Report from Dagstuhl Seminar 23492

Model Learning for Improved Trustworthiness in Autonomous Systems

Ellen Enkel*[1], Nils Jansen*[2], Mohammad Reza Mousavi*[3], and Kristin Yvonne Rozier*[4]

1 **Universität Duisburg-Essen, DE.** ellen.enkel@uni-due.de
2 **Ruhr-Universität Bochum, DE.** n.jansen@science.ru.nl
3 **King's College London, GB.** mohammad.mousavi@kcl.ac.uk
4 **Iowa State University – Ames, US.** kyrozier@iastate.edu

Abstract

The term of a model has different meanings in different communities, e.g., in psychology, computer science, and human-computer interaction, among others. Well-defined models and specifications are the bottleneck of rigorous analysis techniques in practice: they are often non-existent or outdated. The constructed models capture various aspects of system behaviours, which are inherently heterogeneous in nature in contemporary autonomous systems. Once these models are in place, they can be used to address further challenges concerning autonomous systems, such as validation and verification, transparency and trust, and explanation. The seminar brought together the best experts in a diverse range of disciplines such as artificial intelligence, formal methods, psychology, software and systems engineering, and human-computer interaction as well as others dealing with autonomous systems. The goal was to consolidate these understanding of models in order to address three grand challenges in trustworthiness and trust: (1) understanding and analysing the dynamic relationship of trustworthiness and trust, (2) the understanding of mental modes and trust, and (3) rigorous and model-based measures for trustworthiness and calibrated trust.

Seminar December 3–8, 2023 – https://www.dagstuhl.de/23492
2012 ACM Subject Classification General and reference → Reliability; General and reference → Validation; General and reference → Verification; Computing methodologies → Artificial intelligence; Applied computing → Psychology
Keywords and phrases artificial intelligence, automata learning, autonomous systems, cyber-physical systems, formal methods, machine learning, safety, safety-critical systems, self-adaptive systems, software evolution, technology acceptance, trust
Digital Object Identifier 10.4230/DagRep.13.12.24

1 Executive Summary

Ellen Enkel
Nils Jansen
Mohammad Reza Mousavi
Kristin Yvonne Rozier

This report documents the program and the outcomes of Dagstuhl Seminar 23492 "Model Learning for Improved Trustworthiness in Autonomous Systems". Autonomous systems increasingly enter our everyday life. Consequently, there is a strong need for safety, correctness,

* Editor / Organizer

Model Learning for Improved Trustworthiness in Autonomous Systems, *Dagstuhl Reports*, Vol. 13, Issue 12, pp. 24–47
Editors: Ellen Enkel, Nils Jansen, Mohammad Reza Mousavi, and Kristin Yvonne Rozier
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

trust, and explainability. Well-defined models with clear semantics pose a convenient way to address these requirements. The area of model learning provides a structured way to obtain models from data. However, autonomous systems operate in the real world and pose challenges that go beyond the state-of-the-art in model learning. The technical challenges addressed in the seminar are system evolution and adaptations, learning heterogeneous models (addressing aspects such as discrete and continuous behaviours, stochastic, and epistemic uncertainty), and compositional learning. Our vision is that model learning is a key enabler solving the bottleneck of lack of specifications and models in various typical applications and hence, our seminar addressed fundamental challenges to enable impact in a number of application areas. In the seminar we brought together experts in (1) the domain of trust and technology acceptance, (2) the technical methods of model learning, and (3) the applications of model learning in robotics and autonomous systems. The first area includes domain experts in technology management, psychology, and trust; Technical methods include automata learning, synthesis of logical specifications, statistical model learning, machine learning, system identification, and process mining. Application experts include validation and verification, transparency and trust, and explainability, as well as experts in their application in robotics (planning, physical design and validation) and autonomous systems. With this seminar, we actively encouraged the interaction between experts and young researchers in the interdisciplinary areas of artificial intelligence, software engineering, autonomous systems, and human factors both from academia and industry. Content-wise, we emphasized the following directions: model learning techniques for AI-enabled autonomous systems: This involves recent techniques for learning models of evolving and variability-intensive systems; as well as application of model-learning to increase transparency and trust in robotics and autonomous system.

We discussed the following technical research questions during the seminar:

- How can we efficiently learn about system evolution and adaptation?
- How can we learn heterogeneous models, possibly by separating orthogonal concerns?
- How can we scale the mode learning?

Additionally, we discussed the following multi-disciplinary research questions:

- How can adaptive model learning be used to focus the validation and verification effort in evolving systems?
- How can learn model contribute to trust in autonomous systems?
- What types of models can be used to provide understandable explanations for AI-enabled and autonomous systems?

During the discussion four research questions and working groups emerged, that captured their discussion in scientific papers. The following is a short abstract of each paper that is currently in development.

Working group 1: Foundations of Learned Model Validation in the Age of AI

Models serve as the fundamental basis for the design, synthesis, verification, and implementation of software systems, yet before we can use the model for any of these, we must validate the model against the expectations on the system and/or against the real behavior of the system. In many development paradigms, an emerging trend is to move away from entirely human-designed models to models learned using automated techniques. We contribute a concrete roadmap for validating learned behavioral models comprised by a range of popular components. We pinpoint the current limits of model validation, provide insight into the reasons behind these limitations, and identify challenges that should serve as targets for

future research. By means of a running example of a cruise controller with different techniques for model learning, we show how guarantees derived from these techniques interplay with the validation challenges.

