yamamotolab.science, [5]myasugi@fpu.ac.jp, [5]yoshiur@fpu.ac.jp

Abstract. The situation surrounding dairy farmers is changing rapidly due to soaring feed costs and declining demand for dairy products, and an increasing number of farmers are being forced to leave the dairy industry. In order to resolve this situation and to improve Japan's food self-sufficiency ratio, there is a need for higher value-added dairy products. However, there are many problems with Japanese breeding methods, such as increased production costs, reproductive performance and health management of dairy cows. Pasturage is one way to solve these problems, but it has not been widely used because it requires pasture management according to the foraging conditions. In the past, we have been constructing a sensor network that enables health management by attaching acceleration sensors and pH sensors to cows, and this paper describes the results of our study on a technology for automatic classification of cow behavior during grazing.

Keywords: Grazing, AI, Behavior estimation.

1 Behavior Classification

This paper describes a technology for understanding cattle behavior during grazing, which is necessary information for feed management and health management in grazing, using IoT.

First, before going out to pasture, cows are fitted with equipment to log their acceleration and angular velocity. Then, we follow the cows while they are grazing and take video of their behavior. Currently, we are using smart phones. After grazing, the video is watched and labeled on a second-by-second basis. Using the data, we will build a machine learning model and evaluate its accuracy.

In this experiment, the labeling was done according to the following labels(Walking, Stopping, Eating, Chewing, Body licking, Licking other cows, Rubbing body against tree), and the data was used for learning.

2 Related Works

There have been many studies on cattle rearing methods outside of Japan. For example, in a study by Jessica Werner of the University of Hohenheim, Germany [1], cattle behavior was measured using RumiWatch and MooMonitor+, which are said to have higher accuracy than visual observation. In viewing cattle behavior as an important indicator of feed quantity, animal performance, health, and welfare, the study states that proper allocation of feed is important for achieving maximum pasture utilization and seeks ways to automate this process.

3 Result

In this study, we categorized the above data into three types of behavior: walking, resting, and foraging. We retrieved behavioral data from two cows while grazing and developed a process to analyze them. However, the resulting recognition rate was 70-74%. Since the recognition rate was 80% or higher when learning within the data from one cow, it is thought that there were large individual differences or that some noise was included, which lowered the recognition rate when integrated. We believe that it is possible to analyze cow behavior using accelerometers not only to compare data of one behavior with data of another behavior, but also to compare data of one behavior under a certain condition with data of the same behavior under another condition. For example, we foresee that there may be differences in data between walking on a hot day and walking when the temperature is low. In fact, there was a noticeable difference in the condition of cattle when grazed in August and when grazed in September when the weather was relatively cool, including fatigue due to the heat. If we can understand the differences in behavior under different temperatures by comparing such data, we will be one step closer to devising ways to prevent a decrease in milking yield. For example, we can estimate the number of hours after grazing that fatigue appears based on differences in acceleration while walking, and we can estimate the optimal amount of water intake based on comparisons of the degree of drinking.

References

1. [1] Jessica Werner, “Measuring grazing behavior of dairy cows”, Thesis for Doctoral, University of Hohenheim, 2019, https://www.researchgate.net/publication/332092488_Measuring_grazing_behaviour_of_dairy_cows

Programming Learning through the Use of ChatGPT

Ami Otsuka[1], Akira Sasaki[2], Atsushi Ito[1][0000-0002-9686-4019]

[1] Chuo University, Hachioji, Tokyo, Japan
[2] GClue Inc., Aizuwakamatsu, Fukuhima, Japan

Email: amiotsuka.se@gmail.com, akira@gclue.jp, atc.00s@g.chuo-u.ac.jp

Abstract. Programming skills are increasingly required in a variety of professions, but in general, learning programming is challenging, yet as the field of application expands, software engineers are in short supply. This paper describes a study we conducted with the goal of building a means to help beginning programmers learn coding and related theory in a systematic way by using ChatGPT as a new approach, thereby enabling a large number of people to acquire programming skills. This can be called a shift from coding-centered to documentation-centered learning. The purpose of this study is to improve individual skills, not to contribute to group programming projects.

Keywords: Programming education, ChatGPT, Software engineering.

1 Problems with Current Programming Education

(1) Focus of learning is biased toward code creation
Insufficient understanding of upstream processes
Lack of learning opportunities for V-model and reverse engineering

(2) Issues with traditional learning methods
Issues with book-based learning
Difficulty in confirming true understanding
Lack of educational materials on design and requirements definition
Code should be written by AI, not by humans

2 Proposal of New Educational Methods

(1) Need for a new educational method based on the problem
(2) Shift focus of learning to requirements definition
(3) Importance of developing ChatGPT-based educational materials

3 Proposed programming education flow

1. Learn the technology in classes and seminars, or use tools such as OpenAI. 2. think about what you want to make with the technology you learn.
2. Think about what you want to make using the technology you have learned, and consult with OpenAI.
3. Ask OpenAI to create a basic program for what you want to create (this program is called alpha version)
4. Create several versions of the program based on the alpha version, and create a simple requirements document based on the versions. Since there is a possibility of modification at this stage, a detailed definition is not necessary.
5. In order to create a program according to the requirements definition document, a basic design document is created with reference to the alpha version of the program. In some cases, a more detailed design document is also required. For example, if a game is to be created, parts such as the screen display and the system for playing against the CPU.
6. Create a program design document based on the basic design document.
7. Based on the program design document, create the program in cooperation with OpenAI.
8. Combine all the parts and complete the program.

4 Merit of this method

The advantages of this learning method are the ease of learning requirements definitions and the resulting benefits of learning a software engineering approach.

(1) Advantages of studying programming using ChatGPT
(2) Ease and simplicity of learning requirements
(3) Understanding the usefulness of the software engineering approach
(4) Effects of learning from interest and fascination and its impact on long-term memory

5 Conclusions

Programming education using ChatGPT has tremendous potential. The requirement-centered approach to learning allows students to realize their ideas. The approach lowers the barriers to concrete projects such as game development, and allows learners to develop their creativity and abilities. Facilitates self-directed learning and promotes deep understanding and long-term memory. However, appropriate questions, use of AI, and security risks must be considered. The importance of acquiring expertise and education will remain unchanged.

ChatGTP language model as a support for teaching and self-learning in a field of engineering and technical sciences

Anna Czemplik[0000-0002-0979-7322] and Iwona Karcz-Duleba [0000-0003-4946-2225]

Department of Control Systems and Mechatronics
Faculty of Information and Communication Technology
Wrocław University of Science and Technology
anna.czemplik@pwr.edu.pl, iwona.karcz-duleba@pwr.edu.pl

Keywords: Chat GPT, Teaching Support, Engineering, Technical Sciences.

Introduction. Artificial Intelligence (AI) based systems have recently sparked many heated discussions about their potential opportunities and threats. The subject of our analysis is not to assess the capabilities of artificial intelligence in general, but only the AI interface, which is ChatGTP, and its use to acquire and verify knowledge in an engineering education process.

With the release of ChatGTP by OpenAI [1c], many ideas for the use of AI by ordinary users, including students, appear in the public space. The key feature of ChatGTP is the use of a large language model (LLM) and an ability to have a conversation in a natural language – asking questions, generating texts on a given topic. It is known that ChatGTP is good at writing essays (much to the chagrin of teachers), writing simple computer programs, solving medical exams [2], [3], etc. It is also known that it often cheats, 'confabulates' or 'hallucinates' and makes up stories, so expert knowledge is needed to judge a correctness of answers. Therefore, we would like to find out if ChatGTP is suitable for supporting learning in engineering and technical sciences, which require more rigorous and precise language and where uncertain information can even lead to life and health threatening situations. How to prepare students for the effective use of this technology? How to use it wisely? How to ask questions? How to check the correctness of answers? We tested ChatGTP-3 because it is publicly available and free, so it is easy to use by students and people who want to learn more.

Characteristics of the research subject. The ChatGTP model are the subject of numerous studies and tests, but many results and conclusions are recently published on various forums dedicated to AI technologies because scientific publications have their limitations, such as a longer publication cycle. On the basis of the available sources, it is possible to state the main properties of ChatGTP:

- it does not quote, but constructs an answer – LLM forecasts a subsequent word (so-called token) based on a text observed so far;
- it does not determine the probability of word collocations,
- it does not provide sources of information (because it does not know them),

- it does not evaluate information, it is a language model, not a knowledge model,
- it was trained mainly on English texts (other languages are worse represented),
- it does not have internet access (it was trained on texts available until 2022),
- it has a tendency to 'hallucinate' or 'confabulate' factual information,
- it does not analyze images, charts or patterns.

Many of these features are serious barriers to using chat in technical education.

Methodology. The subject under consideration is the use of ChatGTP in the process of education and self-study in engineering fields. We are conducting research in areas such as classical control theory (well established) and metaheuristic optimisation methods (constantly evolving). Instructions (prompts) were given in English. We are asking about consolidated knowledge that were definitely known by 2022. We do not ask for formulas, derivations, diagrams, standards, etc.

The problem is how to examine and evaluate the usefulness of ChatGTP so that it is not user's opinion (impression). In literature, for example, a perceived usefulness and satisfaction of students were examined [4]. We have objections about such an assessment, because it does not measure an effectiveness of learning (e.g. exam pass rate). Typically, students perceive their knowledge as adequate due to their unawareness of what they ought to know and their ability to distinguish between what is important and less important. Therefore, we assumed that ChatGTP will be assessed in the same way as a student in an exam and grades will be given by specialists in a given field. We choose different types of tasks (formulating questions): single and multiple answer tests, open questions, criticism of a given text are among them. Questions will be performed repeatedly. Some research has already been done in a field of cybersecurity [5] but with different methodology. We hope that an analysis of received data will help find the right way to use AI to support teaching in engineering and technical sciences for the benefit of students and teachers.

References

1. OpenAI Homepage, https://platform.openai.com/
2. Kung T., Cheatham M., Medenilla A., et al.: Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. medRxiv. 2022.12.19.22283643. doi: 10.1101/2022.12.19.22283643 (2022)
3. Gilson, A, Safranek, C., Huang, T., Socrates, V., Chi, L. Taylor R., Chartash, D.: How Does ChatGPT Perform on the Medical Licensing Exams? The Implications of Large Language Models for Medical Education and Knowledge Assessment, medRxiv 2023, preprint doi: https://doi.org/10.1101/2022.12.23.22283901 (2023)
4. Boubker, O.,: From chatting to self-educating: Can AI tools boost student learning outcomes? Expert Systems with Applications 238(3):121820, (2023)
5. Malinka, K., Perešíni, M., Firc, A., Hujnák, O., Januš, F.: On the educational impact of ChatGPT: Is Artificial Intelligence ready to obtain a university degree? arXiv 2023, arXiv:2303.11146 (2022)

Analytical conversion of the Strejc model to the first order with time delay (FOTD) model

Anna Czemplik[0000-0002-0979-7322]

Wroclaw University of Science and Technology, Wroclaw, Poland
anna.czemplik@pwr.edu.pl

Abstract. The article presents straightforward formulas for transforming the Strejc model into an equivalent inertial model with time delay (FOTD). These formulas rely on expanding the exponential function into a Laurent series.

Keywords: Strejc Model, First Order Time Delay Model

1 Introduction

Control system design methods typically rely on a mathematical model as their foundation. A classic example is the tuning of PID controllers. Engineering methods predominantly employ simple transmittances that can be derived from the process curve using experimental techniques. The handbook [1] offers an extensive overview encompassing 1,731 design methods for various controller types based on diverse object model variations. Among these, 641 methods utilize the first-order inertial model with time delay (FOTD). The widespread acceptance of this model arises from its blend of simplicity and effectiveness — containing solely three parameters (k – gain, T_1 – time constant, T_0 – transport delay) to approximate any aperiodic process curve. Notably, none of the discussed tuning methods utilize the Strejc model, an equally straightforward model defined by three parameters (k – gain, T – time constant, n – model order). However, when the object is represented by the Strejc model, performing process curve-based identification can determine the parameters of the FOTD model [2].

The chosen modeling approach influences the model parameters, which turn affect the controller values resulting from the tuning rules. Fundamental model identification methods rely on inflection point parameters. (see Fig.1).

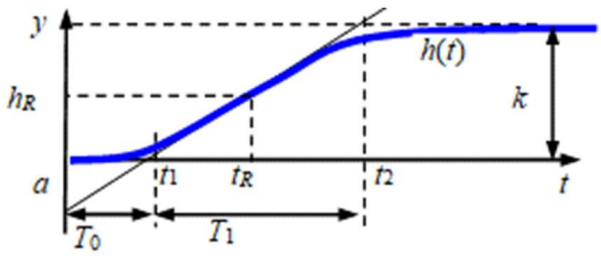


Fig. 1. Identification of FOTD model and the Strejc model based on the inflection point.

The parameters t_R and h_R are read from the process curve, and the slope R at the inflection point can be calculated from the following relationships

$$R = k / T_1 = a / T_0 = h_R / (t_R - T_0). \tag{1}$$

On this basis, the parameters of the FOTD and the Strejc models can be calculated as

$$G_1(s) = ke^{-sT_0} /(T_1 s + 1), \text{ gdzie: } T_1 = t_2 - t_1,\ T_0 = t_1, \tag{2}$$

$$G_n(s) = k /(Ts + 1)^n, \text{ gdzie: } n = 1 + 2\pi t_R^2 R / k^2,\ T = t_R / (n - 1). \tag{3}$$

2 Strejc model conversion

The alternative analytical conversion of T and n parameters to T_1 and T_0 is derived from established formulas used for calculating the step response parameters of the Strejc model. By utilizing the values of T and n, one can compute the time t_R, the value of the function h_R, and the slope R at the inflection point. Subsequently, these parameters then allow the determination of T_1 and T_0 in the FOTD model in the following form

$$T_1 = k / R = k_{T1} T, \text{ where } k_{T1} = (n-1)!e^{(n-1)} /(n-1)^{n-1} \approx a_0 + a_1\sqrt{n-1} \tag{4}$$

$$T_0 = t_R - \frac{h_R}{R} = k_{T0} T, \text{ where } k_{T0} = \frac{(n-1)^n - (n-1)!e^{n-1} + (n-1)!\sum_{i=1}^{n} \frac{(n-1)^{i-1}}{(i-1)!}}{(n-1)^{n-1}} \tag{5}$$

The Maclaurin series of the e^{n-1} function enables the simplification of formula (5),

$$T_0 = k_{T0} T, \text{ where } k_{T0} = (n-1) - (n-1)!k_n R_n /(n-1)^{n-1} \approx b_0 + b_1\sqrt{n} \tag{6}$$

and R_n – the Lagrange rest, k_n – the author's correction of the residual to ensure the accuracy of the formula for small n order. The direct application of these simple formulas is to facilitate the conversion from the Strejc model to the FOTD model. The presented solutions are also being analyzed in terms of developing the inverse conversion method, that is, from the FOTD model to the Strejc model, as an alternative to the Padé approximation [3].

References

1. O'Dywer, A.: Handbook of PI and PID controller tuning rules. Imperial College Press, London (2009)
2. Strejc V.: Naherungsverfahren für aperiodische Übertragungscharakteristiken. Regelungstechnik 1959/4
3. Åström K. J., Hagglund T.: PID Controllers, Theory, Design and Tuning, ISA – The Instrumentation, Systems and Automation Society, 2nd Edition, 1995.

Neural analysis of parameters occurring in a smart building.

Extended Abstract

Andrzej Stachno

Department of Control Systems and Mechatronics,
Wroclaw University of Science and Technology, Wroclaw, Poland
{andrzej.stachno@pwr.edu.pl}

Keywords: Artificial Neuron Networks, Successive Values of a Time Series, Environmental Measurements in an Smart Building.

Extended Abstract

Building automation, in addition to standard control systems, allows for the acquisition and archiving of measurements of parameters of devices operating in it. Sensory elements provide opportunities to represent the world around us in digital form. A perfect example is the KNX automation system, which was developed by electrical companies from various parts of the world. A key aspect of this system is the ability to exchange information between all devices via a communication bus and a unified communication protocol. The mutual compatibility of all devices enables the aggregation of various measurements, such as electricity consumption, temperature, level of volatile organic compounds (VOCs), CO2 concentration, atmospheric pressure, air humidity, and light intensity [4].

One of the most frequently used measurement functions in an intelligent building is monitoring the presence of people. This is accomplished using PIR sensors. However, there are more and more other methods that allow you to determine whether there is a person in a given room. For example, technologies using high frequencies, ultrasound and vision sensors such as cameras. The latter, due to the very high computational complexity of image analysis, are equipped with artificial neural networks. The use of these computational mechanisms opens up new detection possibilities by identifying human silhouettes and counting people in specific zones. Additionally, integrating them with measurements from other sensors allows you to obtain comprehensive measurement information from the space in which people are present. Integrated building automation systems, such as KNX, also use devices for archiving data sent via the communication bus [7]. Thanks to them, the system can collect large amounts of data. A thorough analysis of this data allows for drawing additional conclusions on related topics, e.g. related to the health of people staying in the room, the way they behave in various situations, their impact on the use of energy sources or the way of work and failure of devices affecting work comfort. An innovative approach to the analysis of measurement data is the selection and use of devices that can select the necessary measurements and, based on them, determine the values sought in

accordance with a specific criterion, for example, identifying a person in a room or determining the number of people present in it [6].

The second important aspect in the field of building automation is the analysis of electrical parameters. This mainly involves measuring the current consumed by electrical devices in the building. These measurements are carried out using dedicated measuring devices within the KNX system. Thanks to its distributed structure, this system enables the construction of even the most complex measurement networks.

The next step in the analysis algorithm is classification [1]. This process is a key step in preparing time series for neural analysis [2]. Correct classification of time series allows for a preliminary assessment of whether adequate conclusions can be drawn from the data contained therein. The artificial neural network initially performs classification based on previously defined patterns. Using the classification capabilities of artificial neural networks usually brings very good results while minimizing the computational burden [5]. The effect of correct classification is the correct identification of the phenomenon being searched for. The experiments performed showed that artificial neural networks also bring very good results here, minimizing the computational complexity of the algorithm.

References

1. Stanley N., Eugene H.: An Introduction to Econophysics. Correlations and Complexity in Finance, PWN, 2001.
2. Peitgen H.O., Jurgens H., Saupe D.: Introduction to Fractals and Chaos, PWN, 2002.
3. Technical Analysis of the Financial Markets, WIG PRESS, 1999.
4. Jabłoński A.**:** Intelligent buildings as distributed information systems. CASYS: International Journal of Computing Anticipatory Systems. vol. 21, s. 385-394, Liege,2008.
5. Stachno Andrzej, Jabłoński Andrzej: Hybrid method for forecasting next values of time series for intelligent building control / Andrzej Stachno and Andrzej Jabłonski [i in.]. W: Computer aided systems theory - EUROCAST 2015 : 15th International Conference, Las Palmas de Gran Canaria, Spain, February 8-13, 2015 : revised selected papers. Pt. 1 / Roberto Moreno-Díaz, Franz Pichler, Alexis Quesada-Arencibia (eds.). Berlin; Heidelberg: Springer, cop. 2015. s. 822-829,
(Lecture Notes in Computer Science, ISSN 0302-9743; vol. 9520)
6. Stachno Andrzej: Neural Personal Profile Identyfication Based on Environmental Performance Analysis of Inteligent Building, Computer aided systems theory - EUROCAST 2019 : 17th International Conference, Las Palmas de Gran Canaria, Spain, February 17-22, 2019 : revised selected papers. Pt. 2 / Roberto Moreno-Díaz, Franz Pichler, Alexis Quesada-Arencibia (eds.). Berlin; Heidelberg: Springer, cop. 2020. s. 388-395,
(Lecture Notes in Computer Science, ISBN 978-3-030-45095-3)
7. KNX Association: KNX Advanced Course Documentation – Januar 2019 ISBN-13: 979-8635617861

Scheduling dynamic set of computational tasks in multiprocessor embedded systems

Dariusz Dorota[1][0000−0002−7596−3552] and Czeslaw Smutnicki[2][:0000−0003−4640−5364]

[1] Cracow University of Technology, Cracow, Poland
[2] Wroclaw University of Science and Technology, Wrocław, Poland
ddorota@pk.edu.pl, czeslaw.smutnicki@pwr.edu.pl

Abstract. The paper studies so called multiprocessor scheduling problem which arises in the embedded systems, [5], dedicated among others for IoT (*Internet of Things*), IoV (*Internet of Vehicles*) with an augmented level of security. Enlarged reliability can be achieved through hardware as well as software redundancy. The decision is made by "voting" duplicated, identical devices (microprocessors) which execute the same programs for identical data. Applications one can find in aviation, space vehicles, military equipment, rockets, means of transporting hazardous materials, nuclear installations, chemical installations, mining, drones. The considered case refers to tasks, which are processed independently and in synchronous way on two processors. If a tasks provide contradictory results, we need to re-calculate them on three devices. Then the set of tasks for processing changes during the scheduling in an uncertain and dynamic manner. We provide a probabilistic analysis.

Keywords: Multiprocessor tasks · Scheduling · Dynamic change of tasks

The essentials are introduced as follows: set of tasks $T = \{1, 2, 3, ..., n\}$ has to be processed by identical machines (processors, devices) $M = \{1, 2, 3, ..., m\}$. Any task $i \in T$ needs simultaneous access to a_i, $1 \leq a_i \leq m$, processors, corresponds to redundant processing of the same code, and has duration $p_i \geq 0$. The goal is to find the schedule of tasks, namely allocation to processors and time events, which minimizes certain optimization criteria. We use to this aim the length of the schedule (makespan). Stated the above problem is known as "deterministic scheduling multiprocessor tasks", see e.g. [1, 2] Starting from this problem we pass to our main case "problem with uncertain change of task set" not considered so far in the literature, but considered recently by us, [3].

We lead further considerations for NoC (Network on Chip) embedded architecture with $m \geq 3$ and $a_i \leq 3$, $i \in T$, which is the most common in practice. If $a_i = 1$ then there are no formal objections to the result of the task. If $a_i = 3$ then the final result can be established by "voting of majority". If $a_i = 2$ and the task finishes with conflicting results, we have to re-run on-line an extra three-processor task or single-processor arbitration task to resolve ambiguous result. This implies dynamical change (enhancement) of the set T in non-forecasting

way. Let us assume that each task $i \in T$ with $a_i = 2$ provides identical result with probability $(1 - s)$, while ambiguous – with probability s. The latter case causes creation of additional task with $a_i = 3$ and the same as previously processing time p_i. We call this case as *fault*. It simply implies re-allocation of tasks on processors. Let us denote the set of tasks which require two processors by $Q = \{i \in T : \ a_i = 2\}$. We assume that tasks in T are indexed so that $Q = \{1, 2, \ldots, q\}$, $q = |Q|$, $q \leq n$. Let us define the vector of random variables $X = (x_1, x_2, \ldots, x_q)$ where x_i denotes the final result of $i \in Q$. We set $x_i = 0$ if both processors provided the same result, and $x_i = 1$ if both processors provided different results. In the former case, task i is not repeated, in the latter case we have to extend task set T by adding the new three-processor task i' with the processing time p_i. It is clear that X takes 2^q different values. Thus this describes the set of events for X. Since x_i takes value 0 or 1 independently, then the probability that X takes a particular value $x = (x_1, \ldots x_q)$ is equal

$$P(X = x) = \prod_{i \in Q; x_i = 1} s \prod_{i \in Q; x_i = 0} (1 - s) = s^{q-r}(1 - s)^r, \tag{1}$$

where $r = |\{i \in Q : \ x_i = 0\}|$. Note that x influences not only on the set T but also on the schedule as well as on the makespan value in an in-predictive way. We employ hereinafter denotation $C_{\max}(x)$ for the makespan for realization x of vector X. Special algorithms were already designed, [3], which work for fixed x. They have been designed through an extension of Muntz-Coffman algorithm, see [4] for details, towards the case of dynamic scheduling.

Further practical observations we found in a few numerical experiments. For example, we analyze the instance: $n = 13$; $p_i \in \{5, ...30\}$ selected randomly; $a_i \in \{1, 2, 3\}$, $i \in T$; $q = 4$; $(1 - s) = 0.9$; $s = 0.1$. Note that $P(X = x)$ depends on x, r, and s. Single *fault* occurs seldom $s = 0.1$. Next we checked all possible sequences of *faults*, namely $r = 0, 1, 2, 3, 4$. For $x = (0, 0, 0, 0)$ we obtain $C_{\max}(x) = 90$ with $P(X = x) = 0.6561$. On the opposite side for $x = (1, 1, 1, 1)$ we get $C_{\max}(x) = 140$ with $P(X = x) = 0.0001$. This means that "the bad luck" appears rarely. From the distribution of $P(x)$ and $C_{\max}(x)$ one can easily find the average value, which support us to evaluate quality of the algorithms.

References

1. Blazewicz, J., Ecker, K.H., Schmidt, G., Weglarz, J.: Scheduling in computer and manufacturing systems. Springer Science & Business Media (2012)
2. Drozdowski, M.: On the complexity of multiprocessor task scheduling. Bulletin of the Polish Academy of Sciences Technical Sciences, 43(3) (1995)
3. Dorota, D.: Scheduling multiprocessor task in conditions of uncertainty (Polish), PhD. Thesis, Wroclaw University of Science and Technology, Wroclaw, Poland, (2023)
4. Muntz, R. R., Coffman, E. Optimal preemptive scheduling on two–processor systems. IEEE Transactions on Computers 100, 1014–1020 (1969).
5. Sriram, S., Bhattacharyya, S.S.: Embedded multiprocessors: Scheduling and synchronization. CRC press (2018)

Efficient Manipulation of Control Flow Models in Evolving Software

Tomáš Fiedor[1], Jiří Pavela[1], Adam Rogalewicz[1], and Tomáš Vojnar[1]

Brno University of Technology, Faculty of Information Technology, Czech Republic
{ifiedortom,ipavela,rogalew,vojnar}@fit.vutbr.cz

1 Introduction

Over the last years, the software development community has seen a massive rise in the popularity of *Continuous Integration and Continuous Delivery (CI/CD)* practices. In particular, CI/CD aims to automate building, testing, and deploying software projects during their entire development lifecycle, naturally leading to *evolving software* that is constantly updated with new features without breaking the build.

In their current state, most CI/CD test environments focus on finding functional bugs only, i.e., bugs that lead to software crashes and/or incorrect results. The other types of bugs, such as *performance bugs*, are often overlooked. However, a study of Mozilla Firefox [9] concluded that, compared to any other type of bugs, performance bugs 1) take more time to fix, even up to 2.8 times more than security fixes; 2) require more experienced developers to fix; and 3) the fixes are spanning up to 2.6 times more files. Morever, if such bugs remain in the software unnoticed for a long time (thus becoming the so-called *dormant* bugs), the time needed to fix them almost doubles [3]. In extreme cases, this can even lead to hundred-million dollar projects being abandoned after several years of intense development [5].

Recently, there have been attempts to integrate performance testing to CI/CD [6,2]. However, these solutions usually provide only basic performance testing support (e.g., regression testing, micro benchmarking, or load testing) and focus only on the latest software version. In our previous work [4], we have introduced *Perun*: a novel tool-suite for performance testing and analysis. Perun supports more complex performance analyses that require precise profiling and can possibly take into account multiple previous versions of the software. Such larger perspective is generally necessary to find root causes of hard-to-detect performance issues, especially in the case of so-called *creeping degradations*: performance degradations of software caused slowly over the time.

In order to integrate more complex performance analyses, e.g., as a part of a CI/CD pipeline, we need them to be efficient. One possible approach is to analyze only source files that have been changed since the last analysis. This is successfully used in practice, e.g., by Facebook/Meta Infer [1] in the area of static analysis. In our ongoing work on efficient performance analyses, we build on this principle and focus our analysis to, among other, code fragments with control flow changes. However, this poses numerous challenges in terms of efficient representation, versioning, storage, and difference analysis of control flow models in evolving software that we briefly discuss in Section 2.

In this work, we propose a solution to the above challenges which we have adopted in the particular context of Perun. However, we believe that the data structures and algorithms that our solution is based on are general enough to be useful in other tools that leverage control flow information, allowing them to be used efficiently in CI/CD.

2 Control Flow Models Representation and Analysis

Obtaining precise call and control flow graphs is notoriously difficult, mainly due to *dynamic dispatch calls* found in many modern programming languages. Control flow models are generally unsound and incomplete when reconstructed using static analysis; and sound but still incomplete when obtained by observing dynamic runs. Combining both approaches leads to unsound and incomplete, yet generally more precise control flow models. Hence, we propose an efficient *layered* representation of control flow that composes static and dynamic models while allowing the analyses (e.g. difference analyses, reachability, etc.) to be run on any individual layer or on their various combinations.

Both CI/CD tools and Perun operate on code repositories managed by *Version Control Systems (VCS)*, e.g., Git. However, associating reconstructed control flow models solely with VCS versions (e.g., commit hashes in Git) is in practice not enough for tools that support both standalone and CI/CD workflow, such as Perun. For this reason, we propose to use a hierarchical version labeling system (using, among others, hashes of VCS versions and local repository changes) along with accompanying algorithms for storage, retrieval, and intelligent lazy deletion of individual versions.

Although numerous approaches for difference analysis of call and control flow graphs exist, they are mostly tailored for security-oriented analyses. In our experience, these approaches are not well suited for large evolving codebases (e.g., CPython[1]) as they produce a lot of false (dis)similarity matches. Our solution builds on existing works in this area [8,7] and further adapts them specifically for evolving software to improve the analysis accuracy. Our experiments with the CPython codebase show that our solution does indeed provide better results than the other approaches we are aware of.

Acknowledgments. The work was supported by the project GA23-06506S of the Czech Science Foundation, FIT-S-23-8151 of the BUT FIT, Brno Ph.D. Talent Programme, and Red Hat Inc.

Disclosure of Interests. Jiří Pavela and Tomáš Fiedor are, or have been in the past, funded by Red Hat, Inc. research scholarships.

References

1. Infer: Differential Workflow, https://fbinfer.com/docs/steps-for-ci/
2. Jenkins CI performance plugin, https://jenkinsci.github.io/performance-plugin/RunTests.html
3. Chen, T.H., Nagappan, M., Shihab, E., Hassan, A.E.: An empirical study of dormant bugs. In: Proc. of MSR'14. p. 82–91
4. Fiedor, T., Pavela, J., Rogalewicz, A., Vojnar, T.: Perun: Performance version system. In: Proc. of ICSME'22. pp. 499–503
5. Fukami, C., Mccubbrey, D.: Colorado benefits management system (c): Seven years of failure. Communications of the Association for Information Systems **29**, 97–102 (2011)
6. Javed, O., Dawes, J.H., Han, M., Franzoni, G., Pfeiffer, A., Reger, G., Binder, W.: Perfci: A toolchain for automated performance testing during continuous integration of python projects. In: Proc. of ASE'20. p. 1344–1348
7. Lee, Y.R., Kang, B., Im, E.G.: Function matching-based binary-level software similarity calculation. In: Proc. of RACS'13. p. 322–327
8. Nagarajan, V., Gupta, R., Zhang, X., Madou, M., de Sutter, B., de Bosschere, K.: Matching control flow of program versions. In: Proc. of ICSM'07. pp. 84–93
9. Zaman, S., Adams, B., Hassan, A.E.: Security versus performance bugs: A case study on firefox. In: Proc. of MSR'11. p. 93–102

[1] https://github.com/python/cpython

Textile Sensor Surrogate Modelling using Sparse Identification

Martin Steiger[1 [0009−0005−2227−2124]], Phillip Petz[1 [0000−0002−7943−7226]], and Stephan Schuler[1]

University of Applied Sciences Upper Austria, Hagenberg, Austria
{martin.steiger,phillip.petz,stephan.schuler}@fh-hagenberg.at

Abstract. In this paper we present a new methodology to model hysteretic dynamical systems based on sparse identification of non-linear dynamics (SINDy) by Brunton et al. For this purpose we adapt the SINDy method to approximate coefficients of a discrete Preisach hysteresis and furthermore introduce an algorithm to increase the hysteresis accuracy in regions of interest based on the De Boor algorithm for optimal spline knot distribution. This method is applied to model the force response of a textile-based shear force sensor as introduced by Schuler et al. since hysteretic effects have been observed due to the material properties.

Keywords: Veritifation and Simulation · Surrogate Modelling · Sparse Identification

Schuler et al.[4] conducted a preliminary study to investigate the fabrication and calibration of a textile-based shear force sensor. This research provided a heuristic model to approximate sensor performance metrics, whereby the empirical validation revealed a major accuracy issue. Main cause are the hysteretic properties of the sensor material also investigated in [5]. Compensating this error is especially difficult when integrating the sensor in resource-limited embedded systems. Given the difficulties using surrogate models as a predictive tool is advantageous. We propose a hysteretic approach based on the well-known dircrete Preisach models [3] due to its trivial implementation in hardware. Furthermore a sparse identification approach similar to SINDy by Brunton et al. [2] is used to determine the model parameters. SINDy has sparked attention for being able to discover governing system equations from possibly noisy measurement data using sparsity-promoting regression algorithms. Let be given time series measurement data of two signals $u(t) \in \mathbb{R}^m, y(t) \in \mathbb{R}^n$ as $U = (u(t_\ell))_{\ell=0}^{L-1}$ and $Y = (y(t_\ell))_{\ell=0}^{L-1}$ at L collocation points t_ℓ. Furthermore let be given an ansatz function library $\mathcal{A}$ with K elements $a_k : \mathbb{R}^m \to \mathbb{R}$. We propose that any memoryless $f : \mathbb{R}^m \to \mathbb{R}^n$ can be approximated with a weighted sum of appropriate ansatz functions in the library $\mathcal{A}$ as $f(u) \approx \hat{f}(u) = \mathcal{A}(u)\Xi,\ \Xi \in \mathbb{R}^{K\times n}$. Therefore we are presented with a regression problem to determine the weight coefficients Ξ of the proposed model

$$Y = \mathcal{A}(U)\Xi,\ \hat{\Xi} = \underset{\Xi}{\operatorname{argmin}}(||Y - \mathcal{A}(U)\Xi||_2 - \lambda||\Xi||_p)$$

including a regularization term according to Brunton et al. [2] to promote sparsity. We utilize this method to model the force based on the sensor deformation by selecting so-called Preisach relays as ansatz functions, the weighted superposition of which yields a Preisach operator $\mathcal{P}$ [3]. Let $u(t) \in \mathbb{R}$ and $y(t) \in \mathbb{R}$ be scalar. The discrete formulation of $\mathcal{P}_d$ is advantageous in embedded systems environments

$$\mathcal{P}_d[u, \rho_0](t) = \sum_{k=1}^{K} \xi_k \mathcal{R}_{\alpha_k, \beta_k}[u, \rho_0] \tag{1}$$

$\mathcal{R}_{\alpha_k, \beta_k}[u, \rho_0]$ resembles a relay operator with initial state ρ_0 and (α_k, β_k) are unique combinations of its threshold parameters. We approximate

$$Y \approx \mathcal{P}_d[U, \rho_0] = \sum_{k=1}^{K} w_k \mathcal{R}_{\alpha_k, \beta_k}[U, \rho_0] = \mathcal{R}_{\alpha, \beta}[U, \rho_0]\Xi \tag{2}$$

whereby Ξ is a $\mathbb{R}^K$ weight column vector and $\mathcal{R}_{\alpha,\beta}$ a $\mathbb{R}^{L \times K}$ matrix of evaluated relay elements with different threshold parameter pairs. A problem of similar form has been introduced by Brunton [2]. Consequently we apply the SINDy method

$$\hat{\Xi} = \underset{\xi}{\operatorname{argmin}} \left(||Y - \mathcal{R}_{\alpha,\beta}[U, \rho_0]\Xi||_2 + \lambda ||\Xi||_p \right) \tag{3}$$

whereby $\lambda||\Xi||_p$ is a regularization term to encourage sparsity. We can improve the accuracy of the Preisach model by selecting appropriate threshold parameters (α_k, β_k) with a tighter mesh of thresholds in regions of interest. An efficient way to distribute the thresholds is a modified version of the De Boor algorithm for optimal spline knot distribution as introduced in [1]. The optimization results based on the textile sensor measurement data are matching the expected system behaviour and can be successfully used to predict system outputs.

References

1. de Boor, C.: A Practical Guide to Spline, vol. Volume 27 (01 1978). https://doi.org/10.2307/2006241
2. Brunton, S.L., Proctor, J.L., Kutz, J.N.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the National Academy of Sciences **113**(15), 3932–3937 (2016). https://doi.org/10.1073/pnas.1517384113, https://www.pnas.org/doi/abs/10.1073/pnas.1517384113
3. Preisach, F.: Über die magnetische nachwirkung. Zeitschrift für Physik **94**, 277–302 (1935)
4. Schuler, S., Petz, P., Eibensteiner, F., Langer, J.: A textile based capacitive pressure and shear force sensor. In: 2023 IEEE International Conference on Flexible and Printable Sensors and Systems (FLEPS). pp. 1–4. IEEE (2023)
5. Wohlrab, S., Petz, P., Eibensteiner, F., Langer, J.: Influences of coating and spandex compositions of conductive textiles used as strain sensors using an automated test system. In: Science and Information Conference. pp. 306–321. Springer (2022)

Design and Control of Robot Gripper Hand Using Hierarchical Systems Technology

Kanstantsin Miatliuk[1], Tomasz Kordziukiewicz[1], Rabab Benotsmane[2]

[1]Mechanical Engineering Faculty, Bialystok University of Technology (BUT), Poland
[2]Inst. of Automation and Info-communication, University of Miskolc (UM), Hungary
k.miatliuk@pb.edu.pl, tomekkordziukiewicz@gmail.com,
rabab.benotsmane@uni-miskolc.hu

The design and control method of robot gripper hand using Hierarchical Systems (HS) technology [1,2] is presented in the paper. According to the phases of mechatronic objects' life cycle [2,3], conceptual design (CD) of the gripper was performed first using HS technology and detailed design (DD) of the gripper was realized after that. At the CD phase the conceptual model of the gripper was created.

To define conceptual model of a robot gripper-hand it was necessary to describe in the common HS formal basis gripper dynamics and its structure, gripper environment, coordinator and its coordination (design & control) processes. Furthermore, conceptual model takes into account the connected formal descriptions of a gripper mechatronic subsystems, i.e. mechanical (gripper fingers), electromechanical (gripper finger drives), electronic and computer (gripper design, control and sensors subsystems). HS technology [1,2] was chosen in the work as the formal basis to create the model.

According to HS formal model [1,2] robot gripper being designed ${}_{o}S^{l}$, its subsystems of *l-1* level, environment system ${}_{\varepsilon}S^{l}$ and their processes ${}_{\pi}S^{l}$, are described in aggregated dynamic form ω^{l} by dynamic systems $(\rho,\varphi)^{l}$ [2,5]. Robot gripper structure includes dynamic realizations $(\rho,\varphi)^{l-1}$ of subsystems and their connections γ_{σ}. Structural σ^{l} and dynamic ω^{l} presentations are connected by coordinator S_0 which realizes interlevel relations by performing design (*synthesis* and *analysis*) and control tasks on its selection, learning and self-organization strata at both CD and DD phases.

The result of the *synthesis task* performing in SolidWorks environment for the case of the robot gripper-hand design is presented in Fig.1a [6]. Here the gripper structure σ^{l} is synthesized from its elements ω^{l-1} by realizing structural connections γ_{σ}. This task is described at CD phase as the following coordinator S_0 task:

$$S_0:\{\omega^{l-1}, \gamma_{\sigma}\} \rightarrow \sigma^{l} \qquad (1)$$

Geometric parameters of the robot gripper are described by constructive dimension and connections defect, introduced by numerical positional system [2,4,7].

At the DD phase the *synthesis task* is performed by coordinator (human-computer system) by selecting particular elements of the robot gripper-hand, i.e. fingers, gripper drives (servos), controllers, control glove, sensors, etc. Robot gripper hand designed in BUT, Poland is given in Fig.1b. Similar task is being performed in UM, Hungary.

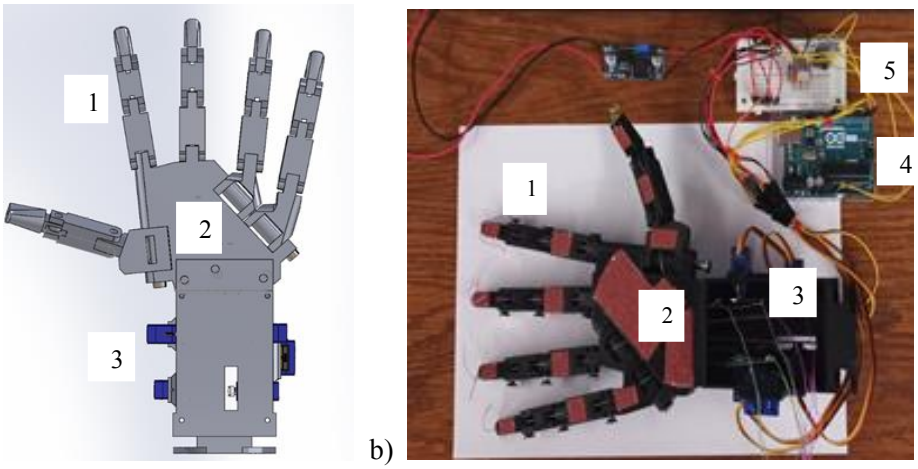


Fig. 1. Gripper hand design: a) SolidWorks system screenshot, b) a gripper constructed in BUT; gripper structure σ^{l} synthesized from its elements ω^{l-1}: 1) gripper fingers, 2) hand, 3) wrist with drives connectors, 4) Arduino UNO Rev 3 controller, 5) a prototype board to which the power cables from the converter are connected**,** etc.

The *analysis tasks* are kinematics and dynamics tasks. At the CD phase kinematics and dynamics models of the robot gripper are presented in *(ρ,φ)* form [2,5]. At the DD phase the models are transformed to concrete equations and program codes [6].

Therefore, the main contribution of the paper is the application of HS technology in design and control of the robot gripper-hand and mechatronic design process formalization. HS technology allows description of mechatronic subsystems of different nature in common formal basis. Including of dynamic systems (*ρ,φ*) [5] as elements of conceptual model allows easy transfer to DD phase and model concretization. Numeric and geometric characteristics introduced [2,4,7] allows coordinated calculations of the robot gripper parameters at both CD and DD phases.

References

1. Miatliuk, K., Kim, Y.H., Kim, K., Siemieniako, F.: Use of hierarchical system technology in mechatronic design. Mechatronics 20, 2, 335–339 (2010).
2. Miatliuk, K.: Conceptual design of mechatronic systems. BUT, Bialystok (2017).
3. Ullman, D.G.: The mechanical design process. 4th edn. McGraw-Hill; Singapore (2010).
4. Miatliuk, K.: Geometrical representation based on the hierarchical systems technology. In: 14th International Conference EUROCAST 2013, pp. 59-60. Las Palmas (2013).
5. Mesarovič, M., Takahara, Y.: Abstract systems theory. Springer Verlag (1990).
6. Kordziukiewicz, T.: Design and control of a gripper-hand of Universal Robot, Master Thesis, BUT, Bialystok (2023).
7. Miatliuk, K., Wolniakowski, A., Kolosowski, P.: Engineering system of systems conceptual design in theoretical basis of hierarchical systems, IEEE Int. Conf. on Systems, Man and Cybernetics, Oct.11-14, Toronto, Canada, pp. 2850-2856, (2020).

Discrimination Criteria for Modeling Association, Aggregation, and Composition in UML Class Diagrams

Miguel Alemán-Flores

Centro de Tecnologías de la Imagen (CTIM), Instituto Universitario de Cibernética, Empresa y Sociedad (IUCES), Universidad de Las Palmas de Gran Canaria
Campus de Tafira, 35017, Las Palmas, Spain
miguel.aleman@ulpgc.es

Extended Abstract

The Unified Modeling Language (UML) provides us with class diagrams as a powerful tool to model the structure of a system by means of classes. Associations between classes play a crucial role in software modeling and have a major influence on the subsequent stages of the development. Three main types of association are considered in order to distinguish several scenarios: simple associations, shared aggregations (also referred to simply as aggregations), and composite aggregations (also known as compositions) [6]. Several interpretations of what these three types of relationship mean and how they reflect usual situations in business or software modeling have been proposed, so that misunderstandings can easily arise. The main issue when modeling associations consists in distinguishing shared aggregations from plain associations, and determining the constraints involved in each type of aggregation [2, 5, 3]. Although some general recommendations or guidelines regarding the use of different UML associations have been proposed in previous works [7, 1, 4], a clear guide or reference would be extremely useful.

This work intends to determine the key factors which help us decide what type of association best describes each relationship in the domain we try to model. By means of a set of criteria, the main differences and similarities between these three options are extracted. These criteria consider the semantics as well as the implications in the structure of the model. Moreover, the constraints which appear when using certain additional elements, such as association classes and self-associations, are also taken into account. The set of criteria which is proposed includes the life cycle of the associated instances, the multiplicities of the relationship, the propagation of the deletion, the linking capability and visibility from one end to the other, the transitivity and identity projection which are implicit, as well as the possible object-level reflexivity (see Table 1).

The main contribution of this work consists in presenting certain criteria which combine semantics and structure in order to identify the best type of association for each case. Furthermore, those criteria can be applied in a wide variety of domains and scenarios and allow finding analogies between them.

Table 1. Criteria in associations, shared aggregations, and compositions (W stands for whole-end, P for part-end). Life cycle: I (independent), C (controller), c (controlled). Multiplicity: ∀ (any value is possible), 1 (mandatorily 1), 0..1 (optionally 1). Delete propagation: L (link), C (cascade). Linking capability: E (external), A (active), P (passive). Visibility: Y (visible), N (not visible). Transitivity: O (optional), M (mandatory). Identity projection: Y (projected), N (not projected). Object-level reflexivity: Y (possible), N (not possible).

	Association	Shared Aggregation W-P	Composition W-P
Life cycle	I	I - I	C - c
Multiplicity	∀	∀ - ∀	1 | 0..1 - ∀
Delete propagation	L	L - L	C - L
Linking capability	E	A - P	A - P
Visibility	Y	Y - Y	Y - N
Transitivity	O	M - M	M -M
Identity projection	N	N - N	Y - N
Object-level reflexivity	Y	N - N	N - N

References

1. Albert, M., Pelechano, V., Fons, J., Ruiz, M., Pastor, O.: Implementing UML association, aggregation, and composition. A particular interpretation based on a multidimensional framework. In: Eder, J., Missikoff, M. (eds.) Advanced Information Systems Engineering. pp. 143–158. Springer Berlin Heidelberg, Berlin, Heidelberg (2003)
2. Fowler, M.: https://martinfowler.com/bliki/AggregationAndComposition.html (2003), last accessed November 2023
3. Gómez-Fuentes, M.C., Cervantes-Ojeda, J., García-Nájera, A.: Association and aggregation class relationships: is there a difference in terms of implementation? In: 2021 9th International Conference in Software Engineering Research and Innovation (CONISOFT). pp. 44–53 (2021). https://doi.org/10.1109/CONISOFT52520.2021.00018
4. Maraee, A., Balaban, M.: Inter-association constraints in UML2: Comparative analysis, usage recommendations, and modeling guidelines. In: France, R.B., Kazmeier, J., Breu, R., Atkinson, C. (eds.) Model Driven Engineering Languages and Systems. pp. 302–318. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
5. Martin, R.C.: UML for Java Programmers. Prentice Hall, Upper Saddle River, NJ (2002)
6. Object Management Group (OMG): Meta-Object Facility (MOF) Specification, Version 2.5.1. https://www.omg.org/spec/UML/About-UML/ (2017), last accessed November 2023
7. Parsons, D.: Objects Working Together: Association, Aggregation and Composition, pp. 141–175. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-54518-5_7

Efficient Hardware Architecture for Random Forest Training

Jakob Winkler[0000−0002−5927−8563], David Hackl, and Michael Lunglmayr[0000−0002−4014−9681]

Institute of Signal Processing, Johannes Kepler University, 4040 Linz, Austria
{jakob.winkler,michael.lunglmayr}@jku.at

Efficient machine learning implementations for resource-constrained edge devices are an active research topic. For these edge devices, pure software approaches are often infeasible for real-time use, and hardware based accelerators have to be employed.

Random Forests, which are ensembles of decision trees, can be implemented using particularly few hardware resources while still achieving competitive prediction results in certain applications when compared to other approaches such as neural networks [2–4, 9].

Recent work covers both implementation for low-end devices [7] and solutions [6, 8] for inference on higher-end devices, showing the promise of these approaches for static problems. For dynamic problems, online methods for learning from new data must be considered, allowing to (re-)train models on edge devices.

In this work, we propose an efficient hardware implementation for training a Random Forest fully in digital hardware, serving as a fast, embedded method implementable in low-end FPGAs, and provide insights about performance and resource metrics depending on different levels of quantization.

Hardware Architecture

As previously discussed in literature, sorting the dataset is often the most time-intensive step of building decision trees [1, 5]. Following an approach similar to the one used in [5] while considering the resource constraints of the target hardware, the proposed hardware architecture utilizes a pre-sorted dataset and uses index mapping tables to avoid sorting during the learning process. This way, the data can be processed in two passes over the dataset per tree level, resulting in linear runtime complexity over tree levels (whereas the number of nodes in the tree grows exponentially).

Table 1 shows preliminary synthesis results for a low-end Cyclone V FPGA device over multiple bit width and dataset size configurations. These results show that the proposed approach enables online learning on resource constrained edge devices due to their low resource usage ($\leq 14\%$ DSP blocks, $< 35\%$ on-chip memory, $\leq 7\%$ of other resources).

Acknowledgements. This work is supported by the COMET-K2 "Center for Symbiotic Mechatronics" of the Linz Center of Mechatronics (LCM), funded by the Austrian federal government and the federal state of Upper Austria.

Bit width	16	16	32	32
m	500	5000	500	5000
Total registers	2898	3184	4084	4384
Logic utilization (ALMs)	1731(4%)	1730(4%)	2749(7%)	2916(7%)
Total block memory bits	107284(2%)	627022(11%)	175124(3%)	977486(17%)
Total RAM blocks	26(5%)	122(22%)	37(7%)	191(35%)
Total DSP blocks	14(13%)	22 (20%)	46(41%)	46(41%)
Fmax(Slow 1100mV 85C Model)	115.79 MHz	92.75 MHz	77.83 MHz	77.74 MHz

Table 1. Preliminary synthesis results for four implementation sizes, targeting a Cyclone V 5CSXFC6D6F31C6 FPGA device.

References

1. Amado, N., Gama, J., Silva, F.: Exploiting parallelism in decision tree induction. In: Proceedings from the ECML/PKDD Workshop on Parallel and Distributed computing for Machine Learning. pp. 13–22 (2003)
2. Baumgartner, S., Huemer, M., Lunglmayr, M.: Efficient Majority Voting in Digital Hardware. IEEE Transactions on Circuits and Systems II: Express Briefs **69**(4), 2266–2270 (2022). https://doi.org/10.1109/TCSII.2022.3144047
3. Breiman, L.: Random forests. Machine Learning **45**(1), 5–32 (2001). https://doi.org/10.1023/A:1010933404324
4. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and Regression Trees. CRC Press (1984)
5. Mehta, M., Agrawal, R., Rissanen, J.: Sliq: A fast scalable classifier for data mining. In: Advances in Database Technology—EDBT'96: 5th International Conference on Extending Database Technology Avignon, France, March 25–29, 1996 Proceedings 5. pp. 18–32. Springer (1996)
6. Saqib, F., Dutta, A., Plusquellic, J., Ortiz, P., Pattichis, M.S.: Pipelined Decision Tree Classification Accelerator Implementation in FPGA (DT-CAIF). IEEE Transactions on Computers **64**(1), 280–285 (2015). https://doi.org/10.1109/TC.2013.204
7. Struharik, J.: Implementing decision trees in hardware. In: 2011 IEEE 9th International Symposium on Intelligent Systems and Informatics. pp. 41–46 (2011). https://doi.org/10.1109/SISY.2011.6034358, ISSN: 1949-0488
8. Winkler, J., Lunglmayr, M.: FPGA Processing of Decision Tree Ensembles Stored in External DRAM. In: 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME). pp. 1–6 (2023). https://doi.org/10.1109/ICECCME57830.2023.10252353
9. Zhou, Z.H.: Ensemble Methods: Foundations and Algorithms. Chapman & Hall/CRC, 1st edn. (2012)

Influence of Spike Encoding, Neuron Models and Quantization on SNN Performance

Daniel Windhager[1][0009−0007−6443−6515], Bernhard Moser[2,3][0000−0001−8373−7523], and Michael Lunglmayr[2][0000−0002−4014−9681]

[1] Intelligent Wireless Systems, Silicon Austria Labs, Linz, Austria
Daniel.Windhager@silicon-austria.com
[2] Institute of Signal Processing, Johannes Kepler University, Linz, Austria
{bernhard.moser,michael.lunglmayr}@jku.at
[3] Software Competence Center Hagenberg GmbH, Hagenberg, Austria

Spiking neural networks (SNNs) are interesting alternatives to classical artificial neural networks (ANNs) for efficient implementation of machine learning tasks on Edge devices. As demonstrated in [1, 2], SNNs are capable of high-performance inference with significantly lower power requirements than ANNs. In contrast to ANNs, which process real vectors, SNNs operate on (sparse) spike sequences η_j

$$\eta_j = \sum_i \alpha_i \delta[k - s_i] \tag{1}$$

with $\delta[k]$ as the discrete delta pulse, $\alpha_i \in \mathbb{Z}$, and the position of a spike $s_i \in \mathbb{N}_0$. An SNN processes such spike sequences as schematically shown in Fig. 1.

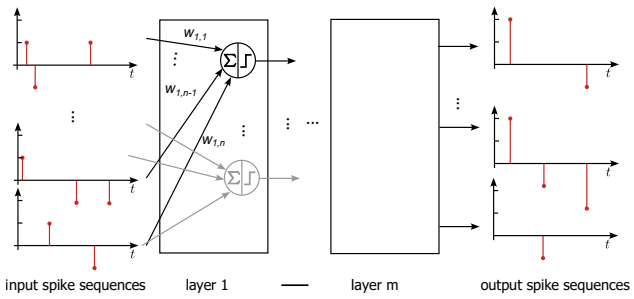


Fig. 1. Processing of spike sequences in an SNN

To encode data in a spike sequence, several methods have been proposed in the literature. Among the most famous are rate encoding, latency encoding, and delta encoding. When processing images, often rate encoding is used. It allows translating pixel values to spike sequences where the rate of spikes per time interval is proportional to the pixel value. When using such an encoding, the number of triggered spikes can become very high which in turn results in a high power consumption when processing an SNN. An interesting alternative for spike encoding is using event-based sampling such as send-on-delta (SOD)

[3]. This delta-encoding allows representing data via sparse graded (i.e. with spike amplitudes not limited to ± 1) spike trains and, as a consequence, allows reducing the energy requirements for processing. In this work, we will investigate the performance of SNNs of different sizes, using different neuron models and different spike encodings on the example of digit recognition. We will put these performance results into perspective by evaluating the number of triggered spikes within the network as a measure of the required energy for processing. We will furthermore investigate how quantization affects the inference performance of SNNs. Tab. 1 shows preliminary results on the inference performance for the MNIST digit recognition dataset for different network structures and different quantization bit widths. These results not only demonstrate the feasibility of using SNNs of moderate sizes for this task, but also the low sensibility of SNNs to quantization effects.

Table 1. Accuracy of inference for MNIST dataset when using different SNN network structures and quantization bit widths.

	Number of neurons per layer	
Bit width	$748-100-10$	$748-200-10$
4	91.18%	93.14%
5	94.96%	95.51%
6	95.31%	96.21%
8	95.49%	96.46%
10	95.61%	96.41%

Acknowledgements The research reported in this paper has been partly funded by BMK, BMDW, and the State of Upper Austria in the frame of SCCH, part of the COMET Programme managed by FFG, as well as by the European Unions Horizon 2020 research and innovation program within the framework of Chips Joint Undertaking (Grant No. 101112268). This work has been supported by Silicon Austria Labs (SAL), owned by the Republic of Austria, the Styrian Business Promotion Agency (SFG), the federal state of Carinthia, the Upper Austrian Research (UAR), and the Austrian Association for the Electric and Electronics Industry (FEEI).

References

1. Davies, M., Wild, A., Orchard, G., Sandamirskaya, Y., Guerra, G.A.F., Joshi, P., Plank, P., Risbud, S.R.: Advancing neuromorphic computing with Loihi: A survey of results and outlook. Proceedings of the IEEE **109**(5), 911–934 (2021). https://doi.org/10.1109/JPROC.2021.3067593
2. Gaurav, R., Stewart, T.C., Yi, Y.: Reservoir based spiking models for univariate time series classification. Frontiers in Computational Neuroscience **17** (2023). https://doi.org/10.3389/fncom.2023.1148284
3. Moser, B.A., Lunglmayr, M.: On quasi-isometry of threshold-based sampling. IEEE Transactions on Signal Processing **67**(14), 3832–3841 (2019). https://doi.org/10.1109/TSP.2019.2919415

Adaptive combination in Frequency Domain: An Approach for Robust Nonlinear Acoustic Echo Cancellation

Alireza Nezamdoust[†][0000−0003−4999−263X], Mario Huemer[*][0000−0002−3164−7232], and Danilo Comminiello[†][0000−0003−4067−4504]

[†] Department of Information Engineering, Electronics and Telecommunications, Sapienza University of Rome, Italy
{Alireza.nezamdoust, Danilo.Comminiello}@uniroma1.it
[*]Institute of Signal Processing, Johannes Kepler University Linz, Austria

Abstract. This paper introduces a novel approach to address the challenges of Nonlinear Acoustic Echo Cancellation (NAEC) by proposing an integration of Frequency Domain Adaptive Filters (FDAF) into a Collaborative Functional Link Adaptive Filters (CFLAF) framework. The research focuses on enhancing the robustness of the adaptive filtering system in the presence of changing linear-to-nonlinear power ratios (LNLR) over time.

Keywords: Nonlinear Adaptive Filters · Frequency-Domain Adaptive Filters · Collaborative Functional Link.

1 Collaborative Filtering for Nonlinear Optimization

Nonlinear adaptive filters have found extensive use across various application domains owing to their versatile capabilities [5, 1]. Many Linear-in-the-Parameters (LIP) nonlinear filters are adapted through time-domain adaptive filters. However, in some real-time applications with extremely long filters this approach can pose a challenge [6]. The drawback lies in the substantial computational resources it demands, typically scaling with the length of the filters [9, 8].

Utilization of Frequency Domain Adaptive Filters (FDAFs), has proven to be a potent tool in signal processing across diverse applications. In contrast to time-domain algorithms, FDAFs exhibit commendable convergence performance even in real-time applications [2, 7]. This study emphasizes the utilization of the Functional Expansion Block (FEB) within FDAF filters, highlighting various expansion methods that allow for an exchange of information between linear and nonlinear filters. Collaborative filtering is introduced as a promising approach to enhance the efficiency and effectiveness of functional links, allowing for a collaborative and adaptive optimization process [3, 4]. The FDAF architecture incorporates a collaborative structure based on the Functional Link Adaptive Filter (FLAF), allowing for the automatic activation or deactivation of the nonlinear filtering path. This dynamic adjustment enables adaptive biasing of the

acoustic channel modeling based on the nonlinearity residual in the echo path. The inclusion of a shrinkage parameter in FDAF facilitates selective retention or removal of the output from the nonlinear filter, providing adaptability to different filtering scenarios. The proposed model demonstrates robustness to varying nonlinearity levels, converging towards a purely linear filter when the echo path is linear and adapting to nonlinearities when present.

- We discuss an innovative framework for nonlinear adaptive filters with a collaborative functional link structure tailored to cost-effective and efficient learning in online applications.
- To achieve a balance between high performance and reduced computational complexity, we introduce a novel approach, defining a partitioned-block frequency-domain Collaborative Functional Link Adaptive Filter (CFLAF). This innovative perspective shifts the adaptation process to the frequency domain, aiming for enhanced efficiency with limited resources.
- The proposed algorithms are systematically assessed in a standard dynamic and in a static scenario.

Acknowledgments This work is supported by the COMET-K2 "Center for Symbiotic Mechatronics" of the Linz Center of Mechatronics (LCM), funded by the Austrian federal government and the federal state of Upper Austria.

References

1. Carini, A., Sicuranza, G.L., Mathews, V.J.: Efficient adaptive identification of linear-in-the-parameters nonlinear filters using periodic input sequences. Signal processing **93**(5), 1210–1220 (2013)
2. Clark, G., Parker, S., Mitra, S.: A unified approach to time-and frequency-domain realization of fir adaptive digital filters. IEEE transactions on acoustics, speech, and signal processing **31**(5), 1073–1083 (1983)
3. Comminiello, D., Nezamdoust, A., Scardapane, S., Scarpiniti, M., Hussain, A., Uncini, A.: A new class of efficient adaptive filters for online nonlinear modeling. IEEE Transactions on Systems, Man, and Cybernetics: Systems **53**(3), 1384–1396 (2022)
4. Comminiello, D., Scarpiniti, M., Azpicueta-Ruiz, L.A., Arenas-Garcia, J., Uncini, A.: Functional link adaptive filters for nonlinear acoustic echo cancellation. IEEE Transactions on Audio, Speech, and Language Processing **21**(7), 1502–1512 (2013)
5. Pan, Y., Yang, G.H.: Event-triggered fault detection filter design for nonlinear networked systems. IEEE Transactions on Systems, Man, and Cybernetics: Systems **48**(11), 1851–1862 (2017)
6. Patel, V., Gandhi, V., Heda, S., George, N.V.: Design of adaptive exponential functional link network-based nonlinear filters. IEEE Transactions on Circuits and Systems I: Regular Papers **63**(9), 1434–1442 (2016)
7. Soo, J.S., Pang, K.K.: Multidelay block frequency domain adaptive filter. IEEE Transactions on Acoustics, Speech, and Signal Processing **38**(2), 373–376 (1990)
8. Uncini, A.: Digital Audio Processing Fundamentals, vol. 21. Springer Nature (2023)
9. Zhao, H., Zhang, J.: Functional link neural network cascaded with chebyshev orthogonal polynomial for nonlinear channel equalization. Signal processing **88**(8), 1946–1957 (2008)

Using of a Robotic Platform to Detect Acoustic Events for Indoor Environments

Stefan Grama[1], Lacrimioara Grama[2[0000−0003−0383−5196]], and Corneliu Rusu[2[0000−0001−5209−2734]]

[1] "Gheorghe Sincai" National College, Cluj-Napoca, Romania
[2] Signal Processing Group, Technical University of Cluj-Napoca, Romania
{Lacrimioara.Grama, Corneliu.Rusu}@bel.utcluj.ro

Abstract. The main topic of this article is to present an audio indoor environment monitoring system for a robotic platform through a graphical user interface. The acoustic event is detected by the use jointly of six distinct models. The final decision is based on the predominant result. The detected event is displayed in the graphical user interface, converted to speech, and based on some pre-learned rules, the robotic platform can choose a required task.

Keywords: robotic platform · indoor monitoring · email alert.

1 The OMNI-Z robotic platform

Communities with modern healthcare systems are rather specific to advanced economies and are often found in societies that tend to age. The percentage of the population with special medical care needs is significant and is anticipated to increase. The involvement of people in these acts is a large financial burden, so a great effort is made to reduce the involvement of the human factor. A versatile and economically viable robotic platform for indoor navigation in cluttered environments with obstacles has been developed at CITST Bucharest to improve the performance of an in-house modified platform, to be suitable for AAL and compatible with e-Health, domotics, and well-being platforms. Robotic platforms that share an everyday indoor environment with people must be able to detect people, track them, and recognize as much of their activities as possible.

1.1 Acoustic Event Monitoring – Implementation and Results

The audio database used for this event detector is the joint TIAGo database version 01 [3] and 02 [2] which consists of 3300 audio signals (meaning 110 different sound classes). In [4], for the TIAGo database version 01, the highest average accuracy of 98.55% was achieved using 34 Mel Frequency Cepstral Coefficients (MFCC-34) and k-Nearest Neighbors (kNN). In the case of version 02, in [1], the highest average accuracy was 99.78% using MFCC-64 and Linear Discriminant Analysis (LDA), while for Support Vector Machines (SVM) was 99.36% for MFCC-34 and 99.74% for MFCC-64.

We have chosen the Python programming language since is not platform-dependent. Based on our previous results, we have considered that MFCC-34 and MFCC-64 should be the most suited features combined with SVM, kNN, and LDA, in the case of the merged databases. Using these combinations we have obtained six distinct models. The audio signals within each class were randomly divided into 80% training and 20% testing, respectively. The overall accuracies (train phase & test phase), computed as the arithmetic mean of five distinct runs, are: MFCC-34 + kNN: 100% & 99.091%; MFCC-64 + kNN: 100% & 98.939%; MFCC-34 + SVM: 99.886% & 99.697%; MFCC-64 + SVM: 100% & 99.545%; MFCC-34 + LDA: 97.652% & 97.576%; MFCC-64 + LDA: 98.561% & 98.333%.

In the monitoring indoor acoustic events part, an arbitrary signal can be chosen from the test signals, or one recorded in real-time by OMNI-Z can be used. Then MFCC-34 and MFCC-64 features are extracted and all six models are applied; the final answer will be given following the highest appearance of the same detected action. The detected event is converted into a vocal audio form. Finally, the proper task will be chosen by OMNI-Z, based on the situation detected: a normal indoor action (moving a chair, opening or closing a door, asking for an ID card, washing machine or microwave working, etc.) or a potentially alarming situation (coughing, asking for some specific types of medicine, etc.).

2 Conclusions

We have discussed an audio indoor environment monitoring system for OMNI-Z through a GUI. The acoustic event is detected by the use of six distinct models, in conjunction. The final decision is based on the predominant result. The detected event is displayed in the GUI, it is also converted to speech, and based on some pre-learned rules, the robotic platform can choose a required task. If the event is considered to be a regular action, the robotic platform sends a notification email to a person in charge of daily monitoring; if it is potentially alarming, it sends an alert email to the person who must intervene and to the caretaker.

References

1. Grama, L., Muscar, L., Rusu, C.: Sound Classification Algorithms for Indoor Human Activities. In: 16th International Conference on Engineering of Modern Electric Systems. pp. 1–4. Oradea, Romania (2021). https://doi.org/10.1109/EMES52337.2021.9484121
2. Grama, L., Muscar, L., Rusu, C., Telembici, T.: Audio Database for TIAGo Service Robot. Carpathian Journal of Electronic and Computer Engineering **14**(1), 1–5 (2021). https://doi.org/10.2478/cjece-2021-0001
3. Grama, L., Rusu, C.: Adding audio capabilities to TIAGo service robot. In: International Symposium on Electronics and Telecommunications. pp. 1–4. Timisoara, Romania (2018). https://doi.org/10.1109/ISETC.2018.8583897
4. Grama, L., Rusu, C.: Extending Assisted Audio Capabilities of TIAGo Service Robot. In: International Conference on Speech Technology and Human-Computer Dialogue. pp. 1–8. Timisoara, Romania (2019). https://doi.org/10.1109/SPED.2019.8906635

Determination of the object volume through processed images from a TOF – camera

F. Wirnsberger[1], P. Fleischanderl[2], T. Thurner[2], and B. Zagar[1][0000−0002−0695−2657]

[1] Montanuniversitaet Leoben, Chair of Electrical Engineering, 8700 Leoben, Austria
[2] Infineon Austria Technologies AG Development Center, 8020 Graz, Austria
franziska.wirnsberger@unileoben.ac.at
https://iet.unileoben.ac.at/

Keywords: time-of-flight camera · volume estimation · calcite · tunnel drainage.

1 Introduction

Calcit deposits in drainage pipes are a concern in the maintenance of tunnels. They arise through washouts and dissolutions of portlandite from cement-containing building material. These drainage pipes must be flushed so that they do not wear out by clogging. In order not to have to flush the entire length of the pipe, knowledge of severe scale's position is helpful. To identify the location of these deposits, ÖBB infra AG prospectively wants to use an autonomously navigating rover that detects deposits and their positions. The rover is supposed to drive through the drainage pipe with a front facing camera whose vantage point is geometrically very limited for an accurately volume scanning camera. A time-of-flight (ToF) camera mounted on the rover allows for real-time distance measurements with high spatial resolution for the target field-of-view of the moving vehicle. By using a TOF – camera, calcit deposits can not only be identified, they can also be quantified. The idea is, to implement an intelligent camera that uses computer vision approaches to, on the one hand, record the thickness of the calcite coating in a spatially resolved manner and, on the other hand, to protect rover from getting stuck in cross-sections that are too small in terms of obstacles.

2 Signal Processing Approach

In the proposed paper we elaborate on the digital signal processing necessary to estimate the calcite volume and thickness that is limiting the free cross section the rover can use to navigate through, by analyzing and combining results from both gray level and time-of-flight image sequences [1, 2] obtained at frame rates of 30 frames per second. This frame rate given the rover's average speed of about 1 m/s results in images separated geometrically by about 30 mm and resulting in a progression of the camera's vantage point within the pipe by a like amount.

Generating volume estimates requires sequences of calcite cross sections generated for secting planes oriented orthogonal to the pipe's axis, the rover's average path of motion. These secting planes are generated for each frame and constitute the cylinder base which is multiplied by the advance of the vantage point of the camera obtained by analyzing the optical flow.

Since the camera's optic is rather simple and the imaged distances are typically less than 1000 mm strong geometric distortions are deteriorating the recorded and processed images. These deteriorations require an online frame rectification before pattern matching can proceed to match corresponding points in 3D. We report on the camera calibration which is necessary prior to the image rectification process, the matching of like object points in 3D based on both gray level and TOF images, and the estimation of the camera's true motion based on deterimining the optical flow [3] of the image sequence. Figure 1 shows a single frame of – in this case – an RGB image and a derived depth image.

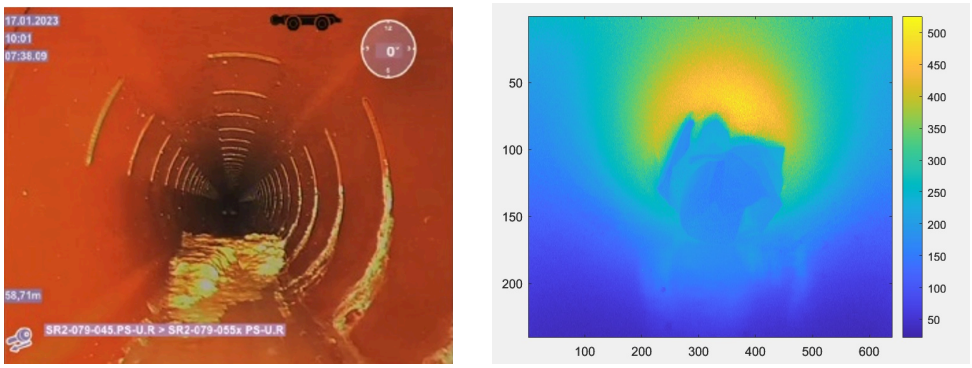


Fig. 1. Left: a single frame taken from the video stream, right: a resultant depth image, the color is indicating distance from the camera

Acknowledgments. The authors gratefully acknowledge the partial financial support for the work presented within the framework of the EU Horizon 2020 Programme under grant GA 101012456, Programme In2Track3 / Shift2-Rail.

References

1. Y. Cui, S. Schuon, D. Chan, S. Thrun, and C. Theobalt: 3D Shape Scanning with a Time-of-Flight Camera, 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 2010, pp. 1173-1180, doi: 10.1109/CVPR.2010.5540082.
2. E. Charbon: Introduction to Time-of-Flight Imaging, SENSORS, 2014 IEEE, Valencia, Spain, 2014, pp. 610-613, doi: 10.1109/ICSENS.2014.6985072.
3. Joel Gibson, Oge Marques: Optical Flow and Trajectory Estimation Methods, Springer Briefs in Computer Science, 2016.

Analysis of the accuracy of electrical resistance tomography measurements

Sabrina Affortunati[2][0000−0002−5068−3132] and Bernhard Zagar[1][0000−0002−0695−2657]

[1] Johannes Kepler University, Institut of Measurement Technology, 4040 Linz, Austria
sabrina.affortunati@jku.at
[2] Montanuniversitaet Leoben, Chair of Electrical Engineering, 8700 Leoben, Austria

Keywords: electrical resistance tomography · system noise analysis

1 Introduction

Electrical resistance tomography (ERT) is a widely used method for measuring and localizing properties attributable to changes in conductivity in a region. Electrodes attached to the boundary of the region are used for excitation and measurement. An alternating current source is used for excitation. The aim is to reconstruct the change in conductivity from the voltage measurements taken during excitation. The accuracy of the electronics is a decisive factor. The quality of the reconstructed images depends on it. A major difficulty is the fact that changes in conductivity often lead to relatively small voltage changes at the boundary [2]. The statistical and deterministic errors of a self-developed measurement electronics were determined by analyzing the measurements taken. This allowed us to determine the measurement uncertainty of the system.

2 Signal Analysis Approach

A measurement system was developed to perform electrical resistance tomography on a layer of conductive polymer for pressure measurement and localization. The measurement system was designed for an application with 16 electrodes. An alternating signal is used for excitation to reduce measurement errors. A digital direct synthesizer is used to generate the voltage signal for the voltage controlled current source. A system of demultiplexers and multiplexers connects the current source to the different electrodes and selects the measuring electrode. Two lock-in amplifiers are used for the measurement. The use of lock-in amplifiers allows the reconstruction of signals with a low signal-to-noise ratio [1].

The proposed work investigated the limits of the measurement system in terms of systematic and deterministic error. The focus is placed on the determination of errors involved in difference ERT. In difference ERT, the voltage differences between two different measurements conducted at two different moments are used for the reconstruction of the conductivity change. Deterministic

errors in the measurement, such as, for example, those caused by the voltage drop across the multiplexer on-resistance, are canceled out and are therefore not relevant. The errors investigated are therefore the standard deviation of the measured voltages (due to system noise) and the mean deviation between the actual applied voltage and the measured voltage (due to the common mode offset adjustment and the differential offset adjustment of the lock-in amplifier). These were determined from an analysis of different measured values.

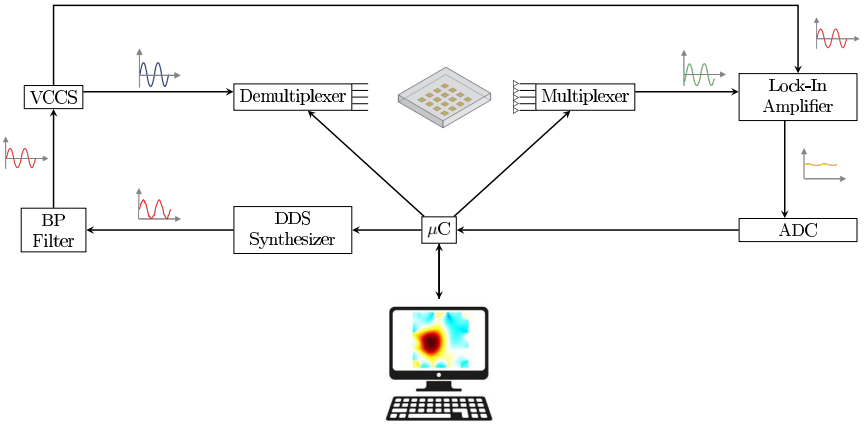


Fig. 1. Overview of the hardware.

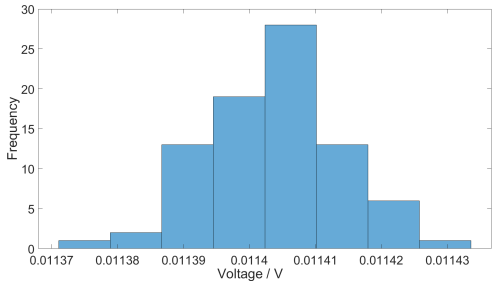


Fig. 2. Histogram of measurements taken on one electrode under the same conditions.

References

1. Horowitz, P., and Hill, W. (1989). The Art of Electronics, 2nd ed., Cambridge University Press, ISBN: 0-521-37095-7.
2. Mueller, J. L., and Siltanen, S. (2012). Linear and Nonlinear Inverse Problems with Practical Applications. (Computational Science & Engineering). Society for Industrial and Applied Mathematics, ISBN: 978-1-611972-33-7.

Modeling Wildlife Accident Risk with Gaussian Mixture Models

Christoph Praschl[0000−0002−9711−4818], David C. Schedl*[0000−0002−7621−3526], and Andreas Stöckl[0000−0003−1646−0514]

University of Applied Sciences Upper Austria, Hagenberg, Austria
*david.schedl@fh-hagenberg.at

Introduction: Traffic accidents involving wildlife pose a widespread problem globally, harming both humans and nature. Additionally, these incidents often result in heavy vehicle damage, leading to expensive repairs and insurance claims [1]. To mitigate these accidents, efforts are underway to better understand wildlife populations near high-risk roads and implement preventive measures such as visual or audible wildlife warning devices, partially utilizing artificial intelligence (AI) [2]. To prevent wildlife accidents, high-risk areas must be identified first. In this work, we propose a model that predicts dangerous areas based on animal sightings and apply it to two road segments in Austria.

Methodology: A wildlife accident risk model is created using Gaussian Mixture Models (GMM) [3]. A GMM is a probabilistic model used to model data. Data are assumed to be generated from a mixture of multiple Gaussian distributions. GMMs are particularly useful when dealing with complex data distributions that a single Gaussian distribution cannot readily describe. The model is created starting from a dataset of observations, with a two-dimensional data point representing the corresponding GPS coordinates of the animals. Depending on the animal species and the respective migration distance, these individual components are weighted differently in the model - a deer moves over larger distances in the same time on average than a hare, for example [4].

The probability density function of the GMM is given by the sum of the individual density functions of the individual Gaussian components, each weighted by its mixing coefficient. Based on such a GMM, spatial risk estimates can thus be calculated depending on the animal positions. The spatial intersection with the adjacent road thus yields risk assessments for wildlife accidents.

Data: Camera drones with RGB and thermal sensors are invaluable for data collection, offering access to remote and challenging locations. Their aerial perspective provides a comprehensive view of roads and surroundings, aiding wildlife population monitoring and risk assessment [5]. Georeferenced data enables precise mapping for preventive measures like wildlife crossings.

Airborne Light-field Sampling (ALFS) [6], based on the light fields introduced by Levoy and Hanrahan [7], utilizes light-field technology for versatile imaging. It creates integral images with adjustable focus, typically requiring specialized hardware. However, a single moving sensor, such as an aircraft camera, can also record light fields. ALFS combines colour and infrared images with position data and a digital elevation model, enabling focused imaging of the ground, and making it useful for wildlife detection in challenging environments like forests.

Results: To assess and enhance the proposed methodology, we chose two road segments in Veichter (Upper Austria) and Gänserndorf (Lower Austria) known for frequent wildlife-related accidents. These areas were meticulously chosen for testing purposes. Comprehensive aerial surveys were conducted along both sides of these road segments at various times of the day, spanning morning, noon, afternoon, and evening. The primary objective of these flights was to capture detailed geo-referenced video data, allowing for the precise identification and tagging of animals linked to accidents, such as deer, rabbits, and pheasants. Using the global positions of the animals, a risk assessment was performed utilizing a GMM, as illustrated in Fig. 1.

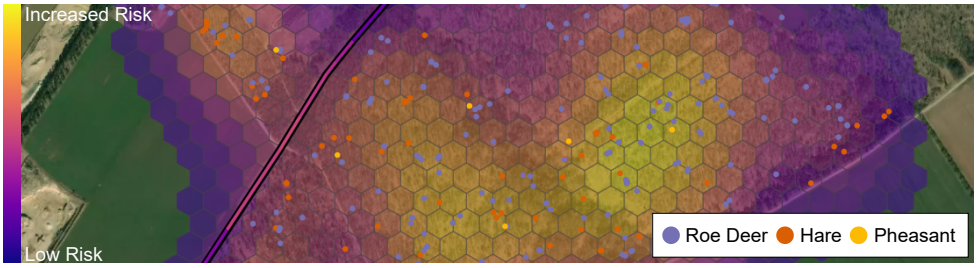


Fig. 1. Application of the risk model employing a GMM based on the precise positions of individual animals, as captured through drone recordings, and the nearby road.

Conclusion: In summary, the integration of visual data from drones and the development of a statistical accident risk model based on a GMM provide a promising foundation for managing wildlife accidents and improving overall road safety. Using this model, it is possible to respond appropriately to the challenges wildlife pose on roadways, saving lives and promoting human-wildlife coexistence. Through continued research, collaboration, and proactive measures, road safety can be further improved, and valuable wildlife can be protected.

References

1. A. Seiler, and B. Manisha "Wildlife and traffic: An inevitable but not unsolvable problem?." In Problematic Wildlife II: New Conservation and Management Challenges in the Human-Wildlife Interactions, 2020.
2. I. Uwanuakwa, et al. "Traffic warning system for wildlife road crossing accidents using artificial intelligence." In Int. Conference on Transport. and Dev., 2020.
3. D. Reynolds "Gaussian mixture models." Encyclopedia of biometrics, 2009.
4. M. Hooten, et al. "Animal movement models for migratory individuals and groups." In Methods in Ecology and Evolution, 2018.
5. C. Praschl and D. C. Schedl, "Towards an Automated Biodiversity Modelling Process for Forest Animals using Uncrewed Aerial Vehicles" In Proceedings of the 11th International Workshop on SESDE, 2023.
6. O. Bimber, I. Kurmi and D. C. Schedl, "Synthetic aperture imaging with drones" in IEEE computer graphics and applications, vol. 39, p. 8–15, 2019.
7. M. Levoy und P. Hanrahan, Light field rendering, in Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, 1996, S. 31–42.

Context-Aware Vehicle Lane Change Prediction based on Knowledge Graphs and Bayesian Inference

Mohamed Manzour[1], Rubén Izquierdo[1], Miguel Ángel Sotelo[1]

[1] Computer Engineering Department, University of Alcalá, Spain
ahmed.manzour@uah.es ruben.izquierdo@uah.es miguel.sotelo@uah.es

Abstract. The prediction of vehicle lane change manoeuvres has witnessed a notable surge in recent years. Some recent works have been directed towards the anticipation of a vehicle's intent by pre-emptively forecasting its trajectory. However, this approach falls short, for it fails to encapsulate the broader contextual intricacies inherent in the vehicle's context. This study is mainly concerned with the prediction of drivers' lane-changing intentions, while considering the Time to Collision (TTC) and all other contextual data. Moreover, this research leverages Knowledge Graphs (KG), which are based on the incorporation of expressive information or linguistically characterized temporal data extracted from the HighD dataset. The KG undergoes training using the 'TransE' model to derive Knowledge Graph Embeddings (KGE). Subsequently, Bayesian Reasoning is applied on top of the Knowledge Graph, using the learned embeddings to undertake the prediction process. The model is able to predict lane changing intentions at 94% f1-score with an average anticipation of three seconds.

Keywords: Lane Change Prediction, Knowledge Graphs, Bayesian Inference.

1 Description of work

Recently, different works focused on predicting vehicle lane changing intention using different inputs and methods, being a remarkable one the research conducted on the grounds of the PREVENTION dataset [1]. In this paper, we propose to develop a context-aware vehicle lane change prediction system enhanced by Knowledge Graphs (KG) [2] by means of graph reification techniques and downstream Bayesian reasoning. A knowledge graph is a directed heterogeneous multigraph. In a KG, a triple < s, r, o > consists of a subject (s) and object (o) entity connected by a relation (r). For example, 'vehicle INTENTION_IS rightLaneChanging'. The subject entity is 'vehicle', the relation is 'INTENTION_IS', and the object entity is 'rightLaneChanging'. The KG is built using our own vehicle ontology (see Fig. 1) describing contextual variables and data. The inputs used by the KG model are: vehicle lateral (velocity and acceleration), TTC to (preceding, left preceding, right preceding, left following, right following) vehicles. Some of these variables are not available in any dataset and, consequently, they have been post-processed based on raw data. Once the KG is built following the vehicle ontology, low dimensional embeddings of all entities and relations are computed using Knowledge Graph Embeddings (KGE) [3]. KGE is a supervised machine learning task

that learns to represent (embed) the knowledge graph entities and relations into a low-dimensional vector space while preserving semantic meaning. In order to avoid the inductive reasoning limitation of KGEs, several graph reification techniques have been applied on the KG prior to training. On top of that, Bayesian inference has been deployed as a downstream mechanism with a view to develop a fully inductive predictive system. After training the system on the HighD dataset [4], and after applying Bayesian reasoning on the learned embeddings, the prediction system achieves an f1-score of 94%, exhibiting the ability to anticipate lane changes with an average value of 3 seconds (a maximum value of 5 seconds). These results are only possible thanks to the way KGs leverage contextual information.

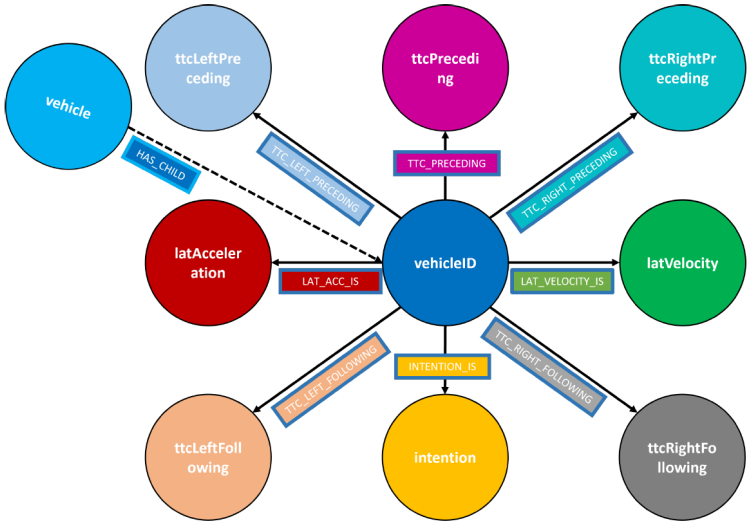


Fig. 1. Vehicle ontology used to build the Knowledge Graph for vehicle lane change prediction.

Acknowledgments. This research has been funded by the HEIDI project of the European Commission under reference number 101069538.

References

1. Izquierdo, R., Quintanar, A., Parra, I., Fernández-Llorca, D., and Sotelo, M. A. "Experimental validation of lane-change intention prediction methodologies based on CNN and LSTM," in IEEE Intelligent Transportation Systems Conference, pp. 3657–3662 (2019)
2. Hogan, A., et al. "Knowledge Graphs". arXiv : 2003.02320v6 [cs.AI] (2021)
3. Martin, A., Hinkelmann, K., Fill, H.-G., Gerber, A., Lenat, D., Stolle, R., van Harmelen, F. "An Evaluation of Knowledge Graph Embeddings for Autonomous Driving Data: Experience and Practice". Proceedings of the AAAI 2020 Spring Symposium on Combining Machine Learning and Knowledge Engineering in Practice (2020)
4. Krajewsfi, R., Bock, J., Kloeker, L., Eckstein, L. "The highD Dataset: a Drone Dataset of Naturalistic Vehicle Trajectories on German highways for Validation of Highly Automated Driving Systems". arXiv: 1810.05642. (2018)

Towards a Unified Incident Detection and Response System for Autonomous Transportation

Liron Ahmeti, Conrad Meyer, Klara Dološ, and Andreas Attenberger

Research Units Big Data Analysis & Digital Forensics, Central Office for Information Technology in the Security Sector (ZITiS), 81677 München, Germany
poststelle@zitis.bund.de
https://www.zitis.bund.de

Extended Abstract

As autonomous vehicles continue to evolve and become more prevalent in everyday private transportation, it is critical that systems automatically detect and handle incidents during vehicle operations. Current systems are limited in scope and functionality, e.g., accident data recorders in standard non-AI vehicles only store driving data locally and must be retrieved using proprietary tools. As a result, it is difficult or impracticable to link data between vehicles involved in an accident without extensive manual data preparation [1]. In addition, data integrity may be increasingly compromised by targeted attacks on connected vehicles [2]. In Europe, EU Regulation 2019/2144 must be implemented by 2024 to establish a standardised data backup system to combat several of the above issues [4]. In this work, we propose a comprehensive data monitoring solution for autonomous vehicles that can be used for storage and forensic analysis of driving data.

The approach is based on closed recording systems in the vehicle, as required by the EU, and extends them with an additional monitoring system outside the vehicle. The continuous availability of the data transmission capability enables implementation of an external system. In doing so, we secure the data inside and outside the vehicle. Autonomous vehicles transmit safety-relevant data to a server located outside the vehicle. We rely on external data backup to ensure integrity and availability, especially in an accident. Currently, external data backup is not covered by the EU regulation, but it would allow law enforcement to access the data more quickly and shorten the time to clarify incidents. In addition, data is protected against loss through vehicle damage in accidents or targeted local attacks. In the Client2X architecture shown in Figure 1, clients consisting of autonomous vehicles and road-side units (RSUs) are connected and can exchange messages. The clients are also connected to a server that can retrieve data from third parties. By combining both internal and external vehicle data including cloud and V2X data, a more comprehensive understanding of vehicle performance, traffic conditions and other relevant factors can be achieved. For example, data recording can be extended during adverse driving conditions

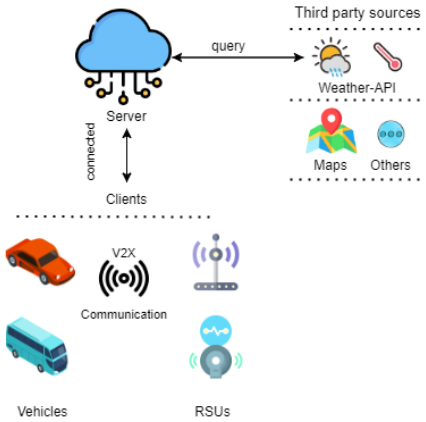


Fig. 1. Client2X Architecture of the Incident Response System.

or cross-checked between vehicles. Furthermore, by introducing machine learning algorithms, driver identification can be leveraged for improved anti-theft protection [3]. Asymmetric encryption is utilized for secure transmission.

Compared to other vehicle data recording approaches, by adding a server component, incidents can be detected more quickly as well as facilitating data access and recovery. Using a Client2X architecture, we can achieve higher performance and increased data availability for specific processing. This goes beyond current implementations and regulatory frameworks. Future work will include testing and implementing the presented approach with data sources external and internal to the vehicle including cloud sources. Additionally, the benefits of a Client2X architecture in incident detection need to be verified with specific use cases.

References

1. Attenberger, A.: Data sources for information extraction in automotive forensics. In: Computer Aided Systems Theory–EUROCAST 2019: 17th International Conference, Las Palmas de Gran Canaria, Spain, February 17–22, 2019, Revised Selected Papers, Part II 17. pp. 137–144. Springer (2020)
2. Dibaei, M., Zheng, X., Jiang, K., Abbas, R., Liu, S., Zhang, Y., Xiang, Y., Yu, S.: Attacks and defences on intelligent connected vehicles: a survey. Digital Communications and Networks **6**(4), 399–421 (2020). https://doi.org/https://doi.org/10.1016/j.dcan.2020.04.007, https://www.sciencedirect.com/science/article/pii/S235286481930197X
3. Dolos, K., Meyer, C., Attenberger, A., Steinberger, J.: Forensic driver identification considering an unknown suspect. International Journal of Applied Mathematics and Computer Science **31**, 587 – 599 (2021)
4. European Parliament and Council of the European Union: Regulation (eu) 2019/2144 of the european parliament and of the council (Dec 2019)

Tackling "Padre Anchieta" Roundabout from New Simulation and Optimization Perspectives*

Oliver Castilla[1], Leopoldo Acosta[2], and Javier J. Sanchez-Medina[3]

[1] Grado en Ingeniería Electrónica Industrial y Automática, Universidad de La Laguna
[2] Departamento de Ingeniería Informática y de Sistemas, Universidad de La Laguna
lacosta@ull.edu.es
[3] IUCES, Universidad de Las Palmas de Gran Canaria
javier.sanchez@ulpgc.es

Abstract. In the Canary Islands there are 813 cars per 1,000 habitants. On the Tenerife island, this fact with other several factors derived from the demographic growth in recent decades have triggered an extreme situation that is approaching the collapse of several main roads of the insular road network. The preliminary status of a project, focused in solving a critical roundabout collapse will be shared, together with the ongoing challenges and solutions tried.

Keywords: Traffic · Control system · Simulation.

During the last years, several traffic studies have been realized by some organizations commissioned by the insular government ("Cabildo de Tenerife") [1, 2]. The results, conclusions and actions derived from that studies are still to be known. Literature is plenty of applicable developments, in similar situations in other cities. There are smaller works which have studied easier traffic situations trying to regulate the vehicles flow witch roundabouts [3] or smart traffic lights [4], but not using both.

The starting project we are sharing has many stages ahead. In a first stage we will study and simulate the access roads to the city of San Cristóbal de La Laguna, centered on the roundabout of "Padre Anchieta" to carry out an action plan to alleviate congestion on the roads around "the first city on the island". This environment has a high population density, as well as commerce, leisure and educational services from elementary to higher education. That is why, facilitating access and exit from its urban center is essential for the correct development of its municipal economy as well as for the lives of its residents and people who come to it daily.

The present project aims at developing software tools and data analyses, fed from real data facilitated by "Cabildo de Tenerife". The offsprings may be

* Supported by Universidad de La Laguna and Universidad de Las Palmas de Gran Canaria.

recomendatory, such as possible modifications in the road network, road extensions, burials, uprisings or the implementation of intelligent traffic lights and vehicle flow control systems.

A central element to be proposed may be developing a ramp metering online system, that can control the incoming volumes to "Padre Anchieta" roundabout, in order to avoid its blocking, and that problem wide ramifications in La Laguna. The intention is to propose a real solution with a model validated in simulation that serves as a basis for the possible modifications that this route may suffer in order that traffic does not continue to be a problem of insular interest.

"Cabildo de Tenerife" has on his public website yearly data collection from static traffic stations, which are measuring the vehicle flows giving a final graphic results an tables where we can take an initial idea about traffic congestion and arrivals in some specific time zones and places.

As an example, in figure 1 we have the averaged hourly load distribution. More data will be compiled in order to have the best possible picture of the phenomenon to be modelled.

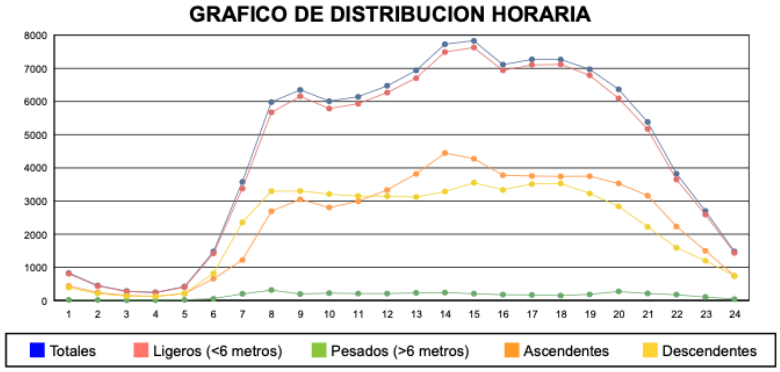


Fig. 1. Graphic flow of vehicles given by time zone on "Padre Anchieta" roundabout, 2022. *Source: Cabildo de Tenerife*

.

References

1. Consultoría Alomon Homepage, https://consultoralomon.com/. Last accessed 6 April 2021
2. Europapress Homepage, https://europapress.es/. Last accessed 19 May 2023
3. Jarillo Rodríguez, J. J. Rotondas Interurbanas. En XIV Symposium Nacional de Vías y Obras de Administración Local (Vyodeal). I Congreso Nacional de Carreteras Locales. 1997.
4. Molina-Navarro, Antonio, et al. Los semáforos inteligentes en la logística urbana sustentable. Revista de Aplicaciones de la Ingenieria, 2016, vol. 3, no 9, p. 26-33.

Edge-Processing of Myoelectric Signals

Klaus Buchenrieder

Institut für Technische Informatik
Universität der Bundeswehr München
Neubiberg, Germany
klaus.buchenrieder@unibw.de

Keywords: Myoelectric Sensor · Edge Processing · Prosthesis Control

Extended Abstract

Myoelectric sensors measure electrical signals which are generated by muscles to control external devices, such as prosthetic limbs or robotic arms. To tap the excitation of a muscle, skin-contact electrodes are placed over a residual muscle pack to capture the signal when it contracts. Figure 1 shows a classic system in which the myoelectric signal is amplified and band-limited with an anti-aliasing filter [4]. Almost without exception, all control devices on the market are designed and built to work in this way [3].

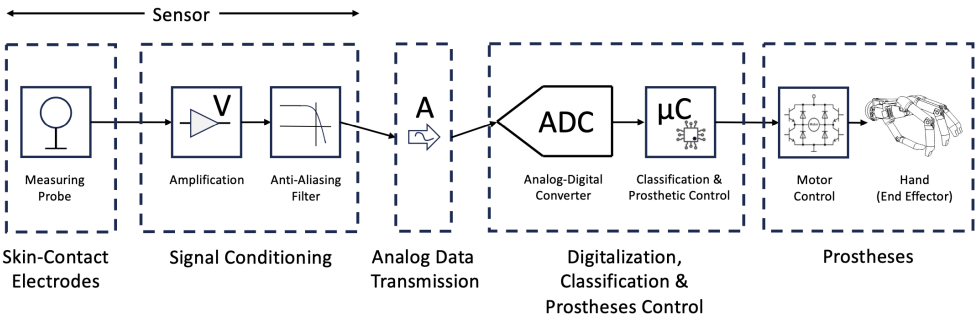


Fig. 1. Traditional sensor-actor arrangement.

Conventional sensors provide either a low-pass filtered, highly amplified raw myoelectric or a rectified narrow-band (RMS) signal [1] to determine the grip or action from the sensor signals. This can be done in two ways: first, by calculating the difference between the sensor outputs and applying a threshold to the result, or second, by classifying features extracted from one or more signals using pre-trained algorithms. Since current classification systems that primarily use artificial intelligence, sensor systems are now expected to provide an adequate number of features to suitably separate the classification space. In addition to features, time- and spike-information is often expected to carry out neuromorphic computations. Established methods for hum and noise suppression are also

no longer satisfactory, as external noise and interference have increased and became a ubiquitous problem. The development of signal processing algorithms and specialized hardware, which can remove unwanted noise while perfectly capturing relevant muscle signals, has become a true challenge. Another long-standing problem is the variability of the signals to be processed due to muscle fatigue, changes in skin impedance and electrode placement. As each user has a unique physiology, muscle structure and wearing preference, modern sensors must also be technically designed to adapt effectively to different users, taking into account factors such as limb size, muscle distribution and user preferences.

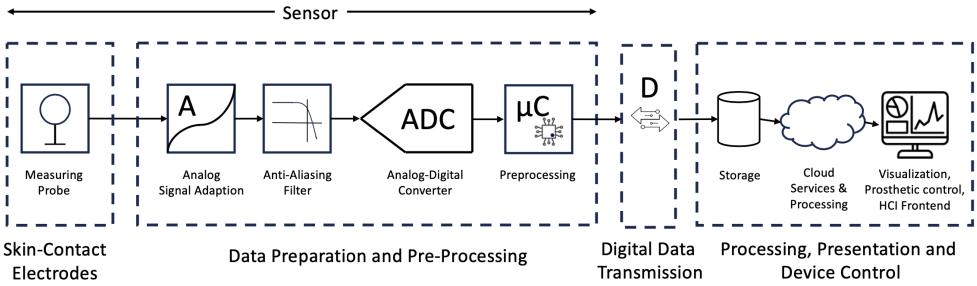


Fig. 2. Myoelectric sensor based on edge-processing principles.

In this paper, we present an integrated sensor that processes cutaneous signals directly at the front-end, Figure 2. Edge processing has several advantages: compact design, hum and interference reduction, adaptive and parameterizable digital filters at the sensor-edge, feature extraction on the device, and inclusion of time-information for better classification [2] with AI algorithms or neuromorphic computing methods.

Acknowledgment

This research is part of the projects: VITAL-SENSE and ROLORAN, funded by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr. dtec.bw is funded by the European Union – NextGenerationEU.

References

1. **A**ttenberger, A., Wojciechowski, S.: A Real-Time Classification System for Upper Limb Prosthesis Control in MATLAB. In: Computer Aided Systems Theory - EUROCAST 2017, Volume 10672 of LNCS. (2017) 359-365.
2. **G**aßner, P., Buchenrieder, K.: Improved Classification of Myoelectric Signals by Using Normalized Signal Trains. In: EUROCAST 2019, Volume 12014 of LNTCS. (2019) 372-379.
3. **O**skoei, M., Hu, H.: Myoelectric Control Systems – A Survey. In: Biomedical Signal Processing and Control, Volume 2, Issue 4, 275-294.
4. **P**rakash, A., Sharma, S., Sharma, N.: A compact-sized surface EMG sensor for myoelectric hand prosthesis. Biomedical Engineering Letters 9, (2019) 467–479.

Enhancing Virtual Reality Medical Training through Gesture Recognition and Action Sequencing

Daniel Arias Ruiz-Esquide[1,2], Jose Juan Reyes Cabrera[1,2], Abian Hernandez-Guedes[1], Agustín Trujillo Pino[4,*], Miguel Angel Rodriguez-Florido[1,2,3]

1. Medical and Audiovisual Technology Group, Institute of Biomedical and Health Research, University of Las Palmas de Gran Canaria
2. Chair of Medical Technology,ctm.ulpgc.es,University of Las Palmas de Gran Canaria
3. Laboratory of Technology-Based Simulation and Education, General and Children's Insular Teaching Hospital, www.motivando.me
4. Imaging Technology Center (CTIM), Institute of Cybernetics, Business and Society (IUCES), University of Las Palmas de Gran Canaria

Abstract. In recent years, Virtual Reality (VR) has become a powerful tool for medical training, allowing students to immerse themselves in simulated dissection rooms and interact with virtual patients. However, a challenge arises when discrepancies occur between the perceived and actual depth or proximity to objects within the simulated environment. This misalignment can lead to unintended interactions, impacting training quality, potentially causing discomfort or inefficiency.

While this is not the main scope of our work, it is important to highlight that issues in VR resulting from these discrepancies are well-studied [1]. Common side effects include motion sickness (or cybersickness) and perceptuomotor after-effects.

In this paper, we propose a solution leveraging insights from the Aloha [2] project's "Action Chunking with Transformers" (ACT) algorithm [3]. The authors used Transformers [4] due to their advantage at modeling non-Markovian sources. Unlike Recurrent Neural Networks (RNN), transformers have a non-sequential nature that allows this and optimizes training times.

While Aloha primarily focuses on precise manipulation tasks, we aim to adapt it to enhance VR medical training. Our goal is to enable an intelligent system to recognize, analyze and learn the gestures of medical professionals and use this to subsequently guide and refine the actions of trainee students.

In our group, we are using a VR Dissection Room [4] that lets the students learn anatomy through this technology while we get a real-time tracking of their actions in the virtual space. It has screens to show instructions to the students, a traditional dissection table and a lifelike body to interact upon. Navigating and accomplishing tasks in this environment are both the key parts of the application and the main source of problems since the target student may have limited familiarity with VR technology and this only makes the already mentioned problem more likely to happen.

Drawing inspiration from Aloha, our approach involves training an artificial intelligence (AI) model to learn sequences of actions or gestures performed by medical professionals during specific procedures. This AI model functions as a virtual mentor that understands the nuanced gestures of experienced doctors and observes the actions of the student within the VR environment. By comparing the student's gestures with learned patterns, the AI subtly ensures that their actions align with professional standards. This correction mechanism operates seamlessly, without overtly assisting or providing hints to the student. Instead, it imbues the system with intelligence, bridging the gap between the student's intended actions and the technology. This minor guidance aims to prevent unintended actions, fostering a more accurate and realistic training exercise, improving the experience of the student.

While the proximity of the student's controls to the intended target is typically a reliable indicator, we explore additional considerations such as distance thresholds to enhance the system's robustness. The AI's suggestions can be selectively incorporated, providing flexibility for educators to tailor the level of guidance based on the learner's proficiency and the specific medical scenario or procedure.

In conclusion, our approach integrates VR technology with gesture recognition and action predictions inspired by the Aloha project, offering a novel solution to address the challenges of depth perception discrepancies in medical training. This is the foundation of our proposal, highlighting the potential to revolutionize the way medical students engage with virtual dissection rooms and simulated patients, ultimately improving the efficacy of VR-based medical education.

Keywords: Virtual Reality, Medical Training, Gesture Recognition, Action Sequencing, Action Prediction, Aloha Project, Action Chunking with Transformers, Artificial Intelligence.

References

1. Baniasadi T, Ayyoubzadeh SM, Mohammadzadeh N. Challenges and practical considerations in applying virtual reality in medical education and treatment. Oman Med J. 2020;35(3):e125. doi: 10.5001/omj.2020.43
2. Zhao, T. Z., Kumar, V., Levine, S., & Finn, C. (2023). Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware. arXiv preprint arXiv:2304.13705. doi: 10.48550/arXiv.2304.13705
3. ACT Algorithm GitHub Repository. (Last update on September 5, 2023). Retrieved from https://github.com/tonyzhaozh/act
4. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30. doi: 10.48550/arXiv.1706.03762
5. AVRIR project description: https://ctm.ulpgc.es/noticia/impacto-mediatico-proyecto-avrir-2022

Advanced Protocols and Integrated Algorithms for Comprehensive Monitoring of Cardiac Implantable Electronic Devices

Marek Kulbacki[4,5][0000−0003−4609−106X], Dariusz Jagielski[1,7][0000−0002−3066−4030], Artur Bąk[4][0000−0003−2165−193X], Jakub Segen[4,5][0000−0003−4148−4817], Michał Kulbacki[6][0000−0001−9113−962X], Zenon Chaczko[2][0000−0002−7348−4128], Jerzy Rozenblit[3][0000−0002−7348−4128], Jan Nikodem[1][0000−0002−3748−9255], and Ryszard Klempous[1][0000−0003−4889−0930]

[1] Wrocław University of Science and Technology, Wrocław, Poland
[2] School of Elect. and Data Engineering University of Technology Sydney, Australia
[3] Depart. of Electrical and Computer Engineering University of Arizona Tucson, USA
[4] Polish-Japanese Academy of Information Technology, Warsaw, Poland
[5] DIVE IN AI, Wrocław, Poland
[6] Mathematical Institute, University of Wrocław, Poland
[7] Center for Heart Diseases, 4th Military Hospital, Wrocław, Poland

1 Extended Abstract

Implantable cardioverter-defibrillators (ICDs) pose potential challenges associated with their presence in patients' bodies [8, 2]. Locating [7] and detecting ICDs may present difficulties [6], as does maintaining the continuity and integrity of electrodes. Breaks or damages to electrodes can lead to disruptions in signal transmission, and electrodes embedding into tissues may cause local irritation and inflammatory responses [1]. There is also a risk of ICD displacement [5], adhesion phenomena, and sensitivity to magnetic resonance imaging (MRI). Regular assessment and monitoring are crucial to ensure patient safety and the effective therapy of implantable cardiac devices. This article systematically explores state-of-the-art approaches and advanced algorithms to monitor cardiac implantable electronic devices (CIEDs) comprehensively. In managing cardiac arrhythmias and enhancing patient outcomes, CIEDs necessitate meticulous monitoring to ensure proper functioning and prompt intervention during malfunctions. The paper introduces the extension of CaRDLA-X [4] detection protocol, leveraging advanced image data analysis techniques for precise localization of CIEDs within the human body. This precision is crucial for an accurate assessment of device performance and the identification of potential issues. In addition to CaRDLA-X, we discuss alternative approaches to CIED detection and localization, encompassing methodologies involving artificial intelligence (AI), radiolocation, and ultrasound techniques. Each approach carries its unique advantages and limitations, with the selection of the most suitable method contingent upon the specific needs and constraints of the application. The research delves into identification algorithms [3] facilitating the unequivocal determination of the type of CIED implanted in a patient. This accurate identification is

imperative for tailoring monitoring strategies and interpreting device-generated data. Furthermore, the paper explores integrity check algorithms, providing mechanisms to assess the status of various CIED functions, including electrode integrity, positioning, defibrillation capabilities, and stimulation functions. The article proposes an integrated solution combining CaRDLA-X detection protocols, CIED identification algorithms, and integrity check algorithms to monitor CIEDs comprehensively. This integrated approach offers a holistic perspective on device performance, facilitating the timely identification of potential problems. Integrating information from CIED monitoring protocols with patient clinical data enables personalized monitoring strategies and informed clinical decision-making. The advanced protocols and algorithms presented in this article aim to substantiate effective monitoring of CIEDs, thereby contributing to enhanced patient management and therapeutic efficacy. The proposal of an integrated solution represents a pioneering step towards the comprehensive supervision of implantable cardiac devices, ensuring patient safety and optimal therapeutic efficiency.

References

1. Caughron, H., Bowman, H., Raitt, M.H., Whooley, M.A., Tarasovsky, G., Shen, H., Matheny, M.E., Selzman, K.A., Wang, L., Major, J., et al.: Cardiovascular implantable electronic device lead safety: Harnessing real-world remote monitoring data for medical device evaluation. Heart Rhythm **20**(4), 512–519 (2023)
2. Forman, J., Baumbusch, J., Jackson, H., Lindenberg, J., Shook, A., Bashir, J.: Exploring the patients experiences of living with a subcutaneous implantable cardioverter defibrillator. European journal of cardiovascular nursing **17**(8), 698–706 (2018)
3. Howard, J.P., Fisher, L., Shun-Shin, M.J., Keene, D., Arnold, A.D., Ahmad, Y., Cook, C.M., Moon, J.C., Manisty, C.H., Whinnett, Z.I., et al.: Cardiac rhythm device identification using neural networks. JACC: Clinical Electrophysiology **5**(5), 576–586 (2019)
4. Jacob, S., Shahzad, M.A., Maheshwari, R., Panaich, S.S., Aravindhakshan, R.: Cardiac rhythm device identification algorithm using x-rays: Cardia-x. Heart Rhythm **8**(6), 915–922 (2011)
5. Kuliński, W.: Problems in physical management in patients with an implanted cardiac pacemaker. Acta Balneologica **173**(2) (2023)
6. Parthiban, N., Esterman, A., Mahajan, R., Twomey, D.J., Pathak, R.K., Lau, D.H., Roberts-Thomson, K.C., Young, G.D., Sanders, P., Ganesan, A.N.: Remote monitoring of implantable cardioverter-defibrillators: a systematic review and meta-analysis of clinical outcomes. Journal of the American College of Cardiology **65**(24), 2591–2600 (2015)
7. Umay, I., Fidan, B., Barshan, B.: Localization and tracking of implantable biomedical sensors. Sensors **17**(3), 583 (2017)
8. Van Gelder, B.M., Bracke, F.A., Oto, A., Yildirir, A., Haas, P.C., Seger, J.J., Stainback, R.F., BOTMAN, K.J., Meijer, A.: Diagnosis and management of inadvertently placed pacing and icd leads in the left ventricle: a multicenter experience and review of the literature. Pacing and Clinical Electrophysiology **23**(5), 877–883 (2000)

Machine Learning Methods for Synovitis Arthritis Analysis in Human Joints: A Survey

Artur Bąk[4 [0000−0003−2165−193X]], Ryszard Klempous[1 [0000−0003−4889−0930]], Jan Nikodem[1 [0000−0002−3748−9255]], Jerzy Rozenblit[3 [0000−0002−7348−4128]], Zenon Chaczko[2 [0000−0002−2816−7510]], Micha Kulbacki[6 [0000−0001−9113−962X]], Konrad Kluwak[1 [0000−0003−0983−2228]], Konrad Wojciechowski[4 [0000−0001−7025−8355]], Wojciech Bożejko[1 [0000−0002−1868−8603]], Dariusz Jagielski[1 [0000−0002−3066−4030]], Jakub Segen[4,5 [0000−0002−8697−6025]], Ito Atsushi[[0000−0002−9686−4019]7], Katarzyna Gruszecka[1,8], Marta Skoczyńska[8 [0000−0001−8921−8919]], Anna Panejko[8], Hubert Kowalczyk[1], and Marek Kulbacki[4,5 [0000−0003−4609−106X]]

[1] Wroclaw University of Science and Technology, Wroclaw, Poland
[2] School of Elect. and Data Engineering University of Technology Sydney, Australia
[3] Depart. of Electrical and Computer Engineering University of Arizona Tucson, USA
[4] Polish-Japanese Academy of Information Technology, Warsaw, Poland
[5] DIVE IN AI, Wroclaw, Poland
[6] Mathematical Institute, University of Wroclaw, Poland
[7] Faculty of Economics, Chuo University, Tokyo, Japan
[8] Depart. of Rheumatology and Internal Diseases, Lower Silesian Specialist Hospital, Wroclaw, Poland

1 Extended Abstract

Synovitis arthritis, characterized by inflammation of the synovial membrane in joints, poses challenges in early diagnosis and effective treatment [8, 3] increasing a mortality risk [9]. Current diagnostic methods for synovitis arthritis, such as physical examinations [5] are only sometimes accurate, but as relevant as the most popular imaging studies [6, 7, 10, 2].

This article critically reviews various machine learning (ML) methods to analyze synovitis arthritis in human joints [1]. The exploration encompasses a range of classic ML techniques, including supervised learning, unsupervised learning, and modern deep learning approaches [14]. It also indicates new directions opened for simulation, like arthritis modeling with joint-on-a-chip [12] or exploring different factors like signaling pathways in rheumatoid arthritis [13, 4].

The paper emphasizes early and precise detection of synovitis arthritis to improve patient outcomes. Furthermore, it discusses the latest progress in deep learning with various neural architectures for automated joint analysis. We also address the challenges and limitations of current ML approaches, like the necessity for large and diverse datasets [11], model interpretability, and generalization of different patient populations. A discussion on potential future directions in ML-based synovitis arthritis detection, such as the integration of multimodal data sources and advancements in transfer learning, concludes the review.

This survey aims to provide a consolidated resource for interdisciplinary researchers, clinicians, and practitioners in rheumatology and medical imaging by synthesizing current research. It aims to enhance understanding of the current ML methods for synovitis arthritis detection, paving the way for improved diagnostic capabilities and patient care.

References

1. Binvignat, M., Pedoia, V., Butte, A.J., Louati, K., Klatzmann, D., Berenbaum, F., Mariotti-Ferrandiz, E., Sellam, J.: Use of machine learning in osteoarthritis research: a systematic literature review. RMD open **8**(1), e001998 (2022)
2. Bk, A., Wereszczyski, K., Segen, J., Mielnik, P., Fojcik, M., Kulbacki, M.: Human finger joint detection on USG image: structure and optimization. IJET (2024)
3. Cush, J.J.: Rheumatoid arthritis: early diagnosis and treatment. Rheumatic Disease Clinics **48**(2), 537–547 (2022)
4. Ding, Q., Hu, W., Wang, R., Yang, Q., Zhu, M., Li, M., Cai, J., Rose, P., Mao, J., Zhu, Y.Z.: Signaling pathways in rheumatoid arthritis: implications for targeted therapy. Signal Transduction and Targeted Therapy **8**(1), 68 (2023)
5. Dougados, M., Jousse-Joulin, S., Mistretta, F., d'Agostino, M.A., Backhaus, M., Bentin, J., Chalès, G., Chary-Valckenaere, I., Conaghan, P., Etchepare, F., et al.: Evaluation of several ultrasonography scoring systems for synovitis and comparison to clinical examination: results from a prospective multicentre study of rheumatoid arthritis. Annals of the rheumatic diseases **69**(5), 828–833 (2010)
6. Farrant, J., OConnor, P., Grainger, A.: Advanced imaging in rheumatoid arthritis: Part 1: Synovitis. Skeletal radiology **36**, 269–279 (2007)
7. Hayashi, D., Roemer, F.W., Katur, A., Felson, D.T., Yang, S.O., Alomran, F., Guermazi, A.: Imaging of synovitis in osteoarthritis: current status and outlook. In: Seminars in arthritis and rheumatism. pp. 116–130. Elsevier (2011)
8. Heidari, B.: Rheumatoid arthritis: Early diagnosis and treatment outcomes. Caspian journal of internal medicine **2**(1), 161 (2011)
9. Van den Hoek, J., Boshuizen, H., Roorda, L., Tijhuis, G., Nurmohamed, M., Van den Bos, G., Dekker, J.: Mortality in patients with rheumatoid arthritis: a 15-year prospective cohort study. Rheumatology international **37**, 487–493 (2017)
10. Jamar, F., van der Laken, C.J., Panagiotidis, E., Steinz, M.M., van der Geest, K.S., Graham, R.N., Gheysens, O.: Update on imaging of inflammatory arthritis and related disorders. In: Seminars in Nuclear Medicine. Elsevier (2023)
11. Kulbacki, M., Segen, J., Habela, P., Janiak, M., Knieć, W., Fojcik, M., Mielnik, P., Wojciechowski, K.: Collaborative tool for annotation of synovitis and assessment in ultrasound images. In: Computer Vision and Graphics: International Conference, ICCVG 2014, Warsaw, Poland. Proceedings. pp. 364–373. Springer (2014)
12. Li, Z.A., Sant, S., Cho, S.K., Goodman, S.B., Bunnell, B.A., Tuan, R.S., Gold, M.S., Lin, H.: Synovial joint-on-a-chip for modeling arthritis: Progress, pitfalls, and potential. Trends in Biotechnology **41**(4), 511–527 (2023)
13. McMaster, C., Bird, A., Liew, D.F., Buchanan, R.R., Owen, C.E., Chapman, W.W., Pires, D.E.: Artificial intelligence and deep learning for rheumatologists. Arthritis & Rheumatology **74**(12), 1893–1905 (2022)
14. Parashar, A., Rishi, R., Parashar, A., Rida, I.: Medical imaging in rheumatoid arthritis: A review on deep learning approach. Open Life Sciences **18**(1) (2023)

6DoF Motion Tracking in Immersive Technologies and Its Usefulness in Medical Application

Jan Nikodem[1 [0000−0002−3748−9255]], Ryszard Klempous[1 [0000−0003−4889−0930]], Jerzy Rozenblit[2 [0000−0002−7348−4128]], Hubert Kowalczyk[1], Konrad Kluwak[1 [0000−0003−0983−2228]], Marek Kulbacki[3,4 [0000−0003−4609−106X]], and Artur Bąk[3 [0000−0003−2165−193X]]

[1] Wroclaw University of Science and Technology, Wroclaw, Poland
[2] Depart. of Electrical and Computer Engineering University of Arizona Tucson, USA
[3] Polish-Japanese Academy of Information Technology, Warsaw, Poland
[4] DIVE IN AI, Wroclaw, Poland

Extended Abstract

Augmented Reality and Mixed Reality are both considered as immersive technologies. Augmented Reality (AR) most often uses virtual 3D graphics overlay on a camera image. The compatibility of the real and virtual worlds is considered as static, in which this digitally produced 3D graphics enhanced the real world. With respect to virtual pointView and projectionType variables, AR uses visual markers [7] or Simultaneous Localization and Mapping (SLAM) techniques [5, 2] without pre-trained markers, to realize the 6DoF tracking [8], [4].

Mixed Reality (MR) environment is characterized by the capability of interaction between any of pair from three; user, real and virtual world. To provide a more immersive and realistic experience, MR introduces interactions which results in significantly higher requirements for compatibility of the coordinate system of the real and virtual worlds, at least in the Field of View area. The virtual soft surface is required to give way under pressure of a real finger in the exact place user touch. To provide a more immersive and realistic experience, MR introduces interactions which results in significantly higher requirements for compatibility of the coordinate system of the real and virtual worlds, at least in the Field of View area.

Based on AR and MR medical applications require high accuracy of mapping virtual and real world, both in terms of space and time. In space, because in a simulator the results of the interaction of the real laparoscopic instrument on the patient's virtual tissues must reproduce real interactions (deflection, shift, splitting, cutting) as faithful as possible. In time, because in medical applications the basic feedback is the eye-hand coordination. In such a feedback mechanism, any image delay results in the more difficult task performance.

In this presentation, we will dissect intricately a process of 6DoF object tracking in 3D space, discussing the methodology of calculations and focusing on its stability/repeatability, accuracy, precision and short delay performance [3,

6]. In our tests, we worked with Microsoft HoloLens - computerized holographic head-up device running Windows 10, since it promises to be a mobile and cost-effective solution for argumented and mixed realities [1] where the use-case's accuracy and precision requirements are very stringent.

We will describe the tests performed and discuss the obtained results The spatial accuracy as the distance between the reported (virtual) and physical (real) position, respectively denoting horizontal, vertical, combined spatial accuracy and their precision will be consider. We will discuss the distribution of spatial accuracy across all collected mesh data in the dataset due to different view angles, distances and projections (perspectives). Temporal precision of the Hololens2 device we discuss based on the variability of inter-sample intervals between consecutive timestamps.

References

1. Guo, H.J., Prabhakaran, B.: Hololens 2 technical evaluation as mixed reality guide (2022)
2. Hübner, P., Clintworth, K., Liu, Q., Weinmann, M., Wursthorn, S.: Evaluation of hololens tracking and depth sensing for indoor mapping applications. Sensors **20**(4) (2020). https://doi.org/10.3390/s20041021, https://www.mdpi.com/1424-8220/20/4/1021
3. Kinnen, T., Blut, C., Effkemann, C., Blankenbach, J.: Thermal reality capturing with the microsoft hololens 2 for energy system analysis. Energy and Buildings **288**, 113020 (2023). https://doi.org/https://doi.org/10.1016/j.enbuild.2023.113020, https://www.sciencedirect.com/science/article/pii/S0378778823002505
4. Rabbi, I., I., Ullah, S.: Extending the tracking distance of fiducial markers for large indoor augmented reality applications. Advances in Electrical and Computer Engineering **15**(pp.59-64) (2015). https://doi.org/10.4316/AECE.2015.02008, https://doi.org/10.4316/AECE.2015.02008
5. Taketomi, T., Uchiyama, H., Ikeda, S.: Visual slam algorithms: a survey from 2010 to 2016. IPSJ Transactions on Computer Vision and Applications **9**(1) (2017). https://doi.org/10.1186/s41074-017-0027-2, https://doi.org/10.1186/s41074-017-0027-2
6. Tavakoli, H., Walunj, S., Pahlevannejad, P., Plociennik, C., Ruskowski, M.: Small object detection for near real-time egocentric perception in a manual assembly scenario. CoRR **abs/2106.06403** (2021), https://arxiv.org/abs/2106.06403
7. Thompson, S., Omori, Y., Komuro, T., Okayama, K., Kato, N., Tanaka, H.: Simulation and evaluation of lentimark markers for accurate pose estimation. In: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 4112–4117 (2018). https://doi.org/10.1109/SMC.2018.00697
8. Weinmann, M., Jäger, M., Wursthorn, S., Jutzi, B., Hübner, P.: 3d indoor mapping with the microsoft hololens: Qualitative and quantitative evaluation by means of geometric features. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences **V-1-2020**, 165–172 (08 2020). https://doi.org/10.5194/isprs-annals-V-1-2020-165-2020

Exploration of Medical Human Digital Twin Applications for Cerebral Palsy

Marek Kulbacki[4,5][0000−0003−4609−106X], Jakub Segen[4,5][0000−0002−8697−6025], Michał Kulbacki[6][0000−0001−9113−962X], Zenon Chaczko[2][0000−0002−2816−7510], Wojciech Bożejko[1][0000−0002−1868−8603], Patrycja Mrozek[7][0009−0009−3655−2687], Carmen Paz Suárez Araujo[8][0000−0002−8826−0899], Ryszard Klempous[1][0000−0003−4889−0930], and Marcin Bonikowski[7][0000−0002−2028−7024]

[1] Wrocław University of Science and Technology, Wrocław, Poland
[2] School of Elect. and Data Engineering University of Technology Sydney, Australia
[3] Polish-Japanese Academy of Information Technology, Warsaw, Poland
[4] DIVE IN AI, Wrocław, Poland
[5] Mathematical Institute, University of Wrocław, Poland
[6] Mazovian Neuropsychiatry Center, Zagórze, Warsaw, Poland
[7] Instituto Universitario de Ciencias y Tecnologías Cibernéticas, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

1 Extended Abstract

Cerebral Palsy (CP) is a multifaceted neurological disorder [2, 13, 12] demanding individualized and multidisciplinary medical interventions [1] to address its impact on movement, posture, and coordination [11]. In recent years, the concept of a "digital twin" has emerged as a promising tool in healthcare, allowing for the creation of virtual replicas of individuals [10] to simulate and monitor their health conditions. This paper provides a comprehensive exploration of the applications, methodologies, and challenges associated with the implementation of medical human digital twins [14] specifically tailored for individuals with cerebral palsy.

The exploration encompasses an overview of existing research and developments in the field of digital twins, emphasizing their utility in understanding the dynamic and heterogeneous nature of cerebral palsy [5]. The paper explores various imaging modalities, sensor technologies [4], and data fusion techniques employed in acquiring the necessary input for constructing accurate digital representations of individuals with CP [3]. Additionally, it delves into the computational models and simulation frameworks utilized to simulate the biomechanics [6], neural activity, and overall functionality of the affected individuals.

Furthermore, the exploration addresses the integration of digital twins into clinical practice, highlighting their potential contributions to diagnosis, treatment planning, and personalized rehabilitation strategies [8] for cerebral palsy patients. The ethical considerations and privacy concerns associated with the creation and utilization of medical human digital twins are also discussed.

The synthesis of information presented in this paper aims to provide researchers, clinicians, and technologists with a comprehensive understanding of

the current landscape, challenges, and opportunities in developing and implementing medical human digital twins [9] for cerebral palsy. By shedding light on the state-of-the-art technologies and methodologies, this exploration contributes to the advancement of personalized healthcare solutions [7] for individuals with cerebral palsy, ultimately improving their quality of life and therapeutic outcomes.

References

1. Aisen, M.L., Kerkovich, D., Mast, J., Mulroy, S., Wren, T.A., Kay, R.M., Rethlefsen, S.A.: Cerebral palsy: clinical care and neurological rehabilitation. The Lancet Neurology **10**(9), 844–852 (2011)
2. Bax, M., Goldstein, M., Rosenbaum, P., Leviton, A., Paneth, N., Dan, B., Jacobsson, B., Damiano, D.: Proposed definition and classification of cerebral palsy, april 2005. Developmental medicine and child neurology **47**(8), 571–576 (2005)
3. Carcreff, L., Gerber, C.N., Paraschiv-Ionescu, A., De Coulon, G., Newman, C.J., Armand, S., Aminian, K.: What is the best configuration of wearable sensors to measure spatiotemporal gait parameters in children with cerebral palsy? Sensors **18**(2), 394 (2018)
4. Chang, Y.J., Han, W.Y., Tsai, Y.C.: A kinect-based upper limb rehabilitation system to assist people with cerebral palsy. Research in developmental disabilities **34**(11), 3654–3659 (2013)
5. Gámez Díaz, R., Yu, Q., Ding, Y., Laamarti, F., El Saddik, A.: Digital twin coaching for physical activities: A survey. Sensors **20**(20), 5936 (2020)
6. George, A., Bharat, C., Singh, T., Sahil, C.S.: Digital twin of a musculoskeletal system. In: 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1). pp. 1–4. IEEE (2023)
7. Kamel Boulos, M.N., Zhang, P.: Digital twins: from personalised medicine to precision public health. Journal of personalized medicine **11**(8), 745 (2021)
8. Khan, M.M.R., Sunny, M.S.H., Ahmed, T., Shahria, M.T., Modi, P.P., Zarif, M.I.I., Sanjuan, J., Ahamed, S.I., Ahmed, H.U., Ketchum, E.M.M., et al.: Development of a robot-assisted telerehabilitation system with integrated iiot and digital twin. IEEE Access (2023)
9. Lauer-Schmaltz, M., Cash, P., Hansen, J., Maier, A.: Designing human digital twins for behaviour-changing therapy and rehabilitation: a systematic review. Proceedings of the Design Society **2**, 1303–1312 (2022)
10. Miller, M.E., Spatz, E.: A unified view of a human digital twin. Human-Intelligent Systems Integration **4**(1-2), 23–33 (2022)
11. Papavasiliou, A.S.: Management of motor problems in cerebral palsy: a critical update for the clinician. european journal of paediatric neurology **13**(5), 387–396 (2009)
12. Patel, D.R., Neelakantan, M., Pandher, K., Merrick, J.: Cerebral palsy in children: a clinical overview. Translational pediatrics **9**(Suppl 1), S125 (2020)
13. Sadowska, M., Sarecka-Hujar, B., Kopyta, I.: Cerebral palsy: current opinions on definition, epidemiology, risk factors, classification and treatment options. Neuropsychiatric disease and treatment pp. 1505–1518 (2020)
14. Shengli, W.: Is human digital twin possible? Computer Methods and Programs in Biomedicine Update **1**, 100014 (2021)

Usability testing of virtual reality software in digital pathology

Miklos Kozlovszky[1][0000-0001-8096-9628], Bence Biricz[2], Miklós Vincze[2], Abdallah Benhamida[2], Viktor Jónás[2], Róbert Paulik[2], Boldizsár Szócska[2] and Levente Kovács[3]

[1] BioTech Research Center, EKIK Óbuda University Budapest, Hungary
[2] John von Neumann Faculty of Informatics, Óbuda University, Budapest, Hungary
{kozlovszky.miklos, abdallah.benhamida}@nik.uni-obuda.hu
{biricz.bence, miklos.vincze, jonas.viktor, paulik.robert, szocska.boldizsar, kovacs}@uni-obuda.hu

Abstract. Virtual reality (VR) is becoming increasingly popular in industrial applications and research fields. While the development of virtual reality applications is progressing rapidly, it is still a challenge to define precisely the aspects that make virtual reality software high quality, user-friendly, and usable in the long term. The biggest problem is that traditional software testing standards are not fully applicable for virtual reality software, and can only be partially followed by developers. The use of virtual reality for scientific and industrial purposes can offer new opportunities for researchers and change the way humans interact with computers. For virtual reality software various usability metrics can be defined, such as: reduction of negative environmental effects, discomfort in virtual environments, content consistency and intuitive control. The aim of this paper is to enumerate and define the aspects that can be used to design a user-friendly virtual environment for examining digitalized pathological sections, and to understand the parameter set that can be used to test the usability of the software. Using the defined metrics, our goal is to show and describe our software framework, that enables automatic usability testing of virtual reality software. The developed software can be used to objectively measure the user's head and hand motion and the positional difference between the physical and virtual space of VR glasses and controllers.

Keywords: Usability testing, VR, Virtual Reality, Medical software, digital pathology.

1 Usability testing in VR environment

Usability testing aims to determine how comfortable, fast, and learnable a particular software is. To test usability, we have developed a testing process, and established a VR testing environment to record objective data during testing process.

1.1 Head and arm movement tracking

We found that in order to create a truly user-friendly interface, it is important to measure and analyze the user's movement and physical activities while using the system. Our team's goal was to develop a VR software intended for pathological use, where users must look at specific patterns for hours.

1.2 Jitter evaluation in VR environment

We tested the VR jitter difference among different virtual reality glasses. We have successfully done comparison of the position tracking accuracy of differ-ent VR headsets. Testing was done with HTC Vive Pro and Oculus Rift virtual reality goggles. During testing, we aimed for the best settings and calibration procedures recommended by the manufacturers.

2 Discussion

The results of this study provide support that the measurement of hand and head movement and rotation define objective information in measuring the usability of a virtual reality application. From the jitter focused tests it can be clearly see that the slight tremor of the body increases with the variation in the virtual reality position and significantly more increases in the rotation deviations. The values around the rotation axis exceed the maximum of 0.5°.

In conclusion, the development of virtual reality (VR) technology has led to new opportunities in scientific research and industrial applications. While VR software are rapidly evolving, defining the precise aspects that make it user-friendly and functional remains a challenge. The use of VR in scientific research can revolutionize the way humans interact with computers, as it pro-vides a more immersive and interactive environment.

Our team defined objective usability factors, such as head and hand movement and rotation, and tested the problems of VR devices. We developed software which measures the users' head and hand movements and assesses the positional accuracy of VR glasses and controllers in relation to the physical space.

Acknowledgments. The research received funding from the 2019-1.3.1-KK-2019- 00007 “Innovációs szolgáltató bázis létrehozása diagnosztikai, terápiás és kutatási célú kiberorvosi rendszerek fejlesztésére” national project. Additionally, the work was supported in part by the Eötvös Loránd Research Network Secretariat (Development of cyber-medical systems based on AI and hybrid cloud methods) under Agreement ELKH KÖ-37/2021. Authors would like to thank for all the financial support hereby.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

In vitro to *in vivo* extrapolation of pharmacokinetic model of doxorubicin treated mammary tumor ⋆

Borbála Gergics[1 [0009−0001−2607−1179]], Dániel András Drexler[1 [0000−0001−6655−4354]], and Levente Kovács[1 [0000−0002−3188−0800]]

Óbuda University, Budapest, Hungary
gergics.borbala@uni-obuda.hu
drexler.daniel@uni-obuda.hu
kovacs@uni-obuda.hu

Keywords: IVIVE · PKPD modeling · parameter identification.

1 Extended Abstract

Cancer is considered a prominent health problem worldwide due to the increasing number of cases and high mortality rates [1]. The complexity and heterogeneity of cancer cause limited treatment options. The effectiveness of treatment can vary based on the type and stage of cancer and the individual characteristics of the patients [2]. Well-established and relatively cost-effective curing methods such as chemotherapy are designed for the average, which can lead to several side effects and resistance [3]. Personalized chemotherapy aims to modify therapy to the specific features of the parameters of a patient, allowing for a more effective and less toxic approach.

Tumor spheroids are an essential and widely used tool in cancer research as they fill the gap between two-dimensional *in vitro* cell cultures and *in vivo* models. With the help of spheroids, tumor progression, and therapy optimization can be established cost-effectively and more ethically [4]. However, interspecies scaling and extrapolation of the *in vitro*-based mathematical models of tumor growth, pharmacodynamics (PD), or pharmacokinetics (PK) are required in order to gain clinically relevant predictions. *In vitro* to *in vivo* extrapolation (IVIVE) offers a solution to calibrate human models with the use of *in vitro* tumor spheroids and preclinical datasets even if the clinical data is rarely available [5, 6].

⋆ Project no. 2019-1.3.1-KK-2019-00007. has been implemented with the support provided by the National Research, Development, and Innovation Fund of Hungary, financed under the 2019-1.3.1-KK funding scheme. This project has been supported by the Hungarian National Research, Development, and Innovation Fund of Hungary, financed under the TKP2021-NKTA-36 funding scheme. The work of Dániel András Drexler was supported by the Starting Excellence Researcher Program of Óbuda University, Budapest Hungary. Borbála Gergics is with the Óbuda University, Applied Informatics and Applied Mathematics Doctoral School, Budapest, Hungary.

Here, we present a tumor growth and PD/PK model that was calibrated on *in vivo* data (mouse) and *in vitro* data (tumor spheroids). In order to describe the pathophysiological phenomena of tumors using differential equations, we applied formal reaction kinetics [7–9]. The *in vivo* values were gained from cytotoxicity measurements of doxorubicin on $Brca1^{-/-}$; $p53-/-$ mouse mammary tumor model and the *in vitro* spheroids consisted of cell line previously isolated from the same mouse mammary tumor [10]. Parameter identification of the models was done independently based on *in vitro* and *in vivo* experiment data set, then we investigated the model fit with a hybrid strategy: we integrated both the *in vivo* and *in vitro* measurement values and the parameter estimation was conducted with allowing exactly one parameter to be scale-specific (*in vitro* or *in vivo*-based) while the other parameters remain *in vivo*-based. With this method, IVIVE of the PD/PK model of doxorubicin-treated mammary tumors can be carried out. Nevertheless, this approach is based on only *in vivo* preclinical studies; the integration of human datasets is necessary, and other species experiments are strongly recommended to achieve more precise IVIVE broadening on clinical prediction and interspecies scaling.

References

1. International Agency for Research on Cancer Homepage, https://gco.iarc.fr/tomorrow/en/dataviz/isotype. Last accessed 29 Nov 2023
2. Arneth, B.: Tumor microenvironment. Medicina, **56(1)**, 15 (2019)
3. Love, R. R., Leventhal, H., Easterling, D. V., and Nerenz, D. R.: Side effects and emotional distress during cancer chemotherapy. Cancer, **63(3)**, 604-612. (1989)
4. Weiswald LB, Bellet D, Dangles-Marie V.: Spherical cancer models in tumor biology. In: Neoplasia **17(1)**, 1-15 (2015)
5. Jaroch, K., Jaroch, A., and Bojko, B.: Cell cultures in drug discovery and development: The need of reliable in vitro-in vivo extrapolation for pharmacodynamics and pharmacokinetics assessment. Journal of Pharmaceutical and Biomedical Analysis, **147**, 297-312. (2018)
6. Wetmore, B. A.: Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. Toxicology, **332**, 94-101. (2015)
7. Drexler, D. A., Ferenci, T., Füredi, A., Szakács, G., and Kovács, L.: Experimental data-driven tumor modeling for chemotherapy. IFAC-PapersOnLine, **53(2)**, 16245-16250. (2020)
8. Gergics, B., Vajda, F., Ládi, A., Füredi, A., and Drexler, D. A.: Pharmacodynamics modeling based on in vitro 3d cell culture experiments. In 2023 IEEE 17th International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 000499-000504). IEEE. (2023, May).
9. Gergics, B., Vajda, F., Puskás, M., Füredi, A., and Drexler, D. A.: Mathematical modeling of phototoxicity during fluorescent imaging of tumor spheroids. In 2023 IEEE 27th International Conference on Intelligent Engineering Systems (INES) (pp. 000291-000296). IEEE. (2023, July)
10. Hámori, L., Kudlik, G., Szebényi, K., Kucsma, N., Szeder, B., Póti, Á., ... and Szakács, G.: Establishment and characterization of a $Brca1^{-/-}$, $p53^{-/-}$ mouse mammary tumor cell line. International Journal of Molecular Sciences, **21(4)**, 1185. (2020)

Transformation of IEC 61131-3 onto an Embedded Platform Using LLVM

Nils Frank[1], Markus Friedrich[1] and Benjamin Kormann[2]

[1] Krones AG, Böhmerwaldstraße 5, 93073 Neutraubling, Germany
{nils.frank, markus.friedrich}@krones.com
[2] Munich University of Applied Sciences, Lothstr. 64, 80335 Munich, Germany
benjamin.kormann@hm.edu

Industrial processes are getting more complex and production lines become more connected. The deprecated architecture of traditional automation technology does not meet today's requirements for modern software engineering in order to solve future problems. Programmable Logic Controllers (PLCs), introduced in the late 1960s as successors to hard-wired systems, use standard-defined programming languages specified in IEC 61131-3 [2]. However, their basic architecture has remained largely unchanged and is no longer sufficient for modern industrial needs [3]. These requirements include flexible hardware availability, short development and shipping times, cost pressure, and a modular range of products and connectivity. Traditional automation technology is encountering constraints due to its aging architecture, limiting its capabilities. PLC development lacks dynamic processes and relies on live on-site changes, incurring costs and downtime. Proprietary systems among PLC suppliers create incompatibilities despite the IEC 61131 standard. The IEC 61499 standard advocates for event-based systems for future PLCs, but faces obstacles in industry adoption [4]. The need for modernization is recognized, but a consensus on implementation remains elusive.

This paper proposes the use of an embedded platform as the future platform for automation controls and a transpilation system to port existing automation technology applications onto an embedded target system, see Figure 1. This system consists of two main parts, an LLVM[3] compiler tool-chain for automated application logic transformation and an execution engine for runtime requirements. Although the IEC 61131-3 standard defines multiple programming languages, the proposed system focuses on Structured Text (ST). PLCs typically include other application elements like user interaction, which is out of the scope of the presented approach.

The LLVM compiler front end parses the PLC application (ST) and generates a technology independent LLVM Intermediate Representation (IR) of it. In [1] the authors also propose an IEC 61131-3 compiler front end using LLVM. The generated IR is compiled to an object file by the back end containing the transformed functions of the application. At this stage the source application is entirely transformed into the execution engine using its defined interfaces to base functions. The execution engines creates a PLC-typical cyclic execution model

[3] "LLVM" is not an acronym but rather the project's full name. See: https://llvm.org/

using the PLC configuration like cycle time and bus configuration. The two parts are linked together creating a target executable. Additional runtime configuration and application parameters at run time is transferred onto the embedded system. The entire transpilation process with the transpilation system onto the embedded system performs automatically.

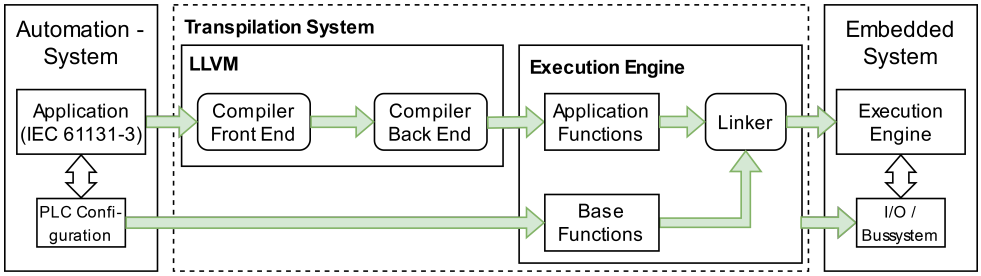


Fig. 1. Transpilation of the Application from an Automation to an Embedded System

A working prototype demonstrates the feasibility of the concept using an industrial closed-loop real-time application from the beverage industry without being limited to this particular industry. The hardware of the prototype consists of real automation components: Beckhoff CX2040 as industrial PC running a real-time patched (PreemptRT) Linux operating system including EtherCAT communication to sensors and actuators commonly used in the food and beverage industry.

This forms a base for a new, future-proof automation architecture which can integrate existing IEC 61131-3 source code with newly-developed code onto a common and modern embedded platform. More work has to be done to support all IEC 61131 elements as well as more analysis to meet all requirements of high-end real-time applications.

References

1. Catalão, T., De Sousa, M.: IEC 61131-3 Front-End for the LLVM Compiler Family. In: IEEE International Conference on Emerging Technologies and Factory Automation, ETFA. vol. 2020-September, pp. 1191–1194 (2020). https://doi.org/10.1109/ETFA46521.2020.9211921, iSSN: 1946-0740
2. Internation Electrotechnical Commission: Programmable controllers - Part 3: Programming languages. Standard IEC 61131-3:2013, Internation Electrotechnical Commission (2013)
3. Sehr, M.A., et al.: Programmable Logic Controllers in the Context of Industry 4.0. IEEE Transactions on Industrial Informatics **17**(5), 3523–3533 (May 2021). https://doi.org/10.1109/TII.2020.3007764, https://ieeexplore.ieee.org/document/9134804/
4. Thramboulidis, K.: IEC 61499 vs. 61131: A Comparison Based on Misperceptions. Journal of Software Engineering and Applications **06**(08), 405–415 (2013). https://doi.org/10.4236/jsea.2013.68050, http://arxiv.org/abs/1303.4761, arXiv:1303.4761 [cs]

Retail 4.0 as a Booster for Industry 5.0

M. Horn[1], S. Kranzer[1], J. Nöbauer[1], L. Portenschlager[1], T. Neureiter[1], P. Pankonin[1], V. Schleifer[1], R. Zniva[1]

Salzburg University of Applied Science, Salzburg, Austria

1 Extended Abstract

In their 2022 paper titled Industry 5.0: improving humanization and sustainability of Industry 4.0 the authors highlight the limitations of the Industry 4.0 concept, particularly focusing on the role of humans in smart factories and the aspects related to sustainable development [1]. After a strong focus on solely technology, a more human centered approach is needed. On the other hand, in brick and mortar retail, technology such as robotics, sensor technology and the collection and use of data from a wide range of systems are solutions to current problems relating to staff availability and cost efficiency. In retail in particular, however, the focus of all efforts has always been on the individual customer. The customer journey takes center stage. This work compares the respective domains of industry and commerce with regard to the use of modern technologies and the role of humans and shows possibilities for a mutual transfer of knowledge.

The knowledge transfer between the domains industry and retail can offer a decisive advantage over competitors in the future. Established technologies in Industry 4.0, such as the Internet of Things (IoT), Cloud Computing, Big Data Analytics (BDA), Augmented Reality (AR) and Artificial Intelligence (AI), can also enhance efficiency in the retail sector. Specific use cases include the deployment of in-store sensors for product tracking and customer monitoring, coupled with advanced AI solutions for analytics and image processing [2]. Additionally, the use of robots in the industry is now indispensable and will also play a crucial role in retail in the future. Possible applications in Retail 4.0 for robots include cleaning tasks, detection of empty shelves and providing information for customers [3]. The transition of industrial technologies into the retail sector presents new challenges and issues that must be considered in their development and implementation.

At the Point of Sales in brick and mortar retail, conditions are less controlled and regulated than in industrial environments. Changing settings and layouts as well as mostly unrestricted customer access create difficulties. This means that not all technologies can be directly transferred. Challenges exist, among others, in the area of product recognition using deep learning due to the high number of articles [4], in mobile robotics with respect to navigation, path planning and collision avoidance [5], in the use of collaborative robots alongside untrained individuals [6], or in the context of data protection, data security, and the corresponding legal regulations [7]. Further developments in these and other fields are a prerequisite for a smooth implementation of Retail 4.0 and will solve future problems of Industry 5.0.

Retail 4.0's integration of AI, IoT, AR and BDA significantly aligns with Industry 5.0's focus on augmenting human capabilities and experiences. Its application of collaborative robotics in dynamic, human-interactive spaces shows potential industrial adaptations. Furthermore, Retail 4.0's utilization of digital twins for inventory and customer service, alongside its strategies for addressing data privacy, security, and ethical AI concerns, offers valuable guidance for Industry 5.0 [8]. This knowledge and technological transfer, especially in human-robot collaboration, not only advances technological capabilities but also ensures they align with human values and needs [9].

The technologization towards Retail 4.0 is well advanced and leads to various challenges due to the human-centered, uncontrolled, and varying environment at the Point of Sales. Here, numerous research fields emerge which subsequently benefit the automation of retail and further can be considered as a boost for Industry 5.0.

Acknowledgments. The results and findings presented here were partly developed in the Retailization 4.0 [10] and 5G Exploration Space Salzburg [11] projects. Special thanks to the funding bodies The Austrian Research Promotion Agency (FFG) funding Retailization 4.0 and Land Salzburg (WISS) funding 5GEXPS.

References

1. Grabowska, S., Saniuk, S. & Gajdzik, B. Industry 5.0: improving humanization and sustainability of Industry 4.0. *Scientometrics*. **127**, 3117-3144 (2022,6)
2. Har, L., Rashid, U., Chuan, L., Sen, S. & Xia, L. Revolution of Retail Industry: From Perspective of Retail 1.0 to 4.0. *Procedia Computer Science*. **200**
3. Guha, A. & Grewal, D. How robots will affect the future of retailing. *AMS Review*. **12**, 245-252 (2022,12)
4. Wei, Y., Tran, S., Xu, S., Kang, B. & Springer, M. Deep Learning for Retail Product Recognition: Challenges and Techniques. *Computational Intelligence And Neuroscience*. **2020** pp. e8875910 (2020,11)
5. Lewandowski, B., Wengefeld, T., Müller, S., Jenny, M., Glende, S., Schröter, C., Bley, A. & Gross, H. Socially Compliant Human-Robot Interaction for Autonomous Scanning Tasks in Supermarket Environments. *2020 29th IEEE International Conference On Robot And Human Interactive Communication (RO-MAN)*.
6. Valori, M., Scibilia, A., Fassi, I., Saenz, J., Behrens, R., Herbster, S., Bidard, C., Lucet, E., Magisson, A., Schaake, L., Bessler, J., Prange-Lasonder, G., Kühnrich, M., Lassen, A. & Nielsen, K. Validating Safety in Human–Robot Collaboration: Standards and New Perspectives. *Robotics*. **10**, 65 (2021,6)
7. Martin, K., Kim, J., Palmatier, R., Steinhoff, L., Stewart, D., Walker, B., Wang, Y. & Weaven, S. Data Privacy in Retail. *Journal Of Retailing*. **96**, 474-489 (2020,12)
8. Noble, S. M., Mende, M., Grewal, D., & Parasuraman, A. (2022). The Fifth Industrial Revolution: How harmonious human–machine collaboration is triggering a retail and service [r]evolution. *Journal of Retailing*, 98(2), 199–208.
9. Adel, A. (2022). Future of industry 5.0 in society: human-centric solutions, challenges and prospective research areas. *Journal of Cloud Computing*, 11(1), 40.
10. Retailization Homepage, https://retailization.at/, last accessed 2023/11/13
11. 5G Exploration Space Salzburg, https://www.5g-explorationspace.net/category/5g-exps/, last accessed 2023/11/13

Applicability of Chatsbots in Closed-Source Projects

Alfred Feldmeyer[1] and Kai Höfig[2]

[1] Siemens AG, Department ACC-DE, Otto-Hahn-Ring 6, 81739 Munich, Germany
alfred.feldmeyer@siemens.com
[2] Technical University Rosenheim, Hochschulstraße 1, 83024 Rosenheim, Germany
kai.hoefig@th-rosenheim.de

1 Motivation

The increased complexity of modern software development projects, either due to a complex domain such as industrial automation or a large codebase, leads to a need for more understanding during the software development activities. To get familiar with the domain and source code of a software project, most developers read the documentation and the source code, do minimal implementation tasks, join pair programming sessions [2], or ask a given project mentor. Additionally, quality attributes like code readability are often sacrificed while implementing software for embedded systems [1].

AI-based chatbots can support software development by providing code fragments, uncovering bugs, solving domain-specific questions, and conveniently enabling a conversation. Domain knowledge is also part of the source code within the same project or in different projects of the same domain. Since current chatbots such as ChatGPT are primarily trained on open-source information, the user cannot interact with its close repository or use internal documents' domain knowledge.

Therefore, we discuss in this abstract a locally installed chatbot system that can incorporate internal documents and closed source codebases to decrease the learning effort and increase the interactive chatbot's answers, especially for complex domains such as industrial embedded systems for industry automation.

2 Tool Architecture & Setup

For our experiments, we used CodeLlama-13b [3] with 8k maximum input length and 16k token result overall. The setup was done using the Hugging Face Transformer Framework [4]. The initial approach to concatenate several code files and send them as context for a smaller project was given up because the input limit was exceeded in too many cases. Therefore, the experiment was limited to asking questions to single code files.

The overall architecture is depicted in Figure 1. A single code file is provided as a prefix for a prompt to CodeLlama, and it was told to respond as a software

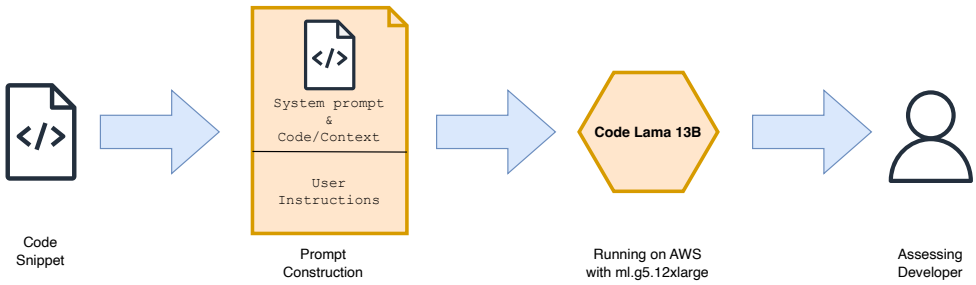


Fig. 1. Architectural Flow Chart

developer. CodeLlama currently runs on AWS but can be installed locally or in a private cloud environment to protect the users' intellectual property.

In our experiment, we use a large software project that develops industrial firmware for an IoT gateway (around 13GiB in size, mainly C++ and C). We did ask rudimental questions such as the amount of instructions in a file, the overall behavior of a class, or the immutability of a class.

3 Results & Outlook

The answers to our questions were promising, yet not idempotent. Also, similar questions about open source software asked to chatbots such as ChatGPT were more precise. Further training and finetuning is required. Additionally, we encountered the problem of a vanishing context since we used a prefix to include context in the form of a code file. As a result, the context did overgrow and was quickly forgotten by the chatbot system (approximately after three or four subsequent questions). In future developments, we thus want to tackle these challenges by implementing a fine-tuned approach and providing a reference dataset.

References

1. Feitosa, D., Ampatzoglou, A., Avgeriou, P., Nakagawa, E.Y.: Investigating Quality Trade-offs in Open Source Critical Embedded Systems. In: Proc. 11th Int. ACM SIGSOFT Conf. Qual. Softw. Archit. pp. 113–122. QoSA '15, Association for Computing Machinery, New York, NY, USA (May 2015). https://doi.org/10.1145/2737182.2737190
2. Hannay, J.E., Dybå, T., Arisholm, E., Sjøberg, D.I.K.: The effectiveness of pair programming: A meta-analysis. Information and Software Technology **51**(7), 1110–1122 (Jul 2009). https://doi.org/10.1016/j.infsof.2009.02.001
3. Rozière, B., Gehring, J., Gloeckle, F., et al.: Code Llama: Open Foundation Models for Code (2023), https://ai.meta.com/research/publications/code-llama-open-foundation-models-for-code/
4. Wolf, T., Debut, L., Sanh, V., et al.: Transformers: State-of-the-Art Natural Language Processing. In: Liu, Q., Schlangen, D. (eds.) Proc. 2020 Conf. Empir. Methods Nat. Lang. Process. Syst. Demonstr. pp. 38–45. Association for Computational Linguistics, Online (Oct 2020). https://doi.org/10.18653/v1/2020.emnlp-demos.6

Machine Learning based Parameter Estimation of Energy Models in Digital Production Environments

Fabian Fichtl, Marco Ullrich, Frieder Heieck, and Bernd Lüdemann-Ravit

Institute for Production and Informatics (IPI) Sonthofen; Mittagstrasse 28A, 87527 Sonthofen, Germany
fabian.fichtl@stud.hs-kempten.de, marco.ullrich@hs-kempten.de, frieder.heieck@hs-kempten.de bernd.luedemann-ravit@hs-kempten.de

Keywords: Machine Learning · Model Optimization · Energy

1 Introduction

In order to exploit the full added value of a holistic industrial metaverse in the context of energy modelling, it is necessary to generate energy models of single components. The energy behaviour of individual systems can be mathematically constructed with corresponding physical parameters. Influential parameters such as efficiency, inertia behavior and friction can be taken from the data sheets of the component manufacturers. The calculated absorbed power of the mathematical model differs from the measured power of the real system. In this abstract, optimization methods are developed and compared with each other in order to optimize the mathematical energy model by means of a parameter estimation. For the energy model based on a physical model which gets control inputs through a trajectory in combination with the energy input/output function it is possible to derive insights to the present energy usage.

The optimized energy model then serves as the basis for energy forecasts and optimizations in industrial plants and entire factories.

2 Concept

By breaking down the assembly into submodels (Fig. 1) with transmission functions and individual efficiency factors, an overarching equation for subsequent optimization can be formulated. The efficiency grade of each component will vary during different operating states for example the servo motor in speed or torque [1], which is correlated to the speed and acceleration of the work piece and all intersecting components.

With an understanding of model dependencies, the task is to validate, test and compare the process and results of different machine learning optimization methods[2]. This requires methods such as physics-informed neural networks,

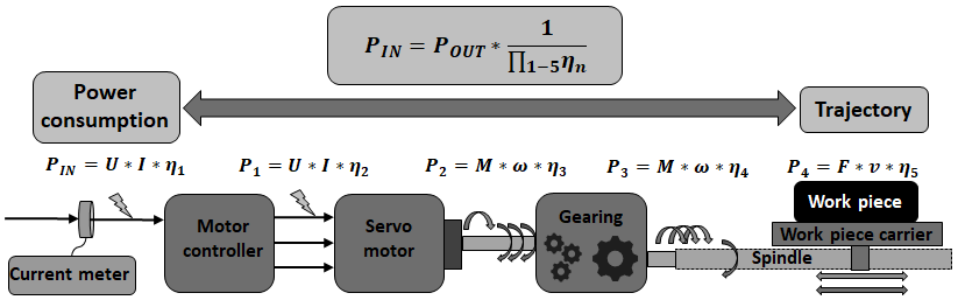


Fig. 1. Schematic representation of the real actuator

polynomial regression, and multiple regression on the generated dataset, followed by validation against the physical model. These methods are part of semi-supervised machine learning and use both labeled data and physical equations to guide the learning process. The variation of the trajectory and testing of the model with different parameters, such as the work piece weight, will expose the need to reduce model complexity and/or establish constraints for optimizing the graphs. This limitation is based on provided manufacturer information and could combined with the reduced complexity resolve the interplay between the optimized parameters so called the "black box problem"[3].

3 Conclusion

The results of this work will establish the foundation for energy prediction and optimization, ensures the increase of production efficiency. Moving forward, the emphasis will be on evaluating the scalability of this approach and its models. The scalability with several degrees of freedom needs to be investigated

References

1. Don Labriola. "The Variation of Servomotor Efficiency in Different Applications." *Power Transmission Engineering*, Chicago Office 1840 Jarvis Avenue Elk Grove Village, IL 60007 USA, 2010. Last visited: 14.11.23. Available: https://www.powertransmission.com/articles/9193-the-variation-of-servomotor-efficiency-in-different-applications
2. Sourav Das, Solomon Tesfamariam. "State-of-the-art review of design of experiments for physics-informed deep learning." arXiv, 13 Feb. 2022, pp. 17, Springer Nature, Canada, 2021 https://arxiv.org/abs/2202.06416v1 https://doi.org/10.1007/978-3-030-85523-3_1.
3. Professor Samir Rawashdeh. "AI's Mysterious 'Black Box' Problem, Explained." University of Michigan-Dearborn, Power Transmission Engineering, 4901 Evergreen Road Dearborn, MI 48128, USA, 2023, (Last accessed: 14 Nov 2023) Available: https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained.

Efficient classification of live sensor data on Low-Energy IoT devices with simple Machine Learning methods

Martin Hanreich[1 [0009−0007−8855−8445]], Oliver Krauss[1 [0000−0002−8136−2606]], and Gerald Zwettler[1 [0000−0002−4966−6853]]

University of Applied Sciences Upper Austria, Hagenberg, 4232, Austria
{martin.hanreich, oliver.krauss, gerald.zwettler}@fh-hagenberg.at

Keywords: Time-Series-Classification · IoT-Devices · Energy-Efficiency · Machine Learning

Introduction Internet of Things (IoT) devices are ubiquitous, but often cannot be permanently connected to the power grid and therefore have to rely on batteries and other energy sources like solar power. These mobile devices should run over a long period without the need for recharging, but still be able to perform comparatively complex tasks from the realm of machine learning like a classification task on live sensor data [7]. The state-of-the-art as far as accuracy is concerned are Neural Network models like LSTM (Long short-term memory) or CNN (Convolutional neural network) [4]. However, these models usually take up more Flash memory, more RAM of the devices and more CPU time than traditional machine learning approaches, which in turn causes a higher energy demand. Another possibility is to transfer the time series data to an external server to do the calculations there and send back the result. For a continuous classification of sensor data, this approach is often not feasible, since the constant transmission of a large quantity of data over a wireless network is required.

We discuss a pipeline for efficient time series classification of live sensor data on low-energy IoT devices with on-device processing. We present our case study of detecting boats docking at an IOT-enabled buoy, which is required to be energy-self-sufficient for several months at a time.

Related work There are numerous publications which are focusing on Deep Learning or Shallow learning for Time Series classification in the context of IoT devices like [5], [3], [6] and [1]. They tackle the hardware constraints with a reduction in the complexity of the networks. A sizeable number of publications also deal with traditional approaches like [2] and [9] or [8]. In recent years, the focus on neural net-based approaches has become even more pronounced due to factors like increased hardware capabilities.

Classification on Low-Energy IoT devices In contrast to other publications like the ones mentioned above, this work focuses on establishing an entire robust pipeline rather than a specific algorithm. It ranges from the initial data collecting, finding the right model and best fitting data preparations, and determining the hyper-parameters, to the classification on the device itself. A variety

of existing approaches, libraries, and tools are used and compared to find the combination that satisfies the requirements the specific task poses. The main underlying concern for all the steps in the pipeline is the trade-off between performance and resource usage, so the entire detection system can be fit easily even on devices with hardware limitations.

The pipeline and all of its steps are shown on the boat docking use case. A microcontroller with an accelerometer and gyroscope is integrated into a buoy. Based on the sensor data, it can now be detected if a boat is docked on the buoy or not using only a few kilobytes of Flash memory and RAM.

Discussion and Outlook We discuss all necessary steps for classifying sensor streams on an IoT device, with a focus on the resources aspect based on a practical example. It is shown that when considering the unique attributes of the task and data, a simple decision tree can be sufficient. That means that even though other approaches like Deep Learning are becoming more and more feasible on microcontrollers, a less complex solution is still preferable if the accuracy demand is met as well. Less complex solutions can work on lower hardware specifications, which makes systems more cost and energy-efficient. However, the approach was conceived with a specific use case in mind at the moment. It needs to be tested how far it can be generalized for other situations. Also, quantifying the exact energy-saving properties in comparison to Neural Network approaches through an empirical study is pending.

References

1. Bhaskar, N., Suchetha, M., Philip, N.Y.: Time series classification-based correlational neural network with bidirectional lstm for automated detection of kidney disease. IEEE Sensors Journal **21**(4), 4811–4818 (2020)
2. Chen, Z., Gu, J.: High-throughput dynamic time warping accelerator for time-series classification with pipelined mixed-signal time-domain computing. IEEE JSSC **56**(2) (2021)
3. Dennis, D.K., Acar, D.A.E., Mandikal, V., Sadasivan, V.S., Simhadri, H.V., Saligrama, V., Jain, P.: Shallow rnns: A method for accurate time-series classification on tiny devices. In: NeurIPS. pp. 12916–12926 (2019)
4. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A.: Deep learning for time series classification: a review. Data mining and knowledge discovery **33**(4), 917–963 (2019)
5. King, T., Zhou, Y., Röddiger, T., Beigl, M.: Micronas: Memory and latency constrained hardware-aware neural architecture search for time series classification on microcontrollers (2023)
6. Lai, L., Suda, N.: Enabling deep learning at the lot edge. In: IEEE ICADD (2018)
7. Saha, S.S., Sandha, S.S., Srivastava, M.: Machine learning for microcontroller-class hardware: A review. IEEE Sensors Journal **22**(22) (2022)
8. Vaughan, N., Gabrys, B.: Scoring and assessment in medical vr training simulators with dynamic time series classification. Eng. Appl. Artif. Intell. **94** (2020)
9. Wang, X., Lin, J., Senin, P., Oates, T., Gandhi, S., Boedihardjo, A.P., Chen, C., Frankenstein, S.: Rpm: Representative pattern mining for efficient time series classification. In: EDBT. pp. 185–196 (2016)

Machine learning using a hybrid quantum classical algorithm with Amplitude Data Encoding

Sebnem Gül-Ficici, Maximilian Spiegel, and Ulrich Göhner

University of Applied Sciences Kempten, Bahnhofstraße 61, 87435 Kempten, Germany sebnem.guel-ficici@hs-kempten.de, maximilian.spiegel@hs-kempten.de, ulrich.goehner@hs-kempten.de

Abstract. Artificial Intelligence (AI) has become an integral part of the industry. As an emerging technology, Quantum Computing has already delivered impressive results. In an earlier paper we compared Quantum Machine Learning (QML) to traditional machine learning algorithms, whereby QML unexpectedly delivered inferior results. In this study, we examine novel encoding methods that extend the feature space to improve the performance of QML algorithms. This work seeks to improve upon the results of the prior paper [6]. The outcomes can then be used to realize the potential of Quantum Machine Learning in industrial applications.

Keywords: Deep Learning · Artificial Intelligence · Hybrid Quantum Machine Learning

1 Introduction and Motivation

The integration of Artificial Intelligence (AI) in the industry is progressing fast and is becoming increasingly important in all areas of science and technology. The resulting time- and cost savings are very promising for all kinds of business. AI models, however, require large quantities of data and a lot of time for training. Quantum computers on the other hand do not require that [2]. Unfortunately, they are also very challenging in their application. The goal of this paper is the development of a generic Quantum Machine Learning (QML) algorithm with minimal technical overhead. Prior tests failed to achieve the expected results. Based on the prior paper, this study examines alternative encoding methods providing a greater number of features and enabling the processing of more complex problems than the previously used PauliFeatureMap. The goal is to improve the performance of QML-algorithms and facilitate more effective solutions.

2 Objectives of the current project

Recent development of quantum computers has the potential to fundamentally change the machine learning sector, particularly the field of deep learning. In order to be trained effectively, deep learning algorithms depend on a large amount

of data. However, the processing of the required amount of data is expensive with regards to time and computing power. The added value of a trained model lies in its capability to detect anomalies as early as possible. Utilization of the hybrid approach of Quantum Machine Learning (QML) is expected to reduce the cost of training and evaluation significantly. For the implementation, this work uses several python libraries like PyTorch and Qiskit among others. For encoding of the hybrid QML approach, the PauliFeatureMap provided by the qiskit library [1] was utilized.
The previous tests, which were run on a physical quantum computer at IBM as well as in a simulation, failed to produce the expected results due to the low complexity of the application and the chosen encoding method. A traditional machine learning algorithm can only have a higher performance when processing more complex problems, which often exhibit a great number of features. The method we employed requires one qubit per feature, which is not practical for more complex tasks. In this work we examine the Amplitude Encoding method to meet these challenges. This method allows to store 2^n features [5] and seeks to expand the feature space as well as the performance of the hybrid QML approach [3, 4]. The results of this study can contribute to an improved integration of quantum computers and machine learning in the industry.

3 Expected Project Outcomes

Utilization of the Amplitude Encoding method opens up new ways to handle significantly more complex problems, where quantum computers are likely to be superior to traditional algorithms. The results of this paper will aid in the adoption of quantum computers across the industry and demonstrate new ways of applying artificial intelligence in complex environments.

References

1. Qiskit pauli feature map (2023), https://qiskit.org/documentation/stubs/qiskit.circuit.library.PauliFeatureMap.html
2. Caro, M.C., Huang, H.Y., Cerezo, M., Sharma, K., Sornborger, A., Cincio, L., Coles, P.J.: Generalization in quantum machine learning from few training data. Nature communications **13**(1), 4919 (2022)
3. Mitsuda, N., Nakaji, K., Suzuki, Y., Tanaka, T., Raymond, R., Tezuka, H., Onodera, T., Yamamoto, N.: Approximate complex amplitude encoding algorithm and its application to classification problem in financial operations. arXiv preprint arXiv:2211.13039 (2022)
4. Schuld, M., Fingerhuth, M., Petruccione, F.: Implementing a distance-based classifier with a quantum interference circuit. Europhysics Letters **119**(6), 60002 (2017)
5. Schuld, M., Killoran, N.: Quantum machine learning in feature hilbert spaces. Physical review letters **122**(4), 040504 (2019)
6. Spiegel, M., Gül-Ficici, S., Göhner, U.: Machine learning using a hybrid quantum-classical algorithm. 14th European LS-DYNA Conference 2023 (2023)

Quantitative Trend Analysis of Reinforcement Learning Algorithms in Production Systems

Georg Schäfer[1,3], Tatjana Krau[2], Maximilian Schirl[3], and Stefan Huber[1,3]

[1] Josef Ressel Centre for Intelligent and Secure Industrial Automation
[2] IPI, University of Applied Sciences Kempten, Germany
[3] Salzburg University of Applied Sciences, Salzburg, Austria
georg.schaefer@fh-salzburg.ac.at

1 Motivation and Methodology

As research in Reinforcement Learning (RL) continues to rapidly grow, keeping up with the latest developments in the field of production systems has become increasingly challenging for researchers and practitioners. An exhaustive literature review, automated and reproducible as presented in our repository[4], helps to address this challenge by summarizing the state-of-the-art, providing insights into current trends in RL algorithms, and may be used in future reevaluations.

To this end, we expand on the work of Panzer and Bender [2], and propose search queries consisting of an algorithmic keyword, "Reinforcement Learning", and general keywords "Assembly", "Automation", "Industry", "Manufacturing", and "Production". Moreover, the keyword "Industrie" is included as well to capture papers related to Industrie 4.0. As Fig. 1 depicts, four major literature databases are chosen to base the trend analysis on. After filtering duplicates and papers published prior to 2013, titles and keywords of the remaining papers are compared with the disciplines presented in [2] (save for "Design") to associate papers with discipline clusters. Algorithm clustering follows the same idea, relying on Khan et al. [1] for the number and denominations of popular RL algorithms. The automated literature review system assigned at least one discipline and one or more algorithms to 250 of 8261 documents.

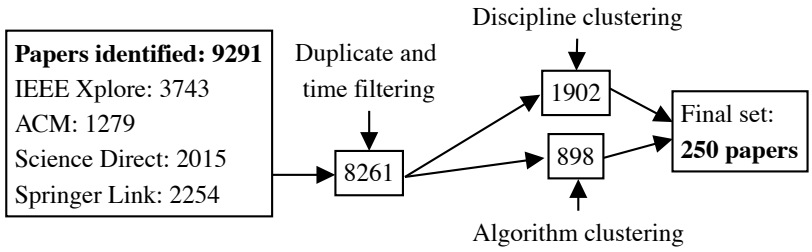


Fig. 1. Streamlined literature review workflow [access date: 22nd Oct 2023].

[4] https://github.com/georgschaefer/trend-analysis-rl-in-production-systems

2 Results and Discussion

Starting in 2017, the number of publications in the studied field of RL in production systems has risen sharply, with the number of papers doubling each year from 2017 to 2019. This sustained upward trend continues with growth factors of 1.62 (2020), 1.32 (2021), and 1.44 (2022) for the following years.

Fig. 2 shows a detailed breakdown of the algorithms used in the queried papers that were assigned at least one discipline and algorithm, split into individual bars for each discipline. Most publications tackle the topics robotics, scheduling and energy management, while Q-learning, DDPG and DQN prove to be the algorithms most mentioned. As such, model-free RL proves to be a promising approach in all disciplines important to the field of production systems.

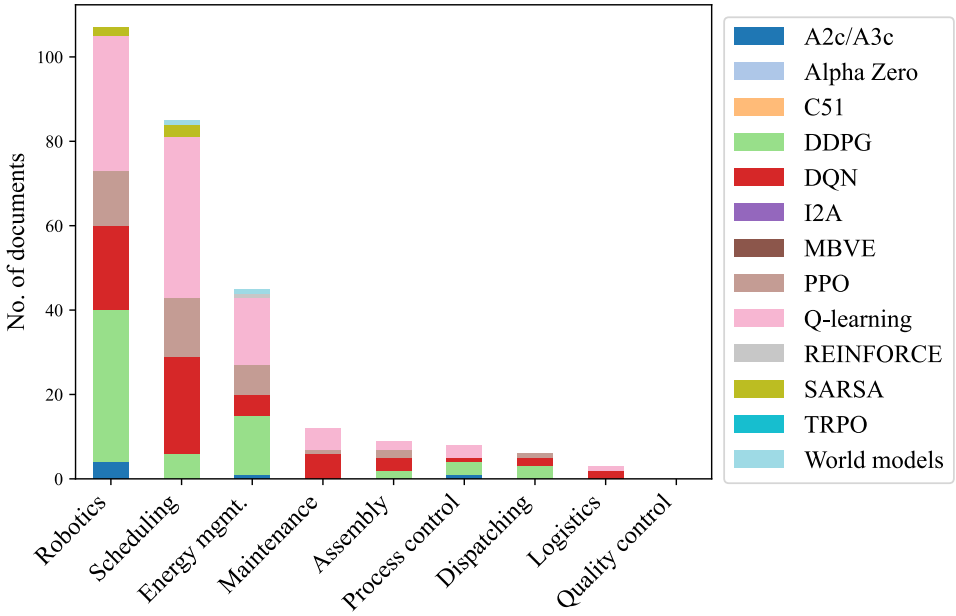


Fig. 2. Reinforcement Learning (RL) algorithms in production systems.

References

1. Khan, M.A.M., Khan, M.R.J., Tooshil, A., Sikder, N., Mahmud, M.P., Kouzani, A.Z., Nahid, A.A.: A systematic review on reinforcement learning-based robotics within the last decade. IEEE Access **8**, 176598–176623 (2020)
2. Panzer, M., Bender, B.: Deep reinforcement learning in production systems: a systematic literature review. International Journal of Production Research **60**(13), 4316–4341 (2022)

Using AutomationML for Advanced Simulation in Industrial Automation*

Robert Fellner[1 [0009−0007−0721−4131]], Ramandeep Pal, and Horst Orsolits[1 [0000−0003−1915−1063]]

University of Applied Scieneces Technikum, Höchstädtplatz 6, 1200 Vienna, Austria
fellner@technikum-wien.at
https://www.technikum-wien.at/

Abstract. This paper explores enhancing manufacturing processes in industrial automation through computer simulations. It integrates AutomationML, an XML-based framework, into Visual Components, optimizing the simulation modeling process. The study adapts a Visual Components XML-Importer add-on to interpret AutomationML data, facilitating efficient updates to an XML file within Visual Components. The findings showcase the practical use of AutomationML in a broader mechatronic production system, contributing to MBSE by improving simulation capabilities and highlighting its potential value in education through simulation-based teaching.

Keywords: Model-Based Systems Engineering (MBSE) · Industrial Automation · Automated Simulation Generation.

1 Introduction

Adopting Model-Based Systems Engineering (MBSE) [4], [2], this study demonstrates AutomationML's feasibility in partially automating simulation models. AutomationML's cutting-edge role in simulation generation, as seen in AML2SIM [1],[3], supports efficient testing and digital twin creation for smart factories and cyber-physical systems.

The study validates AutomationML for partial simulation model automation. It integrates AutomationML seamlessly, adapting Visual Components XML-Importer to accurately interpret and update entry station information. The comprehensive optimization approach ensures accuracy and efficiency in the simulation modeling process.

2 Practical implementation

The study converts and integrates an entry station, part of a broader mechatronic production system, from Visual Components to AutomationML. The meticulously constructed station, assembling axle-bearing blocks with an industrial

* Supported by MA 23 – Economic Affairs, Labour and Statistics) through the research project INVIS (MA23 project 29-11)

robot, forms the basis for a hierarchical structure in AutomationML's Instance-Hierarchy. Semantic roles enhance comprehensibility, while iterative conveyor belt modifications showcase the workflow. An open-source Importer for Visual Components, driven by a python script, adapts the AutomationML instance hierarchy, ensuring seamless loading and facilitating automated simulation generation for efficient virtual prototyping within the MBSE landscape.

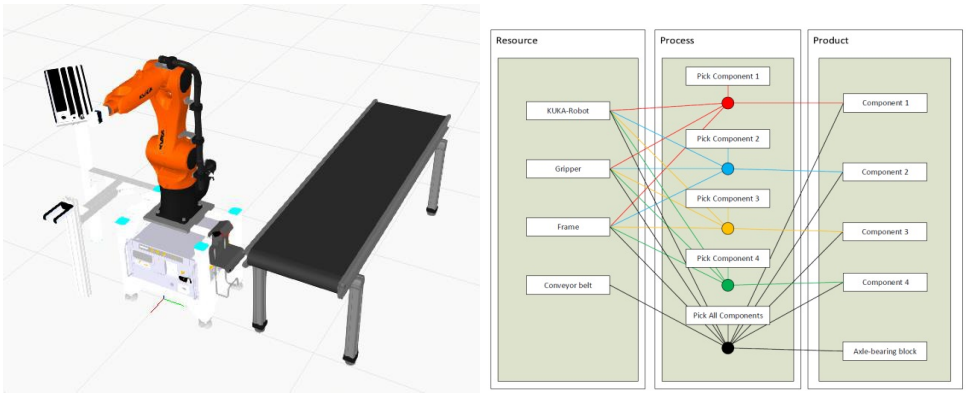


Fig. 1. Entry Station (left) and corresponding PPR roles in AutomationML (right).

3 Results & Outlook

This study reveals pivotal outcomes of the AutomationML-Visual Components interface, advancing MBSE capabilities. Successful initialization improves data storage, exchange, and modular simulation, enhancing efficiency and flexibility in manufacturing processes and in education through simulation-based teaching.

References

1. Behl, H.S., Baydin, A.G., Gal, R., Torr, P.H.S., Vineet, V.: AutoSimulate: (Quickly) Learning Synthetic Data Generation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) Computer Vision – ECCV 2020, vol. 12367, pp. 255–271. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-58542-$6_16, https://link.springer.com/10.1007/978-3-030-58542-6_16, seriesTitle: Lecture Notes in Computer Science$
2. Douglas, B.P., Cakmak, K.: Agile Systems Engineering. In: Advances in Human Factors, Software, and Systems Engineering. vol. 4. AHFE Open Acces. https://doi.org/10.54941/ahfe100126, https://openaccess.cms-conferences.org/publications/book/978-1-4951-2096-1/article/978-1-4951-2096-1_5
3. Drath, R.: AutomationML: A Practical Guide. Walter de Gruyter GmbH & Co KG
4. Manoury, M., Brünnhäußer, J.: Industrie 4.0-Anwendungen: Die Rolle von Simulation und Modellierung (Aug 2022), https://www.digital-engineering-magazin.de/industrie-4-0-anwendungen-die-rolle-von-simulation-und-modellierung/

Development and prototypical implementation of a robot system based on artificial intelligence for gripping and precisely placing unknown objects without a predefined strategy

Joachim Berlak[1[0000-0003-0471-8805]] and Roland Beckert[2] and Stefan Fuchs[2]

[1] FOM University, Campus Munich, Arnulfstr. 30, 80335 Munich, Germany
[2] UNICCOMP GmbH, Bayerwaldstr. 6, 82538, Geretsried, Germany

Keywords: Artificial Intelligence, Machine Learning, Reinforcement Learning, Deep Learning, Robot, Gripping, Unknown Objects

Extended Abstract

Industrial robots have played an important role in reducing production costs and enhancing product quality since the 1960s [1]. The future evolution of industrial robotics necessitates a focus on the development of flexible production systems capable of adapting to changing environmental conditions with minimal effort [2,3]. Current research, exemplified by references [4, 5, 6], is concentrated on the convergence of robotics and artificial intelligence (AI) to empower robots to manipulate unfamiliar objects effectively. This involves equipping robots with cameras and AI systems to enable them to handle objects they have not encountered before. However, this specific challenge has received limited attention in existing literature, with some researchers concentrating solely on well-documented objects [7], while others opt for training AI models using data from simulation systems [8]. Nevertheless, the use of simulation systems presents inherent challenges, including associated costs, the need for complex initial programming, and limitations in real-world applications [9]. The efficient grasping and accurate positioning of unknown objects using AI-equipped robots in an industrial environment constitute the primary motivation for this study.

A fundamental aspect of robot manipulation approaches, especially when dealing with unknown objects, is the choice of AI technology. As illustrated in the work by [10], three primary paradigms are employed: supervised learning (SL), unsupervised learning (UL), and reinforcement learning (RL). While the majority of studies lean towards supervised learning [10, 11, 12], only two out of 28 analyzed approaches adopt an unsupervised learning approach. However, RL offers a fundamentally different perspective, treating the robot as an agent that perceives its environment and interacts with it through actions [13]. An agent encompasses sensors, such as cameras, for environmental perception and actuators, like grippers, for influencing the environment [14]. Utilizing a mathematical function, often referred to as the reward function, the agent receives feedback about the actions it has executed [15].

This paper presents the development of a robot system for handling unknown objects based on RL, applied and tested in a prototypical industrial environment. The evaluation reveals a success rate of 1.8% in the best-case scenario, quantifying the proportion of successful grasping attempts relative to the total number of attempts in the test run. During an exploration phase, where actions were selected randomly, the success rate was recorded at 1.3%. The overall quite low performance of the prototype is attributed

to the limited size of the training dataset. While comparable studies utilize datasets containing several hundred thousand to several million gripping attempts [7], our dataset comprises around 6,000 gripping attempts, with only 77 being successful. Despite these limitations, this research provides valuable insights into the gripping of unknown objects in industrial practice and delineates next steps for further research.

References

1. Pott, A., Dietz, T.: Industrielle Robotersysteme. Entscheiderwissen für die Planung und Umsetzung wirtschaftlicher Roboterlösungen. Springer, Wiesbaden (2019)
2. Lehmann, L., Engelhardt, D., Wilke, W: Kompetenzen für die digitale Transformation 2020. Springer Vieweg, Berlin (2021)
3. Lasi, H., Fettke, P., Kemper, H.-G.,Feld, T, Hoffmann, M.; Industrie 4.0. Wirtschaftsinformatik. 56 (4), 261-264 (2014). doi: 10.1007/s11576-014-0424-4
4. Chen, H., Kiyokawa, T., Wan, W., Harada, K.: Category-Association Based Similarity Matching for Novel Object Pick-and-Place Task. IEEE Robotics and Automation Letters. 7 (2), 2961–2968 (2022). doi: 10.1109/LRA.2022.3145956
5. Liu, R., Nageotte, F., Zanne, P., Mathelin, M., Dresp-Langley, B.: Deep Reinforcement Learning for the Control of Robotic Manipulation: A Focussed Mini-Review. Robotics. 10 (1), 22 (2021). doi: 10.3390/robotics10010022
6. Song, S., Zeng, A., Lee, J., Funkhouser, T.: Grasping in the Wild: Learning 6DoF Closed-Loop Grasping from Low-Cost Demonstrations. IEEE Robotics and Automation (2022) https://arxiv.org/pdf/1912.04344
7. Tremblay, J., To, T., Sundaralingam, B., Xiang, Y., Fox, D., Birchfield, S.: Deep Object Pose Estimation for Semantic Robotic Grasping of Household Objects. In: Aude Billard (Ed.). Conference on Robot Learning (CoRL 2018), Conference on Robot Learning (CoRL 2018), Zürich, 29. - 31.10.2018
8. Gualtieri, M., Platt, R.: Learning 6-DoF Grasping and Pick-Place Using Attention Focus. (2018) http://arxiv.org/pdf/1806.06134v2
9. Kleeberger, K., Huber, M.: Single Shot 6D Object Pose Estimation. In: IEEE (Ed.). 2020 IEEE International Conference on Robotics and Automation, 2020 IEEE International Conference on Robotics and Automation, Paris, 31. Mai - 31. Aug.2020
10. Kleeberger, K., Bormann, R., Kraus, W., Huber, M. F.: A Survey on Learning-Based Robotic Grasping. Current Robotics Reports 1 (4), 239–249 (2020). doi: 10.1007/s43154-020-00021-6
11. Berscheid, L., Meißner, P., Kröger, T.: Robot Learning of Shifting Objects for Grasping in Cluttered Environment. In: IEEE (Ed.). 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 03. - 08. November 2019
12. Wuttke, L.: Künstliche Neuronale Netzwerke: Definition, Einführung, Arten und Funktion. datasolut GmbH (2022). https://datasolut.com/neuronale-netzwerke-einfuehrung/
13. Wuttke, L.: Machine Learning: Definition, Algorithmen, Methoden und Beispiele. datasolut GmbH (2022). https://datasolut.com/was-ist-machine-learning/#Machine-Learning-auf-einen-Blick
14. Russell, S., Norvig, P.: Künstliche Intelligenz. Ein moderner Ansatz. Pearson, Munich, 2012
15. Olesen, A., Gergaly, B., Ryberg, E., Thomsen, M., Chrysostomou, D.: A Collaborative Robot Cell for Random Bin-picking based on Deep Learning Policies and a Multi-gripper Switching Strategy. Procedia Manufacturing 51, 3–10 (2022). doi: 10.1016/j.promfg.2020.10.002

AR digital twin demonstrator for industrial robotics education

Horst Orsolits[1]

University of Applied Sciences Technikum Wien
orsolits@technikum-wien.at

Abstract. The aim of this work is to make a contribution to investigate new approaches for teaching robotics fundamentals using immersive technologies. In this work an existing digital twin for a desktop robot was extended with additional functions.
The performance was then evaluated by means of a pick-and-place application and validated via a test series with an associated questionnaire.

Keywords: Augmented Reality · Digital Twin · Robotics Education.

1 Introduction

The annual World Robotics Report documents 553,052 installations of industrial robots worldwide. This shows a growth rate in robotics of 5% since last year [1]. To ensure that future generations can benefit optimally from the increased robotic demand, intuitive educational programs are being defined. Accordingly, our university enables research for a facilitated entry into robotics through the self-developed Augmented reality Robot Number One (ARNO) platform [2].

2 Problem and task definition

Within the scope of this work, the combined system of desktop robotics and AR will be further developed, and its potential to broaden the access to robotics in an intuitive and easy-to-use way will be explored. The use-case resulting from this, will be validated through a series of tests to assess the system's utility the answer the research question of this work:

- What impact does ARNO's enhanced combination of desktop robotics and augmented reality have on the initial contact in robotics education?
- How do ARNO's extensions ease understanding robotics?

3 Materials and methods

Based on the task and research questions, industrial robotics, robot programming, robot control, direct and inverse kinematics, as well as augmented reality in robotics are considered core topics of this work. As this is an extension of an existing system, the software and hardware components of ARNO are analysed to record the starting conditions.

4 Practical implementation

To benefit form an intuitive and easy-to-use AR-robot-system two main features are added within this work. Firstly the discussed no-code virtual path planning and secondly, to achieve this, the discussed extension of the robotic control in form of inverse kinematics, to obtain the needed joint angles from the Cartesian user input for the final path execution.

5 Results

To assess the system's utility for entry into robotics and education, 15 individuals participated in the described test series, concluding with questionnaire responses. For all participants it was necessary to have successfully finished the robotics fundamentals course at our university to be able to identify potential benefits or drawbacks of the ARNO system. Overall, the combination of desktop robotics and augmented reality gathered positive feedback.

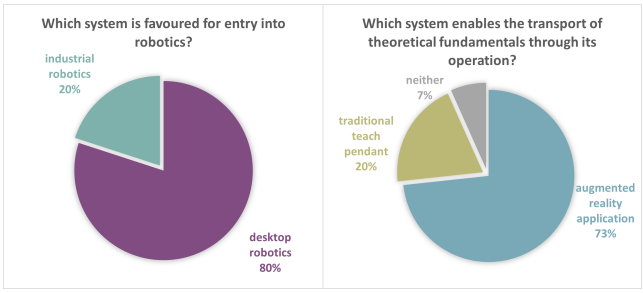


Fig. 1. results of ARNO as entry-level system.

References

1. International Federation of Robotics, 2023. World Robotics 2023 Report: Asia ahead of Europe and the Americas. Available at: https://ifr.org/ifr-press-releases/news/world-robotics-2023-report-asia-ahead-of-europe-and-the-americas (Accessed: 11.01.2024).
2. Orsolits, H., Rauh, S. F. & Estrada, J. G., 2022. Using mixed reality based digital twins for robotics education. In: 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), 17.-21. October 2022, Singapore. New York City, New York: IEEE, pp.56-59.

Accelerating Manual Pick-and-Place Operations with AR-Projected CAD Plans and AI-Assisted Object Recognition

Raphael Seliger, Matthias Micheler, Sebnem Gül-Ficici, and Ulrich Göhner

University of Applied Sciences Kempten, Bahnhofstraße 61, 87435 Kempten, Germany raphael.seliger@hs-kempten.de, matthias.micheler@hs-kempten.de, sebnem.guel-ficici@hs-kempten.de, ulrich.goehner@hs-kempten.de

Abstract. The manual assignment of objects to printed CAD plans by human operators is a labor-intensive and time-consuming task in manufacturing. Here, we present a system leveraging cutting-edge technologies to improve the efficiency of manual pick-and-place operations. Our system projects CAD plans onto a worktable, using deep learning and computer vision to identify objects based on their CAD representation. AR technology is then used to visualize the objects' precise positioning in these projected plans, enabling the human operator to interact seamlessly with them.

Keywords: Deep Learning · Computer Vision · Optical Character Recognition · Contour Matching · Instance Segmentation · Augmented Reality

1 Introduction and Motivation

Artificial intelligence (AI) has an increasing impact on today's workforce and is likely to permanently change the way we work. It holds the potential to even automate areas that have remained largely untouched by digitalization during the ongoing fourth industrial revolution [1]. One such practical industrial scenario is the manual assignment of objects in pick-and-place operations to their correct positioning in printed computer-aided design (CAD) plans. For this process, operators are required to visually identify the part number on an object to manually locate and assign it in the corresponding plan. However, this procedure is labor-intensive, prone to human error, and time-consuming. In the present work, we suggest a novel system to improve these manual pick-and-place operations without changing other manufacturing processes of the same chain.

2 Project Objectives

The primary focus of our work lies on the object recognition system, which consists of two key components to enable unique part identification. Firstly, it involves segmenting the contours of objects with simple geometric shapes using a deep learning system, which is then compared against stored contours within

a graph database. We chose the U-Net architecture [2] for segmentation due to its superior accuracy and speed when compared to other approaches for this specific use case. Secondly, an optical character recognition (OCR) system is employed to extract handwritten part numbers, which are also cross-referenced within the database. For this purpose, a you only look once (YOLO) model is used as the text detector [3], and a convolutional neural network (CNN) was trained with the MNIST dataset and expanded using additional training images showing handwritten digits on actual objects. Based on these two results, the object can be identified. Then, the system visualizes the object's positioning in the projected plan for the operator. The foundation for filling the graph database is done by extracting data from the associated CAD files.

Our projection system is designed to precisely scale and display CAD plans on the worktable. The worktable is also fully adjustable in height to account for different human operators. A linear unit is programmed to move the projector along with the adjusting table height. For this reason, the system requires robust calibration of the projector using a camera and planar patterns mounted physically and projected digitally onto the table. According to these patterns, the radial lens distortion is modeled [4].

3 Expected Project Outcomes

The anticipated outcome of this work is a modular software capable of automating the described tasks. This represents the first functional step towards a fully automated robotic system to replace manual pick-and-place operations. Further research should explore ways to enhance efficiency and usability within the system. For example, the functionality of the calibration camera could be expanded to include the recognition of already placed components. This could lead to a more streamlined pick-and-place process by tracking the work process and calculating data-driven optimizations. Additionally, developing an AI-based system for labeling the PDF or CAD plans could simplify the process of integrating new plans into the system, making the workflow more agile and adaptable in production.

References

1. Sarker, I.H.: AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. SN COMPUT.SCI. **3**, 158 (2022)
2. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention. MICCAI 2015. Lecture Notes in Computer Science, vol **9351**. Springer, Cham (2015)
3. Mondal, R., Malakar, S., Barney Smith, E.H. et al.: Handwritten English word recognition using a deep learning based object detection architecture. Multimed Tools Appl **81**, 975–1000 (2022)
4. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(11), 1330-1334 (2000)

Variety Engineering – A cybernetic concept with practical implications

Markus Schwaninger <markus.schwaninger@unisg.ch>
University of St. Gallen, Switzerland

Stefan Ch. Ott <stefan.ott@ost.ch>
University of Applied Sciences Eastern Switzerland

Management in systemic terms means to cope with complexity. Ross Ashby's Law of Requisite Variety shows the way, - maintaining the varieties of interacting systems in balance. To denote that process, we use the cybernetic concept of « Variety Engineering », which we also formalize. It refers to processes of mutual complexity amplification and attenuation by interacting agents. The purpose of this contribution is to elicit ways of coping with complexity by means of variety engineering. The abstract concepts are illustrated by examples from ecological, social and economic contexts. The cases in point will be used to demonstrate how complexity can be managed, in order to foster sustainability in these three dimensions.

Our research questions are:

1. What is Variety Engineering?
2. Why do we need it?
3. How does it work?

In cybernetic terms, dealing with complexity hinges on Ross Ashby's Law of Requisite Variety: «Only variety can absorb variety». In other words, to maintain a system under control, the control system needs a variety that is (at least) equivalent to the variety of the system controlled. *Variety* is a technical term for *complexity*, which stands for the property of a system to adopt many different states or behaviors. Variety then denotes the (potential or actual) set of states or modes of behavior of a system. A synonym for *variety* is *repertoire of behaviors*.

Ashby noticed that the variety of a system is not an intrinsic property of that system. "The observer and his powers of discrimination may have to be specified if the variety is to be well defined" (Ashby 1956:125). In the context of the management of complexity, different distinctions can apprehend variety, e.g.:

- as a set of components or elements of the system or its environment: V_c
- as a set of states of the system or its environment: V_s
- as a set of modes of behavior of its environment: V_b
- as a set of perturbations coming from the environment: V_p
- as a set of actions or responses of the systems: V_r

In an earlier treatise, variety engineering is conceived as "managing the complexity of a situation", in terms of the mutual adjustment of V_p and V_r (Espejo and Reyes, 2011:50f.). We revert to the more abstract concepts of *VE* and *VH* – eigen-variety and hetero-variety. Any one of the distinctions listed above can be used, if a given situation of variety engineering is analyzed. Two examples:

1. *VEc / VHc*, i.e. eigen- and hetero-variety, each of a set components.
2. *VEb / V*Hb, i.e. eigen- and hetero-variety, each of a set of (modes of) behaviors.

We are going to use these two variants in the following cases. A characteristic of these sets is that their properties can be represented by quantitative data (Ibidem).

Any agent – individual or collective – confronts complexities that are superior to what he or she can attend fully with their variety, i.e., their capacity of absorbing complexity. Variety engineering is the general strategy to offset this deficit.

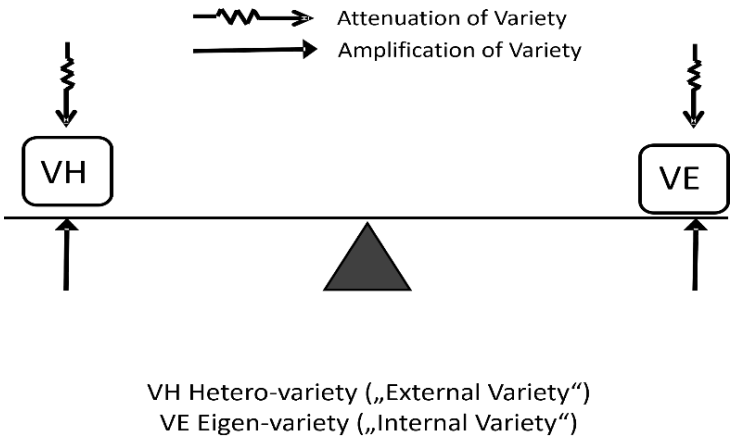


Figure 1: Levers of variety engineering –attenuation and amplification

«Variety Engineering» is a term coined by Stafford Beer (1979:39). It stands for the combination of amplification and attenuation of variety used to bring a system of interactions into balance. Amplification and attenuation are the actions or strategies by which the hetero- or eigen-varieties of a set or agent are changed, according to a quantitative measurement scale.

The primary modes of variety engineering consist in attenuating hetero-variety (also: external variety) and enhancing eigen-variety (also: internal variety), see Figure 1. Let us take, as two interacting agents, a complex market and a company. In this example, the company has two options:

a) To attenuate the complexity / variety of the market, e.g., by modeling, structuring or segmenting it, inventing new ways of clustering market objects, targeting its own activity to parts of that market only (e.g., segments of the market), or reframing the system-in-focus.
b) To amplify its own variety / complexity, e.g., by enhancing its market intelligence, adopting better information systems, increasing the sales force, strengthening its capabilities, etc. Also here, reframing is the most powerful device of variety engineering.

Both approaches, attenuation and amplification can be a blessing: they can lead to an equilibrium of varieties between competing or conflicting agents. At the same time, any one of these strategies can also lead to failure, if it provokes a crass imbalance.

Four cases on variety engineering will be presented, to illustrate how it works and how it can be used in practice to promote the sustainability of the economy and society. The first case is about ecological sustainability. It is followed by one case on each, social and economic sustainability, and finally a case in which the partial perspectives are integrated. Finally, the topic of residual variety is raised, and conclusions are drawn.

References

Ashby WR. 1956. *An Introduction to Cybernetics*. Chapman & Hall: London.

Beer S. 1979. *The Heart of Enterprise*. Wiley: Chichester.

Beer S. 1985. *Diagnosing the System for Organizations*. Wiley: Chichester.

Espejo, R., Reyes, A., 2011. Organizational Systems. Managing Complexity with the Viable System Model. Springer: Berlin.

Using System Archetypes to Explore Business Model Challenges for Digital Textile Microfactories

Meike Tilebein[1,2]

[1] University of Stuttgart, Institute for Diversity Studies in Engineering, Germany
[2] German Institutes of Textile and Fiber Research Denkendorf, Germany

Keywords: System Archetype, Business Model Innovation, Causal Loop Diagram, Fashion Industry, Digitalization

The purpose of this contribution is to further inform the discussion of new, digital technology-based business models in the fashion industry. The new digital technology we refer to is the Digital Textile Microfactory, a digitally networked end-to-end process from bodyscan to product. This new technology is said to be a potential game changer in the industry [1], using digitalization to improve both individualization and sustainability. However, this innovation and related new business models have not yet been much adopted by industry. We want to foster the discussion about new business models using this technology by providing insights from the perspective of systems archetypes in order to explore related risk and success mechanisms in advance.

Digital solutions for value creation networks and production using intelligent networking [2], taking into account the possibilities of flexible production, adaptable processes, customer-centric solutions, optimised logistics, use of data for new services and resource-saving circular economy that have proven to be beneficial to other industries can also be applied to the TCI [3].

Fully integrated small-scale production units bear a vast potential for use in different business models. They are suited also for urban production that, compared to conventional production in global supply chains, can be much faster with less environmental impact [4]. Among these new digital technologies Microfactories have recently been subject to research and development [5]. For the TCI, this is related to the concept of Digital Textile Microfactory (DTMF).

The basic principle of a DTMF is a digitally networked end-to-end digital design and production process for textile and clothing products. Such a Microfactory can cover a complete value creating chain comprising all design and

production steps from the customer to the ready-made product. Compared to conventional processes, production in DTMFs can be more sustainable [6].

In spite of the potential benefits, the industry is still reluctant to adopt DTMFs, due to perceived uncertainty regarding new business models. To reduce uncertainty, System Dynamics can be applied in two ways: with case-specific concrete modeling and simulation [7] or, in a first step, with qualitative causal loop diagrams (CLDs) that are based on system archetypes as a starting point [8].

In this contribution, we apply CLDs of system archetypes to the topic of DTMF-based business models. We describe different business models for DTMFs in the fashion industry and apply system archetypes in order to offer a wider systems perspective on potential dynamics and adversal effects.

References

1. Artschwager, A., Winkler, M., Brunner, L.: Textile Microfactory and Distributed Production – Digitalization as a Game Changer for Backshifting. (2022). https://www.ditf.de/en/index/more-information/interactive-whitepaper-on-textile-microfactory-and-distributed-production.html
2. Kiel, D., Müller, J.M., Arnold, C., Voigt, K.-I.: Sustainable Industrial Value Creation: Benefits and Challenges of Industry 4.0. International Journal of Innovation Management, 21(8), pp. 1–34 (2019)
3. Winkler, M., Moltenbrey, F., Tilebein, M.: Business Model Scenarios For Digital Textile Microfactories. In: Herberger, D., Hübner, M. (eds.) Proceedings of the Conference on Production Systems and Logistics, pp. 574-582. publish-Ing., Offenburg (2022). https://doi.org/10.15488/12141
4. Arfaiee, P., Winkler, M., Tilebein, M.: Smart, Sustainable and Socially Valuable: How Digital Textile Microfactories can Contribute to a Brighter Future. In: Deuse, J. (ed.) How can industrial management contribute to a brighter future? Schriftenreihe der Wissenschaftlichen Gesellschaft für Arbeits- und Betriebsorganisation (WGAB) e. V., vol. 2023, pp. 95–114. GITO mbH Verlag, Berlin (2023). https://doi.org/10.30844/WGAB_2023_6
5. Montes, J., Olleros, F.: Microfactories and the new economies of scale and scope. Journal of Manufacturing Technology Management 31(1), pp. 72–90 (2019)
6. Weiß, M., Winkler, M., Seibold, J., Grau, G.: Welchen Beitrag zur Nachhaltigkeit kann die Digitalisierung liefern? Ein Ansatz zur Bewertung der Digitalisierung in der Textilproduktion hinsichtlich ökologischer und ökonomischer Nachhaltigkeit. Industrie 4.0 Management, 39(2), pp. 25–28 (2023)
7. Davis, J.P., Eisenhardt, K.M., Bingham, C.B.: Developing theory through simulation methods. Academy of Management Review, 32(2), pp. 480–499 (2007)
8. Senge, P.: The Fifth Discipline: the art & practice of the learning organization. Doubleday Currence, New York (1990)

Organisational viability – challenges and approaches in today's Management

Franz Röösli[1] and Kerstin Pichel[2]

[1] Institute for organizational viability, Zurich University of applied Science, Switzerland
[2] Institute for organizational viability, Zurich University of applied Science, Switzerland

Abstract. The concept of organizational viability is highly topical in today's turbulent economy. We would like to revitalize the scientific discourse on this concept on the basis of three dimensions of contingency that we observe time and again in the management practice of viable organizations. We outline challenges and design approaches for positioning in these dimensions of contingency. Our work is based on theoretical research and practical case studies.

Keywords: viable organization, contingency dimensions, Management challenges and approaches.

The current research gap on the concept of organizational viability

. The concept of organizational viability has been known in cybernetics and systems theory for more than 40 years. In view of today's multidimensional challenges, viability is being updated. We see that organizations promote their viability by constantly balancing their position in three dimensions of contingency: Ambidexterity, coherence, adaptability.

Although all three dimensions of contingency are addressed and treated in systems theory [1] and cybernetics [2] and also play an important role in Beer's Viable System Model [3], current management practices for their design are rarely discussed in the context of the viability concept [4]. We venture a current examination of the concept and address existing management challenges and approaches from a theoretical perspective and using case studies.

Contingency dimension ambidexterity

. Organizational ambidexterity describes the ability of an organization to realize opposing organizational activities, such as exploration and exploitation, in a balanced and appropriate manner at both the organizational and individual levels [1]. Three approaches have proven successful in management practice for this challenging combination [5]: exploration and explotation is realized in the individual work process, in specific departments or over time. Our case studies show that organizations change their type of ambidexterity over a period of several years: from structurally organized ambidexterity to contextually and then sequentially designed ambidexterity.

Contingency dimension coherence

. Coherence refers to the formation of a whole [6]. Organizations must find a balance between centralization and autonomy to ensure both flexibility and alignment [1].

From a systemic perspective, the fractal is a solution to break the impasse between hierarchy and autonomy. A fractal - i.e. a structure that manifests itself in both large and smaller units - can be created by management by making key strategic directions, management principles or process flows characteristic of all organizational units [7].

Our case studies show companies that split their organization into two autonomous units as soon as they reach a certain size.

Contingency dimension adaptability

. Organisations must find a balance between change and stability to to enable both predictability and adaptability of plans and resources [8]. Three approaches at the individual and organizational level have proven to be important prerequisites for recognizing and implementing adaptation: Solution-oriented risk management [9], trial and error as a basis for decision-making, combining different perspectives on challenges and adaptation needs [10]. Our case studies illustrate the possibilities of hypothesis-based, collaborative decision-making in strategy development.

Conclusion

. Our conceptual considerations and case studies show the potential contribution of the three contingency dimensions to a viable organization. In this way, we aim to make an initial contribution to the necessary continuous, explicit and discursive interpretation of systems and the corresponding design of a viable organization.

References

1. Hoverstadt, P. (2022). The grammar of systems. From order to chaos & back. SCiO Publications.
2. Ashby, W. R. (1956). An introduction to cybernetics. London: Chapman & Hall.
3. Beer, S. (1984). The Viable System Model: Its Provenance, Development, Methodology and Pathology. *J. Oper. Res. S*oc., 35(1), 7–25.
4. Röösli, F. (2015). Initialisierung musterbrechender Managementinnovation: Eine interdisziplinäre Betrachtung. Lohmar – Köln: Eul Verlag.
5. O'Reilly III, C. A., & Tushman, M. L. (2013). Organizational ambidexterity: Past, present, and future. *Academy of management Perspectives*, *27*(4), 324-338.
6. Laloux. F. (2024). Reinventing Organizations. A guide to creating organizatios inspired by the next state of human consciousness. Herefordshire: Nelson Parker.
7. Hoverstadt, P. (2008). The fractal organization. Creating sustainable organizations with the Viable System Model. Chichester: Wiley.
8. Wolf, C., & Floyd, S. W. (2017). Strategic planning research: Toward a theory-driven agenda. Journal of management, 43(6), 1754-1788.
9. Dalziell, E. P., & McManus, S. T. (2004). Resilience, vulnerability, and adaptive capacity: implications for system performance.Strategy. International Forum for Engineering Decision Making (IFED), Stoos Switzerland, December 2004
10. Hodgkinson, G. P., & Clarke, I. (2007). Conceptual note: Exploring the cognitive significance of organizational strategizing: A dual-process framework and research agenda. Human Relations, 60(1), 243-255.

Interacting with the water cycle - towards an experimental paradigm

Martin Schaffernicht[1][0000-0002-3656-1059] and Stefan Groesser[2][0000-0003-3412-2711]

[1] Facultad de Economía y Negocios, Universidad de Talca, Av. Lircay SN, 3461334 Talca, Chile
martin@utalca.cl
[2] Bern University of Applied Sciences, Quellgasse 21, 2501 Biel, Switzerland

We address the challenge of poor public understanding of human-water interactions. Water is becoming increasingly scarce in a growing number of countries. But despite calls for improved associations between research and water management [1], the problem persisted and worsened, leading to propositions such as a new scientific discipline of socio-hydrology [2]. Scarcity is a relational phenomenon between demand and supply in the short- and mid-term. Most human activities require water, and growing demand for food, material goods and services increase, consequently, the water demand even more. Decreasing precipitations and increasing evapotranspiration due to rising temperatures diminish water supply in addition. The challenge of water scarcity involves producers and consumers, but also regulators, which implies politicians and voters. Changes to regulations, public and private policies, and the consumption/production patterns (partially influenced by regulations) presuppose political will, and therefore, the challenge must be sufficiently well understood by voters and decision-makers [3]. Yet, evidence suggests that the interaction between the water cycle and human activity is opaque for laypeople, and the ever-developing challenge of water scarcity is not well understood. Education and public information regarding the water cycle lacks the human elements and the interaction with them [4].

Laypeople's mental models of the dynamic relationships between human activities and the water cycle have a too narrow boundary and is static in nature. This connects to systems thinking flaws such as the "stock-and-flow error" [5] or "misperception of feedback" [6], that may be amenable to the increasingly problematic relationship between human activity and water [7]. Students underestimate the importance of plants [8] and the invisible parts of the water cycle. They showed a linear and static understanding of the water dynamics: they tend to miss out on cycles and slow processes [9]. Wheeler [10] reported similar results in a study of teachers' beliefs. In addition, misconceptions have also been found in college students [11]; for instance, the concrete existence of groundwater is overlooked or oversimplified [12]. Specific instruments have been developed to make groundwater more perceptible to students [13]. The United Nations engaged in "making the invisible visible" [14]. Adding to the poor understanding of dynamics in the water system, some indirect relationships between human activity and the water cycle are mostly absent from public discussion; namely, the food consumption patterns.

We argue that the inherent dynamic complexity of the systemic interactions between water and human activity is an important factor [15]. Based on literature on the water cycle and on the system dynamics methodology, we propose origins for these misrepresentations and misconceptions:

PC 1: Groundwater is not accounted for because it is invisible.

PC 2: The slowness of water infiltration towards groundwater is not accounted for.

PC 3: In trade-offs between economic interest (short-term benefit) and water footprint (long-term problem), producers prioritize the short-term, possibly because they underestimate the water problem (possibly because of the previous misconceptions PC1 and PC2).

PC 4: Consumers do not account for the water impact of their food consumption choices, possibly because the corresponding water footprint is not perceived, or its consequences underestimated (possibly because of the previous points).

There is currently a lack of data concerning how laypeople understand the problem and how they decide when cast in the role of regulator, producer, or consumer. Since agriculture is the main human direct water user, we focus on the production and consumption of food. We have developed a conceptual model containing the relevant elements of the water cycle: surface water and groundwater with their respective flows of recharge and drain (precipitation, evapotranspiration, infiltration, runoff, groundwater pumping for irrigation and human withdrawals (for agriculture). The model also includes the social roles of producer, consumer, and regulator,

including products with unit margin and unit water consumption. It also allows for diverse scenarios of precipitation reduction.

Currently we emphasize the producer, who can choose the production volume and set a pumping policy for irrigation. We have developed a system dynamics model running over 50 years, where the producer pursues profit maximization through sustainable production (with respect to water reserves). This allows to test PC 1 and PC 2 based on the recorded decisions and simulated values in laboratory experiments, where participants are randomly assigned to a control group with constant precipitation or a treatment group with decreasing precipitations. For PC 1, the question is if the visibility of groundwater data affects the participants' decisions. Referring to PC 2, the infiltration time into groundwater can be varied from 10 to 30 years. The model (structure and optimal decisions for the current version and scenarios) will be presented together with first data at the conference.

References

1. Jackson, R.B., et al., *Water in a Changing World.* Ecological Applications, 2001. **11**(4): p. 1027-1045.
2. Sivapalan, M., et al., *Socio-hydrology: Use-inspired water sustainability science for the Anthropocene.* Earth's Future, 2014. **2**(4): p. 225-230.
3. Fandel, C., D. Bresheards, and E. McMahon, *Implicit assumptions of conceptual diagrams in environmental science and best practices for their illustration.* Ecosphere, 2018. **9**(1): p. e02072.
4. Abbott, B.W., et al., *Human domination of the global water cycle absent from depictions and perceptions.* Nature Geoscience, 2019. **12**(7): p. 533-540.
5. Sterman, J., *Does formal system dynamics training improve people's understanding of accumulation?* System Dynamics Review, 2010. **26**(4): p. 316-334.
6. Sterman, J., *Misperceptions of Feedback in Dynamic Decision Making.* Organizational Behavior and Human Decision Processes, 1989. **43**(3): p. 301-335.
7. Pizarro, R., et al., *Climate Change and Overuse: Water Resource Challenges during Economic Growth in Coquimbo, Chile.* Sustainability, 2022. **14**(6).
8. Ryplova, R., J. Pokorny, and R. Ryplova, *Saving Water for the Future Via Increasing Plant Literacy of Pupils.* European Journal of Sustainable Development, 2020. **9**(3).
9. Ben-Zvi Assaraf, O., et al., *Cultural differences and students' spontaneous models of the water cycle: a case study of Jewish and Bedouin children in Israel.* Cultural Studies of Science Education, 2012. **7**(2): p. 451-477.
10. Wheeler, G. *Water Literacy: Educator Perspectives, Student Knowledge, and Possible Misconceptions*. in *International Symposium on Environmental Water Literacy*. 2012. Japan.
11. Schmid, S. and F.X. Bogner, *Is there more than the sewage plant? University freshmen's conceptions of the urban water cycle.* PLoS One, 2018. **13**(7): p. e0200928.
12. Daughton, C.G., *Ground Water Recharge and Chemical Contaminants: Challenges in Communicating the Connections and Collisions of Two Disparate Worlds.* Groundwater Monitoring & Remediation, 2004. **24**(2): p. 127-138.
13. Nurlaili, S. and D.R. Sari, *Computer assisted instruction (CAI)-based learning: instructional need for pupils' conceptual understanding of water cycle?* Journal of Physics: Conference Series, 2020. **1521**(4).
14. UN-Water. *World Water Day - Making the Invisible Visible*. UN-Water 2022 [cited 2022 10/4/2022]; Available from: https://www.worldwaterday.org/.
15. Özgün, O. and Y. Barlas, *Effects of systemic complexity factors on task difficulty in a stock management game.* System Dynamics Review, 2015. **31**(3): p. 115-146.

Systemic Thinking in IT Management of the Future: Where are the Benefits?

Extended Abstract

Walter Brenner[1][0000-0002-0657-4385], Ralf Schneider[2][0009-0002-2715-3480] and Barbara Brenner[3][0000-0002-1957-9302]

[1] Institute of Information Management, University of St. Gallen, Switzerland
walter.brenner@unisg.ch
[2] Allianz SE, Munich, Germany
ralf.schneider@allianz.com
[3] Institute of Information Management, University of St. Gallen, Switzerland
barbara.brenner@unisg.ch

1 Challenges for IT Management

The ongoing digitalization of the economy and society poses new challenges for IT management: (1) more and more applications and IT infrastructure are being transformed into the public cloud, (2) the use of data analytics is becoming a success factor for companies, (3) IT security has become a central component of IT management and (4) artificial intelligence will enable new business models and applications. In addition to these technical challenges, there are ecological requirements and demands to increase diversity in teams.

Widespread IT management models, such as ITIL [1] or COBIT [2], must be further developed or replaced by new management models. There is little discussion of fundamental questions regarding the design of the management system.

2 Aim of the Article and Structure

The paper will be divided into two central parts: (1) a transcript of a conversation with a Chief Information Officer (CIO) of one of the biggest insurance companies in the world and (2) an analysis of the conversation with the aim of working out which elements of systems thinking should be part of future IT management models and what benefits will arise from it. The CIO with whom the conversation will be conducted and who is coauthor of this paper is one of the few IT managers who has implemented concepts and methods of systems thinking in IT Management.

3 Key Points of Discussion in the Interview and Analysis

In the preparatory work for this article, various starting points have emerged that speak in favor of a stronger systemic orientation of future IT management. For example,

holistic, circular, and loop-oriented thinking and procedures, as called for by Ulrich [3, 4], cybernetic governance, as developed by Bleicher [5], and direct environmental contacts of IT managers, as contained in Beer's model [6, 7], were mentioned in preliminary discussions for this article as important design components of future IT management models. In addition, there is the human-centeredness of systemic thinking, which Ulrich already referred to in his definition of the company as a productive social system [3].

Against this backdrop, the following questions are the focus of the discussion with the already mentioned former Group CIO of a large insurance company: (1) What challenges does the CIO of a large insurance company face because of digitalization, new developments in information and communication technology, regulatory IT compliance and cyber risks? (2) Which concepts and methods of systemic thinking and acting can contribute to overcoming the challenges? (3) Which concepts and methods of systems thinking have been implemented within an insurance company to ensure an effective and efficient operational compliant and secure IT management (e.g. the most important systems thinking principle: everybody knows, that everybody knows, that everybody is knowing)? (4) What does holistic thinking mean when it comes to meeting ecological requirements and increasing diversity? (5) What does systemic thinking mean when dealing with employees? (6) What benefits could arise for companies and their IT management from the increased application of systemic thinking?

4 Limitations and Need for Future Research

The authors are aware of the limitations of their contribution and the unusual structure. Similar to a single case study, it is a conversation with one CIO that serves as a basis. Against this background, this article is able to demonstrate the fundamental usefulness of increased systemic thinking in IT management and thus make a significant contribution to the further development of IT management and provide a powerful impetus for future research.

References

1. ITIL 4 Foundation Homepage, https://www.itsmgroup.com/ch/trainings/itil/foundation, last accessed 2023/11/12
2. COBIT Homepage, https://www.isaca.org/resources/cobit, last accessed 2023/11/12
3. Ulrich, H.: Die Unternehmung als produktives soziales System. Paul Haupt, Bern/Stuttgart (1968)
4. Ulrich, H., Krieg, W.: Das St. Galler Management Modell, 1st edn. Paul Haupt, Bern/Stuttgart (1972)
5. Bleicher, K., Abegglen, C: Das Konzept Integriertes Management. 9th edn. Campus Verlag, Frankfurt/New York (2017)
6. Beer, S.: Brain of the firm - A development in management cybernetics. Herder and Herder, New York (1972)
7. Schwaninger, M.: The Evolution of Organizational Cybernetics. Scientiae Mathematicae Japonica, 13(2), 856 – 880 (2006)

Developing a Digital Twin for Cardiovascular Risk Assessment: A Simulation Model and Dashboard

Stefan N. Grösser[1][0000-0003-3412-2711] and Marco Hänni[2]

[1] Bern University of Applied Sciences, Industrial Engineering, 3012 Bern, Switzerland
[2] 3012 Bern, Switzerland
stefan.groesser@bfh.ch

Abstract. The article discusses the development of a prototype of a digital twin, consisting of an initial cybernetical simulation model and a dashboard, to forecast cardiovascular risk based on an individual's lifestyle and dietary habits. The aim is to provide information on how lifestyle modifications can reduce cardiovascular risk.

Keywords: Digital Twin, Simulation Methodology, Cardiovascular Risk.

1 Introduction and Literature

As time passes, human lifestyles undergo changes, leading to increased sedentary behavior and decreased physical activity. This, along with the consumption of energy-dense foods, has resulted in a rise in obesity rates and associated health issues, such as diabetes, heart disease, and stroke. Unhealthy lifestyles, combined with inadequate nutrition and other factors, contribute significantly to the development of noncommunicable diseases (NDCs), which are commonly linked to unhealthy behavior. The four main types of noncommunicable diseases are cardiovascular diseases, cancers, diabetes, and chronic respiratory diseases. Cardiovascular disease (CVD) is the leading cause of death globally, accounting for a staggering 18.6 million fatalities, with ischemic heart disease being the primary contributor, responsible for 9.44 million deaths. Ischemic stroke was the second-highest cause of death within the CVD category.

The INTERHEART [2] study conducted in 2004 was a global investigation of heart attacks that included 15, 152 cases and 14,820 controls in 52 countries. The study analyzed the impact of nine risk factors on the occurrence of myocardial infarctions and found that these factors accounted for 90% of the population attributable risk in men and 94% in women. PAR is a measure of the incidence that would be eliminated if exposure to the disease was eliminated. A 2020 study focused on the prevalence of cardiovascular disease in relation to behavioral and metabolic factors, as well as socioeconomic and psychosocial factors, in 21 communities over a period of 9.5 years. The study also examined the impact of factors such as solid fuel use for cooking and ambient matter pollution on health outcomes.

Wang et al. [3] investigated 193,846 individuals across 20 Chinese regions between 2011 and 2012, revealing that obesity was a surprising risk factor for cardiovascular disease. Hypertension, chronic kidney disease, diabetes, physical inactivity, depression,

poor diet, low education, smoking, and inadequate sleep duration were also identified as significant risk factors.

We identified current applications and simulation models for human digital twins (HDT) that perform well in calculating cardiovascular risk. However, precision medicine aims to not only calculate risk, but also simulate lifestyle interventions with HDT. Sanchez et al. [1] study suggests that a model that considers causality is necessary for achieving the goal of precision medicine. Currently, there is no such solution in the market.

2 Contribution

The development of models that incorporate personalized medicine, lifestyle adjustments, and environmental factors to calculate cardiovascular risk is crucial. By doing so, the global burden of CVD can be reduced through the application of HDT. Our team undertook this endeavor and modeled the physiological aspects of the matter, which served as the basis for the development. To calculate cardiovascular risk, we chose the Framingham risk calculation with the non-laboratory formula, which, together with body-mass-index (BMI), is the top indicator. The formula includes six variables, with a focus on modeling the BMI due to the influence of lifestyle through diet and exercise. Existing models are already able to simulate body weight in a relatively valid way, but they are oversimplistic in terms of real-time inclusion of decision data other than the initial weight and the calorie reduction. For a classic cardiovascular disease case, we examined three policies: complete lifestyle change, calorie restriction, and increasing physical activity without changing diet. The results showed that the most effective policy was the complete lifestyle change, followed by calorie restriction, and then increasing physical activity without changing diet. The study also showed that diet was the second most impactful factor, followed by exercise.

References

1. Sanchez P, Voisey JP, Xia T, Watson HI, O'Neil AQ, Tsaftaris SA: Causal machine learning for healthcare and precision medicine. R. Soc. Open Sci.**9**: 220638 (2022)
2. Yusuf, S. *et al.:* Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the Interheart study): case-control study', **364**(9438):937-52. doi: 10.1016/S0140-6736(04)17018-9 (2004)
3. Wang, T. *et al.*: Age-specific modifiable risk factor profiles for cardiovascular disease and all-cause mortality: a nationwide, population-based, prospective cohort study', *Lancet Reg. Health - West. Pac.* **17**, p. 100277, doi: 10.1016/j.lanwpc.2021.100277 (2021)

The Human-Centered AI-DATA Model for Digital Customer Journeys in E-Commerce

Heidrun Mühle[1][0009−0005−8111−8899], Oliver Krauss[2][0000−0002−8136−2606], and Andreas Stöckl[2][0000−0003−1646−0514]

[1] 506 Data & Performance GmbH, Linz, 4020, Austria heidrun.muehle@506.ai
[2] University of Applied Sciences Upper Austria, Hagenberg, 4232, Austria
{firstname.lastname}@fh-hagenberg.at

Introduction In an era characterized by rapid technological advancements and increasingly complex socio-technical systems, especially regarding Artificial Intelligence (AI) based E-Commerce Optimization, the demand for a comprehensive method to resolve problems, which could result from the lack of trust in AI-Software-Systems, has never been more significant.

Systems thinking offers a robust framework for understanding the intricate relationships between various components of a system, making it particularly relevant in today's digital and AI-based software landscape. Concurrently, the E-Commerce sector is undergoing a transformation, driven by advancements in AI and data analytics grounded in legal changes regarding the General Data Protection Regulation (GDPR) and the upcoming EU AI Act [2].

We explore the synergies between AI-based E-Commerce and Systems Thinking by introducing the AI-DATA model, a human-centered Customer Journey Optimization approach in the context of E-Commerce marketing strategy based on Trustworthy AI. AI-DATA is a phase model consisting of Awareness, Interest, Desire, Action, Trust and Again, and is an extension of the AIDA model [4], which also considers the customer path to contain activities after a transaction has finished, and a cycle of repeat customers.

Methodological Framework We use the AI-DATA model in a multi-faceted methodology that combines process mining techniques with Large Language Models (LLMs) to capture, analyze, and optimize various phases of the Onsite Customer Journey within Website Analytics. The model uses First Party Data Tracking to gather strategic tracking events, offering a granular view of customer interactions from pre-sales to after-sales. This approach is inherently multidisciplinary, integrating elements from system thinking, data science and behavioral marketing to create a comprehensive framework [3].

We propose an optimal *happy path* guiding users towards a streamlined and satisfying experience identified via Process mining [5]. We also identify *moments of truth* which causes a user to enter the next phase in the AI-DATA model, and *decision points* at which user churn happens. At these points, marketing actions can be taken to move the user to the next phase of the model or reduce churn. The effective arrangement of content by each Journey Phase of the AI-DATA-Model, enabled by this approach, potentially shortens the Customer Journey

and provides customers with precisely the information they need at the moment when they need it. This strategy of timely relevant content placement carries immense potential to sway the customer's decision to continue their journey. The AI-DATA model, characterized by its human-centric design and integration of online marketing consulting insights, tackles the challenges of personalization in E-Commerce marketing strategy creation. Importantly, it acknowledges and addresses specific cognitive biases, considering critical human factors that frequently influence customer behavior and decision-making processes.

Furthermore, the AI-DATA model integrates techniques from behavioral economics to gain a deeper understanding of the customer journey. By analyzing behavioral patterns and cognitive biases, the model can provide accurate insights into customer decision-making processes [1]. This allows businesses to effectively tailor their marketing strategies and create personalized offers based on the specific needs and preferences of their customers.

Conclusion and Future Directions We present the AI-DATA model and methodology which offers a structured AI-based approach to enhancing customer engagement and conversion rates in E-Commerce mixed with process mining to optimize websites. It identifies *moments of truth*, *decision points* and a *happy path* in customer journeys, thereby enabling businesses to place timely and relevant content effectively. Moreover, the model's focus on cognitive biases and human factors makes a significant contribution to the development of Trustworthy Marketing AI.

In a case study, the AI-DATA model will be implemented at an Online Shop for Health Supplements, a large e-commerce company in the dietary supplements sector. Through the implementation of the model, it is our aim to enable companies to increase their conversion rates. Via a socio-technical feedback loop, we enable the adaptation of website content aligned with the specific interests of the website users.

References

1. Adkisson, R.V.: Nudge: Improving decisions about health, wealth and happiness. The Social Science Journal **45**(4), 700–701 (2008). https://doi.org/10.1016/j.soscij.2008.09.003
2. European Commission: Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts (Nov 2023),
3. Mayrhuber, E., Krauss, O., Hanreich, M., Stöckl, A.: Towards an Ontology and Process Mining-based System for Targeted E-Commerce Marketing Strategy Suggestions. In: ICECCME Proc. pp. 1–6. IEEE. https://doi.org/10.1109/ICECCME57830.2023.10252333
4. Strong, E.: The Psychology of Selling and Advertising. McGraw-Hill book Company, Incorporated (1925)
5. Terragni, A., Hassani, M.: Analyzing customer journey with process mining: From discovery to recommendations. In: 2018 IEEE FiCloud (2018). https://doi.org/10.1109/FiCloud.2018.00040

Using LLMs and Websearch in Order to Perform Fact Checking on Texts Generated by LLMs

Simone Sandler[1 [0000−0001−8919−7804]], Oliver Krauss[1 [0000−0002−8136−2606]], and Andreas Stöckl[1 [0000−0003−1646−0514]]

[1] University of Applied Sciences Upper Austria, Campus Hagenberg, Softwarepark 11, 4232 Hagenberg, Austria
{simone.sandler, oliver.krauss, andreas.stoeckl}@fh-hagenberg.at
[2] Advanced Information Systems and Technology
[3] Digital Media Department

Introduction This paper targets the detection of misinformation in GPT3[4] generated texts and the FEVER[7] dataset, using Large Language Models (LLM) and Google search. Given the uncertainty associated with LLM-produced text, a requisite arises for a fact-checking system geared towards non-human-written content under varied contexts. In general, there are two different approaches to perform fact checking: manual checking, favored by entities like politifact[3], and automated or semi-automated checking, like ProoFVer[5]. Our research delves into fully automated checking of LLM-produced texts.

Background Large Language Models (LLMs) like GPT3, BLOOM, and PaLM, trained on diverse data, are especially well suited for unsupervised learning tasks [6]. These LLMs primarily utilize attention-based transformers[8], allowing for global interdependencies between inputs and outputs. They perform tasks through few-shot-learning (with some examples), one-shot-learning (with one example), or zero-shot-learning (without examples)[4]. Some LLMs, such as ChatGPT[1], make use of human feedback. The LangChain tool, which incorporates GPT3 with Google search, employs a technique called ReAct to blend internal (LLM) and external (Wikipedia, Search Results, etc.) knowledge. ReAct (synergizing reasoning and acting) evolves upon existing chain-of-thought methods through a focus on reasoned action[9].

Methods Our fact checking approach is done in two parts. During the first part, claims made in the text are extracted. In the second part, these claims are verified. Besides claims in the text, we decided to also check abbreviations contained in the text. This is done because multiple terms match the same abbreviation, often misleading a language model. We developed prompts to perform both, the extraction and the validation of claims and abbreviations in texts. We created a text validation pipeline that combined our methods and outputs a truthfulness score for a given text.

Results We tested the whole pipeline using LLM-generated texts and reached an accuracy of 0.79. In order to compare our work with other fact-checking approaches, we also tested the validation part of our method with the FEVER

dataset, which already contains extracted claims. The dataset also contains evidences that can be used to evaluate the claims, which we didn't use because we wanted a method that works without prior evidence collection. We were able to reach an accuracy of 0.78 on the FEVER data. According to the Fact Verification on FEVER leaderboard [2] the highest accuracy on FEVER data is achieved by ProoFVer [5] with an accuracy of 0.79. Our accuracy, with 0.78, is slightly lower, but it should be considered that our method is not using any evidences contained in the dataset.

Conclusion and Outlook We present a fact checking method by extracting claims and abbreviations and verifying them individually from texts generated by LLMs. Our method is shown to work with 0.78 accuracy on LLM generated texts, and with 0.78 accuracy on the FEVER dataset. Our method does not require any kind of preprocessing, and can work well in any context where the right answer can be found via web search. This is an improvement from existing comparable methods, where the correct information has to be stored in a knowledge graph or on Wikipedia. Unfortunately, our method is prone to fake news and misinformation information found on the internet.

In the future, we would like to create tests that check how our method reacts to misinformation contained in web search results and develop methods against such cases.

References

1. Introducing ChatGPT (Apr 2023), https://openai.com/blog/chatgpt, [Online; accessed 12. Apr. 2023]
2. Papers with Code - FEVER Benchmark (Fact Verification) (Apr 2023), https://paperswithcode.com/sota/fact-verification-on-fever, [Online; accessed 4. Apr. 2023]
3. PolitiFact (Mar 2023), https://www.politifact.com, [Online; accessed 30. Mar. 2023]
4. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.: Language models are few-shot learners. CoRR **abs/2005.14165** (2020), https://arxiv.org/abs/2005.14165
5. Krishna, A., Riedel, S., Vlachos, A.: Proofver: Natural logic theorem proving for fact verification. CoRR **abs/2108.11357** (2021), https://arxiv.org/abs/2108.11357
6. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al.: Language models are unsupervised multitask learners. OpenAI blog **1**(8), 9 (2019)
7. Thorne, J., Vlachos, A., Christodoulopoulos, C., Mittal, A.: FEVER: a large-scale dataset for fact extraction and VERification. In: NAACL-HLT (2018)
8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
9. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., Cao, Y.: React: Synergizing reasoning and acting in language models (2023)

Data Based Prediction of the Duration of the Postoperative Stay of Patients

Julia Vetter[1], Marina Strobl[1], Gerhard Halmerbauer[2], Tilman Königswieser[3], and Stephan M. Winkler[1,4]

[1] School of Informatics, Communications and Media
University of Applied Sciences Upper Austria, Hagenberg, Austria

[2] School of Business and Management
University of Applied Sciences Upper Austria, Steyr, Austria

[3] Salzkammergut-Klinikum, Oberösterreichische Gesundheitsholding GmbH

[4] Department of Computer Science
Johannes Kepler University; Linz, Austria

Efficient bed management is very important for hospitals. Current bed planning software primarily allows to identify free and occupied beds, but does not include any prediction of future bed occupancies. In this study, our aim is to identify models that predict the postoperative length of stay of patients.

Since 2015, FH Upper Austria and Oberösterreichische Gesundheitsholding (OÖG) have joined forces and develop the platform *LeiVMed* for the monitoring and analysis of data of treatments of patients. In the current research context, our main goal is to identify models that are able to predict the length of stay of patients [1] after surgery for improved bed and resource planning using machine learning (ML). Our data base consists of more than 91,000 surgical procedures storing information about performed services and treatments of 27 different medical case classes (extracapsular cataract extraction, appendectomy, caesarean section, etc.).

In general, machine learning is understood as that branch of computer science that is dedicated to the development of methods for learning knowledge and models from given data. In *supervised machine learning* the modeling algorithm is given a set of samples with input features as well as target / output values - i.e., the data are "labeled". The goal then is to find a model (or a set of models) that is (are) able to map the input values to the designated target variable(s).

In this research we analyze the use of several different ML methods; white box modeling (which produces models whose structure is not hidden, but can be analyzed in detail) is used as well as black box ML techniques (methods that produce models which are functions of the inputs and produce outputs, where the internal functioning of the model is either hidden or too complicated to be analyzed). For white box modeling we use *symbolic regression*, a method for inducing mathematical expressions on data. The key feature of this technique is that the object of search is a symbolic description of a model, whose structure

The work described in this paper was supported by LeiVMed, a joint research project of FH OÖ and OÖG.

is not pre-defined. This is in sharp contrast with other methods of nonlinear regression, where a specific model is assumed and the parameters / coefficients are optimized during the training process. Using a set of basic functions we apply *genetic programming* (GP, [2]) as search technique that evolves symbolic regression models as combinations of basic functions through an evolutionary process. Multi-criterial optimization can be used to guide the search process towards compacter, more concise models. We compare the so achieved results (especially accuracies and variable impact analyses) with those achieved using the following black-box modeling methods: *Artificial neural networks* (ANNs) are networks of simple processing elements, which are called neurons and are in most cases structured in layers. ANNs have been used in computer science for several decades, in recent years they have gained popularity massively due to the development of deep learning algorithms [3]. *Random forests* are ensembles of decision trees, each created on a set of randomly chosen samples and features from the available training data basis. The most frequently used algorithm for inducing random forests combines bagging and random feature selection [4]. These methods have in common that a basic model structure is assumed and then parameters are optimized during the training phase. Additionally, the structure of the model does not give information about the structure of the analyzed system; the models are used as so-called *black box models*.

We used data taken from over 91,000 pre- and postoperative routine records from hospital databases that were classified into eight different classes based on length of stay: 1d (day), 2d, 3d, 4d, 5d, 6d, 1 week to 1 month, and >1 month. Prediction models were generated using the above mentioned ML algorithms; input features include demographic information such as age, sex, and size, risk factors such as smoking or diabetes, and information about the stay (surgery, number of services, etc.), using either only the data up to the day of surgery or additional postoperative data from one to seven days after surgery.

The results show that the prediction of the length of stay works better in the early phases than a few days after the operation. The decisive factor here is the increasing variance in the number of services as well as the continuously shrinking amount of data available for modeling. The models of case categories with a high proportion of acute cases (e.g., appendix) can still be expanded, but these results show the feasibility of predicting the length of stay after surgery.

References

1. Stone, K., Zwiggelaar, R., Jones, P., Parthalain, N.M.: A systematic review of the prediction of hospital length of stay: Towards a unified framework. PLOS Digital Health **1** (2022) 1–38
2. Affenzeller, M., Winkler, S., Wagner, S., Beham, A.: Genetic Algorithms and Genetic Programming - Modern Concepts and Practical Applications. Volume 6 of Numerical Insights. CRC Press, Chapman & Hall (2009)
3. Goodfellow, I., und Aaron Courville, Y.B.: Deep Learning (Adaptive Computation and Machine Learning). MIT Press, Cambridge (USA) (2016)
4. Breiman, L.: Random forests. Machine Learning **45** (2001) 5–32

Building Spanish Trustworthy Question-Answer Datasets for Suicide Information*

Pablo Ascorbe[1], María S. Campos[2], César Domínguez[1], Jónathan Heras[1], Magdalena Pérez[3], and Ana Rosa Terroba-Reinares[1,4]

[1] Departamento de Matemáticas y Computación, Universidad de La Rioja, Spain
{pablo.ascorbe, cesar.dominguez, jonathan.heras}@unirioja.es
[2] Unidad de Salud Mental Espartero, Logroño, La Rioja
mscampos@riojasalud.es
[3] Teléfono de la Esperanza
magdalenaperez@telefonodelaesperanza.org
[4] Fundación Rioja Salud
arterroba@riojasalud.es

Suicide is a public health problem since worldwide more than 800,000 suicides are estimated to occur every year, one every 40 seconds, figures of epidemic proportions [4]. Such is the scale of the problem that specialists tasked with answering questions related to it are overwhelmed and overburdened. Moreover, much of the information that can be found about suicide on the Internet could be more harmful than helpful. These problems could be tackled by means of automatic Question-Answering (Q&A) systems [3]; however, it is necessary to have trustworthy corpora that help to validate, train, and guide the construction of such systems [2]. Since, up to the best of the authors knowledge, there is not a Spanish Q&A corpus for suicide information, the aim of this work is to create such a corpora for suicide information. In particular, we have considered three levels of quality [1]: *bronze-standard*, when the entire data has been generated automatically and has little processing; *silver-standard*, when starting from a bronze corpora, a processing stage is applied to refine the data followed by an annotation and validation stage conducted by experts; and, *gold-standard*, when the corpus has been manually generated and validated by experts.

The starting point to obtain the corpora is a set of trustworthy Spanish documents provided by suicide experts (in our case, it is composed of a total of 151 documents). Then, a bronze-standard Q&A corpora was built using three large language models, two in Spanish (bertin-gpt-j-6B-alpaca and Llama-2-7b-ft-instruct-es) and one in English (t5-base-squad-qag), all of them freely available at Hugging Face platform. We split the documents into chunks, and for each chunk, we asked the language models to generate a Q&A pair. Using this procedure, a total of 22,920 Q&A pairs were obtained, but many generated questions were incomplete, repeated or contained essentially the same information among them. Therefore, some filters were applied to increase the quality of

* This work was partially supported by Grant PID2020-115225RB-I00 funded by MCIN/AEI/ 10.13039/501100011033, and by funds for the 2023 strategies of the Spanish Ministry of Health, which were approved in the CISNS on June 23, 2023, to support the implementation of the Mental Health Action Plan.

the generated corpus. A first filter was basic data processing operations, such as the elimination of empty, duplicated or incomplete pairs. Then, we trained a deep learning classification model (named bertin-roberta-base-spanish-spanish-suicide-intent, again freely available at Hugging Face) using a suicidal behaviour dataset to determine whether a question-answer pair contains information about suicide. This model allowed us to filter out not suicide related pairs. Finally, we removed the pairs that were semantically similar by using an embedding and the cosine distance. After all these steps, we obtained the final version of our bronze-standard corpora, leaving us with 4,901 Q&A pairs.

From that bronze corpora, a manual filtering was conducted by non-experts to eliminate Q&A uninteresting pairs; i.e., they talked about suicide but seemed not to be useful or were too ambiguous. This left 484 pairs to be evaluated by experts. In order to perform such an evaluation, a web application was developed for the validation of the corpus by a group of psychologists and psychiatrists, allowing them to update or remove the pairs that did not pass their filter; thus obtaining the silver-standard corpus with 380 Q&A pairs.

In order to evaluate the models performance to generate interesting Q&A pairs, we describe the number of pairs obtained by each model in each step of the process. As a starting point, 7,806 were generated from Bertin, 4,558 from Llama-2, and 10,557 from t5-base-squad. After automatic filters, the number of pairs were reduced to 760 from Bertin, 1,166 from Llama-2, and 2,977 from t5-base-squad. After manual filtering by a non-expert, we had 337 from Bertin, 55 from Llama-2, and 92 from t5-base-squad. Finally, the silver corpus had 272 from Bertin, 45 from Llama-2, and 63 from t5-base-squad. We can observe that the Bertin model had the best performance.

In addition, a gold-standard corpus directly extracted from the FAQs sections contained in some of the documents provided by experts (i.e. they are not automatically generated) has been also provided. This dataset consists of 118 Q&A items.

The built Q&A corpus and models described previously are freely available at `https://huggingface.co/PrevenIA`. These corpus are the first step towards the validation, training, and construction of automatic Q&A systems that provide information about suicide.

References

1. Casola, S., Lavelli, A., Saggion, H.: Creating a silver standard for patent simplification. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1045–1055 (2023)
2. Das, B., Nirmala, S.J.: Improving healthcare question answering system by identifying suitable answers. In: 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon). pp. 1–6. IEEE (2022)
3. Rogers, A., Gardner, M., Augenstein, I.: QA dataset explosion: A taxonomy of NLP resources for question answering and reading comprehension. ACM Computing Surveys **55**(10), 1–45 (2023)
4. WHO: Suicide worldwide in 2019: global health estimates (2021)

Personalized ML-Assisted Respiratory Muscle Training for Patients with Paraplegia *

Alfredo Lopez[1], Clara Schachner[1], Shahzaib Khan[1,2], Tabarka Rajab[1,2], and Manuela Geiß[1]

[1] Software Competence Center Hagenberg GmbH
[2] International Center for Chemical and Biological Sciences, University of Karachi

Abstract. In this study, we integrate ML algorithms to enhance the new mouth-controlled computer mouse IntegraMouse Air. This device is designed to empower paraplegics, facilitating self-directed digital work, study and communication. Our approach incorporates a ML-based respiratory training component, utilizing both unsupervised and supervised learning methods. We seek to offer patients a personalized training experience, providing guidance and tailored suggestions for optimal training levels, based on measurements, meta-data and training history.

Keywords: Paraplegia · ML-assisted respiratory muscle training

1 Introduction

In 2019, there were 0.9 million incident and 20.6 million prevalent cases of spinal cord injury [1]. A serious issue for those affected is compromised respiratory muscle function, often leading to complications such as pneumonia, the primary cause of death in patients with complete paraplegia [2]. LIFEtool Solutions GmbH [3] is an organization focused on self-driven life for people with disabilities. They developed the mouth-controlled computer mouse IntegraMouse, which enables paraplegics to operate a computer autonomously. However, there is a pressing need to improve the mouse with respiratory muscle training capabilities, as existing respiratory training devices like the Aerofit [4] require caregiver assistance.

2 Methodology

Data has been collected with a prototype of the new IntegraMouse AIR, which is equipped with ten holes of different size, controlling the pressure resistance and therefore training intensity while breathing. Initially, a dataset composed of 40 healthy individuals and 3 paraplegic patients was acquired. Each participant underwent two test series: one involving sequential breathing exercises at each hole

* This work has been supported by the project 'IntegraMouse AIR' (FFG Project No. 891253) as well as by BMK, BMAW, and the State of Upper Austria in the frame of the SCCH competence center INTEGRATE (FFG grant no. 892418) as part of the FFG COMET Competence Centers for Excellent Technologies Program.

and one more extended session with the user selecting the hole size, simulating a real-world breathing training sequence. For each participant the breathing pressure time-series were acquired as well as meta-data (sex, age, fitness level, etc.), blood pressure, maximum inspiratory pressure (PImax) [5] and Borg values (rating scale for subjectively perceived exhaustion) [6]. The left plot in Fig. 1 shows the pressure curves for one participant in the sequential setting. To establish the optimal hole size (OHS) we will combine cluster analysis and expert medical input to define ground truth labels that will subsequently be used to develop supervised ML approaches, like simple neural networks, XGBoost, and Random Forests. We then predict the OHS for a specific person based on a subset of data, e.g. meta-data and one PIMax measurement with corresponding Borg value. The right plot of Fig. 1 shows input features extracted for one participant.

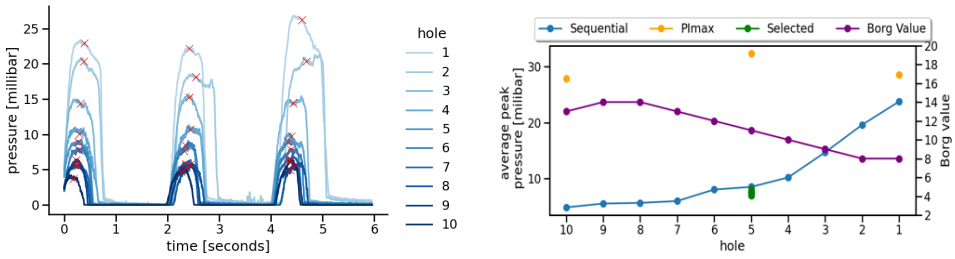


Fig. 1: (left) Pressure curves for one participant in the sequential setting with associated peak values (red crosses); (right) Blue: avg. peak pressure for 10 sequential holes; Orange: PImax at the beginning, middle, and end of the session; Green: avg. peak pressure at the selected hole; Purple: Borg values (right axis).

3 Discussion and Outlook

In the future we will incorporate additional features for prediction, such as optimal training duration and ideal number of repetitions. Furthermore, we aim to integrate patients' training history into our model, enabling continuous feedback during the training process.

References

1. Ding, W., et al. (2022). Spinal cord injury: The global incidence, prevalence, and disability from the global burden of disease study 2019. Spine, 47(21), 1532.
2. DeVivo, M. J., Kartus, et al. (1989). Cause of death for patients with spinal cord injuries. Archives of internal medicine, 149(8), 1761-1766.
3. Lifetool Homepage, www.lifetool.at. Last accessed 13 Nov 2023
4. Aerofit Homepage, https://www.airofit.com/. Last accessed 14 Nov 2023
5. Pessoa, I., et al. (2014). Reference values for maximal inspiratory pressure: a systematic review. Canadian respiratory journal, 21, 43-50.
6. Borg, G. (1970). Perceived exertion as an indicator of somatic stress. Scandinavian journal of rehabilitation medicine.

Modelling the risk of overweight and obesity

J. Ignacio Hidalgo[1,2], Elisabeth Mayrhuber[3,4], Stephan M. Winkler[3], Daniel Parra[1], J. Manuel Velasco[1], José J. Zamorano-León[5], and Oscar Garnica[1]

[1] Faculty of Informatics; Universidad Complutense Madrid, Spain

[2] Instituto de Tecnología del Conocimiento, Madrid, Spain

[3] Bioinformatics Research Group; University of Applied Sciences Upper Austria, Hagenberg, Austria

[4] Faculty of Informatics, plf.doc; Technical University Vienna, Austria

[5] Faculty of Medicine; Universidad Complutense Madrid, Spain

Obesity is an excessive accumulation of body fat that presents a health risk. A widely used measure to define obesity is the Body Mass Index (BMI), which is calculated by dividing the weight of a person in kilograms by the square of their height in meters (kg/m^2). A person is considered overweight if their BMI is higher than 25 and obese if their BMI exceeds 30. Obesity is considered a condition influenced by genetic, environmental, and behavioral factors, and it is associated with numerous health risks, including diabetes, heart disease, and certain types of cancer.

Modeling the risk of being overweight or obese is crucial at the individual and public health levels. With good prediction models, individuals at risk can be identified to prepare early intervention actions to prevent the development of obesity. Medical staff could propose habits and changes in diet, physical activity, and behavior to avoid obesity and its complications. Moreover, predicting the risk of being overweight can improve individual health outcomes, reduce healthcare costs, and inform effective public health strategies [1].

From January 2018 to June 2022, the GenObIA (Genetic and Environmental Obesogenic Factors Interaction for Artificial Intelligence Algorithms Development) project, an ambitious research initiative focused on combating obesity through advanced technology and genetics, collected an extensive dataset that combines sociodemographic, environmental, behavioral, and genetic data. The project aims to develop comprehensive strategies to predict, prevent, and manage obesity and its associated health complications, with particular interest in diabetes. The GenObIA consortium includes ten research groups from public hospitals and universities in Madrid and primary care teams from various areas of Madrid. The associated groups include the Consorcio Regional de Transportes de Madrid, the General Directorate of Youth and Sports of the Community of Madrid, the Spanish Paralympic Committee, and other organizations.

This paper aims to create predictive models, using artificial intelligence, to identify individuals at risk of developing overweight or obesity. The models will

Work supported by The Spanish Ministerio de Innovación Ciencia y Universidad -grants PID2021-125549OB-I00, PDC2022-133429-I00 and RTI2018-095180-B-I00, and Comunidad de Madrid-FEDER Funds (GenObIA-CM) S2017/BMD-3773.

use the sociodemographic, environmental, behavioral data of individuals and genetic analysis. In this sense, we aim to reveal how individual genetic features influence factors like food metabolism and the benefits of specific physical exercises. This genetic analysis is expected to provide a more comprehensive understanding of obesity and its related health risks.

In order to achieve the aforementioned goals, several machine learning (ML) methods will be utilized. These methods include interpretable ML models, like symbolic regression, as well as closed-box ML models, like artificial neural networks (ANNs). The results generated with different ML methods are evaluated based on the model accuracy and the explainability/interpretability of the resulting models.

Interpretable models, also called glass box models, are ML models whose structure can be interpreted and understood by humans and therefore the results generated with such models are explainable. As representatives of this group of models, symbolic regression by genetic programming (GP, [2]) is used in this research. Symbolic regression aims to find a symbolic description of a model by using GP to find suited combinations of operations, variables and constants through an evolutionary process.

As representatives for closed box models, ANNs will be used. ANNs are networks constisting of neurons, which represent simple processing units. These neurons are typically organized, especially in deep learning [3]. The structure of such models can not be interpreted by humans, and therefore posthoc explainability methods like SHAP [4] and LIME [5] are used in order to explain the results generated by such models.

In addition to comparing glass box and closed box models, clustering analysis [6] is conducted based on different features, such as lifestyle factors and health parameters; we will analyze the variance of obesity levels in the resulting clusters.

With the comparison of glass box and closed box models combined with an clustering analysis, health parameters and SNPs, that significantly contribute to obesity can be identified.

References

1. Fruh, S.M.: Obesity: Risk factors, complications, and strategies for sustainable long-term weight management. Journal of the American association of nurse practitioners **29** (2017) S3–S14
2. Goodfellow, I., und Aaron Courville, Y.B.: Deep Learning (Adaptive Computation and Machine Learning). MIT Press, Cambridge (USA) (2016)
3. Affenzeller, M., Winkler, S., Wagner, S., Beham, A.: Genetic Algorithms and Genetic Programming - Modern Concepts and Practical Applications. Volume 6 of Numerical Insights. CRC Press, Chapman & Hall (2009)
4. Lundberg, S., Lee, S.I.: A unified approach to interpreting model predictions (2017)
5. Ribeiro, M.T., Singh, S., Guestrin, C.: "why should i trust you?": Explaining the predictions of any classifier (2016)
6. Rokach, L., Maimon, O. In: Clustering Methods. Springer US, Boston, MA (2005) 321–352

Customization and Analysis of Orthopedic Aids*

Christoph Praschl[1 [0000−0002−9711−4818]], Mert Dalkilic[1 [0009−0001−8542−5176]], Sophie Bauernfeind[1 [0009−0005−5290−3900]], Markus Walkolbinger[2], and Gerald A. Zwettler[1 [0000−0002−4966−6853]]

[1] Research and Development Department, University of Applied Sciences Upper Austria {christoph.praschl, mert.dalkilic, sophie.bauernfeind, gerald.zwettler}@fh-hagenberg.at
[2] WAKO 3D Ltd. m.wakolbinger@wako3d.com

Introduction: In addressing congenital malformations, managing chronic diseases, and tending to musculoskeletal injuries, the provision of appropriate orthopedic aids like prostheses and orthoses stands as a pivotal measure. These aids need to be precisely tailored to the patient's needs and anatomy, a task traditionally undertaken by skilled orthopedists or orthopedic technicians. However, this process has predominantly adhered to analog methodologies. Yet, the landscape is shifting with the emergence of adaptive manufacturing techniques, which is catalyzing the progression of orthopedic technology in the context of Industry 4.0 [4, 7]. This transformative trajectory incorporates digital avenues such as 3D scans of the human body [5], 2D pedobarographic scans that aid in recommending suitable orthopedic correctives [1], and digitally manufactured orthopedic aids [4]. By combining these processes – model creation through scans and patient-specific adaptions – it becomes possible in implementing a (semi-) automated approach to orthopedics [10]. This raises the question: How can orthopedic aids be automatically customized while ensuring their structural integrity?

Related Work: Finite Element Method (FEM) encompasses structural analyses [3] and is widely used within engineering and biomechanics, including the orthopedic realm [8]. However, while FEM has found widespread use in medical contexts, studies have predominantly concentrated on analyzing structures within the human body, such as bones. Scant attention has been afforded to the examination of orthopedic aids, although exceptions exist in studies concerning bio-prosthetic heart valves [2], below-knee prosthetics [6] or prosthetic sockets [9]. However, these studies only focused on manually modelled aids.

Methodology: For the creation of patient-specific orthopedic aids, we propose a two-step procedure including (I) the elastic adaption of a medical aid based on the 3D scan of the affected body part and (II) the structural analysis of the resulting digital aid in comparison to its well known template using FEM. To do so, the aid's template A, the body scan B and a reference skeleton S are used. A as well as B are therefore represented by registered 3D meshes, while S is a 3D poly line within B. By finding the spatially next point of S for every vertex in A, a trace can be defined, that intersects with B, while also allowing to identify inward and outward facing vertices based on its intersections with A. The intersection point with B defines the target position A'_i of every inward-facing

* Supported by Austrian Research Promotion Agency FFG.

vertex A_i in A. For adapting also the outward-facing vertices A_o, a backward mapping is applied based on the already known positions of A'_i allowing to create a patient-specific aid A' in an as-rigid-as-possible manner.

Based on arbitrary sample regions, the template A as well as the patient-specific adaption A' can be compared using FEM analysis. By simulating pressure or heat, its effects are measured at well-defined sample points on both models, allowing a quantitative analysis of A' in reference to the well known A.

Evaluation: For the evaluation of the proposed process, two orthopedic aids, namely a forearm and a foot orthosis, are automatically adapted based on a 3D scan of the corresponding body part and a manually created skeleton. Based on the adapted results, both a pressure as well as a heat conduction analysis are applied using FEM. Based on multiple sample points, it is shown that both forces similarly affect the aid's template and the patient-specific customization, allowing to infer structural similarity and characteristics.

Conclusion and Outlook: In the current work, it is shown that a two-step process can be used for the automated digital customization of orthopedic aids, while ensuring structural integrity using FEM. In the future, this process should be applied to additional aids, also including prostheses next to orthoses.

References

1. Bauernfeind, S., Praschl, C., Wakolbinger, M., Zwettler, G.A.: Classification of footprints for correctives in orthopedics. In: 2023 International Conference on Electrical, Computer, Communications and Mechatronics Engineering. IEEE (2023)
2. Borazjani, I.: Fluid–structure interaction, immersed boundary-finite element method simulations of bio-prosthetic heart valves. Computer Methods in Applied Mechanics and Engineering **257**, 103–116 (2013)
3. Courant, R.: Variational methods for the solution of problems of equilibrium and vibrations. Bulletin of the American Mathematical Society **49**(1), 1 – 23 (1943)
4. Ejnisman, L., Gobbato, B., de França Camargo, A.F., Zancul, E.: Three-dimensional printing in orthopedics: from the basics to surgical applications. Current Reviews in Musculoskeletal Medicine **14**, 1–8 (2021)
5. Haleem, A., Javaid, M.: 3d scanning applications in medical field: a literature-based review. Clinical Epidemiology and Global Health **7**(2), 199–210 (2019)
6. Jweeg, M.J., Hamzah, H., Al-Waily, M., Al-Shammari, M.A.: A finite element simulation of nano effects on stress distribution in a below knee prosthetic. In: Materials Science and Engineering. vol. 1067, p. 012141. IOP Publishing (2021)
7. Molnár, I., Morovič, L.: Design and manufacture of orthopedic corset using 3d digitization and additive manufacturing. In: IOP Conference Series: Materials Science and Engineering. vol. 448, p. 012058. IOP Publishing (2018)
8. Naoum, S., Vasiliadis, A.V., Koutserimpas, C., Mylonakis, N., Kotsapas, M., Katakalos, K.: Finite element method for the evaluation of the human spine: A literature overview. Journal of Functional Biomaterials **12**(3), 43 (2021)
9. Silver-Thorn, B., Childress, D.S.: Parametric analysis using the finite element method to investigate prosthetic interface stresses for persons with trans-tibial amputation. Journal of rehabilitation research and development (1996)
10. Zwettler, G.A., Trixner, M., Schartmüller, C., Bauernfeind, S., Stockinger, T., Praschl, C.: Towards an automated process for adaptive modelling of orthoses and shoe insoles in additive manufacturing. In: Proceedings of the 12th International Workshop on Innovative Simulation for Healthcare, IWISH. CAL-TEK srl (2023)

Shape Classification and Analysis of Shockwave Signals caused by Plasma Bubble Variations in Shockwave Therapy Devices

Hannah Janout[1,2,3][0000−0002−0294−3585], Jonas Flatscher[3], Stephan Winkler[1,2][0000−0002−5196−4294], and Cyrill Slezak[3][0000−0003−4718−9966]

[1] University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria
[2] Johannes Kepler University, 4040 Linz, Austria
[3] Ludwig Boltzmann Institute for Traumatology, 1200 Vienna, Austria

Keywords: Spark Variation · Data Science · Classification · Signal Processing · Shockwave Therapy.

Extracorporeal shockwave therapy (ESWT) uses rapid, but short-duration, high-energy acoustic waves to stimulate healing and tissue regeneration [1][2]. They offer a non-invasive and alternative approach to promote pain relief and tissue repair in cases where conventional treatment methods have been less effective. Thus, ESWT has become a popular therapeutic modality for cartilage and bone diseases, as well as muscoskeletal disorders [2][3].

In this study, electrohydraulic shockwave devices are the main focus. These devices generate the shockwave pulse through an electrode located in the devices applicator head [3]. Once high voltage runs through the electrodes tip, an electric spark is generated and a shockwave forms [3]. The spreading shockwave is focused by the applicator head's reflector and arrives with maximum pressure at a designated focus [3]. Preceding the focused shockwave is a smaller, unfocused shock wave known as the primary wave [3].

In experiments measuring multiple shockwave signals in a water bath, a great variety of shockwave signal forms becomes apparent. Each of the shown signals has been cropped to the region of interest around the shockwave signal. A representative selection of three types is given in Figure 1, these images display the vast variety of signals recorded during identically executed measurements. Out of the shown signals, only image (a) displays a well-defined shockwave, while the other two either form more than one significant peak or display other distinctively novel shapes. For one, a small shift away from the focal point of the spark-generated plasma bubble has large effects on the shockwave shape, causing the observed anomalies. For two, the deviation from the spherical shape of the collapsing bubble further influences the observed wave. To which extend the variation of position and collapse of the bubble effect the final observed shockwave signal and its corresponding efficiency during therapy are currently unclear. Therefore, detailed analysis of shockwave shapes and their point of origin is required to develop new methods to improve the efficiency of shockwave therapy and enable the improvement of current applicator head designs.

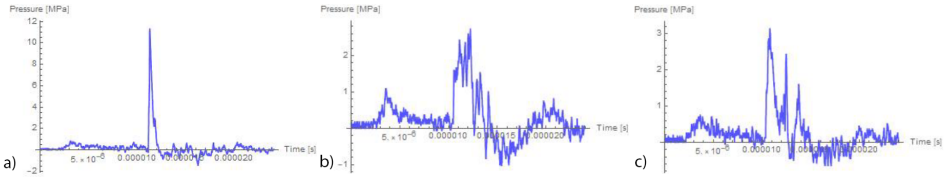


Fig. 1. Observed shapes of wave signals during identical measurements. Only image (a) displays a proper shockwave signal, (b) and (c) display pressure waves generated by misaligned sparks. Images taken from [4].

We aim to implement a workflow for automatic localization and extraction of shockwave signals inside of a recording through methodologies of signal processing and artificial intelligence and compare their suitability for the task. Therefore, the need of manual extraction of the shockwave signal is omitted and the analysis is accelerated.
Furthermore, we aim to classify the varying shapes of shockwave signals and draw a connection between the wave signal and the place of origin of the spark generated plasma bubble. Since the origin's position cannot be observed directly, we use a total of four sensors placed in a water bath to determine it. These four sensors allow the computation of the spark position through the concept of multilateration. An evolutionary strategy was implemented to compute the best possible position of the four sensors in the water bath. This evolutionary strategy varies the sensors' position in a fictive 3D water bath and searches for the best combination to predict the spark origins through multilateration accurately. Each candidate is evaluated by computing multiple origins scattered inside a fictive applicator head while fulfilling hardware constraints set by the used oscilloscope. The best-performing sensor combination was translated into a 3D sensor fixture and printed using a 3D printer. This sensor fixture is then used for in situ measurements inside a water bath and becomes the base for the planned analysis of shockwave signal shapes and spark generated plasma bubbles.

References

1. Glenn Elert. The physics hypertextbook. *Found July*, 1998-, 9:2008.
2. CL Simplicio, J Purita, W Murrell, GS Santos, RG Dos Santos, and JFSD Lana. *Extracorporeal shock wave therapy mechanisms in musculoskeletal regenerative medicine. J Clin Orthop Trauma*, 2020; 11(Suppl 3):S309-S318. doi: 10.1016/j.jcot.2020.02.004.
3. Ogden, J. A., Tóth-Kischkat, A., & Schultheiss, R. (2001). *Principles of shock wave therapy. Clin Orthop Relat Res*, (387), 8-17. doi: 10.1097/00003086-200106000-00003. PMID: 11400898.
4. Böhm, J. *Extrakorporale Stoßwellentherapie Simulation elektrohydraulisch erzeugter Stoßwellen und deren Druckfelder bei Gewebeinteraktion.* 2021

Analysis of C. elegans using Multi Modal Imaging

Jonas Schurr[1,2][0000−0001−8536−8354], Georg Sandner[3][0000−0001−8740−6123], Andreas Haghofer[1,2][0000−0001−6649−5374], Josef Scharinger[2][0000−0001−6502−7501], and Stephan Winkler[1,2][0000−0002−5196−4294]

[1] Bioinformatics Research Group, University of Applied Sciences Upper Austria, Softwarepark 11–13, Hagenberg, Austria
[2] Department of Computer Science, Johannes Kepler University, Altenberger Straße 69, Linz, Austria
[3] Center of Excellence Food Technology and Nutrition, University of Applied Sciences Upper Austria, Stelzhamerstraße 23, 4600 Wels, Austria
jonas.schurr@fh-hagenberg.at

The utilization of Caenorhabditis elegans as an in vivo model organism comprises several noteworthy advantages, including the potential for higher throughput during substance testing, the minimization of substances and cost per experiment. And especially, due to its similarities to human and animals, it can significantly reduce the necessity for animal testing [1]. Additionally, it allows an in-depth examination of the regulation of various genes and transcription factors. An important and omnipresent task is the investigation of substance toxicity. The overall goal is to allow automated and detailed analysis of C. elegans as a model organism to provide reliable data to reveal the effects of substances shown in up or down regulation of gene expression [1]. Especially due to fluorescence microscopy, and its advantage to examine living organisms, gene and protein expression can be extracted in these images. Different modalities, in this case the produced brightfield and fluorescence images, are used to improve the analysis of the effect of different substances. To measure the effects of different substances onto the fitness and survivability of a worm, relevant features need to be extracted and are of high importance to determine the effect of the substances.
To derive insights of substance effects, such as substance toxicity, multiple problems have to be solved. A large amount of experiments produces high amounts of different images. Often, such experiments are performed manually to a high degree while needing expert knowledge for the analysis. The large number of images leads to high interest in automatic methods for extracting and analyzing data [2]. These methods speed up the process of running experiments and increases much needed objectivity, reproducibility and comparability between different experiments. Additionally, the extraction and analysis of features which shows the importance on toxicity outcomes and its effects on the regulation of protein expression are of high interest. To allow the extraction, multiple problems like worm detection, segmentation, separation and the analysis of features itself, need to be solved. The use of manual tasks and a lack of accessible and sufficient analysis tools of fluorescence images of C. elegans for feature extraction lead to our proposed approach. In future the extracted features should also be transferable to videos allowing to track changes in behavior and motility.
The goal is to allow automated segmentation of different C. elegans image modal-

ities (brightfield and fluorescence). Based on the extracted worm instances, features for worm fitness will be extracted to allow conclusion on substance effects. To allow automated analysis of C. elegans an automated approach for the segmentation and localization of single C. elegans is pivotal for further analysis. Currently, this is often done by simple thresholding [2], which is not sufficient in all images (especially on fluorescence images) due to low or not existing gradients, artifacts and especially a certain degree of connected and overlapping worm-shaped objects. Additionally, fluorescence is low in intensity leading to low gradients. Therefore, neural networks with models like mask RCNNs can be used which pay special attention to close, overlapping or occluding objects [2]. Neural networks are trained, with additional images as brightfield images and combined with the corresponding fluorescence images to allow a robust and fully automated extraction of the worms. The effect of different approaches combining information of the fluorescence images and brightfield images by augmentation methods, like masked learning, is going to be investigated to allow accurate subsequent execution of different features. Additional separation of worms in post processing may be required. Based on the accurate segmented and separated

Fig. 1. All images are inverted for better visibility: A) raw brightfield image, B) raw fluorescence image, C) segmented worms by the neural network, D) filtered image allowing feature extraction

worms feature extraction can be performed. Especially, in the fluorescence images intensity features (brightness, gradients, homogeneity, ...) can be of high interest. Additionally, morphological features are extracted like worm length, thickness, area and radius and are of special interest for toxicity. A first result is shown in Figure 1. These features allow the investigation on the effect of different substances. Differences and important features defining the effect of a substance shall be investigated in future with the use of machine learning models.

References

1. Lanzerstorfer, P., Sandner, G., Pitsch, J., Mascher, B., Aumiller, T. & Weghuber, J. Acute, reproductive, and developmental toxicity of essential oils assessed with alternative in vitro and in vivo systems. *Archives Of Toxicology.* **95** pp. 673ff. (2021)
2. Hebert, L., Ahamed, T., Costa, A., O'Shaughnessy, L. & Stephens, G. WormPose: Image synthesis and convolutional networks for pose estimation in C. elegans. *PLoS Computational Biology.* **17**, e1008914 (2021)

Age-Layer-Population-Structure with Self-Adaptation in Optimization*

Kaifeng Yang[1 [0000−0002−3353−3298]], Bernhard Werth[1 [0000−0002−7830−3138]], and Michael Affenzeller[1 [0000−0001−5692−5940]]

Heuristic and Evolutionary Algorithms Laboratory
University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria

Kaifeng.Yang@fh-hagenberg.at

Population diversity in dynamic optimization problems is vital as it enables the algorithm to explore various regions of a changing solution landscape, preventing premature convergence to suboptimal solutions. A diverse population ensures adaptability, resilience, and the ability to track and adapt to shifting optima, essential for effectively addressing the dynamic nature of the problem. Several methods have been devised to sustain diversity in such contexts. Dynamic adaptation of parameters, such as mutation rates or selection pressures, based on the changing landscape, allows the algorithm to dynamically adjust to new optima. Niching methods, including crowding or fitness sharing, explicitly encourage the maintenance of diverse solutions by penalizing similarity. Besides, restart mechanisms, such as Bipop-CMA-ES [2], also keep the population diversity by reinitializing the population. Additionally, hybridization techniques and multi-population strategies introduce diverse genetic pools that can exchange information, facilitating a broader exploration of the altered problem space. These methods collectively enable evolutionary algorithms to effectively navigate dynamic environments by preserving diversity, ensuring adaptability, and facilitating the discovery of high-quality solutions amidst changing conditions.

Age-layered population structure (ALPS) [3] in evolutionary algorithms involves organizing the population into layers based on the "age" of individuals, typically representing the number of generations they have survived or the time of creation within the algorithm. Each layer might have distinct characteristics and roles. Younger layers often represent exploration, bringing in fresh solutions, while older layers emphasize exploitation by retaining promising solutions that have survived multiple generations.

ALPS stands out for its exceptional features, notably its continuous integration of new genetic material without necessitating resource-intensive restarts due to fully converged population states. Furthermore, the system adeptly retains genetic information from earlier problem epochs implicitly within higher layers. These distinctive traits not only foster a harmonious balance between exploration and exploitation but also confer robustness against premature convergence, rendering ALPS highly efficient for navigating dynamic environments.

* This work is supported by the Austrian Science Fund (FWF – Der Wissenschaftsfonds) under the project (I 5315, 'ML Methods for Feature Identification Global Optimization).

Nonetheless, a notable drawback of ALPS lies in its demand for meticulous parameterization. Establishing optimal age gaps, mutation rates, and crossover rates proves challenging, given the inherent limitation of a "one-size-fits-all" paradigm, where constant rates are applied across diverse layers. To address this challenge, we advocate for a more tailored evolution approach across different layers, each representing populations of varied ages and convergence states. To avoid exacerbating the complexity of parameterization with additional factors, we revisit the concept of self-adaptation, originally introduced for evolutionary strategies [1]. This approach aims to enhance adaptability and performance without introducing an undue burden of additional parameters.

In this work, we propose a self-adaptive mechanism inspired by evolutionary strategies [1], wherein hyperparameters (including age gap, aging scheme, mutation rate, crossover rate) dynamically adjust based on the evolving problem landscape. Specifically, we utilize individual stepsize adaption strategy [4] for mixed-integer optimization problems. The age gap, aging scheme, and crossover rate are automatically tuned by population similarity, inherent building blocks, and other evolution behaviors. This self-adaptation ensures that ALPS remains responsive to changes, allowing for an automatic fine-tuning of hyperparameters during runtime. By integrating self-adaptation, ALPS gains the flexibility needed to navigate dynamic environments without the need for constant manual intervention, thereby enhancing its adaptability and overall optimization performance. In addition, we will compare the proposed self-adaptive ALPS with various fixed instances of ALPS, in order to gauge both resilience against missparameterization as well as classic performance, using dynamizised versions of the classic TSPLib [6] instances. We will utilize the empirical attainment function [5] to statistically analyze the dynamic features of algorithm's behavior, such as probabilistic distribution of the outcomes and convergence behavior.

References

1. Beyer, H.G.: Toward a theory of evolution strategies: Self-adaptation. Evolutionary Computation **3**(3), 311–347 (1995)
2. Hansen, N.: Benchmarking a bi-population cma-es on the bbob-2009 function testbed. In: Proceedings of the 11th annual conference companion on genetic and evolutionary computation conference: late breaking papers. pp. 2389–2396 (2009)
3. Hornby, G.S.: Alps: the age-layered population structure for reducing the problem of premature convergence. In: Proceedings of the 8th annual conference on Genetic and evolutionary computation. pp. 815–822 (2006)
4. Li, R., Emmerich, M.T., Eggermont, J., Bäck, T., Schütz, M., Dijkstra, J., Reiber, J.: Mixed integer evolution strategies for parameter optimization. Evolutionary Computation **21**(1), 29–64 (2013). https://doi.org/10.1162/EVCO_a_00059
5. López-Ibáñez, M., Paquete, L., Stützle, T.: Exploratory analysis of stochastic local search algorithms in biobjective optimization. Springer (2010)
6. Reinelt, G.: Tsplib—a traveling salesman problem library. ORSA journal on computing **3**(4), 376–384 (1991)

Machine Learning Update Strategies for Real-time Production Environments*

Philipp Neuhauser[1 [0009−0000−8756−8181]], Philipp Fleck[1 [0000−0001−8290−6356]], Sebastian Leitner[1 [0000−0003−4261−2868]] and Stefan Wagner[1 [0000−0002−0323−9274]]

Josef Ressel Center for Adaptive Optimization in Dynamic Environments
Heuristic and Evolutionary Algorithms Laboratory
University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria

philipp.neuhauser@fh-hagenberg.at

In modern application scenarios real-time capable machine learning solutions based on streaming data are increasingly getting more and more important. Waiting several hours for the results of a batch processing system is inconceivable and in many cases also involves certain economic costs. In fact, decisions and analyses should be made as soon as new data points arrive. Furthermore, it is not only important that the trained models can be used in real-time, but also that they can be quickly adapted to the current data situation in order to maintain a high degree of accuracy. Detecting the right point in time for a model update can be a challenging task, especially for unbounded streaming data.

Over the past few years, Apache Kafka [2] has clearly established itself as the de-facto standard for real-time processing of massive data streams in almost all industrial, economic and financial sectors. The underlying distributed server-client architecture efficiently combines the concepts of a high-throughput message broker system with a reliable underlying data storage. Numerous extensions also ensure comprehensive data stream processing. Messages can be extracted from other common data sources using appropriate plugins, modified and joined accordingly with a stream processing API, and ultimately transferred to other systems. Furthermore, the aspect of central schema management with an integrated version control is also taken into account. All these components around Apache Kafka fit together seamlessly to form a sophisticated Kappa architecture, which serves as a central starting point for further data science tasks and machine learning applications [3].

Extracting new insights from real-time data streams requires the use of a reliable machine learning framework. Besides well-established ML libraries like TensorFlow, H2O.ai or Scikit-learn, Microsoft's ML.NET [1] has proven to be a suitable candidate for solving common ML tasks. It offers a variety of different machine learning algorithms as well as a wide range of functions for data preprocessing and feature engineering. Ultimately, the integrated AutoML API also enables autonomous training of machine learning models due to several exchangeable hyperparameter tuners [4]. Especially for real-time applications, it

* The financial support by the Austrian Federal Ministry for Digital and Economic Affairs and the National Foundation for Research, Technology and Development and the Christian Doppler Research Association is gratefully acknowledged.

is of enormous importance that accurate models are (re-)trained as quickly as possible with a minimum amount of human interaction. Often, there is not much time for comprehensive manual fine-tuning, which is why AutoML can certainly be an advantage.

When applying machine learning to unbounded data streams, the question inevitably arises as to when exactly a model needs to be retrained. The right time has to be chosen carefully, because the training phase can be very time-consuming and resource-intensive. Of course, the designed ML pipeline could be triggered manually at fixed times or after a certain amount of new data points have been collected. The literature also suggests the use of online-capable machine learning algorithms that continuously learn and adapt to specific circumstances. Another interesting approach is the recognition of so-called data and/or concept drift, i.e. those situations in which the data points change over time and thus weaken the validity and performance of a machine learning model. Ideally, these changes should be recognized automatically and not require manual assessment by a domain expert. One possibility would be to continuously monitor the model quality (R2, MSE, F1 score, etc.). Another conceivable approach is to continuously track the similarity of individual streaming data batches. In both cases, i.e. a decreasing quality value or excessive deviations in the data, retraining might be necessary.

Within this paper, a holistic software architecture for machine learning with real-time data streams based on Apache Kafka and ML.NET is presented. Subsequently, the designed ML pipeline serves as a starting point for the integration and evaluation of different model update strategies, especially online machine learning, trigger-based (elapsed time and batch of new received data points) and due to detection of loss in model qualities or data drift. Finally, several experiments and benchmark datasets will show which retraining approaches deliver the best test results under certain circumstances, like abrupt/gradual changes or noise in the data. In addition, it should also be determined whether there is a connection between occurring changes in the online data stream and a possible resulting loss of model quality. Ultimately, in combination with AutoML, the aim is to propose a fully autonomous machine learning infrastructure that not only trains models independently, but also keeps them up to date while attempting to ensure high model prediction accuracy.

References

1. Ahmed, Z., Amizadeh, S., Bilenko, M., Carr, R., Chin, W.S., Dekel, Y., Dupre, X., Eksarevskiy, V., Filipi, S., Finley, T., et al.: Machine learning at Microsoft with ML. NET. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2448–2458 (2019)
2. Apache Software Foundation: Apache Kafka (2023), https://kafka.apache.org/
3. Neuhauser, P.: Real-time Machine Learning for Product Quality Prediction. Master's thesis, University Of Applied Sciences Upper Austria, Hagenberg (2022)
4. Neuhauser, P., Wagner, S.: Performance Comparison of Microsoft's AutoML API. In: 35th European Modeling and Simulation Symposium, EMSS 2023. Proceedings of the 35th European Modeling and Simulation Symposium (EMSS 2023) (2023)

Online Machine Learning for the Estimation of Process Times in Dynamic Scheduling*

Michael Heckmann[1 [0009−0003−7731−9941]], Bernhard Werth[1,2 [0000−0002−7830−3138]], Johannes Karder[1,2 [0000−0002−7006−8663]], and Stefan Wagner[1 [0000−0002−0323−9274]]

[1] Josef Ressel Center for Adaptive Optimization in Dynamic Environments
Heuristic and Evolutionary Algorithms Laboratory
University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria
[2] Symbolic Artificial Intelligence
Johannes Kepler University, 4040 Linz, Austria

michael.heckmann@fh-hagenberg.at

Dynamic Scheduling is a popular research topic and has many applications, such as crane scheduling and production and manufacturing planning [3]. As opposed to static optimization of schedules, dynamic scheduling reacts to changes from its corresponding real world system, that may impact the current solution. As with many real world applications, the processing times of the individual steps may be unknown, but can be estimated by machine learning models. For dynamic scheduling, these machine learning models are preferably online models, as the training data can be continuously updated with the current processing times. Online machine learning is a relatively new area of research that is gaining traction in the machine learning community, but compared to the static machine learning models, the amount of research is still underrepresented.

This paper aims to discuss an application of online machine learning models for the approximation of the processing times for production tasks. These tasks are process steps on various machines, with processing times influenced by the processed materials, setup time, human interaction and additional unidentified components. Having a good approximation of the processing times provides the optimization algorithm with valuable information on how the real world processes will behave for the created solutions.

For many problem instances, the behavior of processing times is not kept track of and may change over time, thus training static models does not provide the necessary flexibility. To remedy the lack of data and potentially changing processing time behavior, online machine learning models can be used. These models are designed to continuously train on new data, while also predicting based on the already observed data. Another crucial advantage of online machine learning models is the ability to phase out old training data in favor of more current observations.

This paper covers two aspects. The first aspect is the evaluation and comparison of several machine learning approaches. The second aspect regards the

* The financial support by the Austrian Federal Ministry for Digital and Economic Affairs and the National Foundation for Research, Technology and Development and the Christian Doppler Research Association is gratefully acknowledged.

correlation between the schedule quality and the model quality. Most of the models evaluated are online models, but as a baseline, some other models are used. An example for a baseline model, is the basic online median. More sophisticated online machine learning models are being taken from, River [2]. River provides online machine learning implementations for decision trees, forests, linear regression, and neural networks. Additionally, an online variant of symbolic regression is tested. This variant uses a genetic algorithm that allows changes to the dataset in between generations. Symbolic regression also provides interpretable approximations of the machine processing times.

To evaluate the model quality and its effect on the optimization, a discrete event simulation of a production process was created. To cover a large spectrum of processing time behaviors, the machines in the simulation are supplied with different functions to determine the processing time. These functions can contain factors for the type of work step, material features, random noise, effects of wear on the tools and human factors. Periodically during the simulation, orders are added to the production process for the optimizer to consider. The simulation sends events to the optimization system which keeps track of the current simulation state and optimizes the order of required steps using the OERAPGA [1] optimization algorithm. The resulting solution is then applied to the simulation, which will follow the new processing plan. When a processing step on a machine is completed, the required processing time and the corresponding processing information are captured and supplied to the online machine learning model.

Based on the previously described simulation, experiments are conducted. For these experiments, machines with different processing time behaviors are used for a production process with several product types. The captured data of the experiments will show the results of online machine learning models for various defined machine processing time behaviors. Additionally, the effect of the selected machine learning models on the optimization of the dynamic scheduling problem is evaluated.

References

1. Karder, J., Werth, B., Beham, A., Wagner, S., Affenzeller, M.: Analysis and handling of dynamic problem changes in open-ended optimization. In: Moreno-Díaz, R., Pichler, F., Quesada-Arencibia, A. (eds.) Computer Aided Systems Theory – EUROCAST 2022. pp. 61–68. Springer Nature Switzerland, Cham (2022)
2. Montiel, J., Halford, M., Mastelini, S.M., Bolmier, G., Sourty, R., Vaysse, R., Zouitine, A., Gomes, H.M., Read, J., Abdessalem, T., et al.: River: machine learning for streaming data in python (2021)
3. Priore, P., Gómez, A., Pino, R., Rosillo, R.: Dynamic scheduling of manufacturing systems using machine learning: An updated review. AI EDAM **28**(1), 83–97 (2014). https://doi.org/10.1017/S0890060413000516

Concurrent Evolution of Dynamic Single- and Dual-Crane Scheduling Scenarios*

Johannes Karder[1,2 [0000−0002−7006−8663]], Bernhard Werth[1,2 [0000−0002−7830−3138]], Stefan Wagner[1 [0000−0002−0323−9274]], and Michael Affenzeller[1,2 [0000−0001−5692−5940]]

[1] Josef Ressel Center for Adaptive Optimization in Dynamic Environments
Heuristic and Evolutionary Algorithms Laboratory
University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria
[2] Institute for Symbolic Artificial Intelligence
Johannes Kepler University, 4040 Linz, Austria

johannes.karder@fh-hagenberg.at

In static optimization problems, both the problem model and the specific problem instance data is fixed, i.e. both do not change over the course of the optimization. This differentiates static problems from dynamic optimization problems where either the problem model or the problem instance data (or both) change over time. Dynamic optimization presents a certain challenge to classic optimization algorithms as changes in either the model or the data can induce changes to solution qualities or – in the worst case – invalidate solutions so that they become infeasible.

A concrete example of such a dynamic optimization problem can be found in the domain of warehouse operations. The dynamic optimization problem that is tackled within this paper is crane scheduling. Boysen et al. [1] define a generalized classification scheme for crane scheduling scenarios with interference. In an academic scenario, two gantry cranes continuously receive instructions from an optimizer in form of *crane schedules* to move containers within a warehouse. These cranes share a single crane lane and thus cannot overtake each other. Containers enter the warehouse at so-called arrival locations. Arrival locations are serviced by transport vehicles that arrive in certain intervals and create *pickup demands*. The two cranes must unload these containers from the transport vehicles and store them inside the warehouse on so-called buffer locations. A single buffer location can hold a specific amount of containers stacked above each other, and is thus limited in capacity. The cranes can only ever access the topmost container of each stack. Furthermore, other transport vehicles arrive at so-called handover locations and request specific *dropoff demands* for containers to be delivered by the cranes, which will leave the warehouse afterwards. Containers therefore enter and leave the warehouse continuously and cranes should be optimally scheduled to enable effective and efficient warehouse operation. Since cranes share a single lane, it is important for an optimizer to generate crane

* The financial support by the Austrian Federal Ministry for Digital and Economic Affairs and the National Foundation for Research, Technology and Development and the Christian Doppler Research Association is gratefully acknowledged.

schedules where one crane does not unnecessarily block the other to minimize idle times.

Multiple aspects create stochasticity and dynamicity within this dynamic crane scheduling problem. The amount of *move requests* (pickups and dropoffs) generated by vehicles at arrival and handover locations varies when the scenario progresses. Arriving vehicles generate new move requests and thus increase the set of open move requests, while cranes fulfill and thus decrease the set of open requests. This dynamic aspect is handled by updating existing solutions by adding or removing new and fulfilled move requests from the schedules and re-evaluating their fitness after dynamic changes occurred. Furthermore, cranes also become unavailable after and for a stochastically determined amount of time to simulate crane operators taking breaks in real life. In the presented scenario, it is ensured that at least one crane is always active and can handle all required moves. The switches from dual- to single-crane scenarios and vice versa induce some time during which an optimizer adapts existing solutions and converges in the new search space. This time should be kept to a minimum in order to efficiently provide adapted solutions that are valid for the current system state, e.g. by concurrently evolving solutions for both scenarios and switching between solutions accordingly.

Within this paper, we investigate ways to concurrently evolve solutions for both scenarios using genetic algorithms (GA) [2], and analyze the algorithmic behavior. As a first step, we implement a GA that evolves a population $pop = pop_1 \cup pop_2$, where pop_1 and pop_2 contain only single- and dual-crane solutions, respectively. Each subpopulation is considered individually throughout the generational cycle, i.e., when creating a new population pop, the GA will treat pop_1 and pop_2 isolated from each other and only cross individuals within the same subpopulation, optionally mutate the resulting children, and apply replacement (e.g. elitism) on the respective subpopulation where the child originated from. This approach was already proposed as future work by Karder et al. [3]. In a second step, we investigate ways to migrate essential *building blocks* that are part of high quality solutions, and therefore relevant for both scenarios, between pop_1 and pop_2. Finally, we consider a bi-objective approach in which each solution, represented genotypically by a permutation as *scheduling vector* and a binary vector as *assignment vector*, is evaluated for its single- and dual-crane scenario performances, mapped to objectives 1 and 2, respectively.

References

1. Boysen, N., Briskorn, D., Meisel, F.: A generalized classification scheme for crane scheduling with interference. European Journal of Operational Research **258**(1), 343–357 (2017)
2. Holland, J.H.: Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press (1992)
3. Karder, J., Werth, B., Beham, A., Wagner, S., Affenzeller, M.: Analysis and handling of dynamic problem changes in open-ended optimization. In: International Conference on Computer Aided Systems Theory. pp. 61–68. Springer (2022)

Learning-based algorithm selection for a multiprocessor scheduling problem

Roland Braune[1,2,3][0000−0003−4086−229X]

[1] University of Applied Sciences Upper Austria, Operations Management, Austria
[2] Department of Business Decisions and Analytics, University of Vienna, Austria
[3] Josef Ressel Center for Adaptive Optimization in Dynamic Environments, University of Vienna, Austria
roland.braune@univie.ac.at

The subject of this contribution is a particular kind of multiprocessor scheduling problem. A set of n tasks $T = \{1, \ldots, n\}$ is to be processed on a set of m identical parallel processors. Each task $i \in T$ consists of n_i subtasks or task "slices". Such a sub-task is identified by a pair (i, j), with $i \in T$ and $j \in \{1, \ldots, n_i\}$. The number of processors to be allocated simultaneously for the execution of a task slice is denoted by $size_{i,j}$, also commonly known as the *width* of a task (slice) [4]. All slices (i, j) that make up a task $i \in T$ have unit processing time, that is, $p_{ij} = 1$, $i \in T$, $j \in \{1, \ldots, n_i\}$, and share the same width, hence, $size_{i,1} = size_{i,2} = \ldots = size_{i,n_i}$. Consequently, it is possible to specify the width on the level of outer tasks $i \in T$, yielding $size_i = size_{i,1} = \ldots = size_{i,n_i}$. The slices of a task are organized into chain-like structures by defining appropriate, generalized precedence constraints. Those constraints are enforced by imposing minimum time lags (delays) on every two sub-tasks of one and the same outer task. Since preemption of tasks is not allowed, all slices of a task are also subject to maximum time lags. Chain precedence constraints have already been discussed in the multiprocessor scheduling literature [1], but not in conjunction with maximum time lags.

Figure 1 shows a multiprocessor schedule for $m = 14$ processors, based on three different task chains, with 6, 4 and 4 slices (sub-tasks), respectively. The widths are given as $size_1 = 5$, $size_2 = 3$, and $size_3 = 7$. It can be seen that the processor allocation changes three times for chain 1, once for chain 2 and twice for chain 3. The number of sub-tasks of a particular chain that are processed simultaneously is not limited, but at least one sub-task has to be processed at any time during the processing time window of a chain because of the implicit non-preemption constraint.

A branch-and-bound algorithm has recently been proposed for this strongly NP-hard problem [3], which is, however, sensitive to the "bulkiness" of the items. In other words, instances with items of very small size are known to be exceptionally hard to solve. On the other hand, instances of the latter type are prone to exhibit a resource allocation profile which is "ideal", meaning that all processors except the right-most are loaded up to their maximum capacity. It is known that the corresponding schedule *must* be optimal for the given objective [2]. However, for finding such an ideal feasible schedule, it is very effective to run a tailored subset sum-based heuristic *prior* to the branch-and-bound, as described in [3].

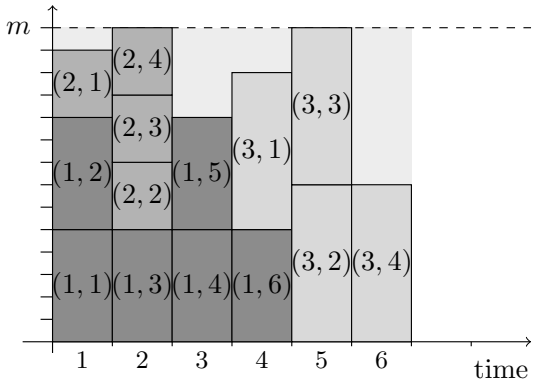


Fig. 1. Scheduling multiprocessor task chains s.t. minimum and maximum time lags.

This heuristic can already solve a very large proportion of instances to optimality and thus avoid running the computationally more expensive branch-and-bound algorithm in these cases.

Unfortunately, the subset sum-based heuristic is itself quite sensitive to the problem structure and in cases where it is not capable of delivering an optimal solution, the branch-and-bound algorithm has to be run anyway. Hence, some computation time is wasted on unavailingly trying to solve the subset-sum problem. If one knew up-front that the heuristic would (likely) fail, this additional computational effort could be eliminated. The goal of the presented work is therefore the development of a predictive learning-based approach for selecting the appropriate algorithm combination (heuristic and branch-and-bound in sequence or just the branch-and-bound) based on structural properties of a problem instance. Promising features are identified based on the problem definition, extracted and used for training various different binary classifiers known from machine learning. First computational experiments show prediction accuracies of around 90% and considerable time savings compared to exact optimization without algorithm selection.

References

1. Blazewicz, J., Liu, Z.: Scheduling multiprocessor tasks with chain constraints. European Journal of Operational Research **94**, 231–241 (1996)
2. Braune, R.: Lower bounds for a bin packing problem with linear usage cost. European Journal of Operational Research **274**(1), 49–64 (2019). https://doi.org/10.1016/j.ejor.2018.10.004
3. Braune, R.: Packing-based branch-and-bound for discrete malleable task scheduling. Journal of Scheduling **25**, 675–704 (2022). https://doi.org/10.1007/s10951-022-00750-w
4. Drozdowski, M.: Scheduling for Parallel Processing. Computer Communications and Networks, Springer Verlag London (2009)

Using the Pilot Method as a Problem-Independent Metaheuristic for Multi-Objective Beam Search *

Oliver Bindreiter, Bernhard Werth[0000−0002−7830−3138], Johannes Karder[0000−0002−7006−8663], and Stefan Wagner[0000−0002−0323−9274]

Heuristic and Evolutionary Algorithms Laboratory
University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria

oliver.bindreiter@fh-hagenberg.at

Multi-objective beam search extends the beam search [2] to be applicable for problems with multiple objectives. Barthelemy et al. [1] presented an exploration of the issues related to problem-dependent heuristic functions used as an oracle to estimate the quality of solutions and the limits associated with those functions. Additionally, the authors analyze how these aspects can directly impact the efficiency and achieved results of the multi-objective beam search for integer multi-objective optimization problems.

Experiments of a preliminary project in the field of scheduling problems have demonstrated the potential viability of applying multi-objective beam search to solve the underlying problem. The approach also showed promising capabilities to generate high quality solutions. However, it's important to acknowledge that the overall effectiveness, the achieved quality, and the diversity within the final set of Pareto-optimal solutions highly depend on the used heuristic functions during the search and construction process. These heuristic functions play an essential role in guiding the search towards Pareto-optimal solutions and therefore, their design becomes vital. Furthermore, rigorous testing and validation, demanding extensive effort and deep understanding of both the problem and the chosen algorithmic approach, are necessary while designing these heuristic functions. Honda [4] proposed a backtracking approach to overcome these issues.

This paper proposes a problem-independent metaheuristic extension for the multi-objective beam search algorithm. Using the pilot method [5] as a look-ahead function, multi-objective beam search becomes applicable to any kind of multi-objective optimization problem without further problem specific adaptions. The look-ahead function then replaces the heuristic functions as an estimator for solution quality. This eliminates the need of an additional heuristic function. Furthermore, the selection of the next layer out of a set of fronts is inspired by NSGA-II [3] and utilizes a crowding factor. While the target functions

* The work described in this paper was done within the project Adaptive scheduling in multi-stage production systems with sensor-based predictions for reducing energy consumption (SCHED-ENERGY), #891173 within the funding program Production of the Future, 41st call - by the Austrian Research Promotion Agency (FFG) and funded by the Republic of Austria.

provide the neccesary orders to calculate a front, the crowding distance makes it possible to establish an order within a single front.

The integration of the pilot method algorithm introduces a forward-looking mechanism that replaces the heuristic functions. To estimate the fitness of a subtree during the search process, a new search process for each objective of the underlying problem is started. The sole purpose of these individual search procedures is to optimize each respective objective individually. By combining the best quality found per objective, the obtained qualities serve as heuristic values for the entire subtree. This leads to the selection of subtrees with the most promising result in any objective. During the evaluation of the next layer some subtrees are revisited but since this time all successors are considered by the pilot method, especially interesting subtrees get searched more intensively after every iteration. Furthermore, the heuristic values represent a lower bound for their respective target function. Therefore, by further exploring the selected subtrees better solutions can be identified.

In each iteration of the beam search the next layer consisting of the best n nodes (i.e. subtrees) is selected. In multi-objective beam search, this is achieved by calculating the fronts of increasing orders until the best n nodes are found. However, should the amount of remaining slots in the next layer be too low as to fully include a front of a given order completely only some nodes of the given front can be selected. Inspired by NSGA-II, the crowding distance is calculated and the front gets sorted by the respective crowding factors. Therefore, the search process prefers underrepresented areas of the search space.

Extending the multi-objective tree search with the proposed aspect results in an increased quality of the final Pareto-optimal solutions. The integrated improvement of using the pilot method instead of heuristic functions facilitate a more effective algorithm that reduces the required beam width required to gather the final set of solutions. While remaining effective and accurate, adopting a smaller beam width enables a more resource-efficient and faster optimization process. Beyond increased efficiency, the NSGA-II inspired extension of the crowding distance also yields a more diverse set of final solutions.

References

1. Barthelemy, T., Parragh, S.N., Tricoire, F., Hartl, R.F.: Beam search for integer multi-objective optimization (2015)
2. Bisiani, R.: Encyclopedia of artificial intelligence. pp. 56–58. Wiley & Sons (1987)
3. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: Nsga-ii. IEEE Transactions on Evolutionary Computation pp. 182–197 (2002)
4. Honda, N.: Backtrack beam search for multiobjective scheduling problem. In: Multi-Objective Programming and Goal Programming. pp. 147–152. Springer Berlin Heidelberg (2003)
5. Voss, S., Duin, C., Fink, A.: Looking ahead with the pilot method. Annals of Operations Research pp. 225–242 (2005)

Solving two-machine sum-cost flow shop problem on D-Wave Quantum Cloud Service

Wojciech Bożejko[1 [0000−0002−1868−8603]],
Mariusz Uchroński[1,2 [0000−0002−9185−1841]], and
Mieczysław Wodecki[1 [0000−0001−8188−4503]]

[1] Wrocław University of Science and Technology
Janiszewskiego 11-17, 50-372 Wrocław, Poland
{wojciech.bozejko, mieczyslaw.wodecki}@pwr.edu.pl
[2] Wroclaw Centre for Networking and Supercomputing
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław
mariusz.uchronski@pwr.edu.pl

Extended abstract

In a two-machine flow problem with profit maximization (total weighted number of on-time tasks), each of n tasks should be performed sequentially on the first and then on the second machine. The execution times of the tasks p_{ij}, the due dates d_i (on the second machine), and the weight w_i resulting from completing the task on time are given. Determine the order of task execution (the same on both machines) that maximizes profit $\sum_i w_i(1-U_i)$, where $U_i = 1$ if the task i is late, 0 otherwise. The problem under consideration belongs to the class of NP-hard problems. There are relatively few papers devoted to even the most studied equivalent problem of the late task minimization problem $F2\|\sum w_iU_i$ and its solution methods. Ruiz and Stützle [3] presented a greedy algorithm, Liao et al. [2] an algorithm based on the tabu search method.

By C_i we mean the moment of completion of the task i on the last (second) machine. Let matrix $\mathbf{x}=[x_{ij}]_{n\times n}$ where $x_{ij} \in \{0,1\}$. We must determine

$$\max \sum_i \sum_j w_j x_{ij}, \quad \text{subject to} \tag{1}$$

$$C_i \leq \sum_{j=1}^{n} d_i x_{ji}, \quad i = 1,2,\ldots,n, \tag{2}$$

$$\sum_{i=1}^{n} x_{ij} \leq 1, \ j = 1,2,\ldots,n, \quad \sum_{j=1}^{n} x_{ij} \leq 1, \ i = 1,2,\ldots,n. \tag{3}$$

$$\left(\sum_{i=1}^{j}\sum_{s=1}^{n} p_{s1}x_{is}\right) + \left(\sum_{i=j}^{k}\sum_{s=1}^{n} p_{s2}x_{is}\right) \leq C_k, \ j = 1,2,\ldots,k, \ k = 1,2,\ldots,n, \tag{4}$$

$$C_i \in \mathbb{N} \cup \{0\}, \ i = 1,2,\ldots,n, \ x_{ij} \in \{0,1\}, \ i,j = 1,2,\ldots,n. \tag{5}$$

The matrix $\mathbf{x}$ is interpreted as follows: if $x_{ij} = 1$, then a task j on the position i in the schedule is on-time.

Computational experiments. The calculations were performed in the D-Wave Leap quantum cloud environment in two configurations. The first one involved using the LeapHybridCQMSampler solver, which uses a hybrid CPU (and GPU) and quantum QPU processors. The second one used the native QUBO (Quadratic Unconstrained Binary Optimization) approach obtained by converting the CQM (Constrained Quadratic Model) problem to BQM (Binary Quadratic Model), which allowed the use of quantum annealing. The results are presented in Table 1 (F – objective function, t_{QPU} – QPU_ACCESS_TIME, t_{RUN} – RUN_TIME).

Table 1. Results of computational experiments (with limited computation time)

		LeapHybridCQMSampler			DWaveSampler	
run	n	F	t_{QPU} [ms]	t_{RUN} [s]	F	t_{QPU} [ms]
1	4	17	32.05	4.97	16	16.02
2	4	17	32.05	5.17	13	16.03
3	4	17	32.04	5.10	12	16.00
4	4	17	32.02	5.08	18	16.02
5	4	17	32.05	5.18	13	15.99
1	5	26	16.03	5.00	20	16.05
2	5	26	32.06	5.35	13	16.05
3	5	26	16.03	5.02	26	16.02
4	5	26	32.06	5.23	20	16.05
5	5	26	16.03	5.00	18	16.04
1	6	45	32.07	5.24	-	-
2	6	45	16.01	5.00	-	-
3	6	45	32.09	5.31	-	-
4	6	45	32.05	5.46	-	-
5	6	45	32.08	5.09	-	-

The solution values (F column) obtained with the LeapHybridCQMSampler solver are generally better, but they use much longer processors time. In turn, the quantum annealing results obtained by DWaveSampler are obtained in (practically) constant time. Directly running quantum annealing (DWaveSampler) is of limited use due to the requirement of a large number of qubits, which limited the size of instances to $n \leq 5$ (for $n = 6$ D-Wave reports embedding error). In our experiment, the D-Wave Advantage_system4.1 environment had 5627 qubits.

References

1. Bożejko W., Pempera J., Uchroński M., Wodecki M.: Determination of the Lower Bounds of the Goal Function for a Single-Machine Scheduling Problem on D-Wave Quantum Annealer. ICCS 2023, LNCS 14077, pp. 201–208 (2023)
2. Liao, C. J., Liao, L. M., Tseng, C. T.: A performance evaluation of permutation vs. non-permutation schedules in a flowshop. International Journal of Production Research, **44**(20), 4297–4309 (2006)
3. Ruiz, R., Stützle, T.: A simple and effective iterated greedy algorithm for the permutation flowshop scheduling problem. European Journal of Operational Research **177**(3), 2033–2049 (2007)

Integrating optimization techniques and live tracking software in maritime logistics

Bruno Lorenzo Arroyo-Pedraza[1], Sergio Leopoldo Benítez-Delgado[1], Airam Expósito-Márquez[1 [0000−0003−2272−175X]], Christopher Expósito-Izquierdo[1 [0000−0002−0846−1675]], and Israel López-Plata[1 [0000−0002−4666−270X]]

Universidad de La Laguna, C/ Padre Herrera s/n, 38200, San Cristóbal de La Laguna, Santa Cruz de Tenerife, Spain http://www.ull.es

Abstract. This paper explores the integration of the Greedy Randomized Adaptive Search Procedure within live tracking software to optimize port logistics. Focused on the Berth Allocation Problem incorporating container movement costs, the study harnesses real-time vessel position data from the Automatic Identification System to predict arrivals at ports. The GRASP dynamically allocates berths, minimizing container movement expenses and orchestrating efficient vessel sequencing within port infrastructures. Results demonstrate the software's effectiveness in proactive planning, resource allocation and congestion mitigation. The synergy between predictive analytics and GRASP-driven optimization maximizes port facility utilization, enhancing operational efficiency.

Keywords: Berth allocation problem · GRASP · Vessel tracking

1 Introduction

The Berth Allocation Problem (BAP) [1] is an outstanding logistical challenge within container port operations, involving the assignment of berths to incoming vessels and the subsequent allocation of containers within the yard. This problem is compounded by the costs associated with moving containers between vessels and the yard, adding a crucial dimension of optimization to the allocation process. In this case, the objective function is to minimize the costs derived from transfering containers between vessels and yard.

2 Optimization approach

A Greedy Randomized Adaptive Search Procedure (GRASP) [2] is here proposed to solve the BAP. This iterative metaheuristic starts by constructing initial solutions through a greedy approach, selecting berth allocations based on a cost function that considers container movement expenses between vessels and the yard. The randomization component introduces diversity in solution space exploration, enabling a more comprehensive search. Adaptation comes into play

through local search procedures that iteratively improve solutions, adjusting berth assignments to minimize container movement costs. This iterative process balances exploration and exploitation, gradually refining solutions until an optimal or near-optimal allocation configuration is achieved, effectively optimizing the logistical challenges posed by the BAP with container movement costs.

The computational experiments have demonstrated that the GRASP is able to obtain high-quality solutions in short computational times when solving realistic scenarios composed of several dozens of vessels over the planning horizon.

3 Live tracking software

The live tracking software harnesses the wealth of real-time data from Automatic Identification System (AIS) sources, allowing comprehensive monitoring of vessel positions globally. Leveraging this AIS information [3], the software employs predictive algorithms to forecast vessel arrivals at various ports, enabling proactive planning and resource allocation. Integrating the GRASP within its framework, the software optimizes the arrangement of vessels within port infrastructures based on predicted arrival times. GRASP dynamically allocates berths, considering container movement costs between vessels and the yard, thereby orchestrating an efficient sequence of vessel arrivals and departures. By iteratively refining berth allocations through GRASP's adaptive and exploratory mechanisms, the software maximizes port infrastructure utilization, minimizes congestion and optimizes the overall logistical flow, ensuring smoother operations and enhanced efficiency within the complex maritime environment.

4 Conclusions

This study underscores the efficacy of integrating advanced optimization methodologies, particularly the GRASP, in addressing intricate challenges within port logistics. The application of GRASP in the BAP with considerations for container movement costs demonstrated its capability to generate optimized solutions by dynamically arranging vessel schedules within ports. The live tracking software proved invaluable in facilitating proactive planning and resource allocation. The synergy between predictive algorithms and GRASP-enabled optimization led to enhanced efficiency and maximized utilization of port facilities.

References

1. Imai, A., Nagaiwa, K., Tat, C. W.: Efficient planning of berth allocation for container terminals in Asia. Journal of Advanced Transportation **31**(1), 75–94 (1997)
2. Feo, T.A., Resende, M.G.C.: Greedy Randomized Adaptive Search Procedures. Journal of Global Optimization **6**(2), 109–133 (1995)
3. Ten, K. H., Kang, H. S., Siow, C. L., Goh, P. S., Lee, K. Q., Huspi, S. H., & Soares, C. G. (2023). Automatic identification system in accelerating decarbonization of maritime transportation: The state-of-the-art and opportunities. Ocean Engineering, 289, 116232.

Incrementally Solving the Dynamic Stacking Problem *

Sebastian Leitner[1,2][0000−0003−4261−2868], Stefan Wagner[1][0000−0002−0323−9274], and Michael Affenzeller[1,2][0000−0001−5692−5940]

[1] Josef Ressel Center for Adaptive Optimization in Dynamic Environments
Heuristic and Evolutionary Algorithms Laboratory
University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria
[2] Institute for Formal Models and Verification
Johannes Kepler University, 4040 Linz, Austria

sebastian.leitner@fh-hagenberg.at

Stacking problems are important everywhere from container terminals to steel plants. Wherever there is a gantry crane there is a desire to operate it as efficiently as possible and to minimize the relocation effort. There is no shortage of static problem variants and corresponding solvers [2]. Sadly, these are insufficient for applications in the real world where dynamic effects and uncertainties must be considered and problems must be solved online.

In this paper, we investigate a simulation-based dynamic stacking problem with uncertainty[3] implemented as part of our benchmarking framework for dynamic online problems [1]. Blocks arrive at a dedicated location in unknown intervals they then need to be relocated to one of the multiple buffer stacks where they await being declared as ready. When ready the blocks must be put onto a transporter in time for further processing. The goal is to optimize crane operations by ensuring that the arrival location is not blocked the blocks are delivered in time and the crane does not perform unnecessary relocations. The challenge is that there are multiple sources of uncertainty. Future blocks have an unknown arrival, ready, and due date. Blocks already in the system have a fixed due date but an unknown ready date. The future availability of a transporter at the handover location is also uncertain.

To solve the dynamic stacking problem we propose a framework for incremental online optimization that can integrate the previously published iterative tree-search-based solvers to significantly improve their performance. The central piece of our incremental online optimizer is the so-called State Manager it is responsible for keeping track of three kinds of state. The known state of the world in the present and past, our plans, and our predictions for uncertain events in the future. This enables the State Manager to react to change events, generated by the so-called Change Detector, by either adapting the existing plan and predictions or reoptimizing in case the existing plan is no longer viable. Adapting an existing plan is considerably faster than re-optimization which leads to increased

* The financial support by the Austrian Federal Ministry for Digital and Economic Affairs and the National Foundation for Research, Technology and Development and the Christian Doppler Research Association is gratefully acknowledged.

solver performance as we show in our experiments. The time saved by not having to reoptimize very similar problems over and over can be used to proactively solve for likely future events. In this case, the State Manager manages multiple solutions for possible futures and uses them appropriately as it becomes clear which one fits reality best. This enables the optimizer to quickly adapt to the changes that occur again improving performance.

Table 1. Preliminary results show incremental solver outperforming.

Type	Blocked arrival			Blocks on time			Moves/Block		
	Rule	Iter	Incr	Rule	Iter	Incr	Rule	Iter	Incr
6A	0.7	0.3	**0.0**	244.8	**273.4**	272.3	3.417	**2.940**	3.407
6B	5.8	0.9	**0.0**	233.7	**304.8**	304.0	3.511	**2.769**	3.160
6C	21.4	1.3	**0.0**	158.6	302.7	**316.0**	4.943	**2.875**	3.039
6D	18.1	0.6	**0.0**	216.4	218.8	**219.0**	**3.030**	3.069	3.531
9A	83.1	16.4	**0.0**	214.7	248.2	**255.0**	2.999	**2.926**	3.123
9B	66.8	4.7	**0.0**	205.2	266.0	**278.6**	3.064	**2.818**	2.910
9C	62.3	44.7	**0.0**	158.6	213.2	**272.9**	3.702	3.548	**2.906**
9D	79.9	37.2	**0.0**	206.1	210.0	**214.8**	**2.618**	2.885	3.165

Using a collection of challenging problem configurations with varying sizes and characteristics we sample problem instances. The new incremental approach is compared to both the rules-based and iterative model-based solver from [3]. Preliminary experiments are summarized in Table 1. It shows improvements in the primary objective called average blocked arrival time across all configurations. As well as improvements in the average number of blocks delivered before their due date in all but the first two configurations. The metric for crane moves per block delivered on time is less clear which makes sense since the number of moves required increases when the arrival is not blocked.

While we used the dynamic stacking problem to demonstrate the viability of our approach we want to emphasize that the general framework is applicable to a wide range of dynamic optimization problems with uncertainty. It can easily incorporate an existing optimizer and improve overall perfomance.

References

1. Beham, A., Leitner, S., Karder, J., Werth, B., Wagner, S.: Dynstack - a benchmarking framework for dynamic optimization problems in warehouse operations. In: Proceedings of the 2022 GECCO Conference. pp. 1984–1991 (Jul 2022). https://doi.org/10.1145/3520304.3533957
2. Lersteau, C., Shen, W.: A survey of optimization methods for block relocation and premarshalling problems. Computers and Industrial Engineering **172**, 108529 (2022). https://doi.org/10.1016/j.cie.2022.108529
3. Raggl, S., Beham, A., Wagner, S., Affenzeller, M.: Solution approaches for the dynamic stacking problem. In: Proceedings of the 2020 GECCO Conference. pp. 1652–1660 (2020). https://doi.org/10.1145/3377929.3398111

Predicting the processing effort for block relocation problems

Roland Braune[1,2,3][0000−0003−4086−229X] and Michael Raunig[2]

[1] University of Applied Sciences Upper Austria, Operations Management, Austria
[2] Department of Business Decisions and Analytics, University of Vienna, Austria
[3] Josef Ressel Center for Adaptive Optimization in Dynamic Environments, University of Vienna, Austria
roland.braune@univie.ac.at

The block relocation problem (BRP) is concerned with minimizing the number of relocations needed to retrieve items from a storage yard in a predefined order. The yard is organized into bays each of which is of two-dimensional type, meaning that the stored items are organized into multiple stacks of potentially different height. The items are assumed to carry integer numbers reflecting their retrieval priority, with lower numbers being more urgent to be retrieved. Figure 1 shows an example problem with 3 stacks, each of height 3, indicating the first item to be retrieved with lightgray background color. It usually has to be expected that the items are not stored in retrieval order, making it necessary to relocate other items on top of them before they can be picked up and moved to a handover stack for further processing or transport. This problem setting most prominently arises in container terminals, but also, for example, in steel production facilities where large quantities of steel slabs are stacked in outdoor storage yards prior to further processing such as hot rolling.

7	4	9
1	5	3
8	2	6

Fig. 1. A block relocation problem involving 9 items in a 3×3 stack layout.

The latter scenario serves as the background and motivation for the research presented in this paper. In this specific context, the block relocation problem occurs as a subproblem of a pickup-and-delivery vehicle routing problem (PDVRP) [1]. The primary goal is to fulfill transportation requests for steel slabs in the most efficient manner. Hence, besides the routing aspect itself that focuses on the minimization of the total travel time, the actual pickup of specific slabs from storage yards may consume a substantial amount of time and is therefore also subject to optimization. Although the block relocation problem encountered in this real-world setting differs from the academic one in that usually only a subset of slabs have to be retrieved from a bay, the underlying goal remains the same, namely minimizing the number of necessary relocations.

The nesting of the block relocation relocation problem within a vehicle routing problem is currently addressed in a hierarchical fashion. In fact, a large neighborhood search (LNS) [4] algorithm is employed for the VRP part of the problem. At any pickup node, the algorithm has to solve a corresponding block relocation problem to be able to get accurate timing results for the visit time window. Since the LNS evaluates thousands of different solutions during its trajectory through the search space, the computation time needed to solve the embedded block relocation problems is of crucial importance. During the neighborhood evaluation phase, a quick and accurate estimate of the number of relocations needed to clear the slab(s) to be retrieved would be sufficient to quickly assess the quality of the generated solutions. To match both the swiftness and accuracy requirements, we devise an approach based on predictive analytics.

Besides the above mentioned goals, our approach focuses on simplicity of usage and spares complex feature engineering. The main idea is to provide the configuration of the bay to a deep neural network for regression *as is*. This means that, for example, the BRP shown in Figure 1 is directly taken in its (natural) matrix representation and fed into the neural network's input layer. This way of input processing captures the BRP as if it was an image. Such kind of input structure can most directly be handled by convolutional neural networks (CNNs). However, the configuration of such a neural network, including the determination of the number of layers and their concrete form, can have a considerable impact on the quality of the prediction. Therefore, a major aspect of our research was to find an effective layout for the employed CNN.

The training data was generated by solving BRP benchmark instances as introduced by Caserta et al. [2]. The considered instances involve 3 to 7 stacks with heights ranging from 3 to 5. To obtain the exact number of relocations for each of those instances, we used a tailored version of the iterative integer programming-based approach initially proposed by Liu et al. [3]. This method allows to generate optimal solutions even for the larger instances under consideration. Extensive computational experiments give evidence of the accuracy of our regressor, also bearing the comparison with conventional regression techniques and various bound-generating approaches.

References

1. Battarra, M., Cordeau, J.F., Iori, M.: Pickup-and-Delivery Problems for Goods Transportation, chap. 6, pp. 161–191. https://doi.org/10.1137/1.9781611973594.ch6
2. Caserta, M., Voß, S., Sniedovich, M.: Applying the corridor method to a blocks relocation problem. OR SPECTRUM **33**(4), 915–929 (2011)
3. Liu, S., Liu, S., Lu, C.: An integer programming formulation and iterative scheme algorithm for the block relocation problem. Journal of control and decision pp. 1–10 (2022)
4. Pisinger, D., Ropke, S.: Large Neighborhood Search, pp. 399–419. Springer US, Boston, MA (2010). https://doi.org/10.1007/978-1-4419-1665-5_13

Modelling Electric Vehicle Routing Problem with Heterogeneous Fleet for Simultaneous Pickup and Delivery

Prateek Gupta[1], Devanand[1], Antonio Ken Iannillo[1], Jorge Augusto Meira[1], State Radu[1], and Danilo D'Aversa[2]

[1] SnT, University of Luxembourg, 29 Av. John F. Kennedy, 1855 Luxembourg
{prateek.gupta,devanand.devanand,antonioken.iannillo,
jorge.meira,state.radu}@uni.lu
[2] Gulliver Luxembourg S.à r.l. danilo.daversa@gullivernet.lu

Abstract. Vehicles that run on fossil fuel emit numerous air pollutants, and transitioning from conventional vehicles to their electric counterparts within a transportation fleet significantly reduces overall emissions. Electric vehicles are now an economically viable mode of transportation with decreasing cost of batteries, increasing power output, and rapid expansion of charging infrastructure. The Electric Vehicle Routing Problem (EVRP) aims to create routing plans for customers using a fleet of electric vehicles with diverse battery constraints. This paper proposes a new mathematical model for solving simultaneous pickup and delivery vehicle routing using a heterogeneous electric and fossil fuel vehicle fleet. The model is tested on logistics partner data using a standard CPLEX solver, and we observe promising results.

Keywords: simultaneous pickup and delivery · EVRP · transportation · heterogeneous fleet

1 Motivation for a new model for mixed fleets with EVs

There is growing interest in the logistics sector to utilize electric vehicles for transportation, driven by incentives promoting green technologies and reducing reliance on fossil-fuel-based vehicles. With improved battery performance and reduced battery maintenance, electric vehicles can be cost effective in the long term. However, electric vehicles have high initial costs and limited driving range due to frequent charging. In addition, they are susceptible to battery degradation on prolonged usage and have substantial battery drainage when driven at maximum speed. With these challenges, fossil-fuel-based vehicles need to gradually transition away instead of being completely replaced with electric vehicles. Such a heterogeneous fleet of fossil-fuel-based vehicles and EVs would have to address the complexities of the standard operating behaviour of EV batteries. While there is extensive research on heterogeneous fleets with fossil-fuel vehicles, there is limited study on heterogeneous fleets that include electric vehicles.

Papers closely related to our work focus on the impact of load on energy consumption [1] and allowing partial recharging by multiple chargers [4]. We focus on simultaneous pickup and delivery to allow better utilization of the available single charge. We refer the reader to [2] for further related work on EVRPs.

2 Formulation using heterogeneous fleet

Employing a heterogeneous fleet with simultaneous pickup and delivery in vehicle routing provides flexibility in resource utilization with varied customer requirements. We consider constraints on the speed and the capacity of each vehicle. We also consider changes in traffic during rainfall and peak hours. An increased travel time can make the remaining route infeasible for an EV with insufficient charge. We further allow adherence to government regulations by imposing restrictions on overall journey time and distance for a driver. We assume that EVs are fully charged at the depot before the start of the journey and a buffer time for charging is included. To solve this scenario, we propose a new formulation for the capacitated vehicle routing problem with a heterogeneous fleet of electric and fossil fuel vehicles for simultaneous pickup and delivery. We apply custom cuts that improve the solver's performance, such as lazy constraints similar to subtour elimination [3]. This problem aims to find the minimum distance to assign routes from the depot to all customers' locations with simultaneous pickup and delivery of items using a fleet of vehicles with constraints on capacity and speed. Some well-known green energy-based objective functions can easily replace our existing distance-based objective.

3 Experiments

For our experiments, we partnered with a logistics company that provides real world data to test our model. The model is a mixed-integer linear program (MILP). To solve it, we use CPLEX 22.1.1.0. The initial experimental results for a realistic number of customers are promising. We expect further improvement in performance with refined custom cuts using lazy evaluation. Future extensions to this formulation can include time windows, split deliveries, allowing multiple attempts and multiple charging stations for longer journeys with EVs.

References

1. Goeke, D., Schneider, M.: Routing a mixed fleet of electric and conventional vehicles. European Journal of Operational Research **245**(1) (2015)
2. Kucukoglu, I., Dewil, R., Cattrysse, D.: The electric vehicle routing problem and its variations: A literature review. Computers & Industrial Engineering (Nov 2021)
3. Laporte, G., Mercure, H., Nobert, Y.: An exact algorithm for the asymmetrical capacitated vehicle routing problem. Networks **16**(1), 33–46 (1986)
4. Sassi, O., Cherif, W.R., Oulamara, A.: Vehicle Routing Problem with Mixed fleet of conventional and heterogenous electric vehicles and time dependent charging costs (2014)

A Predictive Maintenance Concept for Sustainable Lubricant Oil Usage based on Federated Learning

Hadi Ghaeni[1], Ferdinand Heinrich[2], Franz Wenninger[2],
Tim Egger[1] and Benjamin Kormann[1]

[1] Munich University of Applied Sciences, Lothstraße 64, 80335 München, Germany
{hghaeni,tegger,kormann}@hm.edu
[2] Fraunhofer-Institut für Elektronische Mikrosysteme und Festkörper-Technologien (EMFT), München, Germany
{ferdinand.heinrich,franz.wenninger}@emft.fraunhofer.de

Introduction Good machine lubrication is essential for the smooth operation of production facilities. Regular analysis of the lubricant can provide early insights into possible suboptimal machine running conditions. These can occur trough oil aging at the end of the lubricant lifetime or through spontaneous changes in the tribological system. Therefore, online monitoring of the tribological system using sensors could help identify failures in machine parts. Furthermore, normal aging processes such as additive depletion and lubricant oxidation could be quantified and used to predict the remaining useful lifetime and therefore the optimal lubricant change. The SmartGear project pursues a novel approach using sensor and machine learning technology to aim towards a predictive maintenance solution. Real time analysis of sensor data is used to monitor the oil condition and to predict remaining useful life (RUL), which enables efficient maintenance plans.

Federated Learning Model for Lubricant Oil Condition In a federated learning setup machine learning is performed on separate nodes with individual data-sets (local or global systems) [1]. To aggregate the knowledge from the nodes, the model parameters are sent to a global model, but no sensitive raw data is transferred, see Fig. 1. This approach enables to train individualised models for each node and to aggregate the knowledge of each node without transferring sensitive raw data [2].

The design is based on a real-world test setup used to assess the aging of lubricant oil [3]. This architecture includes multiple local training nodes and a central global node serving as an aggregator. Each local training node is equipped with lubricant oil sensor unit which acquires real time sensor data like temperature, humidity, permittivity and conductivity. These need to be translated into DIN conform measures of lubricant condition, such as Total Acid Number (TAN), additive content, water content etc [4]. The goal of each local node is now to predict the lubricant oil analysis values from the local sensor values and past lubricant oil analysis values.

The sensor data from each local training node is labelled using the laboratory results at the time stamps when an oil sample was taken. This combination of

data are the respective local training data-sets with consistent feature spaces. Each local training node trains its local model independently. The global aggregator collects and aggregates model parameter updates from local nodes with a federated averaging (FedAvg) strategy [1]. The resulting global model parameters are transferred back to the local nodes, with no direct exchange of information between the local nodes.

First results performed on real data of this federated learning approach resemble those of a traditional machine learning system.

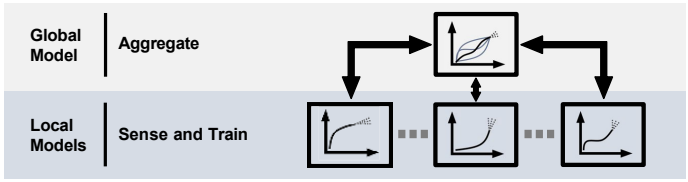


Fig. 1. Centralized Federated Machine Learning Architecture

Summary and next steps Due to the high quality of lubricants aging mechanisms yield data sets with stable conditions for extended periods of time before slowly deteriorating at the end of the lubricant lifetime. Therefore, data on changes in lubrication conditions is scarce and needs to be augmented. An objective is to improve the accuracy and efficiency by a federated machine learning simulation environment, where plausible additional data can be generated. Furthermore, it needs to be investigated whether the integration of existing federated learning frameworks with traditional machine learning methods will improve the quality of training and produce more accurate results. Preliminary results already show robustness when dealing with noisy data.

Acknowledgments This work has been supported by Josef Bernbacher & Sohn GmbH & Co. KG, Fraunhofer-Institut (EMFT), Klüber Lubrication München GmbH & Co. KG, Munich University of Applied Sciences funded by Bayerische Forschungsstiftung (BFS).

References

1. C. Nguyen et al.: Federated Learning for Internet of Things: A Comprehensive Survey IEEE Communications Surveys & Tutorials. Institute of Electrical and Electronics Engineers (IEEE) (2021). https://doi.org/10.1109/comst.2021.3075439
2. V. Pruckovskaja et al.: Federated Learning for Predictive Maintenance and Quality Inspection in Industrial Applications (2023). https://doi.org/10.48550/arXiv.2304.11101
3. E. Kimpl and M. Stroi: The 3 columns of gear condition monitoring (2019), OilDoc Conference Rosenheim, Rosenheim, 2019
4. J. Bhat and B. Sonawane and M. Desai: Condition Monitoring of Worm Gearbox through Oil Analysis (2020). https://doi.org/10.1007/978-981-16-1079-0_30

Introduction to Circular System Design and First Use Cases for Sustainable Product Development in Smart Meter Remanufacturing and Robotics

Nathanael Nafz[1[0009-0009-5420-7625]] , Bernhard Höfig[1[0009-0007-6646-9089]], Markus Glück[1[0000-0002-7631-5389]] and Niclas-Alexander Mauß[2]

[1] Aalen University of Applied Sciences, Beethovenstr. 1, 73430 Aalen, Germany
[2] CIRCULAR REPUBLIC / UnternehmerTUM GmbH, Freddie-Mercury-Str. 5, 80797 München, Germany
bernhard.hoefig@hs-aalen.de

Abstract: In this paper, we present the concept of "Circular System Design" as a new method for a holistic and sustainable design process. Selected examples from production technology and robotics are used to illustrate the concept and explain the future research framework.

Keywords: Sustainable Design, Circular Economy, Product Design, Robotics, Systems Engineering

1 Motivation

Resource efficiency and circularity are the key success factors for sustainable products. Product design as well as the methods and tools used in the design departments play a central role here. New approaches are pursuing the goal of a circular economy, with which economic development is largely decoupled from the need for new natural resources. In addition to the issues of economic efficiency, the material use and the business model, new methods and tools are needed for product design in order to focus stringently on circularity and resource efficiency. [1]

2 Circular System Design Approach

It is estimated that 80 % of costs and efforts associated with sustainability aspects of a product or a service are determined in the first 20 % of product design. This means that the key contributions and prerequisites for a circular design are laid at the beginning of the product development. The earlier a circular concept becomes the basis for system development, the greater the leverage in terms of resource efficiency and circularity. Circular thinking and action must be the central guideline for product development right from the beginning. We therefore propose a new approach for holistic product design that is based on a "*Circular System*" and moves away from the pure focus on a specific

product. The "Circular System" represents an agile design model which, in addition to the pure focus on functional requirements, also includes life cycle aspects (the "*Life Cycle System*") and interactions with the environment (the "*Context System*"). The result is a paradigm shift in business model generation and product development. The associated potential is illustrated in two selected use cases.

3 Use Cases

The first use case presents the remanufacturing concept of Lorenz, a leading supplier of water meters. These are state-of-the-art mechatronic systems that are regularly calibrated and enhanced measuring capabilities are added. Through consistent circular system design, the product and the underlying business model have been successfully improved in terms of sustainability and circularity. [2] In the second use case, we discuss approaches for "green" robotics. Today, robots are often not optimized with respect to sustainability aspects. The potential for extending their life cycle by innovative system design and smart operation is still underestimated. [3] Furthermore, the components and materials that are used, are in general not optimized for circularity. As a consequence, a paradigm shift towards "green" robotics is necessary based on the new circular system design approach. This will lead to substantial contributions to minimize energy consumption and zero emission production.

4 Conclusions

The new approach of "Circular System Design" replaces the traditional view on a product by a holistic, sustainable system design in a wider context. In addition to traditional economic design objectives, ecological and social aspects are addressed throughout the entire product life cycle. The development of new methods and digital tools is required to cope with the increasing system complexity in the design process.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. *S. Kadner et al., Circular Economy Roadmap für Deutschland*, acatech - Deutsche Akademie der Technikwissenschaften, 2021. doi: 10.48669/CEID_2021-4.
2. Herrmann, C.; Vetter, O.: *Ökologische und ökonomische Bewertung des Ressourcen-aufwands – Remanufacturing von Produkten*. VDI Zentrum Ressourceneffizienz GmbH (VDI ZRE), VDI ZRE Publikationen: Studien, 2021
3. D. Gritzner et al., *Synthesizing Energy-Efficient Production System Controllers that Exploit Braking Energy Recuperation*, IEEE Int. Conf. on Automation Sci. and Engg. (CASE), S. 519-528, 2018

Sustainable Architectures: Energy Efficiency in Technology and Applications

Enrique Ismael Mendoza Robaina[1,2][0000−0002−9892−0942], José María Silva Bravo[1,3], Alberto García Cabrera[1,4], and Diego Ambite Varona[1,5]

[1] IBM Global Services España, Madrid, Spain https://www.ibm.com
[2] enrique.ismael.mendoza.robaina@ibm.com
[3] josem.silva@ibm.com
[4] jose.alberto.garcia@ibm.com
[5] diego.ambite@ibm.com

Abstract. This paper compares technologies for energy efficiency, aiming to establish sustainable software and architectural patterns, and highlights a methodology for optimizing IT energy use.

Keywords: GreenIT · Sustainability · Energy Consumption · IT Architecture

1 Introduction

In the evolving landscape of sustainable technology, our study introduces an innovative method to measure and analyze the energy efficiency of various technological solutions. We focus on evaluating the impact of different databases, programming languages, and communication protocols on the energy consumption of IT architectures. This method involves a comprehensive approach to quantify energy usage, aiming to guide decisions towards more sustainable IT practices.

2 Expected benefits

The anticipated benefits of our research are multifold. Primarily, it aims to provide a deeper understanding of the energy implications of different technological choices. By identifying more energy-efficient technologies, organizations can reduce their carbon footprint and align with global sustainability goals. Additionally, our findings are expected to aid in cost reduction through energy savings and improve overall system performance, offering a dual advantage of environmental and economic sustainability.

3 Experiments Description

Our experimental approach involved a series of tests to compare the energy consumption across different technological dimensions. The technical experiments include comparisons among relational and non-relational databases, different programming languages, and communication protocols. The key findings are:

3.1 Database Comparison

It was found that NoSQL databases (MongoDB) consume significantly less energy compared to SQL databases (MySQL). This suggests that the choice of database type can have a considerable impact on the carbon footprint of an architecture.

3.2 Programming Languages

Among the evaluated languages (Java, Python, NodeJS), Java proved to be the most energy-efficient. Python was the least efficient, consuming almost double the energy compared to the others.

3.3 Communication Protocols

Comparing GRPC with HTTP/REST protocols, GRPC was found to be much more energy-efficient, indicating that the choice of communication protocol can also significantly contribute to energy efficiency.

4 Future work

Moving forward, our research will expand to include more diverse technologies and scenarios. Future work aims to develop a benchmarking framework for evaluating the energy efficiency of emerging technologies in IT. Additionally, we plan to explore the integration of AI and machine learning techniques to predict and optimize energy consumption in various IT architectures. This ongoing research will contribute to the development of guidelines and best practices for sustainable IT development. Furthermore, the study suggests continuing technical experiments to establish a baseline that helps define efficient and environmentally friendly architectures

References

1. Cormenzana, B., Marinescu, M.-C., Marrero, M., Mendoza, S., Rueda, S., Uceda-Sosa, R.: Models for sustainability. In: 2015 IEEE First International Smart Cities Conference (ISC2), pp. 1–6 (2015). https://doi.org/10.1109/ISC2.2015.7366221
2. Calero, C., Platinni, M.: Green in Software Engineering. Springer, Castilla-La Mancha (2015) ISBN:978-3-319-08580-7
3. Journey to Cloud Tales IV: ¿Arquitecturas elegantes como paradigma GreenIT?, https://www.muycomputerpro.com/2023/01/30/cloud-tales-greenit, last accessed 2023/4/10
4. How to Architect for Sustainability in a Cloud Native Environment, https://www.contino.io/insights/cloud-native-sustainability, last accessed 2022/12/11

Sky condition classification by sky-camera images

Ruymán Hernández-López[1[0000-0003-4611-1913]] and Carlos M. Travieso-González[1,2[0000-0002-4621-2768]]

[1] Institute for Technological Development and Innovation in Communications (IDeTIC). University of Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, 35001 Canary Islands, Spain
ruyman.hernandez@ulpgc.com
[2] Signals and Commuications Department. University of Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, 35001 Canary Islands, Spain
carlos.travieso@ulpgc.es

Abstract. Cloudy conditions on a local scale are a key issue for forecasting renewable energy generation using photovoltaic panels. Knowing the state of the sky is a valuable piece of information that can be used in the decision-making process involved in the operation of systems. For example, it can determine whether conditions are favorable for activating a stand-alone system that requires a minimum level of radiation or whether it may lead to higher consumption than the energy generated if the cloudy conditions are adverse. In this research, the performance of *EfficientNet* models for classifying cloudy conditions is studied.

Keywords: Cloudiness Classification, Photovoltaic Power, Renewable Energy, Deep Learning, Transfer Learning, EfficientNet Models, Sky Images.

Extended Abstract

The production of electric power from solar energy is affected by the changes caused by solar occlusion caused by clouds. Detecting and understanding cloud cover have been investigated for estimating and forecasting solar irradiance to predict photo-voltaic power generation [1]. Different techniques can be applied in the prediction of solar radiation but, nowadays, the techniques that are commonly applied are based on deep learning due to its good results and the rise of this type of technology. In the current state of the art, multiple studies using sky cameras can be found [2, 3, 4, 5, 6, 7]. This paper presents a study in which EfficientNet architectures are tested for the classification of sky images obtained with a hemispherical sky camera. The work developed makes use of the Transfer Learning classification technique where the pre-trained EfficientNet models are implemented as base models in the neural network architecture used in the experiments. “Fig. 1” shows a representative diagram of the complete architecture of the classification system. The implementation of this system has been carried out using the TensorFlow platform and the Keras library.

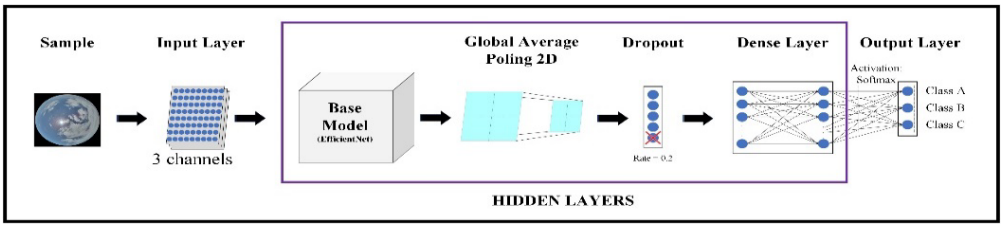


Fig. 1. Representative diagram of the architecture.

The following list presents the different base models used in this study: EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, EfficientNet-B3, EfficientNet-B4, EfficientNet-B5, EfficientNet-B6 and EfficientNet-B7. On the other hand, the database consists of a total of 4500 images showing different cloud conditions and is labelled with the following classes: clear sky, cloudy sky and partly cloudy sky.

The best result of all implementations was obtained with the EfficientNet-B1 base model, achieving an average accuracy of **98.07%** with a standard deviation of 0.32. In the final paper, more information about the capture of the data and the used architecture for the identification.

Acknowledgments. This work is supported under the Grant CEI 2021-06, from direct agreement SD-21/08 by Consejería de Economía, Industria, Comercio y Conocimiento from Gobierno de Canarias to ULPGC.

References

1. C.-L. Fu y H.-Y. Cheng, «Predicting solar irradiance with all-sky image features via regression,» *Solar Energy,* vol. 97, p. 537–550, November 2013.
2. S. Park, Y. Kim, N. J. Ferrier, S. M. Collis, R. Sankaran y P. H. Beckman, «Prediction of Solar Irradiance and Photovoltaic Solar Energy Product Based on Cloud Coverage Estimation Using Machine Learning Methods,» *Atmosphere,* vol. 12, p. 395, March 2021.
3. Q. Paletta, G. Arbod y J. Lasenby, «Omnivision forecasting: Combining satellite and sky images for improved deterministic and probabilistic intra-hour solar energy predictions,» *Applied Energy,* vol. 336, p. 120818, April 2023.
4. R. A. Rajagukguk, R. Kamil y H.-J. Lee, «A Deep Learning Model to Forecast Solar Irradiance Using a Sky Camera,» *Applied Sciences,* vol. 11, p. 5049, May 2021.
5. Q. Paletta, G. Arbod y J. Lasenby, «Benchmarking of deep learning irradiance forecasting models from sky images – An in-depth analysis,» *Solar Energy,* vol. 224, p. 855–867, August 2021.
6. P. Manandhar, M. Temimi y Z. Aung, «Short-term solar radiation forecast using total sky imager via transfer learning,» *Energy Reports,* vol. 9, p. 819–828, March 2023.
7. Y. Karout, S. Thil, J. Eynard, E. Guillot y S. Grieu, «Hybrid intrahour DNI forecast model based on DNI measurements and sky-imaging data,» *Solar Energy,* vol. 249, p. 541–558, January 2023.

Enhancing Solar Energy Forecasting Precision: A Comprehensive Analysis Using an Ensemble of Neural Networks

Carlos M. Travieso-González[1] and Sergio Celada-Bernal[1]

[1] Signals and Communications Department (DSC), Institute for Technological Development and Innovation in Communications (IDeTIC), University of Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, Spain
carlos.travieso@ulpgc.es, sergio.celada101@alu.ulpgc.es

Abstract. This study explores the application of an ensemble of 1000 GRU neural networks to forecast solar energy generation. The results reveal significant improvements in precision by reducing prediction variability. Uncertainty analysis indicates that approximately 85% of predictions fall within one standard deviation. These findings endorse the potential of this approach to enhance the management of solar energy systems and drive the adoption of renewable energy. This study provides a robust foundation for future research and advancements in the field of solar energy.

Keywords: Ensemble Methods, Neural Networks, Solar Energy.

1 Introduction

The research emphasizes the pivotal role of solar energy in fostering a sustainable energy system. With the exponential growth of solar power, particularly through photovoltaic panels, there is a pressing need to optimize efficiency by accurately predicting power values from solar inverters [1]. The study explores the potential of neural network technology, specifically the Gated Recurrent Unit (GRU), known for its effectiveness in time series prediction and renewable energy system optimization [2].
The central question addressed is whether an ensemble of 1000 GRU neural networks can significantly improve prediction accuracy compared to conventional methods or individual networks. The goal is to maximize the ability to capture the variability and complexity of solar data, potentially influencing the efficiency and profitability of solar energy systems. The research relies on three years of data (2016-2018) from a 3.3 kW peak inverter connected to 22 150 W panels, sampled every 15 minutes.

2 Methodology

This study introduces a predictive method for solar power inverter output utilizing an ensemble of 1000 neural networks. The choice of an ensemble arises from the variability observed with a single neural network, acknowledging that the optimal network may not fully capture all training data information. The ensemble, achieved through result averaging, exhibits decreased error with an increasing number of contributing networks.

3 Results

The study assesses the model's stability concerning the ensemble's number of neural networks. Table 1 displays nRMSE values for 2016 and 2017, varying the number of networks. Additionally, Fig. 1 visually depicts the impact of changing the ensemble size on nRMSE. Fig. 2 utilizes 2018 inverter data to showcase model uncertainty and prediction error distribution, providing insights into the model's reliability [3].

Table 1. Table captions should be placed above the tables.

	1 network	10 networks	100 networks	300 networks	1000 networks
nRMSE (%) 2017	28,22	28,11	28,10	28,09	28,09
nRMSE (%) 2018	28,23	28,12	28,07	28,06	28,06

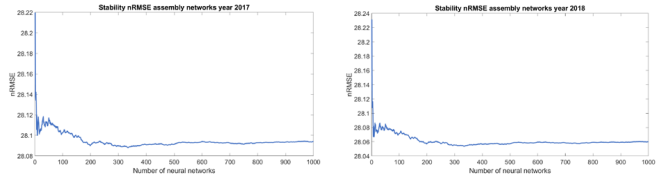


Fig. 1 Analysis of the stability of the nRMSE as a function of the number of nets

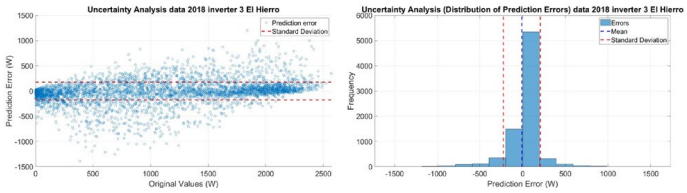


Fig. 2. Model uncertainty analysis.

Acknowledgments. This work is supported under Grant MAC/1.1.b/278 (MICROGRID-BLUE), from the Framework of the 2nd Call of the INTERREG V-A MAC 2014-202, co-financed with FEDER funds. Too, this work is supported under the Grant CEI 2021-06, from direct agreement SD-21/08 by Consejería de Economía, Industria, Comercio y Conocimiento from Gobierno de Canarias to ULPGC.

References

1. J. A. García-Sánchez, J. M. Sánchez-Lozano, and A. García-Cerrada, "Deep neural networks for wind and solar energy prediction," Neural Processing Letters, vol. 47, no. 3, pp. 1095-1109, 2018. doi: 10.1007/s11063-017-9613-7.
2. Y. Zhang, Y. Liu, and J. Li, "Short-term solar energy forecasting: Integrated computational intelligence of LSTMs and GRU," PLOS ONE, vol. 18, no. 10, 2023. doi: 10.1371/journal.pone.0285410
3. D. Chakraborty, J. Mondal, H. B. Barua, and A. Bhattacharjee, "Computational solar energy – Ensemble learning methods for prediction of solar power generation based on meteorological parameters in Eastern India," Renewable Energy Focus, vol. 44, pp. 277-294, 2023. doi: 10.1016/j.ref.2023.01.006.

Surrogates for Fair-Weather Photovoltaic Module Output

Dominik Falkner[1,2], Michael Bögl[1], and Michael Affenzeller[2,3]

[1] RISC Software GmbH, Hagenberg

[2] Institute for Formal Models and Verification, Johannes Kepler University, Linz

[3] University of Applied Sciences Upper Austria, Hagenberg

dominik.falkner@risc-software.at

Background and Motivation The success of the energy transition and the future stability of energy networks are deeply tied to the output of renewable energy like photovoltaic (PV) modules. International Renewable Energy Agency reports the need for a rapid increase in the deployment of renewables to meet the 1.5 ℃ climate goal, they predict a seven-fold increase by 2030 [1]. The federal association for photovoltaics in Austria, project that PVs produce 15 percent of the total consumption by 2030 [2]. As PVs gain prominence in the energy mix, so does their impact on the power grid's balance, which is especially significant for energy providers. Furthermore, compared to large PV power plants, the increasing number of small, privately owned PV systems has relatively greater installation diversity per kilowatt produced. Therefore, there is a need to create tools that can simulate PV module output across a diverse set of parameters. In this paper, we tackle the issue by investigating the construction of surrogate models for fair-weather PV output and provide a benchmark for their performance as well as a process to build surrogates for PV modules.

Method At present, the PVGIS [3] tool is available for calculating fair-weather PV output. It is an online platform which offers hourly data on solar radiation and photovoltaic system efficiency for locations across Europe, Africa, Asia, and America. It estimates the output based on satellite image analysis, taking into account various installation parameters such as slope, azimuth angle, peak installed power, system loss, and the type of PV technology used. Nevertheless, PVGIS has a rate limit in place. Moreover, as a web-based application, it comes with inherent delays due to network transmission, which makes it unsuitable for larger-scale simulations. In the first step, we aim to build a crawler for PVGIS to amass data for training and evaluating surrogate models. The crawler facilitates downloading and storing data for parametrized modules. The resulting software tool will be made freely available via Github. For the purpose of building surrogates, state of the art models like NHITs [4] and PatchTST [5] are employed. NHITs is based on multilayer perceptrons, and PatchTST represents a transformer-based approach. Both have been shown to be effective in modeling time series and are suitable for the purposes proposed in this study. Moreover,

both support transfer learning, which is used to train from multiple datasets [6]. Additionally symbolic regression, a white-box model, is considered. It has shown proficiency in identifying nonlinear associations and generalizes well [7].

The study investigates the construction of surrogate models for different parameterized PV modules. For data collection, the crawler gathers fair-weather output data from locations sampled from predetermined areas around the world. The initial experiment evaluates the models' capability to accurately estimate the power output of these modules without varying parameters. In the subsequent experiment, parameters are varied for each location. This assesses the possibility of establishing a surrogate for different parameter configurations. The final experiment gauges the models' proficiency to generalize. It involves methodically enlarging the primary sampling region and subsequently testing the models using data from the newly included zones, thereby establishing a standard for accuracy within a specific region. Our research lays the groundwork for the development of surrogate models for PV modules, which offer a good approximation to the PVGIS service. Moreover, surrogates are fast enough to be used in real-time contexts. This study acts as a precursor for future studies which incorporate weather to mirror real-world PV outputs. For this purpose, a detailed description of the methodology and data will be made available.

Acknowledgements The work was done within the project "Secure Prescriptive Analytics", funded by the state of Upper Austria as part of the program "#upperVISION2030".

References

1. International Renewable Energy Agency: World Energy Transitions Outlook 1.5C Pathway 2022 edition. IRENA (2022), ISBN: 978-92-9260-429-5
2. Bundesverband Photovoltaic Austria: Die Österreichische Photovoltaic und Speicherbranche 2022 (visited on 2023-10-23), https://pvaustria.at/wp-content/uploads/2023_Fact_Sheet_PV_Branche.pdf
3. Joint Research Centre: Photovoltaic geographical information system PVGIS (visited on 2023-10-23), http://re.jrc.ec.europa.eu/pvgis
4. Challu, C., Olivares, K.G., Oreshkin, B.N., Garza, F., Mergenthaler-Canseco, M., Dubrawski, A.: N-hits: Neural hierarchical interpolation for time series forecasting (2022). https://doi.org/https://doi.org/10.1609/aaai.v37i6.25854
5. Nie, Y., H. Nguyen, N., Sinthong, P., Kalagnanam, J.: A time series is worth 64 words: Long-term forecasting with transformers. In: International Conference on Learning Representations (2023). https://doi.org/https://doi.org/10.48550/arXiv.2211.14730
6. Oreshkin, B.N., Carpov, D., Chapados, N., Bengio, Y.: Meta-learning framework with applications to zero-shot time-series forecasting (2020). https://doi.org/10.1609/aaai.v35i10.17115
7. Affenzeller, M., Winkler, S.M., Kronberger, G., Kommenda, M., Burlacu, B., Wagner, S.: Gaining deeper insights in symbolic regression. In: Genetic and Evolutionary Computation, pp. 175–190. Springer New York (2014). https://doi.org/10.1007/978-1-4939-0375-7_10, https://doi.org/10.1007/978-1-4939-0375-7_10

How can digital twins help to accelerate the transition to a carbon-neutral energy system?

Tobias Rodemann [1] and Christiane Attig [2]

[1] Honda Research Institute Europe, Germany
[2] University of Lübeck, Germany
tobias.rodemann@honda-ri.de, christiane.attig@uni-luebeck.de

Good investment decisions into green energy systems are hard due to high budget requests, complex systems, and multiple objectives to consider. Simulation models of current and potential future energy systems (so called digital twins) can help to generate a deeper understanding and better recommendations. We developed detailed simulation models for our R&D facility (see Fig. 1).

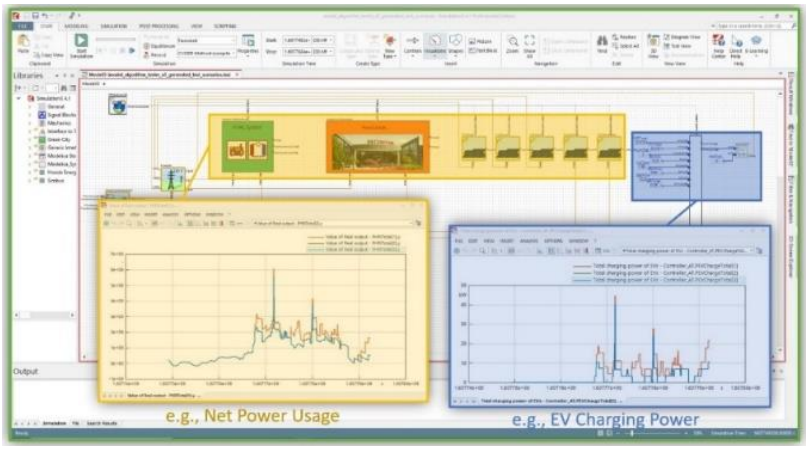


Fig. 1. Digital Twin framework of our facility including PV system, charging stations and co-generator.

The simulation model was coupled to different optimization approaches [1,2]. We could identify optimal energy system configurations based on current and expected demand patterns and possible (hardware) extensions of the current system. The information provided by the digital twin was generally considered as novel and valuable, but often did not trigger a direct action. We identified two aspects that delay or prevent these investment decisions: First, the precision of the simulation is limited, due to various factors such as model simplifications, or unclear future conditions. As a result, the accuracy of simulation results is very hard to assess. Second, there is a general lack of trust in complex, automated simulation and optimization tools [3]. In combination this might lead to no decisions being made even if the analysis shows a clear benefit of an investment. We therefore propose to model these factors explicitly in a psychological model of the decision maker. As a main contribution we introduce a conceptual framework that might serve as a groundwork for modeling the psychological processes of a decision maker as part of the simulation and optimization process. This framework is

based on a control-theory model of self-regulation [4], which has been applied to other areas of energy efficiency behavior (e.g. [5]). According to psychological control theory, behavior can be understood as resulting from a negative feedback loop that aims at diminishing any discrepancy between a goal status and the perceived status [4]. This means that the user perceives the current simulated state of the energy system (i.e., input), which is compared to a pre-defined energy efficiency goal (i.e., reference value). However, a second feedback loop has to be considered: The user also compares expected and perceived system performance. The result of this comparison is the user experience. Consequently, if there is a discrepancy between the perceived actual and the expected system performance, frustration might follow. Moreover, the perceived system performance also determines the perceived traceability of the system (i.e., transparency, understandability, predictability [6]), which serves as a prerequisite for trust in the system. Hence, even if the optimization tool suggests practicable solutions for enhancing energy efficiency, the decision to follow the system's suggestions can be impeded by aspects of user experience. From this conceptual framework we derive future research directions which ultimately lead to the question: How can digital twins support a human decision maker to trigger the right investment for a greener energy system?

References

1. Pour, P.A., Rodemann, T., Hakanen, J., Miettinen, K.: Surrogate assisted interactive multiobjective optimization in energy system design of buildings. Optim. Eng. **23**, 303–327 (2021)
2. Liu, Q., Lanfermann, F., Rodemann, T., Olhofer, M., Jin, Y.: Surrogate-assisted many-objective optimization of building energy management. IEEE Comput. Intell M. **18**, 14–28 (2023)
3. Muir, B.M. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics, **37**, 1905–1922 (1994)
4. Carver, C.S., Scheier, M.F. On the structure of behavioral self-Regulation. In Boekaerts, M., Pintrich, P.R., Zeidner, M. (eds.). Handbook of Self-Regulation, pp. 41–84. Academic Press, Cambridge (2000)
5. Arend, M.G., Franke, T., Stanton, N.A. (2019). Know-how or know-why? The role of hybrid electric vehicle drivers' acquisition of eco-driving knowledge for eco-driving success. Appl. Ergon. **75**, 221–229 (2019)
6. Schrills, T., Franke, T. How do users experience traceability of ai systems? examining subjective information processing awareness in Automated Insulin Delivery (AID) systems. ACM Trans. Interact. Intell. Syst. (2022)

Energy Management System - Design and Implementation

Lukasz Korus[1 [0000−0001−7184−679X]] and Andrzej Jablonski[2 [0000−0001−9442−7678]]

Wroclaw University of Science and Technology,
Faculty of Information and Communication Technology,
Janiszewskiego st. 11/17, 50-372 Wroclaw, Poland.
lukasz.korus@pwr.edu.pl
andrzej.jablonski@pwr.edu.pl

Abstract. The article addresses the issue of designing and implementing an Energy Management System in small and medium-sized enterprises, starting from the high-level business needs, functional and non-functional requirements, through the architectural aspects of infrastructure and software, and concluding with proposals for analysis of the collected data. Furthermore, the article provides information about practical challenges encountered during the implementation of such complex systems and potential directions for their further development. This topic is becoming increasingly essential today due to the rising energy costs, particularly for small and medium-sized companies that rarely use multi-level Decentralized Control Systems (DCS) to optimize energy consumption.

Keywords: Data acquisition and presentation systems · IIoT systems · EDGE computing · Energy management

1 Extended Abstract

This article tackle the issue of energy management, which is understood as the effort to minimize energy consumption in industrial facilities through a continuous improvement approach [3]. Energy management is crucial, not only from a financial perspective, but also in light of the growing problem of climate warming. This challenge requires significant changes for companies [4], that have not previously prioritized energy consumption reduction. This transformation demands an unavoidable investment, which must be carefully planned and gradually implemented to avoid disrupting ongoing operations. The paper presents the entire process of designing and implementing a complex data acquisition and presentation system. This system, in conjunction with a Deming-like cycle approach, will assist companies in reducing their energy consumption. The article provides information on high-level business goals and both functional and non-functional requirements. Furthermore, it offers detailed insights into the architectural aspects of the infrastructure and software, as well as the algorithms that support efficient data analysis [1]. Additionally, the article discusses the most significant

issues related to selected non-functional requirements [2]. The final section delves into the matter of further system development, addressing some of the problems mentioned in the preceding section. The initial step involves a comprehensive analysis of detailed requirements, beginning with high-level business needs and encompassing both functional and non-functional requirements. This phase is fundamental in modern project management to gain a deep understanding of the business challenges and to ultimately choose the most suitable technical solutions for addressing them. Gathering certain requirements at the project's outset is crucial, even in Agile-based methodologies such as Scrum. This early requirement gathering helps to comprehend the project's scope, limitations, and potential obstacles from the very beginning of the process. In accordance with the specified requirements, it is possible to define the architectural aspects of both hardware (infrastructure) and software. The most advantageous approach is to present several solutions to the business stakeholders, outlining their respective advantages and disadvantages, as well as their financial implications. This approach allows them to select a solution that is not only well-suited for their needs, but also aligns with their financial considerations.This article introduces two alternative options: a straightforward architecture centred around an RPI computer, and a more sophisticated solution built upon IIoT infrastructure. In the following step, authors provide a comprehensive description of all the details and issues that arose during the implementation phase, with a primary focus on those related to non-functional requirements. This article addresses challenges pertaining to Big Data topics and considers solutions within this domain. Additionally, authors explore EDGE computing as a potential solution to address a significant portion of these issues. Some of the algorithms that support users in gaining a better understanding of energy consumption structure are presented along with references. Additionally, a compound method of data analysis is proposed.

All the primary issues encountered during the implementation phase are outlined, and some potential solutions are suggested for the further development of the presented system. In summary, the results of the proof of concept (PoC) and real-world business cases are provided to demonstrate the effectiveness of such solutions in practical situations.

References

1. Korus, L., Piorek M.: Compound method of time series classification. Nonlinear Analysis: Modelling and Control. **20(4)**, 545–560 (2019), doi: 10.15388/NA.2015.4.6.
2. Korus, L., Andrzej J.: Aspects of Measurement Data Acquisition and Optimisation in the Energy Transformation of Industrial Facilities, Advanced, Contemporary Control, PCC2023, 2023. 39–48 (2023).
3. Prashar A.: Adopting PDCA (Plan-Do-Check-Act) cycle for energy optimization in energy-intensive SMEs. Journal of Cleaner Production. **145**, 277–293 (2017)
4. Frasner J., Morea F., Krenn C., Uson J.A., Tomasi F.: Energy efficiency in small and medium enterprises: Lessons learned from 280 energy audits across Europe. Journal of Cleaner Production. **142(4)**, 1650–1660 (2017)

Performance and computation time gains caused by sampling rate reduction in time series deep anomaly detection

Dominik Wiesner [1][0009−0009−7689−3283] and
Frank Schirmeier[1][0009−0002−2842−0376]

[1]University of Applied Sciences Kempten, Bahnhofstraße 61, 87435 Kempten, Germany
dominik.wiesner@hs-kempten.de, frank.schirmeier@hs-kempten.de

Abstract. **Keywords:** Green AI · Deep Anomaly Detection · Sampling Rate Reduction · Time Series · Machining.

1 Introduction

Green AI is endeavouring to increase the efficiency of deep learning models. This can be achieved by reducing the cost of the result (R) in equation (1) [1].

$$Cost(R) \propto E \cdot D \cdot H \tag{1}$$

The cost (R) depends on the cost of running a single training example (E), the size of the data set (D) and the number of hyperparameter experiments (H). Consequently, among other things, the efficiency of an AI model can be increased by reducing the time required for a training example, and also by reducing the amount of data required for the models. This paper examines the impact of these measures on the performance of deep anomaly detection (DAD) in the context of semi-supervised methods on CNC-Machining time series. The results are discussed in the broader context of the research project DIONE-X, in which the impact of novel data reduction and obfuscation schemes on innovative data-driven business use cases in the milling industry is studied.

2 Methodology

The machining data set introduced in [2] is used for the experiments. The data set covers acceleration data in the X, Y and Z directions on 3 CNC machines over 3 years in production. The time series data is divided into chunks that have binary labels (OK/NOK) according to process status. To reduce the amount of training data (D), the data set is decimated by means of sampling rate conversion. For the analyses, the entire data set is reduced to 4 further iteration levels with 1000 Hz, 500 Hz, 250 Hz and 125 Hz in addition to the full resolution of 2000 Hz. This halves the size of the data set with each iteration. As a result, the

data volume in the last iteration stage is 1/16 of the original data volume. For the DAD methods used, the recorded sensor data is segmented into windows. To capture an identical semantic context window, the temporal context of 2 seconds is used for each clock rate reduction stage. As a result, the number of data points of the segmented windows and thus the execution and training time of a single training example is halved for each iteration. The DAD methods analysed are 2 reconstruction methods and 1 semi-supervised prediction method. Semi-supervised methods use only "OK" data as training data. The reconstruction methods calculate an anomaly score using convolutional autoencoders, which are trained on the raw data (1D) and on spectrograms (2D). For the prediction method, LSTM networks are trained on raw data. To ensure a sufficiently large latent space, the number of layers within the methods is varied. The training and the analyses are carried out at the corresponding resolutions. The AUC-ROC of the respective inference models is calculated for evaluation.

3 Conclusion

In summary, the detection performance of the models improves in initial data decimation steps (with decreasing sampling rate and thus lower data volume), while strong decimation leads to a loss of anomaly detection capability. The AUC-ROC increases from 2000 Hz to 1000 Hz to 500 Hz. This increase can arguably be attributed to the beneficial effect of noise smoothing. A large performance drop can be recognised between 500 Hz and 250 Hz. From 250 Hz onwards, classification is essentially indistinguishable from a random decision. These phenomena are equally recognisable in both the reconstruction and prediction methods. The evaluations show the effects of the measures for a greener AI on deep anomaly detection; the results allow to minimise calculation efforts by reducing sensor data sampling rate. In addition, the smaller amount of data allows to reduce training and inference time, implying that the search for suitable hyperparameters is faster or more extensive for machine learning models. In addition, less storage space is required for the converted data set.

Acknowledgments. This research has received funding from the German Federal Ministry of Education and Research under grant agreement 02J21D130 in the scope of the research project DIONE-X.

References

1. Schwartz, R., Dodge, J., Smith, N.A., Etzioni, O.: Green AI. arXiv e-prints arXiv:1907.10597 (Jul 2019). https://doi.org/10.48550/arXiv.1907.10597
2. Tnani, M.A., Feil, M., Diepold, K.: Smart data collection system for brownfield cnc milling machines: A new benchmark dataset for data-driven machine monitoring. Procedia CIRP **107**, 131–136 (2022). https://doi.org/https://doi.org/10.1016/j.procir.2022.04.022, leading manufacturing systems transformation – Proceedings of the 55th CIRP Conference on Manufacturing Systems 2022

Exploring the green AI potential of Adapter Tuning for Language Models

Dennis Mustafić[1 [0009−0003−0772−612X]], Frank Schirmeier[1 [0009−0002−2842−0376]]

University of Applied Sciences Kempten, Bahnhofstr. 61, 87435 Kempten, Germany
dennis.mustafic@hs-kempten.de, frank.schirmeier@hs-kempten.de

Abstract. Large Language Models are of fundamental importance for future AI usage in industry and society. The environmental impact of training and inference of such models however is devastating, which necessitates research into more efficient model mechanisms. In the simple context of the well-established BERT architecture, this study performs a comprehensive comparison between classical fine-tuning, which involves most or all of the model weights, and parameter-efficient adapter tuning, in which shallow trainable layers are introduced throughout the model while the vast majority of original weights remain unchanged. A series of experiments were carried out in which the $\text{BERT}_{\text{BASE}}$-architecture was trained on the LexGLUE benchmark with fine-tuning and adapter tuning for each data set. Through extensive comparisons in the experiments, it was found that adapter tuning is advantageous from a green AI perspective. We observed that the usability of the Adapter Tuning methodology is related to the complexity of the downstream fine-tuning task.

Keywords: Adapter Tuning · Parameter Efficiency · Language Models

Introduction and Motivation Fine-tuning of pre-trained Transformer models is ineffective, as all weights are completely changed when adapting to a new task. This leads to parameter-inefficiency, as the pre-trained model parameters cannot be reused and therefore a completely new model is required for each task. Adapter tuning by Houlsby et al. [1] is an alternative approach to allow for more efficient parameter training. Here, the pre-trained model is freezed, and adaptation to the downstream task occurs by training so-called adapter modules instead, which are incorporated into the Transformer layers. By design, the theoretic efficiency is significantly increased, as the adapter modules only comprise 0.5-8% of the original parameters and can be replaced depending on the task. GPU resources are also saved during the training process, as only a fraction of the gradients needs to be calculated during backpropagation. This means that language models can be trained on less powerful infrastructures and still achieve similar performance, which also gives an indication of the efficiency of adapter tuning. [1,2] Adapter Tuning can be used alternatively or additionally to other approaches like Low Rank Adaptation [4] and derivations thereof.

To assess the usability of Adapter Tuning in practice, this study carries out a series of tests in order to specialize the underlying BERT architecture to different

downstream tasks. Four different downstream tasks as specified in the LexGLUE benchmark dataset [3] are evaluated. In these tests, the performance (evaluated by task-specific means) of classical fine-tuning serves as a base-line to which the performance of the models trained with Adapter Layers is compared.

Conclusion From a Green AI perspective, other studies [5–7] showed that Adapter Tuning can result in significant efficiency gains. Through this study, we were able to confirm the results of [5–7] and also to make statements about the relationship between data set complexity and the usability of adapter modules.

The results show that adapter tuning achieves comparable performance to fine-tuning for less complex tasks and is therefore a legitimate alternative training method. The used adapter tuning with bottleneck layers generated smoother loss curves which arguably contributed to better generalization capabilities for the $\text{BERT}_{\text{ADAPTER}}$-models in some cases, similar to [2]. Nevertheless, the results showed that adapter tuning as parameterized in this study was not able to achieve sufficiently good results for the more complex tasks. For future studies, we hence propose – while assessing the useability of adapter tuning and similar methods like LoRA [4] – to simultaneously and explicitly keep track of the complexity of the downstream task and to develop a model that maps complexity to adapter tuning performance.

References

1. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S. (2019). Parameter-efficient transfer learning for NLP. In International Conference on Machine Learning (pp. 2790-2799). PMLR.
2. He, R., Liu, L., Ye, H., Tan, Q., Ding, B., Cheng, L., Low, J.-W., Bing, L., Si, L. (2021). On the effectiveness of adapter-based tuning for pretrained language model adaptation.
3. Chalkidis, I., Jana, A., Hartung, D., Bommarito, M., Androutsopoulos, I., Katz, D. M., Aletras, N. (2021). LexGLUE: A benchmark dataset for legal language understanding in English.
4. Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models. arXiv preprint arXiv:2106.09685.
5. Karimi Mahabadi, R., Henderson, J., Ruder, S.: Compacter: Efficient Low-Rank Hypercomplex Adapter Layers. In: Advances in Neural Information Processing Systems (NeurIPS), Vol. 34, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, J. Wortman Vaughan, pp. 1022-1035. Curran Associates, Inc. (2021)
6. Rücklé, A., Geigle, G., Glockner, M., Beck, T., Pfeiffer, J., Reimers, N., Gurevych, I.: AdapterDrop: On the Efficiency of Adapters in Transformers. (2021)
7. Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., Hu, S., Chen, Y., Chan, C.-M., Chen, W., et al.: Parameter-efficient fine-tuning of large-scale pre-trained language models. Nature Machine Intelligence 5(3), 220–235 (2023). Nature Publishing Group UK London.

Robustness of scale free behavior in generalized preferential attachment networks *

Laura Sacerdote[0000−0003−2527−5935]

Dept. Mathematics "G. Peano", University of Torino,
Via C. Alberto 10 10123 Torino, Italy

Abstract. Scale free behavior of the degree distribution characterizes a number of real world phenomena. The seminal work by Barabási and Albert [1] explains the high frequency of this feature in different networks, showing how the preferential attachment rule determines the scale free property. However, in many instances, preferential attachment coexists with other types of attachments or with detachment. Here, some recently results by the author and some collaborators are collected, discussing whether the scale free behavior is robust with respect to such variants of the attachment rules and showing some cases where this feature is destroyed.

Keywords: Preferential attachment · Degree distribution · Discrete and continuous time models.

1 Robustness of scale free behavior

Barabási and Albert [1] suggested the use of the preferential attachment rule to model the World-Wide-Web (WWW) but the same rule regulates the growth of many other networks: genetic, social, financial and citation are some examples of such networks. In the preferential attachment framework, new and existing nodes are connected according with probabilities proportional to the number of existing links characterizing the considered node. Such attachment rule seems reasonable for different modeling instances but it is often too simple for more complex behaviors. Indeed, this attachment rule can coexist with others and this can change the resulting degree distribution. Sometimes, one may expect changes of the parameter of the degree distribution or the removal of the scale free property.
In a social network new links arise when new members join a node. Often, each member connects to the most followed participants, according with a preferential attachment rule. However, new connections also arise when a member connects to a particular group such as work colleagues or schoolmates and these cases do not follow the preferential attachment paradigm. This situation determines the coexistence of preferential and uniform attachment rules and we

* Work partially supported by the MIUR-PRIN 2022 project "Non-Markovian dynamics and non-local equations", no. 202277N5H9 and by INDAM (GNAMPA)

expect changes in the degree distribution. Other phenomena may interfere with a preferential attachment rule in a social network. For example, the popularity of some individuals may decrease after a while and determine the detachment of some followers. Alternatively, someone may be impressed by the popularity and follow a preferential attachment rule, others may prefer connections with not so popular individuals. For example, pages offering health services may be visited thanks to their popularity or because they offer faster service having a limited number of customers. Similar types of dynamics may appear in many other contexts, from genetic to the WWW. We plan to present the effects of these different types of perturbations of the preferential attachment rule with the aim of understanding the robustness of the scale free behavior of the degree distribution in the following modeling cases of interest:

1. co-presence of detachment and attachment phenomena;
2. co-presence of different attachment rules;
3. co-presence of preferential attachment and anti-preferential attachment.

With reference to the first instance, in [2], we studied the use a birth and death process to develop a generalization of the Yule model to describe the appearence/disappearence of nodes in the WWW. We showed the lack of scale free behavior when the detachment phenomenon dominates the attachment one. When this is not the case, scale free behavior remains but its parameter changes. In [3] we considered two different causes for the appearance of new links in a network, i.e. the preferential or the uniform rules showing that the scale free behavior is robust and the presence of the uniform attachment does not play any role on the asymptotic degree distribution. The third case is discussed in [4] where it was shown the persistence of the scale free behavior, despite the presence of the anti-preferential attachment rule. However, the exponent results modified. Interestingly, this model is able to recover any power law with any exponent less than -3 if we suitably modify the parameter proportion of links following the preferential or the anti-preferential attachment rules.

Finally, we will present some new results for a preferential attachment model in which, at each step i, the number of added links is X_i, $i = 1, 2, \ldots$, with the X_i i.i.d. random variables, not necessarily bounded.

References

1. Barabasi, F., Albert, : Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
2. Lansky, P., Polito, F. and Sacerdote, L.: The role of detachment of in-links in scale-free networks. Journal of Physics A: Mathematical and Theoretical **37**(345002), (2014)
3. Pachon, S., Sacerdote, L and Yang, S.: Scale-free behavior of networks with the copresence of preferential and uniform attachment rules. Physica D,**371**, 1-12 (2018)
4. De Ambroggio, U., Polito, F. and Sacerdote, L.: On dynamic random graphs with degree homogenization via anti-preferential attachment probabilities On dynamic random graphs with degree homogenization via anti-preferential attachment probabilities. Phys.D **414** (132689), 12, (2020).

Optimal Boundary for the First-Passage-Time problem

Lubomir Kostal[1][0000−0002−2708−6268], Laura Sacerdote[2][0000−0003−2527−5935], and Cristina Zucca[2][0000−0003−2144−4201]

[1] Institute of Physiology, Czech Academy of Sciences, Prague, 142 20, Czech Republic kostal@biomed.cas.cz
http://https://www.biomed.cas.cz/ kostal/
[2] Dept. of Mathematics "G. Peano", University of Torino, 10123 Torino, Italy
{laura.sacerdote,cristina.zucca}@unito.it

Abstract. Brownian Motion with drift μ is often used to describe systems acting as integrator of received external inputs. In this context the First Passage Time of the process through a suitable boundary is interpreted as the output of the system. Estimate of the value of the not observed input is then performed through the observed output. Here we investigate the amount of information about the input carried by the observed output. In particular, we determine particular shapes of the boundary allowing better estimates of the input. The basic tools for this study are the mutual information and the input-specific information that are used to investigate the existence of optimal boundaries within some classes of interest. The study is then extended to systems modeled by Ornstein-Uhlenbeck process.

Keywords: Brownian Motion · Input Detection · Mutual Information · Ornstein-Uhlenbeck process.

1 Optimality of the boundary shape

In various modeling instances, one deals with systems receiving external inputs whose summation gives rise to a final input that determines the global evolution of the system. It is often difficult to observe such input while it is natural to observe the generated output that in many situations corresponds to the observation of the First Passage Time (FPT) of the process describing the evolution of the system through a boundary. Typical frameworks for this type of problems are Integrate-and-Fire neuronal models or reliability problems.

Here, we first consider a system described by a Brownian Motion with drift,

$$W_t = \mu t + \sigma B_t \tag{1}$$

with $\mu > 0$ that can not be observed, while we can observe T, its FPT through a known boundary S, that can be constant or time-dependent. Our goal is to find the boundary that is optimal to determine μ from the observed FPTs. The unknown μ should be read as a random variable M with a given probability

density function $f_M(\mu)$. We employ the conditional probability density function $f_{T|M}(t|\mu)$ of T when M is known, to evaluate μ in terms of the mutual information $I(M;T)$ that can be written as

$$I(M;T) = \int_{\mathcal{M}} i(\mu;T) f_M(\mu) d\mu, \tag{2}$$

where $i(\mu;T)$ is the input-specific information [1]

$$i(\mu;T) = \int_T f_{T|M}(t|\mu) \log \frac{f_{T|M}(t|\mu)}{f_T(t)} \tag{3}$$

and

$$f_T(t) = \int_M f_{T|M}(t|\mu) f_M(\mu) d\mu. \tag{4}$$

Since the problem is too general we will restrict it by specifying particular classes of time dependent boundaries.
The first class of boundaries of interest contains the constant ones, i.e. $S(t) = S$. In this case the response random variable T has distribution

$$f_{T|M}(t|\mu) = \frac{S}{2\pi\sigma^2 t^3} \exp\left\{-\frac{(S-\mu t)^2}{2\sigma^2 t}\right\} \tag{5}$$

and

$$E(T|M=\mu) = \frac{S}{\mu}. \tag{6}$$

Other boundaries are also considered, using the numerical evaluations of the FPT distribution when the analytical expression is not available [2].

Beside the description of the system as a Brownian Motion we also look for optimal boundaries in the case of a system modeled through the Ornstein-Uhlenbeck process.

Acknowledgments. Work partially supported by MIUR-PRIN 2022 project "Non-Markovian dynamics and non-local equations", no. 202277N5H9, by INDAM (GNAMPA) and by by the Spoke 1 "FutureHPC BigData" of ICSC - Centro Nazionale di Ricerca in High-Performance-Computing, Big Data and Quantum Computing, funded by European Union - NextGenerationEU.

Disclosure of Interests. The authors have no competing interest.

References

1. D'Onofrio, G.,Kostal: Coordinate Invariance as a Fundamental Constraint on the Form of Stimulus-Specific Information Measures. Biol. Cybern. **112**, 1–2 (2018)
2. Sacerdote, L., Giraudo, M.T.: Stochastic Integrate and Fire Models: a Review on Mathematical Methods and their Applications Stochastic Biomathematical Models with Applications to Neuronal Modeling. In: Bachar, Batzel and Ditlevsen (eds) LNM (Biosciences subseries), Vol. 2058: pp. 99–142 Springer, Heidelberg (2013)

Extreme values and hitting probabilities in LD-QBD processes

A Di Crescenzo[1][0000−0003−4751−7341], A Gómez-Corral[2][0000−0003−4057−7763], and D Taipe[2][0000−0002−2207−4129]

[1] Università degli Studi di Salerno, Italy
adicrescenzo@unisa.it
[2] Universidad Complutense de Madrid, Spain
{agcorral,dtaipe}@ucm.es

Abstract. We focus on distributional properties of the maximum level visited by a level-dependent quasi-birth-death process, and related hitting probabilities of the phase process. We exemplify the solution with two epidemic models: the SIS model with vertical and horizontal disease transmission; and the SIR model with constant population size.

Keywords: Epidemic model · First-passage time · Hitting probability · Quasi-birth-death process.

1 LD-QBD processes and extreme values

A level-dependent quasi-birth-death (LD-QBD) process can be seen as a time-homogeneous continuous-time Markov chain $\mathcal{X} = \{(I(t), J(t)) : t \geq 0\}$ that takes values in a countable state space $\mathcal{S} = \{(i, j) : i \in \mathbb{N}_0, j \in \{0, ..., M_i\}\}$, for integers $M_i \in \mathbb{N}_0$, which is partitioned as $\cup_{i=0}^{\infty} l(i)$ with ith level $l(i) = \{(i, j) : j \in \{0, ..., M_i\}\}$, for $i \in \mathbb{N}_0$. The q-matrix of $\mathcal{X}$ is assumed to be conservative, irreducible and regular, and has the structured form

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & & & & \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & & & \\ & \ddots & \ddots & \ddots & & \\ & & Q_{i,i-1} & Q_{i,i} & Q_{i,i+1} & \\ & & & \ddots & \ddots & \ddots \end{pmatrix},$$

where the sub-matrices $Q_{i,i'}$ are of dimension $(1 + M_i) \times (1 + M_{i'})$, for $i' \in \{\max\{0, i - 1\}, i, i + 1\}$, and have entries $q_{(i,j),(i',j')}$, for integers $j \in \{0, ..., M_i\}$ and $j' \in \{0, ..., M_{i'}\}$.

In this talk, we focus on the extreme values of process $\mathcal{X}$ before its first entrance to $l(0)$, and hitting probabilities of the phase process $\{J(t) : t \geq 0\}$ as the maximum level of $\mathcal{X}$ is visited for the first time. Specifically, we define the random vector $(\tau_{\max}, I_{\max}, J(\tau_{\max}))$, where $I_{\max}$ is the maximum level visited by $\mathcal{X}$ before the first entrance to $l(0)$, $\tau_{\max}$ is the first time to reach the maximum

number $I_{\max}$, and $J(\tau_{\max})$ is the phase at time $\tau_{\max}$. Under the assumption that the first passage from any initial state $(i_0, j_0) \in \mathcal{S} \setminus l(0)$ to level $l(0)$ is certain, efficient algorithms for computing the marginal distribution of $I_{\max}$, and restricted Laplace-Stieltjes transforms of $\tau_{\max}$ on the sample paths of process $\mathcal{X}$ such that $\{I_{\max} = i, J(\tau_{\max}) = j\}$, are developed as an extension of the well-known block-Gaussian elimination technique. Hence, the algorithmic complexity of the solution is similar to that in [2, Algorithm A] and [3, Algorithms 1A-3A] in the finite case; for details, see [1].

The approach can be seen as an alternative one to that of Refs. [5, 6] for the running maximum $\bar{I}(T) = \sup\{I(t) : t \in [0, T]\}$ attained by $\mathcal{X}$ during a time interval $[0, T]$ where T is either an independent and exponentially distributed time, or a fixed time. More concretely, time T is described here as the first time that process $\mathcal{X}$ makes a transition to any state in level $l(0)$, so T is closely linked to the dynamics of $\mathcal{X}$ and, for practical use, amounts to the duration of a life cycle in a population, of an outbreak in epidemics, and of a busy period in a queueing model. The utility of the solution is demonstrated in two epidemic models: the SIS model with vertical and horizontal disease transmission, which is linked to a LD-QBD process with infinitely many states; and the SIR model with constant population size, which is formulated as a finite QBD process.

Acknowledgments. This research was funded by the Government of Spain, Ministerio de Ciencia e Innovación (Project PID2021-125871NB-B-I00). A Di Crescenzo was partially supported by the GNCS-INdAM, and by PRIN 2022 PNRR project P2022XSF5H. D Taipe also acknowledges the support of Banco Santander–Universidad Complutense de Madrid (Pre-doctoral Researcher Contract CT63/19-CT64/19).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Di Crescenzo, A., Gómez-Corral, A., Taipe, D.: Extreme values and related hitting probabilities in level-dependent quasi-birth-death processes with applications to epidemics. Under review, 1–27 (2023)
2. Gaver, D.P., Jacobs, P.A., Latouche, G.: Finite birth-and-death models in randomly changing environments. Adv Appl Probab **16**, 715–731 (1984)
3. Gómez-Corral, A., López-García, M.: Perturbation analysis in finite LD-QBD processes and applications to epidemic models. Numer Linear Algebra Appl **25**, 2160 (2018)
4. Gómez-Corral, A., Palacios-Rodríguez, F., Rodríguez-Bernal, M.T.: On the exact reproduction number in SIS epidemic models with vertical transmission. Comp Appl Math **42**, 291 (2023)
5. Javier, K., Fralix, B.: On the study of the running maximum and minimum level of level-dependent quasi-birth-death processes and related models. J Appl Probab **60**(1), 14–29 (2023)
6. Mandjes, M., Taylor, P.: The running maximum of a level-dependent quasi-birth-death process. Probab Eng Inf Sci **30**(2), 212–223 (2016)

Disease incidence in a stochastic SVI model with waning immunity*

M.J. Lopez-Herrero[1 [0000−0003−2835−2940]] and D. Taipe[1 [0000−0002−2207−4129]]

Complutense University of Madrid, Madrid, Spain
{lherrero, dtaipe}@ucm.es

Abstract. This paper deals with the long-term behaviour and incidence of a vaccine-preventable contact disease, under the assumption that both vaccine protection and immunity after recovery are not lifelong. The mathematical model is developed in a stochastic markovian framework. The evolution of the disease in a finite population is thus represented by a three-dimensional continuous-time Markov chain, which is versatile enough to be able to compensate for the loss of protection by including vaccination before the onset of the outbreak and also during the course of the epidemics.

Keywords: Epidemic model · Temporary immunity · Vaccine failures.

1 Introduction

An essential tool to represent the evolution of an infectious process through a population is mathematical modeling. Many epidemic models involve a compartmental division of individuals which take into account their health status with respect to the infection. It is also a common assumption that individuals are randomly in contact with each other and have no preferences for relationship.

In this paper, we consider an infectious disease taking place in a finite and isolated population. We assume that the disease is transmitted by direct contact with an infectious individual. After recovery, individuals show temporary immunity. In addition, we assume that the disease is a vaccine-preventable infection, but vaccinated individuals are not lifelong fully protected, either because of vaccine failures or because of vanning immunity. Given the latter fact, re-vaccination of susceptible individuals is planned as the epidemic progresses.

Consequently, the involved compartmental model divides the population in four classes: Susceptible to the infection (S), vaccine protected (V), infectious (I) and recovered-temporary immune individuals (R). Once temporary immunity is loss, the individuals become again susceptible to the disease The fundamental compartmental model for studying an infectious disease that confers immunity

* Supported by Ministry of Science and Innovation (Government of Spain), Project PID2021-125871NB-I00. D. Taipe acknowledges the support of Banco Santander and Complutense University of Madrid, Pre-doctoral Researcher Contract CT63/19-CT64/19.

is the SIR proposed by Kermack & McKendrick [1], which can be modified to take into account the control of disease spread (e.g., vaccination [2, 3]) or the effects of loss of immunity [2, 4].

In this paper, we focus on the study of the evolution of a communicable disease by incorporating vaccination and loss of immunity into the epidemiological model, considering both waning vaccines and temporary immunity after recovery. Specifically, we consider a markovian model, representing the spread of a pathogen in a finite population of constant size N.

Mathematical description involves a continuous time Markov chain (CTMC), which provides he evolution of the disease in terms of the number of individuals present in each compartment at any time t. Hence, we record the number of unprotected susceptible, $S(t)$, vaccinated, $V(t)$, infected, $I(t)$, and recovered-temporary immune individuals, $R(t)$. The hypothesis of constant size for the population gives, for any $t \geq 0$, the following relationship among the sizes of the allowed compartments: $N = S(t) + V(t) + I(t) + R(t)$. In consequence, the evolution of the epidemics is represented by a three-dimensional CTMC

$$\mathcal{X} = \{(I(t), S(t), V(t)) : t \geq 0\}$$

with state space $\mathcal{S} = \{(i, s, v) : 0 \leq i, s, v \leq N, 0 \leq i + s + v \leq N\}$, containing a possibly large, number of $(N + 1)(N + 2)(N + 3)/6$ states.

The aim of the present research is, in one hand, to observe the long term behaviour of the Markov chain and the epidemics itself. And secondly, to study the total number of infections taking place in an outbreak of the disease (i.e., the period of time starting when the first case of infection appears and ending when there are not infectious individuals in the population).

Theoretical derivations related to the stationary distribution and the probabilistic description of the disease incidence during an outbreak will result from the application of the first-step methodology [5, 6].

References

1. Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London. Series A. **115**, 700-721 (1927)
2. Turkyilmazoglu, M.: An extended epidemic model with vaccination: Weak-immune SIRVI. Physica A. Jul 15;598:127429 (2022)
3. Zhou, A., Sattayatham, P., Jiao, J.: Dynamics of an SIR epidemic model with stage structure and pulse vaccination. Advances in Difference Equations. 2016:140 (2016)
4. Batistela, C.M., Correa, D.P.F., Bueno, Á.M., Piqueira , J.R.C.: SIRSi compartmental model for COVID-19 pandemic with immunity loss. Chaos Solitons Fractals. Jan;142:110388 (2021)
5. Artalejo, J.R., Lopez-Herrero, M.J.: Stochastic epidemic models: New behavioral indicators of the disease spreading. Applied Mathematical Modelling, **38**, 4371-4387 (2014)
6. Amador, J., Lopez-Herrero, M.J.: Cumulative and maximum epidemic sizes for a nonlinear SEIR stochastic model with limited resources. Discrete & Continuous Dynamical Systems-Series B, **23**, 3137-3157 (2018)

A model for the spreading of fake news and a related two-dimensional finite birth-death process

Spina Serena[0000−0001−6408−7596] and Paraggio Paola[0000−0002−3308−7937]

University of Salerno, Via Giovanni Paolo II, 132, Italy.
{sspina,pparaggio}@unisa.it

Abstract

The attention on mathematical models for the propagation of fake news in social networks is increasing nowadays, since it is becoming a crucial topic for information security. Some recent papers in this area have been devoted to construct discrete-time stochastic models for the spreading of fake news in a community (see [5], [2]). Other approaches are finalized to describe the growth of the number of individuals reached by rumors through deterministic or stochastic models (cf., for instance, [6], [1]).

In this work we propose a new model for rumor spreading among the individuals of a restricted population divided into compartments. The proposed model, like many similar ones, is inspired by epidemiological issues. Indeed, although the contexts of application are different, the dynamics which drive the time-evolution of the compartments are similar. More in detail, we suppose that a population of size N is divided into three categories: ignorants, spreaders and skeptics. A spreader is a person that knows the fake news and transmits it when he/she meets an ignorant or a skeptic. An ignorant is a person who does not know the rumor and then, when he/she is contacted by a spreader, becomes a spreader. A skeptic, when contacted by a spreader, is able to discover the evidence of the received fake news and thus persuades the spreader to not transmit it anymore.

We denote respectively with $X(t), Y(t), Z(t)$ the number of spreaders, ignorants and skeptics at time $t \geq 0$, such that $X(t) + Y(t) + Z(t) = N$, with N a fixed natural number. We suppose that the initial condition at time 0 is $X(0) = j,\ Y(0) = k,\ Z(0) = N - j - k$, with $0 < j, k < N$. We assume that the rate of contacts between a spreader and an ignorant is $\lambda > 0$, and between a spreader and a skeptic is $\xi > 0$. Following the described rules of the rumor spread we obtain the following system of ordinary differential equations, for $t \geq 0$:

$$\begin{cases} \dfrac{dX(t)}{dt} = X(t)\left(\lambda\, Y(t) - \xi\ (N - X(t) - Y(t))\right) \\ \dfrac{dY(t)}{dt} = -\lambda\, X(t)Y(t). \end{cases} \tag{1}$$

For the system (1), we investigate the time-evolution of the population compartments. We also analyze the effect of varying the parameters, the equilibrium points and the asymptotic behavior. One can observe that, for $t \to +\infty$ all the spreaders become skeptic, so that the diffusion of the fake news is stopped. We remark that a similar model has been studied in [4].

As in many similar researches (see, for instance, [3]), we study also the stochastic counterpart of the proposed model. Specifically, we consider a two-dimensional birth-death process $\{(X(t), Y(t)),\, t \geq 0\}$ such that:

$$\begin{aligned}
&\mathsf{P}\begin{pmatrix} X(t+\Delta t) = n+1 \\ Y(t+\Delta t) = m-1 \end{pmatrix}\Bigg| \begin{matrix} X(t) = n \\ Y(t) = m \end{matrix}\Bigg) = \lambda n m \Delta t + o(\Delta t), \\
&\mathsf{P}\begin{pmatrix} X(t+\Delta t) = n-1 \\ Y(t+\Delta t) = m \end{pmatrix}\Bigg| \begin{matrix} X(t) = n \\ Y(t) = m \end{matrix}\Bigg) = \xi n (N-n-m)\Delta t + o(\Delta t), \qquad (2) \\
&\mathsf{P}\begin{pmatrix} X(t+\Delta t) = 0 \\ Y(t+\Delta t) = m \end{pmatrix}\Bigg| \begin{matrix} X(t) = 0 \\ Y(t) = m \end{matrix}\Bigg) = 1,
\end{aligned}$$

for any $n \geq 1$, $m \geq 0$. It is worth noting that Eqs. (2) describe effectively the stochastic counterpart of the deterministic dynamic described before The state space of the process $\{(X(t), Y(t)); t \geq 0\}$ is given by

$$S = \{(n, m) : 0 \leq n \leq j + k,\ 0 \leq m \leq \min\{k, j + k - n\}\}\,.$$

The states $(0, m)$, for any $m \geq 0$, are absorbing. Indeed when the number of spreaders becomes 0 the diffusion of the fake news stops, as can be also obtained from system (1). We determine the partial differential equation for the probability generating function. We also construct and employ a simulating algorithm in order to analyse the process evolution.

Keywords: Rumor spread · Two-dimensional stochastic process · Birth-death process · Probability generating function.

References

1. Di Crescenzo, A., Paraggio, P., Spina, S.: Stochastic growth models for the spreading of fake news. Mathematics **11**(16), 3597 (2023)
2. Esmaeeli, N., Sajadi, F.: On the probability of rumour survival among sceptics. Journal of Applied Probability, **60**(3), 1096–1111 (2023)
3. Giorno, V., Nobile, A.: Time-inhomogeneous finite birth processes with applications in epidemic models. Mathematics **11**(21), 4521 (2023)
4. Giorno, V., Spina, S.: Rumor spreading models with random denials. Physica A: Statistical Mechanics and its Applications, **461**, 569–576 (2016)
5. Mahmoud, H.: A model for the spreading of fake news. Journal of Applied Probability **57**(1), 332–342 (2020)
6. San Martín, J., Drubi, F., Rodríguez Pérez, D. Uncritical polarized groups: The impact of spreading fake news as fact in social networks. Mathematics and Computers in Simulation **178**, 192–206 (2020)

A Sojourn-Based Approach to Discrete-time Semi-Markov Reinforcement Learning

Giacomo Ascione[1] [0000−0003−3860−7097]

Scuola Superiore Meridionale, Napoli 80138, Italy
g.ascione@ssmeridionale.it

Abstract. In this talk we discuss a newly introduced approach to discrete-time semi-Markov decision processes based on the sojourn time process. More precisely, we embed the semi-Markov process into a bivariate Markov one. With this new approach, the agent is allowed to consider different actions depending also on the sojourn time of the process in the current state. A numerical method based on Q-learning algorithms for finite horizon reinforcement learning and stochastic recursive relations is investigated. To show the effectiveness of the method, we discuss two toy examples. Finally, future aims will be presented.

Keywords: Semi-Markov decision processes · Q-learning algorithms · Deep reinforcement learning.

With the wider use of Artificial Intelligence in everyday life, Reinforcement Learning plays a prominent role in Machine Learning nowadays. This is, in a certain sense, the adaptation of the notion of Operant conditioning (see, for instance, [4] for a review) to artificial intelligence. In this context, Reinforcement Learning is usually formalized by means of Markov Decision Processes, where the best expected cumulative reward is obtained by solving a set of equations, commonly known as Bellman's Equation. The Markov property of the environment is crucial, as the Dynamic Programming Principle on which the equations are based strongly rely on it. Furthermore, since the environment is usually unknown, the value function has to be approximated. This is done by means of Watkins and Dayan's Q-learning method in the infinite horizon case, see [6, 7]). The method is, however, based on the fixed point form of the solution in the infinite horizon case, hence it cannot be directly adapted to the finite horizon one, unless the problem can be transformed into an infinite horizon one. Let us stress that the Q-learning method has been adapted to the finite horizon setting in [5].

In general, we cannot assume the environment to be Markov. Here we discuss a generalization of the model to the case of a semi-Markov environment, hence using semi-Markov decision processes in place of a Markov ones. Semi-Markov processes were introduced by Lévy in [2] and they are defined as processes exhibiting the Markov property only on specific stopping times. In the context of decision processes, in literature one usually consider the underlined process to be a stepped continuous-time semi-Markov processes, i.e. processes that can

change among a certain number of discrete states via a Markov chain and such that the inter-jump times are independent of each other, but not necessarily exponentially distributed. Furthermore, for decision processes, one assumes that the action is taken *at each change of state*, and that it has effect directly on the distribution of the sojourn time.

However, to model discrete observations of the environment, it could be useful to consider discrete-time semi-Markov processes, also called semi-Markov chains. Such kind of processes can be characterized in terms of their sojourn time process, by embedding them in a bivariate Markov chain. This represents the main improvement we will present. Indeed, the Markov embedding not only lets us construct discrete-time semi-Markov decision processes, but extends the *range* of actions that the agent can take, which can be then applied also while waiting for the state to change, affecting the distribution of the residual sojourn time in place of the full one. To use these processes in Reinforcement Learning, we need to prove the effectiveness of the Q-learning algorithm to this specific case, arguing as in [5]. Two toy examples in which this method applies will be then presented. The first one is related to the idea of getting an Artificial Intelligence that feels much more natural to a human observer with a gambler's fallacy (see [8]). The second one is a different take on a classical problem in the context of semi-Markov reinforcement learning. Precisely, we consider a preventive maintenance model in which the agent is partially informed, on precise time instants, of the stress level of the system and thus can decide its regime accordingly.

Finally, future aims will be addressed, with particular attention to the adaptation of this approach to Deep Reinforcement Learning [3] (and thus on handling the Replay Memory) and on the infinite horizon case.

This talk is based on [1], which is a joint work with Salvatore Cuomo from University of Naples Federico II.

References

1. Ascione, G., Cuomo, S.: A sojourn-based approach to semi-Markov Reinforcement Learning. Journal of Scientific Computing **92**(2), 36 (2022).
2. Lévy, P.: Processus semi-markoviens. In: Proceedings of the International Congress on Mathematics III, Amsterdam, 1954 (1954).
3. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al.: Human-level control through deep reinforcement learning. nature **518**(7540), 529—533 (2015).
4. Staddon, J. E. R., Cerutti, D. T.: Operant Conditioning. Annual Review of Psychology **54**(1), 115–144 (2003)
5. Vivek, V., Bhatnagar, S.: Finite horizon Q-learning: Stability, convergence and simulations. arXiv preprint arXiv:2110.15093 (2021).
6. Watkins, C.J.: Learning from delayed rewards. PhD thesis, Cambridge University (1989)
7. Watkins, C.J., Dayan, P.: Q-learning. Machine Learning **8**(3–4), 279–-292 (1992)
8. Warren, P.A., Gostoli, U., Farmer, G.D., El-Deredy, W., Hahn, U.: A re-examination of "bias" in human randomness perception. Journal of Experimental Psychology: Humuan Perception and Performance **44**(5), 663 (2018).

On simulation strategies for first passage times of fractional time-changed diffusions ⋆

Maria Francesca Carfora[1][0000−0002−4570−1690]
Enrica Pirozzi[2][0000−0001−8692−2759]

[1] Istituto per le Applicazioni del Calcolo "Mauro Picone", Consiglio Nazionale delle Ricerche, Napoli, Italy
mariafrancesca.carfora@cnr.it
[2] Dipartimento di Matematica e Applicazioni "Renato Caccioppoli", Università degli Studi di Napoli Federico II, Napoli, Italy
enrica.pirozzi@unina.it

Many real phenomena, from finance to physics and biology, can be modeled as diffusing particles so that the probability of reaching a given state for the first time can be expressed as the First Passage Time (FPT) of a stochastic process through an assigned boundary. Among the different mathematical methods for solving FPT problems there are analytic methods for Volterra integral equations or corresponding differential problems (iterated kernels, Laplace transforms, etc.) as well as numerical algorithms to approximate their solutions. In recent years, the study of complex systems such as porous media, biological systems, complex fluids, networks (just to mention a few) has also revealed the occurrence of anomalous diffusions, where the mean square displacement of the diffusing particle deviates from the traditional linear relationship with time. The interest in such processes is closely related to their capability to delay in some sense the time flowing, so to preserve more information about the past evolution (a sort of memory) of the modeled phenomenon. The FPT problem can be also formulated for anomalous diffusions by fractional operators; the resulting FPT models are able to describe phenomena evolving on different time-scales. Starting from theoretical known results, such as the probability density function of the FPT of a time-changed Brownian motion [1], we propose some different strategies for the simulation the FPT of such a process. We show, discuss and compare the results by graphics, tables and derived approximations. The study has the aim to provide a fast and accurate numerical algorithm for simulating FPT of these fractional diffusions through linear boundaries, by specializing for the time-changed case some simulation algorithms already designed for the classical case.

In particular, we consider the time-changed Brownian motion, that is the stochastic process obtained by composing a Brownian motion and the right-continuous inverse of an α-stable subordinator process ([2]). The new process is neither Gaussian nor Markovian. However, its "subordinated transition pdf" can be defined ([3]), which satisfies a suitable "fractional Fokker-Planck" equation involving a Caputo fractional derivative of order α.

⋆ Partially supported by the project PRIN-MUR 2022XZSAFN.

In principle, it is possible to investigate the FPT problem for fractional diffusions by simulating a large number of trajectories of the process and register their crossing times, in order to approximate the FPT density by a histogram. However, as the fractional exponent α decreases, the time before crossing tends to increase, and very long simulations are needed, so that this approach becomes sometimes unfeasible. Then, relying on previously obtained results [1] for the FPT density of such a time-changed Brownian motion in presence of a linear boundary, we propose a numerical algorithm to approximate this distribution and investigate its accuracy for varying fractional indices.

In particular, we focus on the simulation method based on the hazard rate function [4], here specialized for the case of a time-changed Brownian motion. Note that this first case can be easily generalized, by using some transformation formulae, in order to consider other time-changed diffusions. We present results of our simulations for different values of the fractional index $\alpha \in (0, 1)$ and different distances between the starting point of the process and the boundary. Comparisons are made with the simulation algorithm for the trajectories of the time-changed Brownian motion ([5]), in order to confirm the agreement of the two approaches. The related computational times are also reported, to confirm the feasibility of the proposed algorithm for different values of α.

References

1. Leonenko, N. and Pirozzi, E. : First passage times for some classes of fractional time-changed diffusions. Stochastic Analysis and Applications **40**(4), 735–763 (2022)
2. Meerschaert, M. M. and Straka, P.: Inverse stable subordinators. Math. Model. Nat. Phenom. **8**(2):1–16 (2013)
3. Bertoin, J.: Lévy Processes, Vol. 121. Cambridge, UK: Cambridge University Press, (1996)
4. Ross, S. M.: Introduction to Probability Models. 10th edn. Academic Press, (2010)
5. Meerschaert, M. M. and Sikorskii, A.: Stochastic Models for Fractional Calculus. De Gruyter Studium, (2019)

On two coupled fractional stochastic differential equations for neuronal dynamics and simulations*

Enrica Pirozzi[0000−0001−8692−2759]

Dipartimento di Matematica e Applicazioni "Renato Caccioppoli",
Università degli Studi di Napoli Federico II, Napoli, Italy
enrica.pirozzi@unina.it

EXTENDED ABSTRACT

By substituting the well-known fractional Caputo-derivative to the classical derivative, it is possible to construct the fractional version of the stochastic Leake Integrate-and-Fire (sLIF) model for neuronal dynamics, [1], [2]. The immediate advantage of a such mathematical generalization of the sLIF model is that it makes possible to adopt a different time-scale. Indeed, the fractional derivative depends of an fractional order $\alpha \in (0,1]$: for $\alpha = 1$, we recover the classical (integer) case; for positive α but less than 1, we will be integrate the process on a finer time-scale or equivalently on a dilation of the time, in some sense. By means of statistical techniques applied to available data, it is possible to obtain useful estimates for tuning the values of fractional order α. As an addition feature of a such approach, we remark also that the non-local fractional derivative constitutes the suitable differential operator for describing dynamics with memory effects.

Beyond the need to construct a model by using the fractional approach, we also consider the phenomenological requirement to insert a correlated input in the neuronal model. This was consider previously ([3],[4]) by substituting a "colored"noise in place of the usual delta-correlated white noise. By following a similar strategy, here we consider a fractional correlated input.

Specifically we consider the two following stochastic equations for the stochastic process $V(t)$ describing the membrane voltage of the neuron and for the stochastic input $\eta(t), \forall t > 0$:

$$\mathcal{D}^{\alpha} V(t) = F(t, V(t), V_L) + \sigma\eta(t), \qquad V(0) = V_0, \tag{1}$$

$$\mathcal{D}^{\beta} \eta(t) = G(t, \eta, I) \quad + \gamma dW(t), \qquad \eta(0) = \eta_0, \tag{2}$$

with $\alpha, \beta \in (0,1]$. Here, $\mathcal{D}^{\nu}$ is the fractional Caputo derivative ([5],[6]) with $\nu = \alpha, \beta \in (0,1]$, respectively. Here, V_L stand for the resting value of V, and I is the term of a synaptic or other injected current.

Such a model allows to describe the neuronal dynamics by means of a α-fractional stochastic process $V(t)$ obtained as the fractional integral of a β-fractional stochastic correlated input $\eta(t)$.

* Partially supported by the project PRIN-MUR 2022XZSAFN.

About the study of such a model, the purpose is to verify the mathematical assumptions under which the existence of solutions is guaranteed ([7]) and to investigate about possible simulation techniques ([8], [9],[10]) for generate samples of first passage times of $V(t)$ through a constant level, i.e. the firing threshold.

We revise some existing simulation techniques, we specialize them for our case and provided simulation results in order to validate the theoretical advantages of the proposed model. For different values of fractional orders α and β, of involved parameters and functions F, G, we investigate the proposed model by means simulations, by providing statistical estimations and useful comparisons.

References

1. Burkitt, A. N., A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input, Biol Cybern (2006) DOI 10.1007/s00422-006-0068-6
2. Teka W., Marinov T.M. and Santamaria F., Neuronal Spike Timing Adaptation Described with a Fractional Leaky Integrate-and-Fire Model, PLoS Comput Biol., 10 (2014).
3. Bazzani A., Bassi G. and Turchetti G., Diffusion and memory effects for stochastic processes and fractional Langevin equations, Phys. A Stat. Mech. Appl., **324**, 530-550, (2003)
4. Pirozzi E., Colored noise and a stochastic fractional model for correlated inputs and adaptation in neuronal firing. Biological Cybernetics, **112** (1-2), 25–39, (2018)
5. Samko S.G., Kilbas A.A., Marichev O.I., Fractional Integrals and Derivatives: Theory and Applications, Gordon and Breach Science Publishers, (1993)
6. Ascione, G., Mishura, Y. and Pirozzi, E. Fractional Deterministic and Stochastic Calculus, De Gruyter, 2024.
7. Anh,, P.T., Doan, T.S., Huong, P.T., A variation of constant formula for Caputo fractional stochastic differential equations, Statistics and Probability Letters, Volume 145, Pages 351-358, (2019)
8. T.S. Doan, P.T. Huong, P.E. Kloeden, A.M. Vu, Euler–Maruyama scheme for Caputo stochastic fractional differential equations, Journal of Computational and Applied Mathematics, Volume 380, 112989, (2020)
9. Abundo M., Pirozzi E., Abundo, M., Pirozzi, E., Fractionally integrated Gauss-Markov processes and applications, Communications in Nonlinear Science and Numerical Simulation, Volume 101, 2021, 105862, ISSN 1007-5704.
10. Pirozzi, E., On the Integration of Fractional Neuronal Dynamics Driven by Correlated Processes, LNCS volume 12013, Computer Aided Systems Theory – EUROCAST 2019, pp- 211–219

The Dissipative Contact Process: a model for epidemics

Paolo Dai Pra[*], Elisa Marini[†]

Extended Abstract

The contact process is the microscopic counterpart of one of the most basic epidemiological models, the SIS model. Suppose the individuals of a population are placed on the vertices of a graph (V, E); the individual in the vertex $i \in V$ can be either *susceptible* ($x_i = 0$) or *non susceptible* ($x_i = 1$). The configuration $x = (x_i)_{i \in V}$ evolves as a continuous time Markov chain with the following rates:

- each non susceptible individual becomes susceptible with rate 1;
- each susceptible individual becomes non susceptible with a rate equal to a given constant λ times the fraction of non susceptible neighbors.

This definition suffices in finite graphs, while further care is needed in countable graphs (see [1]). In finite graphs the chain is absorbed in the null state ($x_i \equiv 0$) in finite time (we say that the epidemics *dies out*). The *time to absorption* could however vary from a *fast absorption* (at most polynomial in $|V|$) to a slow absorption (exponential in $|V|$). For many (sequences of) graphs the transition from fast to slow absorption occurs as the *infection constant* λ crosses a critical value λ_c. The simplest nontrivial example is the case in which (V, E) is the *complete graph* of N vertices: $V = \{1, 2, \dots, N\}$, $E = V \times V$. The study of this model reduces to the analysis of the one-dimensional Markov process

$$m^N := \frac{1}{N} \sum_{i=1}^{N} x_i,$$

and the critical infection constant is $\lambda_c = 1$. Infinite countable graphs require a more detailed analysis, leading to results in several directions, including ergodicity, convergence, local and global survival of the epidemics; such rich behavior could imply the existence of more than one critical value for the infection constant.

In this paper we introduce a modification of the contact process. The state x_i of each individual takes values in the interval $[0, 1]$, and is interpreted as her *viral load*; we say that an individual is susceptible if $x_i = 0$. The dynamics goes as follows:

- each non susceptible individual becomes susceptible with rate $r > 0$;
- each susceptible individual becomes non susceptible with a rate equal to $\lambda > 0$ times the arithmetic mean of the viral loads of her neighbors, and her viral load jumps to the value 1;
- between jumps, the viral load decays exponentially with rate $\alpha > 0$.

Note that the model is overparametrized, as we could rescale the time to have $r = 1$. We keep however all parameters as they have a useful interpretation: r^{-1} is the time scale at which infected (i.e. non susceptible) individuals lose their immunity, while α^{-1} is the time scale at which infected individuals remain contagious. In many real epidemics $\alpha^{-1} \ll r^{-1}$: for a relatively long time non susceptible individuals are *immune*, and do not contribute to the propagation of the disease. Note that many of the useful mathematical properties of the contact process are lost; for instance, monotonicity breaks due to the presence of immune individuals that block contagion.

Here we study this modified contact process, that we call *dissipative*, in the complete graph. The microscopic dynamics of the system is sketched in Fig. 1.

[*]Dipartimento di Informatica, Università di Verona, Italy
[†]Dipartimento di Matematica, Università di Padova, Italy

Unlike the corresponding classical contact process, the dissipative version does not admit a finite dimensional reduction: any finite-dimensional functional of the empirical measure is non Markovian for large N. We particularly deal with the thermodynamic limit of the process ($N \to +\infty$), and we obtain a law of large numbers (propagation of chaos) and a central limit theorem. Despite the lack of finite dimensional reduction, both the *limit process* corresponding to the Law of Large Numbers and the limit *fluctuation process* corresponding to the Central Limit Theorem are two-dimensional, as they can be expressed in terms of the fraction m of susceptible individuals, and the average viral load v per individual, and the corresponding normal fluctuations. In particular the pair (m, v) solves the deterministic ODE

$$\begin{aligned} \dot{m}(t) &= (1 - m(t))v(t) - rm(t) \\ \dot{v}(t) &= -\alpha v(t) + c(1 - m(t))v(t) - rv(t). \end{aligned} \tag{1}$$

Under suitable conditions on the parameters, including the case $\alpha^{-1} \ll r^{-1}$, the limit process has a unique stable fixed point, which is approached by damped oscillations. The limit fluctuation process reveals, via a Fourier analysis, that noise induces persistent oscillations, despite of the damping exhibited by the limit process. The nature of these oscillation are then investigated by letting the parameters α and λ to diverge (slowly) with N: if properly rescaled, the dynamics of the fraction of non susceptible individuals and of the mean viral load converge to a harmonic oscillator affected by noise:

$$\begin{aligned} dm(t) &= \rho v(t)dt \\ dv(t) &= -\frac{r}{\rho}(1 - \rho)m(t)dt + r(1 - \rho)dW(t) \end{aligned} \tag{2}$$

where $\rho = \alpha/\lambda$ and W is a standard Brownian motion.

In their epidemiological interpretation these results suggest that pandemic waves are induced by finite-size effects, allowing the random noise to excite frequencies close to a characteristic value. The resulting motion is however not strictly periodic: though the period is nearly constant, the amplitude of each oscillation is significantly affected by noise, in agreement with what observed in real epidemics.

References

[1] T. M. Liggett, Interacting particle systems. Springer, New York (2008).

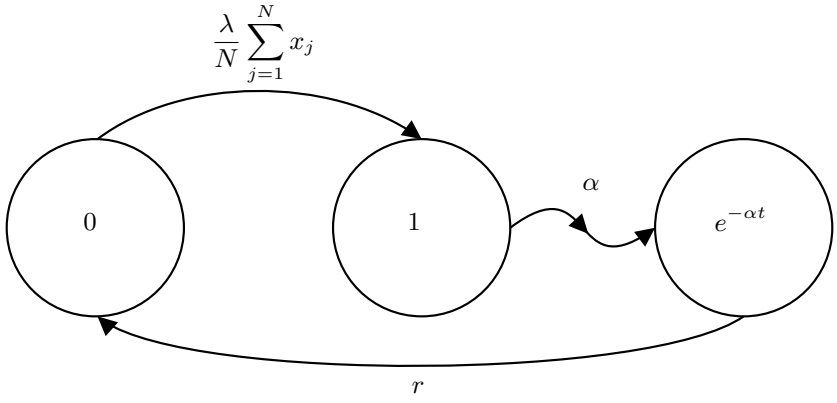


Figure 1: An individual i who is initially susceptible ($x_i = 0$), can become infectious with rate $\frac{1}{N}\sum_{j=1}^{N} x_j$ (mean-field interaction). When this happens, her viral load x_i jumps to the value 1. When infectious, individual i can recover with rate r and, in this case, she becomes immediately susceptible again. Namely, her state jumps to ($x_i = 0$). In the meantime, the viral load of the individual decreases deterministically in time with rate α. Notice that this dynamics is invariant under permutations of the particles. Last, notice that this model is a piecewise deterministic Markov process, in that the x_is have a deterministic continuous dynamics with discontinuities detmined by random jump events.

Handling uncertainties on the right-hand side of a classical transportation model by stochastic optimisation and a matheuristic approach

Abtin Nourmohammadzadeh[0000−0003−0383−0379] and Stefan Voß[0000−0003−1296−4221]

Institute of Information Systems, University of Hamburg, Hamburg, Germany
{abtin.nourmohammadzadeh,stefan.voss}@uni-hamburg.de

Keywords: Right-hand-side uncertainty · Stochastic optimisation · Matheuristics

1 Introduction

The uncertainty on the right-hand side of mathematical models is a topic of interest in non-deterministic optimisation, which has not yet been sufficiently investigated. This kind of uncertainty refers to the lack of precise knowledge or variability in the values of the parameters, which appear on the right-hand side of mathematical models.

In the literature, [4] focuses on the complexity of robust linear programming with right-hand-side uncertainty, while [5] specifically investigates the complexity of two-stage robust linear programming with ellipsoidal right-hand-side uncertainty. These works provide proofs on the NP-hardness of this kind of problems. A related study is [7], which examines the bilevel knapsack problem with stochastic right-hand sides. [6] applies a robust optimisation model with right-hand-side uncertainties to the bus driver rostering problem and proposes a matheuristic solution approach.

We employ the classical transportation problem because it offers a straightforward model featuring two sets of constraints with parameters on their right-hand side, which, in real-world scenarios, exhibit non-deterministic characteristics. It involves determining the optimal allocation of some products from multiple sources to multiple destinations. The classical transportation problem is closely related to other well-known problems within the domain of vehicle routing problems (VRP) ([3]).

2 Mathematical Model and Methodology

The classical transportation problem is a linear programming problem used to minimise the cost of transporting a certain quantity of some products (P is the set of products) from multiple suppliers (I is the set of suppliers) to multiple consumers (J is the set of consumers). $x_{pij} \geq 0$ is the quantity of product p to be transported from supplier i to consumer j. c_{pij} is the cost of transporting one unit of product p from supplier i to consumer j. s_{pi} is the supply capacity of supplier i for product p and d_{jp} is the demand of consumer j for product p. In the mathematical model of the problem, we seek to minimise $Z = \sum_{p\in P}\sum_{i\in I}\sum_{j\in J} c_{pij}x_{pij}$ subject to $\sum_{j\in J} x_{pij} \leq s_{pi}$ for all $p \in P, i \in I$, which ensures that the total amount of each product supplied from each supplier does not exceed its capacity, and $\sum_{i\in I} x_{pij} \geq d_{pj}$ for all $p \in P, j \in J$, which guarantees that the total consumer demand for each product is satisfied. In reality, the capacity of suppliers may fluctuate due to production or product availability issues, and consumer demand can vary for various reasons. Therefore, we treat these two right-hand side parameters, i.e. s_{pi} and d_{pj}, for all products, suppliers and consumers as non-deterministic. For each of these parameters, we consider ten possible values with their corresponding probability and use a stochastic optimisation approach. The probabilities are distributed such that the highest probability is found at the median, with probabilities decreasing as we move away from it in both directions.

We employ a variable neighbourhood search (VNS) algorithm [1] wherein a matheuristic approach, based on the fixed set search method (FSS) [2], is incorporated after the VNS operators in every iteration. As such, the algorithm operates by initialising a random population of solutions. Subsequently, local searches are conducted around each solution utilising the 2-opt and 3-opt operators. Should a neighbouring solution prove superior to the current one, we transition to it and continue the local searches from that point onward. In the event that no further improvement is attainable, a shaking operation is executed whereby a specific portion of the solution is randomly modified, thereby escaping the confines of the local optimum. As the final step of each iteration, a selection of the best solutions from the population are chosen as the fixed set of FSS. Subsequently, the members of the fixed set are subjected to crossovers with each other and iterative processing using mathematical optimisation until no further enhancement can be achieved. At the conclusion of each iteration, the best solutions yielded by all operations are carried forward to the subsequent iteration, where the same operations are repeated on them. The termination condition of the algorithm is the occurrence of stagnation over a number of consecutive iterations.

3 Conclusions

We address 30 instances of varying sizes, generated randomly. These instances involve a range of 3 to 15 suppliers and 10 to 200 consumers. We compare our algorithm's performance against that of a pure genetic algorithm, the pure VNS, and the genetic algorithm with the same matheuristic extension. Our algorithm surpasses the results obtained by these methods. This demonstrates that the VNS algorithm has the potential to outperform a widely used metaheuristic such as the genetic algorithm. Furthermore, the inclusion of the matheuristic extension plays a crucial role in the successful execution of our solution methodology.

References

1. Hansen, P., Mladenović, N., Pérez, J.A.M.: Variable neighbourhood search: methods and applications. Annals of Operations Research **175**(1), 367–407 (2009). https://doi.org/10.1007/s10479-009-0657-6
2. Jovanovic, R., Tuba, M., Voß, S.: Fixed set search applied to the traveling salesman problem. In: Hybrid Metaheuristics, pp. 63–77. Springer (2018). https://doi.org/10.1007/978-3-030-05983-5_5
3. Macharet, D.G., Campos, M.F.M.: A survey on routing problems and robotic systems. Robotica **36**, 1781–1803 (2018). https://doi.org/10.1017/s0263574718000735
4. Minoux, M.: On 2-stage robust LP with RHS uncertainty: complexity results and applications. Journal of Global Optimization **49**, 521–537 (2011). https://doi.org/10.1007/s10898-010-9645-2
5. Minoux, M.: Two-stage robust LP with ellipsoidal right-hand side uncertainty is NP-hard. Optimization Letters **6**, 1463–1475 (2011). https://doi.org/10.1007/s11590-011-0341-z
6. Nourmohammadzadeh, A., Voß, S.: An effective matheuristic approach for robust bus driver rostering with uncertain daily working hours. In: Lecture Notes in Computer Science, vol. 14239, pp. 365–380. Springer (2023). https://doi.org/10.1007/978-3-031-43612-3_23
7. Özaltın, O.Y., Prokopyev, O.A., Schaefer, A.J.: The bilevel knapsack problem with stochastic right-hand sides. Operations Research Letters **38**, 328–333 (2010). https://doi.org/10.1016/j.orl.2010.04.005

Unified Formulations of Entropy and Extropy

N. Balakrishnan[1], F. Buono[2,3], Y. Deng[4], and **M. Longobardi**[5]

[1] McMaster University, Canada
[2] Institute of Statistics, RWTH Aachen University, Aachen, Germany
[3] University of Napoli Federico II
[4] Institute of Integrated Automation, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xian, Shaanxi, China
[5] Department of Biology, University of Napoli Federico II, Italy

Abstract. The measures of uncertainty are a topic of considerable and growing interest. In this talk a general formulation of entropy is proposed and studied. It depends on two parameters and includes Shannon, Tsallis and fractional entropy, all as special cases. Recently, the introduction of the extropy as a measure of uncertainty dual of Shannon entropy opened up interest in new aspects of the subject. Since there are many versions of entropy, a unified formulation for it has been introduced to work with all of them in an easy way. Here, we consider the possibility of defining a unified formulation for extropy by introducing a measure depending on two parameters. For particular choices of the parameters, this measure will give back the well-known formulations of extropy. Moreover, the unified formulation of extropy is analyzed also in the context of Dempster-Shafer theory of evidence.

Keywords: Entropy · Extropy · Dempster-Shafer Theory of evidence.

1 Unified entropy and extropy

We have defined the fractional Tsallis entropy, or unified formulation of entropy, given as

$$S_\alpha^q(X) = \frac{1}{\alpha - 1} \sum_{i=1}^{N} p_i (1 - p_i^{\alpha-1})(-\log p_i)^{q-1}, \tag{1}$$

where $\alpha > 0$, $\alpha \neq 1$ and $0 < q \leq 1$. Later we have introduced the unified formulation of extropy in the context of classical probability theory. We refer to this formulation as fractional Tsallis extropy and, for a discrete random variable X, it is defined as

$$JS_\alpha^q(X) = \frac{1}{\alpha - 1} \sum_{i=1}^{N} (1 - p_i)[1 - (1 - p_i)^{\alpha-1}][-\log(1 - p_i)]^{q-1}, \tag{2}$$

where $\alpha > 0$, $\alpha \neq 1$ and $0 < q \leq 1$. The purpose of giving this definition is based on the fact that this expression includes the classical extropy, Tsallis extropy and fractional extropy all as special cases.

Acknowledgments. Francesco Buono and Maria Longobardi are members of the research group GNAMPA of INdAM (Istituto Nazionale di Alta Matematica) and are partially supported by MIUR - PRIN 2017, project "Stochastic Models for Complex Systems", no. 2017 JFFHSH.

References

1. N. Balakrishnan, F. Buono, M. Longobardi, A unified formulation of entropy and its application, Physica A: Statistical Mechanics and its Applications **596** (2022)
2. F. Buono, Y. Deng, M. Longobardi, The unified extropy and its versions in classical and Dempster-Shafer theories, Journal of Applied Probability, (2023)

Diffusion approximation of time-inhomogeneous SI epidemic model*

Virginia Giorno[0000−0001−6474−3952] and Amelia G. Nobile[0000−0003−2363−7496]

Dipartimento di Informatica, Università di Salerno, Via Giovanni Paolo II, n. 132, 84084 Fisciano (SA), Italy, {giorno,nobile}@unisa.it

Abstract. We consider a time-inhomogeneous diffusion process useful to model the evolution of the infected population in the susceptible-infectious epidemic model in a random environment. We assume that there are no removals, no immunes, and no recoveries from infection. The susceptible individuals become infected through contact with infectious individuals. A such model is suitable for describing some classes of micro-parasitic infections to which individuals never acquire a long lasting immunity and over the course of the epidemic everyone eventually becomes infected. We determine the expression of the transition probability density function and of its conditional moments. Particular attention is dedicated to the first-passage time problem, by deriving closed form results for the first-passage time density through a constant boundary. For the time-homogeneous process, the behavior of the mean and of the variance of the first-passage time is analyzed and asymptotic results are derived for large population size. Some comparisons between the deterministic model and the obtained diffusion process are provided.

Keywords: Population dynamics · First-passage time densities· Asymptotic behaviors.

1 The model

Deterministic and stochastic epidemic models are used in the literature to analyze the behaviors of biological diseases and how they spread (cf. Allen [1, 2], Bailey [3, 4]). The simplest epidemic models assume that an individual can be in one of only two states, either susceptible (S) or infectious (I). For $t \geq t_0$, we denote by $S(t)$ the number of susceptible individuals, by $I(t)$ the number of individuals infected and by N the total population size, where $N = S(t) + I(t)$ is constant. In Giorno and Nobile [5], several time-inhomogeneous deterministic models, useful to describe the evolution of a finite population constituted by susceptible and infectious individuals have been considered. Moreover, stochastic models based on finite birth processes have been analyzed. For these processes, we have determined the explicit expression of the transition probabilities and of

* (Workshop "Stochastic Models, Statistical Methods, and Applied Systems Simulations")

the first-passage time densities. Particular attention has been dedicated to the SI model and its counterpart stochastic based on finite birth processes.

In the deterministic SI model, the dynamics of the population of the infected $I(t)$ can be described by the differential equation:

$$\frac{dI(t)}{dt} = \frac{\lambda(t)}{N}[N - I(t)]\, I(t), \qquad t > t_0, \tag{1}$$

where the transmission intensity function $\lambda(t)$ is a positive, bounded and continuous function of t. The solution of (1) is

$$I(t) = \frac{N\, I(t_0)}{I(t_0) + [N - I(t_0)]\, e^{-\Lambda(t|t_0)}}, \qquad t \geq t_0, \tag{2}$$

with

$$\Lambda(t|t_0) = \int_{t_0}^{t} \lambda(\theta)\, d\theta. \tag{3}$$

In [5], for the SI birth process $\{M(t), t \geq 0\}$ we have assumed that the births at time t occurs with intensity functions given by $\lambda_n(t) = \lambda(t)(N - n)n/N$ for $n = j, j + 1, \ldots, N$.

The aim of the present paper is to build a time-inhomogeneous diffusion process for the infected population as an alternative to the birth process considered in [5]. We interpret $\Lambda(t|0)$ as the mean of a time inhomogeneous Wiener process $\{Z(t), t \geq 0\}$ with covariance function $c(s,t) = \int_0^s \sigma^2(\theta)\, d\theta$ for $0 < s \leq t < +\infty$. Starting from (1) and (2) and making use of the assumption of random environment, we obtain the time-inhomogeneous diffusion process $\{X(t), t \geq t_0\}$, describing the evolution of the infected population, characterized by infinitesimal drift and infinitesimal variance

$$A_1(x,t) = \frac{\lambda(t)}{N}(N - x)x + \frac{1}{4}\frac{\partial A_2(x,t)}{\partial x}, \qquad A_2(x,t) = \sigma^2(t)\, \frac{(N - x)^2 x^2}{N^2}\cdot \tag{4}$$

The space-state of $X(t)$ is $(0, N)$, with 0 and N unattainable end-points.

We determine the transition probability density function of $X(t)$ and we analyze the first-passage time problem through constant boundaries. Moreover, we compare the results of the deterministic SI epidemic model with those obtained via the finite birth process, discussed in [5], and via the diffusion process (4).

References

1. Allen, L.J.S.: An Introduction to Stochastic Processes with Applications to Biology. Chapman and Hall/CRC, Lubbock, Texas, USA (2010)
2. Allen, L.J.S.: Stochastic Population and Epidemic Models. Persistence and Extinction. Springer International Publishing, Switzerland (2015)
3. Bailey, N.T.J.: The Mathematical Theory of Epidemics. Charles Griffin and Co. Ltd. (1957)
4. Bailey, N.T.J.: The Elements of Stochastic Processes with Applications to the Natural Sciences. John Wiley & Sons, Inc. New York (1964)
5. Giorno, V., Nobile A.G.: Time-inhomogeneous finite birth processes with applications in epidemic models. Mathematics **11**, Article number: 4521 (2023)

An Ornstein Uhlenbeck type process with structural breaks in the drift coefficient*

Giuseppina Albano[1][0000−0002−2317−0331], Virginia Giorno[2][0000−0001−6474−3952]

[1] Dip. di Studi Politici e Sociali, Università di Salerno, 84084 Fisciano (SA), Italy, pialbano@unisa.it
[2] Dip. di Informatica, Università di Salerno, 84084 Fisciano (SA), Italy, giorno@unisa.it

Abstract. We consider a time non homogeneous Ornstein Uhlenbeck type diffusion process in which the drift parameters present several jumps. Such a process is able to model the evolution of systems in which structural breaks occur at fixed times due to sudden changes in dynamics.
In a first step, the probability distribution of the process is analyzed by obtaining its transition probability density function and the related moments.
Then, we focus on the statistical analysis of the process. In particular, we provide an inference procedure to address the estimation of the drift and of the infinitesimal variance, generally both depending on time. Starting from a discrete sampling, such procedure is essentially based on the maximum likelihood estimation.
Finally, in order to validate the proposed procedure, a simulation study is made by considering several infinitesimal moments of the process and different sample size. In such simulation study, by means of suitable statistical tests we also discuss the presence of structural breaks in the drift parameters.

Keywords: Non homogeneous process · Inference · Tests for structural breaks.

1 The model

The Ornstein Uhlenbeck (OU) process, originally introduced by Langevin in 1908, to describe the velocity of a particle moving in a fluid, has also used in biological context as model of a single neurons' activity. Further, it is widely used in economics and in financial literature under the name of Vasicek process as a model of interest rates, currency exchange rates, and commodity prices (cf. [1]). We consider a generalization of the classical OU process by including in the infinitesimal moments suitable deterministic time dependent functions and some discontinuities to describe the presence of structural breaks in the dynamics. In

* (Workshop "Stochastic Models, Statistical Methods, and Applied Systems Simulations")

these cases, the problem of estimating the drift and the infinitesimal variance of the process becomes of interest both of theoretical and practical contexts.

Let $\{X(t),\ t \geq \tau_0\}$ be a stochastic process defined in $\mathbb{R}$ described by the following stochastic differential equation (SDE):

$$dX(t) = \big[-\alpha(t)X(t) + \beta(t)\big]dt + \sigma(t)dW(t), \tag{1}$$

where $W(t)$ is a standard Brownian motion, $\beta(t)$ and $\sigma(t) > 0$ are real continuous functions. The function $\alpha(t)$ is a piecewise constant function, that has a finite number L of pieces, i.e.

$$\alpha(t) = \sum_{k=1}^{L} \alpha_k I_k(t) \tag{2}$$

with

$$I_k(t) = \begin{cases} 1\ \tau_{k-1} \leq t < \tau_k, \\ 0\ otherwise \end{cases} \tag{3}$$

denoting the indicator function of the interval $[\tau_{k-1}, \tau_k)$. Here, τ_k, $k = 1, 2, \ldots, L$ are the points of discontinuity of the function $\alpha(t)$ and they describe the times in which the structural breaks occur. The process $X(t)$ is a time inhomogeneous OU process with drift and infinitesimal variance $A_1(x,t) = -\alpha(t)x(t) + \beta(t)$, $A_2(t) = \sigma^2(t)$. We note that τ_0 is the time in which the phenomenon begins to be observed, whereas, $\tau_1, \tau_2, \ldots, \tau_L$ can be interpreted as times in which the observations exhibit some types of anomalies similar to structural breaks. We assume that the observations highlight L structural breaks. After a structural break the process restarts according a different law. Specifically, after the time τ_{k-1} the process the process restarts from $X(\tau_{k-1})$ with a drift $A_1^{(k)}(x,t) = -\alpha_k x + \beta(t)$ and infinitesimal variance $A_2(t)$ until the next break occurs. Hence, $\alpha(t) = \alpha_k$ for $\tau_{k-1} \leq t < \tau_k$ and $\alpha(t) = \alpha_L$ when $t \geq \tau_L$.
We point out that the sample-paths of $X(t)$ are not discontinuous in the points τ_k, but they present trend variations in such points.
We provide the probability distribution of the process and related moments. Moreover, we propose an inference procedure to address the estimation of the parameters α_i for $i = 1, 2, \ldots, L$ and of function $\beta(t)$ and $\sigma(t)$. Such procedure can be seen as an extension of those ones discussed in [2–4].

References

1. Leung, Tim; Li, Xin: Optimal Mean-Reversion Trading: Mathematical Analysis and Practical Applications. World Scientific Publishing Co. ISBN 978-9814725910. (2016)
2. Albano, G., Giorno, V.: Inference on the effect of non homogeneous inputs in Ornstein–Uhlenbeck neuronal modeling. Math. Biosci. Eng. **19**(1), 328–348 (2020)
3. Albano, G., Giorno, V.: Inferring time non-homogeneous Ornstein Uhlenbeck type stochastic process. Computational Statistics and Data Analysis **150**, 107008 (2020)
4. Albano, G., La Rocca, M., Perna, C.: Small sample properties of ML estimator in vasicek and CIR models: a simulation experiment. Decis. Econ. Finance **42**(1), 5–19 (2019)

Modeling growth curves through diffusion processes*

Giuseppina Albano[1 [0000−0002−2317−0331]], Antonio Barrera[2,5 [0000−0001−8686−6232]], Virginia Giorno[3 [0000−0001−6474−3952]], and Francisco Torres-Ruiz[4,5 [0000−0001−6254−2209]]

[1] Dipartimento di Studi Politici e Sociali, Università degli Studi di Salerno, Via Giovanni Paolo II, Fisciano(SA), Italy. pialbano@unisa.it
[2] Departamento de Análisis Matemático, Estadística e I. O. y Matemática Aplicada, Bulevar Louis Pasteur 31, Universidad de Málaga, Spain. antonio.barrera@uma.es
[3] Dipartimento di Informatica, Università degli Studi di Salerno, Via Giovanni Paolo II, Fisciano(SA), Italy. giorno@unisa.it
[4] Departamento de Estadística e I.O., Campus de Fuentenueva, Universidad de Granada, Spain. fdeasis@ugr.es
[5] Instituto de Matemáticas de la Universidad de Granada, IMAG, Calle Ventanilla, 11, Granada, Spain.

Abstract. Stochastic models generalizing deterministic growth curves play an important role in modeling several dynamic phenomena. Recently a general growth curve including the well known growth equations, such as Malthus, logistic, Bertalanffy, Gompertz, was studied. In order to model the variability in the dynamics, two stochastic formulations of this growth equation were then introduced starting from a suitable parametrization of the deterministic model and by adding an additive and multiplicative noise, respectively. The resulting processes were lognormal and Gaussian. The practical application of such stochastic models relies on the determination of parameters, hence the problem of estimating the parameters of the model becomes of great interest in real contexts. This problem is here analyzed by means of the maximum likelihood method. However, the resulting systems of equations are complex and their solutions required techniques ad hoc. In the present study we focus on the numerical solutions and related problems, applying different metaheuristic approaches.

Keywords: Growth curves · Inference in diffusion processes · Metaheuristic procedures.

1 Introduction

The construction and development of mathematical models to describe dynamic phenomena associated with growth curves are topics that have caught the eyes of many researchers. The introduction of regulatory effects into the Malthusian

* (Workshop "Stochastic Models, Statistical Methods, and Applied Systems Simulations")

model gave rise to classical models such as the logistic or the Gompertz curves. Generalizations were then required in order to obtain curves with greater flexibility. The techniques employed are the introduction of suitable parameters and the use of functions with flexible behaviors. The fact that deterministic models cannot control the variability of the described phenomena motivated randomization procedures into the ordinary differential equations that govern such models. This led to the consideration of stochastic differential equations, whose solutions are, under certain conditions, diffusion processes (see, for instance, [1]-[5]). We focus on the general growth equation in [5]-[6], that was originally obtained from the family of functions

$$g_\theta(t) = \left(\eta + \left(1 + \eta^{1-p}\,(1-p)\,t\,\log\alpha\right)^{\frac{1}{1-p}}\right)^{\frac{1}{n}}$$

indexed by the parametric vector $\theta = (n, \alpha, \eta, p)^T$, where n and η are positive real numbers, $0 < \alpha < 1$ and $1 < p < 1 + 1/n$. Initial conditions $x_0 \in (0, \infty)$ at time $t_0 \geq 0$ lead to deterministic models of the form $x_\theta(t) = x_0\, g_\theta(t_0) g_\theta(t)^{-1}$ for $t \geq t_0$. Different asymptotic behaviors of such model regarding parameters n and p are classical models such as the hyper-logistic ones, or even the Richards or hyper-Gompertz curves. The introduction of multiplicative and additive noise gives rise to the models considered here. The former is a non-homogeneous log-normal diffusion process while the second is a Gaussian process. The practical application of such stochastic models relies on the determination of their parameters. This may lead to scenarios where the estimation is hard or even impossible to accomplish. For this reason, the study of the deterministic model, trying to understand the connection between their parameters, allows us to consider bounded parametric regions in order to apply different metaheuristic approaches.

Acknowledgments. Work supported in part by PID2020-1187879GB-100 and CEX2020-001105-M grants, funded by MCIN/AEI/10.13039/501100011033, Spain.

References

1. Rajasekar, S.P., Pitchaimani, M., Quanxin Zhu: Ergodic stationary distribution and extinction of a stochastic SIRS epidemic model with logistic growth and non-linear incidence. Appl. Math. Comp., **377**(125143) (2020)
2. Lo, C.F.: A modified stochastic Gompertz model for tumour cell growth. Comp. Math. Methods Med., **11**(1), 3–11 (2010)
3. Albano, G., Giorno, V., Román-Román, P., Román-Román, S., Torres-Ruiz, F.: Inference on an heteroscedastic Gompertz tumor growth model. Math. Biosci., **328**(108428) (2020)
4. Barrera, A., Román-Román, P., Serrano-Pérez, J.J., Torres-Ruiz, F.: Two Multi-Sigmoidal Diffusion Models for the Study of the Evolution of the COVID-19 Pandemic. Mathematics, **9**(2409) (2021)
5. Albano, G., Giorno, V., Román-Román, P., Torres-Ruiz, F.: Study of a general growth model. Commun. Nonlinear Sci. Numer. Sim. **107**(106100) (2022)
6. Albano, G., Barrera, A., Giorno, V., Román-Román, P., Torres-Ruiz, F.: First Passage and First Exit Times for diffusion processes related to a general growth curve. Commun. Nonlinear. Sci. Numer., **126**(107494) (2023)

Elliptic Curve Kleptography Study

Oscar Cigala-Alvarez[0000−0002−8609−9325], Jezabel Molina-Gil[0000−0001−7702−9264], and Pino Caballero-Gil[0000−0002−0859−5876]

Department of Computer Engineering and Systems, University of La Laguna, 38200, Tenerife, Spain
ocigalaa@ull.edu.es; jmmolina@ull.edu.es; pcaballe@ull.edu.es;

Extended Abstract

Elliptic Curve Cryptography (ECC) [1] represents a significant evolution in asymmetric cryptography, which takes advantage of mathematical properties of elliptic curves over finite fields to provide greater security with shorter keys, compared to traditional public key systems like RSA. This advantage translates into increased efficiency and speed.

In particular, elliptical curves over finite fields are defined by equations in the form of $y^2 = x^3 + ax + b$, where 'a' and 'b' are coefficients that meet certain conditions to ensure the non-singularity of the curve. In the context of ECC, finite fields, which are typically prime, are used to define a finite set of points that satisfy the curve equation.

In relation to key exchange, ECC is applied to define an elliptic version of the Diffie-Hellman algorithm that provides a secure and efficient method of exchanging keys over insecure channels [2]. The security in this process is based on the complexity of the Elliptic Curve Discrete Logarithm Problem (ECDLP), which involves determining the secret number from the base point and the generated public point. [3].

In particular, the Elliptic Curve Diffie–Hellman (ECDH) key exchange is a process whereby two parties can securely generate a shared key over an insecure channel. Each participant begins by selecting a base point on the curve and a secret number. Then, each generates a public point by multiplying their secret number by the base point. These public points are exchanged between the parties, and each uses their secret number to compute the shared point, which will be the same for both. This key exchange process is fundamental in numerous modern security protocols and cryptography, providing a secure and efficient foundation for encrypted communication.

This paper examines the concept of the SETUP attack [4], which is a covert strategy that allows an attacker to insert a hidden vulnerability in cryptographic systems. This attack is designed to allow the attacker to surreptitiously filter out secret key information without detection.

This type of attack is particularly concerning in the context of ECC and Diffie-Hellman, as it allows the attacker to covertly recover private keys, underscoring the need for robust methods to protect the integrity of modern cryptographic systems. Thus, this attack exploits the inherent complexity of the

ECDLP and the ECDH problem, so that when it is carried out, it becomes almost impossible for the end user to detect, making the compromised device appear to function legitimately.

One of the most concerning applications of SETUP attacks is in cryptographic devices such as smart cards, where the manufacturer, by applying this method, can gain exclusive access to the user's private keys. The stealthy nature of the SETUP attack makes it a particularly insidious threat, as it compromises the integrity and trust in modern cryptographic systems.

References

1. N. Koblitz, *Elliptic curve cryptosystems*, Mathematics of Computation, vol. 48, no. 177, pp. 203–209, 1987.
2. W. Diffie and M. E. Hellman, *New Directions in Cryptography*, IEEE Transactions on Information Theory, vol. 22, no. 6, pp. 644–654, 1976.
3. A. Menezes, *Evaluation of security level of cryptography: the Elliptic Curve Discrete Logarithm Problem (ECDLP)*, University of Waterloo, vol. 14, 2001.
4. E. Mohamed and H. Elkamchouchi, *Elliptic Curve Kleptography*, International Journal of Computer Science and Network Security, vol. 10, no. 6, pp. 183–185, 2010.

Privacy-Preserving Machine Learning based on Federated Learning and Secure Computation

L. Syne[1 [0009−0005−3805−1430]], P. Caballero-Gil[1 [0000−0002−0859−5876]], and C. Hernández-Goya[1 [0000−0002−9468−708X]]

Department of Engineering and Systems. University of La Laguna. 38271 La Laguna. Tenerife. Spain

alu0101533254@ull.edu.es pcaballe@ull.edu.es mchgoya@ull.edu.es

Extended Abstract

In recent years, Artificial Intelligence (AI) solutions have been extensively deployed in various sectors, including industry, agriculture, education, finance, and health. This use has become a reality thanks to the availability of large quantities of data used to train AI algorithms. However, in some cases, this data may be personal or sensitive, particularly regarding health or finance data, particularly in Machine Learning. This paper presents a practical exploration of two techniques that are instrumental in achieving privacy-preserving machine learning: Federated Learning (FL) and Secure Multi-Party Computation (SMPC).

On the one hand, Federated Learning was introduced by researchers at Google and represents a collaborative way of training neural network models in a decentralized manner. Its creation was motivated by concerns regarding the confidentiality of users' data in the context of data collection for training machine learning models. FL proposes a way of training these models such that clients' data never leaves their devices.

On the other hand, Secure Multi-Party Computation is a cryptographic mechanism that leverages advanced protocols to perform private computations on data without exposing the underlying information.

FL has gained interest due to privacy concerns surrounding conventional centralized machine learning. However many research works propose additional secure protocols to guarantee sufficient privacy within federated learning. The work [3] presents a hybrid approach that combines SMPC with differential privacy. That proposal improves model accuracy while preserving provable privacy guarantees and protecting against extraction attacks and collusion threats. The protocol proposed in [1] allows a server to compute the sum of large, user-held data vectors from mobile devices securely (i.e. without learning each user's contribution). The paper [2] shows a practical implementation of federated learning with SMPC protecting private data using Tensorflow 2 and MNIST dataset.

In order to preserve the confidentiality of data in the machine learning context we propose a decentralized and collaborative training approach based on federated learning. The individual data are not collected, they remain stored on local devices, and at each training iteration, only individual parameters are

sent to a server for aggregation and global model construction. Additionally, a secure multi-party mechanism is leveraged to guarantee the privacy of parameter aggregations and model updates. These secure computation techniques ensure the aggregator cannot access individual updates and intermediate global models. The proposal here described is a practical combination of Federated Learning with Secure Multiparty Computation. Several experiments have been based on open-source tools and privacy-sensitive datasets. By leveraging these technologies significant contributions are brought to the privacy-preserving machine learning field. Furthermore, emphasizing privacy sensitive datasets demonstrates a commitment to ethical data handling, validating the practicality of the approach in real-world scenarios.

References

1. Keith, B., Vladimir, I., Ben, K., Antonio, M., Brendan, M., Sarvar, P., Daniel, R., Aaron, S., Karn, S.: Practical Secure Aggregation for Privacy-Preserving Machine Learning. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pp 1175–1191, Association for Computing Machinery New York, NY, USA, (2017).
2. Lia, D., Togan, M.: Privacy-preserving machine learning using federated learning and secure aggregation. In 12th IEEE International Conference on Electronics, Computers and Artificial Intelligence, pp. 1-6, (2020).
3. Stacey, T., Thomas, S., Nathalie, B., Heiko, L., Yi, Z., Ali, A., Rui, Z.: A hybrid approach to privacy-preserving federated learning. In: Proceedings of the 12th ACM workshop on artificial intelligence and security, pp. 1-11. Association for Computing Machinery, New York, NY, United States (2019).

Privacy-preserving Malware Detection through Federated Machine Learning: A proposal

Fabio Martinelli[1], Francesco Mercaldo[2,1], and Antonella Santone[2]

[1] Institute for Informatics and Telematics, National Research Council of Italy (CNR), Pisa, Italy
[2] {fabio.martinelli, francesco.mercaldo}@iit.cnr.it
[3] University of Molise, Campobasso, Italy
{francesco.mercaldo, antonella.santone}@unimol.it

Federated machine learning (FML) is a decentralized approach to machine learning where the training data remains on edge devices or local servers rather than being centralized on a single server or in the cloud. The goal of FML is to train a global machine learning model while keeping the training data distributed, typically for privacy, security, or efficiency reasons [2].

In the traditional machine learning pipeline, data is collected from various sources, centralized in a data center or cloud server, and a global model is trained on this aggregated data. This approach can raise privacy concerns, as sensitive data is moved to a central location. FML addresses these concerns by performing training on the local devices or servers where the data is generated and stored.

FML is particularly useful in applications where privacy and data security are paramount, such as computer security, healthcare, finance, and other domains with sensitive data [4].

Malware detection is an active research field where typically from user applications are extracted features aimed to understand whether there is the presence of malicious behaviour [3]. For this task typically is employed machine learning with a set of features obtained from the applications installed by the user: this is the reason the (centralised) model must know the applications installed by the user for feature extraction (or, alternatively, the feature set aimed to identify the application) to perform the detection. This is the reason why malware detection machine learning is not able to preserve user privacy [1]. In this paper, we propose to adopt FML in malware detection, with particular regard to the mobile context. Considering that usually malware is aimed at exfiltrating sensitive and private information, mobile devices represent the most interesting target from the attacker's point of view, considering the huge amount of information stored in our smartphones, tablets, and IoT devices. To the best of the author's knowledge, this paper represents the first attempt to introduce FML in the malware detection context. Figure 1 shows an overview of the following proposal.

As shown in Figure 1, we propose to adopt a global, centralized model, which is released to the various terminals (i.e. mobile devices), so as to be able to carry out on-device malware detection on the local models. The knowledge accumulated by the various devices will be sent to the global model in the form of an update, in this way the knowledge collected by each mobile device will be encapsulated in the global model and therefore shared with all mobile devices, without

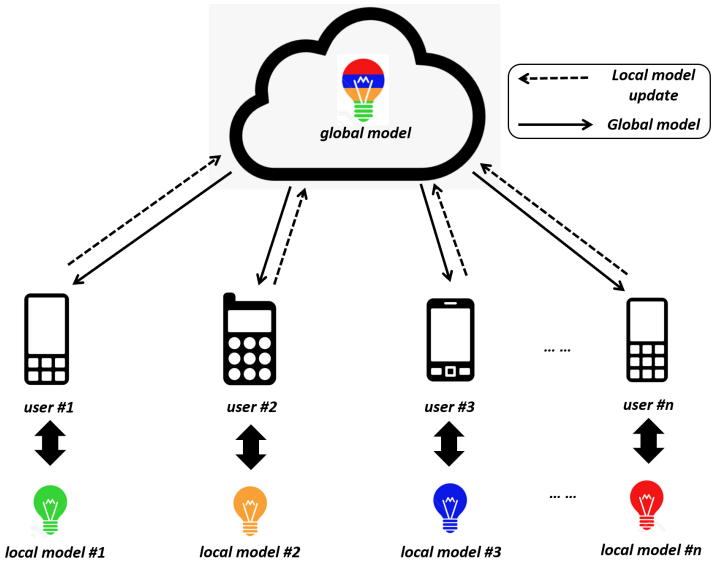


Fig. 1. The workflow of the proposed method for explainable PD detection through spiral draw testing.

the global model or the various mobile terminals knowing information relating to the applications installed by other users or the particular use made, thus preserving privacy and enabling a privacy-preserving mobile malware detection.

Acknowledgement

This work has been partially supported by EU DUCA, EU CyberSecPro, SYNAPSE, PTR 22-24 P2.01 (Cybersecurity) and SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the EU - NextGenerationEU projects.

References

1. Bharadiya, J.: Machine learning in cybersecurity: Techniques and challenges. European Journal of Technology **7**(2), 1–14 (2023)
2. Li, L., Fan, Y., Tse, M., Lin, K.Y.: A review of applications in federated learning. Computers & Industrial Engineering **149**, 106854 (2020)
3. Ye, Y., Li, T., Adjeroh, D., Iyengar, S.S.: A survey on malware detection using data mining techniques. ACM Computing Surveys (CSUR) **50**(3), 1–40 (2017)
4. Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., Gao, Y.: A survey on federated learning. Knowledge-Based Systems **216**, 106775 (2021)

A Generative Model Based Honeypot for Industrial OPC UA Communication

Olaf Sassnick, Georg Schäfer, Thomas Rosenstatter, and Stefan Huber

Josef Ressel Centre for Intelligent and Secure Industrial Automation
Salzburg University of Applied Sciences, Urstein Süd 1, 5412 Salzburg, Austria
{firstname.lastname}@fh-salzburg.ac.at

1 Motivation

Operational technologies (OTs) in industrial production environments are increasingly targeted by cyber attacks, making it necessary to develop new security measures to keep the risks at a manageable level. In this context a honeypot can serve as a decoy and warning system. Once an intruder is identified via anomaly detection, instead of ceasing the communication channel, the intruder is placed inside a honeypot, isolated from the production network. Consequently, all actions and observations by the intruder are performed on the honeypot. The interaction with the honeypot gives also information regarding the attack behavior, which in return can be utilized to strengthen the own security.

Most publications on honeypots for cyber physical systems (CPSs) used parametrized simulations. Franco *et al.* [2] found in their survey on honeypots in 2021 only one single approach [1] that employed machine learning algorithms out of the 44 CPS-related works. Setting up a honeypot for a brownfield OT installation by means of a simulation, however, is a challenging task, as physical system characteristics and corresponding parameters need to be determined beforehand. The recent advances in machine learning enable us to replace conventional simulations with self-learning models, potentially resulting in a more widespread use. A work published in October 2023 by Shan *et al.* [3] shows overall promising results, while focusing on a water treatment plant which communicates via the Modbus communication protocol.

In this work we explore a generative model approach, where a long short-term memory (LSTM) network learns the characteristics of a highly dynamic system based on recorded state space trajectories. The Open Platform Communications Unified Architecture (OPC UA) is used as it is an established communication standard in the automation industry, enabling different manufacturers to operate seamlessly together. By fully utilizing the OPC UA information model, even a completely autonomous training and deployment process seems feasible.

2 Approach

In our approach we assume a scenario as shown in Fig. 1. A threat actor is attacking a production environment remotely via the Internet. By exploiting a

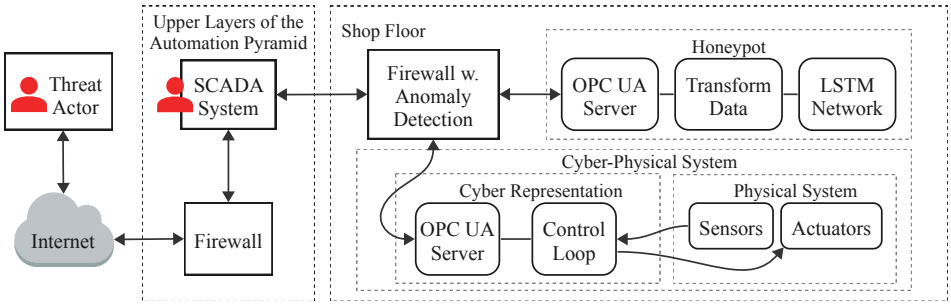


Fig. 1. Overview of the honeypot architecture

system vulnerability or via social engineering the threat actor gained control of the Supervisory Control and Data Acquisition (SCADA) system. As a result, the intruder can interact with the CPS located on the shop floor, communicating via the firewall. The CPS implements an OPC UA server and a control loop, while it has sensors and actuators interacting with the environment on a physical level. Once an anomaly is detected, the firewall redirects all network communication of the intruder to the honeypot. At this point, the honeypot is initialized with the latest known state space vector of the CPS. It consequently predicts the state space trajectory at a predefined sampling rate, matching the sampling rate of the OPC UA server. In our case the CPS is represented by a machine with two fans mounted on a balancing beam resulting in two degrees of freedom. A control loop is maintaining a configurable target yaw and target pitch by generating airflow via the two fans. This setup fits well, as it provides system characteristics with high dynamic range and small physical time constants. For the training of the LSTM network, only data recorded during normal operation of the CPS is used.

Overall, this work contributes towards the research on how LSTM networks can be employed as honeypot systems imitating OPC UA communication. By comparing a wide range of predicted state trajectories with real data we explore the effects of a reduced state space vector, determine the required variety and minimum training data size to sufficiently mimic the CPS and give computing performance results. For further research the full dataset will be published.

References

1. Cao, J., Li, W., Li, J., Li, B.: Dipot: a distributed industrial honeypot system. In: Qiu, M. (ed.) Smart Computing and Communication, vol. 10699, pp. 300–309. Springer International Publishing, Cham (2018)
2. Franco, J., Aris, A., Canberk, B., Uluagac, A.S.: A survey of honeypots and honeynets for internet of things, industrial internet of things, and cyber-physical systems. IEEE Communications Surveys & Tutorials **23**(4), 2351–2383 (2021)
3. Shan, Y., Yao, Y., Zhao, T., Yang, W.: Neupot: a neural network-based honeypot for detecting cyber threats in industrial control systems. IEEE Transactions on Industrial Informatics **19**(10), 10512–10522 (Oct 2023)

Quantum circuit optimization through algebraic techniques

J. García-Díaz, D. Escanez-Exposito, P. Caballero-Gil

University of La Laguna
{alu0101439679, jescanez, pcaballe}@ull.edu.es

1 Extended Abstract

Quantum computing has emerged as a revolutionary paradigm with the potential to solve complex problems that are practically intractable for classical computers. However, the realization of scalable and fault-tolerant quantum computers remains a formidable challenge [1] because quantum circuits, the fundamental building blocks of quantum algorithms, are susceptible to errors and inefficiencies that hinder the practical implementation of quantum algorithms [2].

Algebraic techniques provide a powerful framework for analyzing and manipulating quantum circuits, offering a systematic approach to identify redundancies, exploit symmetries, and streamline quantum computations, which will help in the crucial optimization of quantum circuits to overcome the inherent limitations of current quantum hardware. This work explores the application of algebraic techniques to optimize quantum circuits, aiming to enhance their efficiency, reduce error rates, and pave the way for the development of robust quantum computing systems.

One key focus of this research is the development of novel algebraic notation tailored to address specific challenges in quantum circuit optimization. This notation leverages algebraic structures inherent in quantum circuits to identify and eliminate redundancies, thereby reducing the overall gate count and mitigating the impact of errors. This paper also explores the implementation of some well-known quantum algorithms such as Shor's algorithm for factorization [3].

Currently, the state of experimental implementations of Shor's algorithm still has important technical challenges in realizing a fault-tolerant quantum computer with a sufficient number of qubits for factoring large integers. While large-scale, fault-tolerant quantum computers are not yet fully accomplished, recent advancements in quantum hardware and error correction techniques bring Shor's algorithm closer to practical applicability [4].

This paper presents a comprehensive exploration of quantum circuit optimization through algebraic techniques. By leveraging the inherent mathematical structures of quantum circuits, these techniques contribute to the advancement of quantum computing by addressing key challenges in error mitigation, gate count reduction, and overall performance improvement. The findings of this research have direct implications for the practical implementation of relevant quantum algorithms such as Shor's algorithm for factorization, which will contribute to the ongoing efforts to realize the full potential of quantum computing.

References

1. Nielsen, M.A., Chuang, I.L. (2011). Quantum Computation and Quantum Information: 10th Anniversary Edition. Cambridge University Press.
2. Watrous, J. (2023). Fundamentals of quantum algorithms. `https://learning.quantum-computing.ibm.com/course/fundamentals-of-quantum-algorithms`. [Accessed: 14- 11-2023]
3. Shor, P.W. (1994). Algorithms for quantum computation: discrete logarithms and factoring. Proceedings 35th Annual Symposium on Foundations of Computer Science. IEEE Computer Society Press: 124–134.
4. Skosana, U., Tame, M. (2021). Demonstration of Shor's factoring algorithm for N= 21 on IBM quantum processors. Scientific reports, 11(1), 16599.

Simulated Execution Through Quantum Algebraic Circuit Notation

Daniel Escanez-Exposito[0000−0003−4215−0501]
Pino Caballero-Gil[0000−0002−0859−5876]

University of La Laguna, Tenerife, Spain
{jescanez,pcaballe}@ull.edu.es

Extended Abstract

Effectively addressing the significant task of classically simulating quantum systems with exponentially increasing complexity is of utmost importance, especially in experiments involving a limited number of qubits (less than 50) [5]. This paper introduces the implementation of an independent execution engine capable of performing simulation of quantum algorithms with and without noise. This platform does not use at all the services offered by any of the leading quantum computing platforms (Qiskit [3], QSharp [4], Pennylane [6], Cirq [2], Amazon Braket [1]), since it is an original Python module that carries out its own approach regarding the interpretation and execution of algorithms from the point of view of quantum circuit algebra.

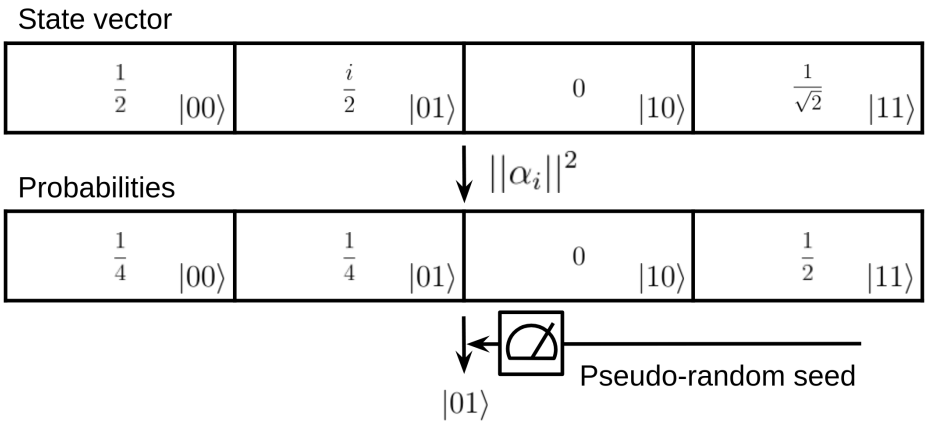


Fig. 1. Conversion scheme between state vector and associated probabilities, including subsequent measurement using a pseudo-random seed.

The system currently being implemented is based on the generation of results through the conversion of the final state vector of the circuit to a range of proba-

bilities. By projecting a pseudo-random seed over this range of probabilities, the simulation of the execution of this circuit can be achieved. In this way, results without quantum noise are obtained, since they start from the ideal probabilities obtained from the state vector (see Fig. 1 and 2). However, by weighting the probability coefficients according to each qubit and its value, it is possible to replicate noise in the system by considering properties such as decoherence or asymmetric noise [7].

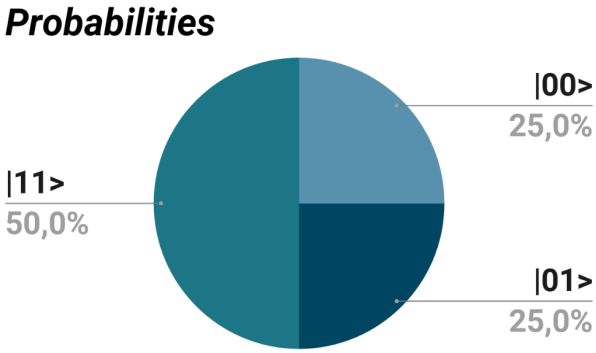


Fig. 2. Circular diagram illustrating the probabilities associated with the state vector.

References

1. Amazon: (2023), https://aws.amazon.com/braket/, accessed: 2023-11-13
2. Google Inc.: (2023), https://quantumai.google/cirq, accessed: 2023-11-13
3. IBM Research: (2023), https://qiskit.org/, accessed: 2023-11-13
4. Microsoft: (2023), https://learn.microsoft.com/en-us/azure/quantum/user-guide/, accessed: 2023-11-13
5. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press (2010)
6. Xanadu: (2023), https://pennylane.ai/, accessed: 2023-11-13
7. Zeh, H.D.: On the interpretation of measurement in quantum theory. Foundations of Physics **1**(1), 69–76 (Mar 1970). https://doi.org/10.1007/BF00708656

Theoretical Backdoor Attack on CRYSTALS-Dilithium

E. Pérez-Ramos[1 [0009−0008−0409−3079]] and P. Caballero-Gil[1 [0000−0002−0859−5876]]

Department of Computer Engineering and Systems. University of La Laguna. 38271 La Laguna. Tenerife. Spain
alu0101207667@ull.edu.es pcaballe@ull.edu.es

Extended abstract

Post-quantum cryptography has gained prominence in recent years as a response to the quantum threat. The National Institute of Standards and Technology (NIST) initiated an algorithm selection process in 2015 to address this issue. After a thorough search and evaluation, encryption algorithms such as CRYSTALS-Kyber and digital signature algorithms such as CRYSTALS-Dilithium, among others, have been selected.

CRYSTALS-Dilithium is a lattice-based algorithm, which follows the well-known Fiat-Shamir scheme. Specifically, its security is based on the difficulty of the problem of finding shortest vectors in lattices. A comprehensive overview of the algorithm can be found at [1].

This paper focuses on a theoretical study of a possible backdoor attack based on [2] that can be inserted in the signature creation process in CRYSTALS-Dilithium. The aim of this attack is to exploit the properties of lattices, modular algebra and the theory of the kleptography to extract the bit-encoded message or document. This scenario occurs in situations where the signature is sent over one channel and the signed message or document is transmitted over another channel.

The attack strategy proposed for CRYSTALS-Dilithium relies on a black-box cryptographic module, which introduces a backdoor into Dilithium's Sign procedure. This vulnerability allows the attacker to extract information about the message being processed. Similar to the approach presented in [2], the attacker can carry out this exploit with just access to the algorithm's public outputs, without necessitating any other forms of interaction or access. Furthermore, it is assumed that the attacker possesses knowledge of both the user's public key pk and the user's secret key sk. In particular, the attack is composed of the following two algorithms: Alg. 1 and Alg. 2.

In the first algorithm, Alg. 1, the function IntervalsKernel is used to assign each bit of M_{bits} to an element within the associated interval. Furthermore, if the length exceeds the available capacity, the StoreIndexes function is used to store the positions of the bits with value 1, allowing the message to be reconstructed later.

In conclusion, this paper introduces a theoretical backdoor attack on the post-quantum digital signature scheme CRYSTALS-Dilithium, assuming the attacker

Algorithm 1 Intervals

$x = \textbf{LightHouse}(M_{bits},\ q)$
if $\theta \leq 256 \cdot k$ **then**
    $\textbf{IntervalsKernel}(M_{bits}, \theta, x)$ ▷ We allocate the bits of M_{bits} according to the x indications.
else
    $\textbf{Break}(M_{bits}) = (M_1, M_2)$
    $z_1, z_2 = \textbf{IntervalsKernel}(M_1, \theta_{M_1}, x),\ \ \textbf{StoreIndexes}(M_2)$
    $z = (z_1, z_2)$
end if
$\textbf{Rebuild}(z)$ ▷ z is restructured as a matrix
return z, x, θ

Algorithm 2 Lighthouse

$\alpha := \textbf{CountOnes}(M_{bits})$ ▷ We count the number of ones
$\beta := \textbf{CountZeros}(M_{bits})$ ▷ We count the number of zeros
$\theta = \alpha + \beta$ ▷ Note $\theta = len(M_{bits})$
if $\alpha \geq \beta$ **then**
    $x := [\frac{\alpha}{\theta} \cdot q] \mod q$
else
    $x := [\frac{\beta}{\theta} \cdot q] \mod q$
end if
return x

knows the user's private and public keys. The attack focuses on the signature creation phase, particularly targeting z. Ongoing research includes exploring attack implementation and the impact on other variables when modifying z.

References

1. Bai, S., Ducas, L., Kiltz, E., Lepoint, T., Lyubashevsky, V., Schwabe, P., Seiler, G., Stehlé, D.: "CRYSTALS-Dilithium, Algorithm Specifications and Supporting Documentation (Version 3.1)", NIST Post-Quantum Cryptography Standardization Round, vol.3, 2021.

2. Ravi, P., Bhasin, S., Chattopadhyay, A., Roy, S.S.: Backdooring Post-Quantum Cryptography: Kleptographic Attacks on Lattice-based KEMs. Cryptology ePrint Archive, 2022.

Detecting SQL Injection Attacks in Aarhus Wi-Fi using an Advanced Intrusion Detection System

Sonia Díaz-Santos[1 [0000−0002−8482−3957]], Pino Caballero-Gil[2 [0000−0002−0859−5876]], and Cándido Caballero-Gil[3 [0000−0002−6910−6538]]

University of La Laguna, St. San Francisco de Paula, 19, 38200, San Cristóbal de La Laguna, Tenerife, Spain
sdiazsan@ull.edu.es, pcaballe@ull.edu.es, ccabgil@ull.edu.es

Extended Abstract

As the reliance on wireless networks continues to grow, the need for robust cybersecurity measures becomes paramount. This paper addresses the challenge of enhancing security in Aarhus Wi-Fi networks by proposing an Intrusion Detection System (IDS) [1] focused on detecting SQL injection attacks with Artificial Intelligence (AI). Leveraging the Aarhus Wi-Fi IDS (AWID) dataset [2], which includes both normal Wi-Fi traffic and simulated attacks, the study specifically targets SQL injection intrusions.

In particular, the proposed system utilizes advanced Machine Learning techniques [3] to categorize and identify instances of SQL injection attacks within the Wi-Fi network. Through a comprehensive analysis of the AWID dataset, we aim to develop an intelligent defense mechanism capable of distinguishing between normal network activity and malicious SQL injection attempts.

The study contributes to the field by offering insights into the application of AI for securing wireless communication [4]. The results and methodologies presented herein provide a foundation for the development of effective intrusion detection systems tailored to the unique challenges posed by SQL injection attacks in Wi-Fi networks. Ultimately, this research endeavors to fortify Aarhus Wi-Fi networks against emerging cybersecurity threats, safeguarding the integrity and confidentiality of wireless communication.

To undertake this project, a preliminary data preprocessing phase is essential to derive a refined dataset. This involves scrutinizing and handling various aspects such as handling null data, scaling, feature encoding, and removing redundant information. Once the data has been cleansed and prepared for analysis, the next crucial step involves a thorough examination to distinguish between normal Wi-Fi traffic and potential attacks. The ultimate goal is to develop an intrusion detection system capable of classifying network activity accurately.

The preprocessing phase is indispensable as it lays the foundation for a robust analysis. Addressing null data, scaling for uniformity, encoding features for effective interpretation, and eliminating redundancies contribute to the creation

of a high-quality dataset. This meticulously curated dataset becomes the cornerstone for training a machine learning model to discern patterns indicative of normal or malicious Wi-Fi traffic.

It is crucial to highlight the necessity of a balanced dataset. The system's ability to accurately differentiate between normal network traffic and potential attacks hinges on having a representative distribution of both cases. Imbalances, where there are too many instances of normal network traffic and too few of attacks, can lead to a skewed model. Therefore, careful consideration is given to ensure that the dataset is well-balanced, preventing the model from leaning heavily towards categorizing all instances as normal network traffic.

Moreover, it is imperative to avoid providing the system with all instances of normal Wi-Fi traffic followed by instances of attacks during the training phase. This approach may lead the model to generalize and treat all instances as normal, resulting in poor performance during real-world application. Instead, a balanced mix of both normal and attack instances is essential for effective learning, enabling the system to make accurate distinctions in identifying potential security threats in Wi-Fi networks.

References

1. H. Liao, C. Richard Lin, Y. Lin, K. Tung, "Intrusion detection system: A comprehensive review," in Journal of Network and Computer Applications, vol. 36, Issue 1, 2013, pp. 16-24, ISSN 1084-8045, doi: 10.1016/j.jnca.2012.09.004.
2. AWID3 Dataset, https://icsdweb.aegean.gr/awid/awid3. University of the AEGEAN.
3. P. Mishra, V. Varadharajan, U. Tupakula and E. S. Pilli, "A Detailed Investigation and Analysis of Using Machine Learning Techniques for Intrusion Detection," in IEEE Communications Surveys & Tutorials, vol. 21, no. 1, pp. 686-728, Firstquarter 2019, doi: 10.1109/COMST.2018.2847722.
4. M. Waqas, S. Tu, Z. Halim et al. "The role of artificial intelligence and machine learning in wireless networks security: principle, practice and challenges," in Artif Intell, Rev 55, pp. 5215–5261, 2022, doi: 10.1007/s10462-022-10143-2.

An IoT Lab dedicated to cybersecurity – teaching, learning and research

Georg Hackenberg[1] [0000-0003-3913-4148] and Mario Jungwirth[2] [0000-0002-1652-8745]

[1] University of Applied Sciences Upper Austria, Wels Campus,
Wels, Austria
[2] Research Group Smart Mechatronics Engineering
University of Applied Sciences Upper Austria, Wels Campus,
Wels, Austria
mario.jungwirth@fh-wels.at

Abstract. An IoT Lab dedicated to cybersecurity, which focuses on teaching and research, development, and testing of secure IoT systems has been established at the University of Applied Sciences Upper Austria, Wels Campus. It will be organized to facilitate the understanding and mitigation of potential security risks associated with IoT devices and networks. The lab will be staffed by faculty members with expertise in cybersecurity, as well as graduate and undergraduate students working on research projects. The goal of the lab will be to advance knowledge in IoT cybersecurity, develop new security technologies and strategies, and train the next generation of cybersecurity professionals.

Keywords: cybersecurity, device analysis, network security, software security, data security, teaching and research in cybersecurity.

1 IoT lab scheme and objectives

The IoT lab focuses on cybersecurity, is designed to explore and address the unique security challenges posed by IoT devices and networks, and has the following areas and components as illustrated in Fig. 1:

a. Device Analysis Area: This area is equipped with various IoT devices, hardware interfaces, and debugging tools. It is used to perform hard- and software security analysis on IoT devices, including firmware reverse engineering, hardware tampering, and side-channel analysis.

b. Network Security Area: This area focus on the security of IoT networks. It includes tools and software for monitoring network traffic, performing penetration testing, and analyzing network vulnerabilities. It could also simulate different network environments to study potential security threats and mitigation techniques.

c. Data Security and Privacy Area: This area focus on the security and privacy of data generated by IoT devices. It includes tools for data encryption, anonymization, and secure data transmission and storage. It is used also to study the implications of data privacy regulations on IoT data.

d. Cyber-Physical Systems Area: Given that many IoT systems interact with the physical world (e.g., industrial control systems, smart home devices, etc.), this area

focuses on the security of cyber-physical systems. It includes setups for studying and demonstrating attacks and defenses in such systems.

Furthermore, an **Education and Training Area** is established, which is used for teaching students about IoT cybersecurity through lectures, workshops, and hands-on exercises. It could also be used for hosting cybersecurity competitions, such as Capture The Flag (CTF) events, to give students practical experience in dealing with cyber threats. Finally, an **Engineering and Research Area** is also part of the concept to work on projects and enable cooperation with industry.

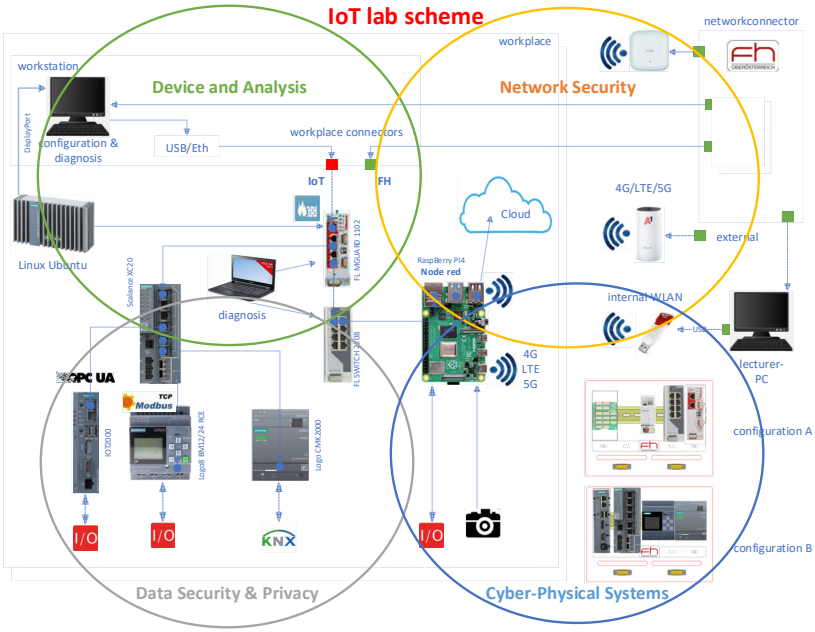


Fig. 1. IoT lab scheme with typical components and areas established at the University of Applied Sciences Upper Austria, Campus Wels.

Acknowledgements

We would like to thank our training partners and sponsors Siemens AG Austria, Phoenix Contact and B&R Industrial Automation for the valuable support. Thanks for their valuable support in setting up the IoT laboratory go to the team from the Department of Electrical Engineering and Electronics.

References

1. Di Martino, Beniamino & Rak, Massimiliano & Ficco, M. & Esposito, Antonio & Maisto, S.A. & Nacchia, Stefania. (2018). Internet of things reference architectures, security and interoperability: A survey. Internet of Things. 1-2. 99-112. 10.1016/j.iot.2018.08.008.
2. O'hearon, Keejoh & Mckee, Michael & Hossain, Nour & Canbaz, M Abdullah. (2021). IoT Privacy and Security in Teaching Institutions: Inside The Classroom and Beyond.

Author Index

Universidad de Las Palmas de Gran Canaria

Johannes Kepler University Linz

University of Applied Sciences Upper Austria

Museo Elder de la Ciencia y la Tecnología

Fundación Universitaria de Las Palmas