Working group 2: Mental Models, Human Models, Trust

Transposing the notion of interpersonal trust into the field of Computer Science, leads to the assumption that a high level of trust might also be pivotal in Human-Computer-Interaction (and even in interactions between two autonomous systems), since it enables the trustor (whether human or non-human) to make better predictions about the trustee. However, whereas humans possess an inherent "trust module," non-human agents lack such a component. Addressing this challenge, the present paper proposes a framework formalizing the trust relationship between a human and an autonomous system, aimed at creating more trustworthy Human-Computer-Interactions.

Working group 3: Research Agenda for Active Automata Learning

We conduct a survey of active automata learning methods, focusing on the different application scenarios (application domains, environment, and desirable guarantees) and the overarching goals that stem from them. We identify the challenges to achieve these goals. We organize a (short) bibliographic study highlighting the state-of-the-art and the technical challenges that are derived from the general goals and give some elements of answers related to these challenges.

Working group 4: Dynamic Interaction of Trust and Trustworthiness in AI-Enabled Systems

Trust is a user-centered notion, while trustworthiness pertains to the properties of the system. They dynamically influence each other, and interact with each other. We focus on AI-enabled systems, where establishing trustworthiness is challenging. In this paper we propose a framework for assessing trust and trustworthiness, and their alignment (calibration). We investigate factors that can influence them. We draw two case studies to illustrate our framework, and derive recommendations based on the insights we gain.

Besides interesting discussions, the four working groups focused on creating scientific articles, capturing their thoughts and insights. As a result, the four articles will be submitted to a special issue on Trust and Trustworthiness in Autonomous Systems International Journal of Software Tools for Technology Transfer (JSTTT) in 2024. Additionally, the organizers of this Dagstuhl Seminar will organize a track at the (A)ISoLa conference of October and November 2024 in Greece to deepen the discussion with the Dagstuhl attendees and with additional experts on this topic. Post-conference proceedings will document the insights gained.

2 Table of Contents

3 Overview of Talks

3.1 Learning-based Testing of Autonomous Systems

Bernhard K. Aichernig (TU Graz, AT)

Joint work of Martin Tappler, Edi Muškardin, Bernhard Aichernig, Bettina Könighofer
Main reference Martin Tappler, Edi Muškardin, Bernhard K. Aichernig, Bettina Könighofer: "Learning Environment Models with Continuous Stochastic Dynamics", CoRR, abs 2306.17204, 2023
URL https://doi.org/10.48550/arXiv.2306.17204

Learning-based testing combines model learning and model-based test-case generation in order to further automate the testing process. In our recent research we are aiming for testing autonomous systems, including autonomous cars (ADAS) and robots. In order to facilitate interaction with machine-learned systems, we have developed the AALpy library, an active automata learning library for Python. Next to standard automata learning algorithms, like L*, it supports the learning of stochastic models, like Markov decision processes or stochastic Mealy machines that are necessary to capture the policies of autonomous systems. Next to the algorithms implemented in AALpy, we also developed methods to learn timed models, like Timed Automata.

Our current research focuses on the scalability of automata learning to large state spaces in order to test autonomous systems. We aim to provide insights into the decisions faced by the agent by learning an automaton model of environmental behavior under the control of an agent. However, for most control problems, automata learning is not scalable enough to learn a useful model. In order to overcome this limitation, we compute an abstract state-space representation, by applying dimensionality reduction and clustering on the observed environmental state space. The stochastic transitions are learned via passive automata learning from observed interactions of the agent and the environment. In an iterative model-based reinforcement learning (RL) process, we sample additional trajectories to learn an accurate environment model in the form of a discrete-state Markov decision process (MDP). We applied our automata learning framework on popular RL benchmarking environments in the OpenAI Gym, including LunarLander, CartPole, Mountain Car, and Acrobot. Our results show that the learned models are so precise that they enable the computation of policies solving the respective control tasks.

3.2 Lifelong Learning of Reactive Systems

Bernhard Steffen (TU Dortmund, DE), Falk Howar (TU Dortmund, DE)

Joint work of Alexander Bainczyk, Bernhard Steffen, Falk Howar
Main reference Alexander Bainczyk, Bernhard Steffen, Falk Howar: "Lifelong Learning of Reactive Systems in Practice", in Proc. of the The Logic of Software. A Tasting Menu of Formal Methods – Essays Dedicated to Reiner Hähnle on the Occasion of His 60th Birthday, Lecture Notes in Computer Science, Vol. 13360, pp. 38–53, Springer, 2022.
URL https://doi.org//10.1007/978-3-031-08166-8_3

The talk presents our lifelong learning framework for continuous quality control that integrates automata learning, model checking, and monitoring into a six-phase continuous improvement cycle that is design to capture entire system life-cycles. Technical backbone of our framework is ALEX, our open source, web-based learning tool for defining adequate test blocks, as well

as for serving as test execution environment and as platfrom for learning Mealy machines. Key to the industrial success of our framework are a) the guarantee that the level of quality can only increase when using our framewok, b) the continuous improvement the originally customer-provided (regression) test suites, c) the maintenance of achieved quality levels even across system changes, and d) the visualization of system changes using automatically generated difference trees and difference automata. All this is illustrated using an adaptive cruise control system (ACC) that has been implemented in a one year students project.

3.3 Neuro-symbolic Model Learning for Controller Synthesis and Verification

Jyotirmoy Deshmukh (USC – Los Angeles, US)

Joint work of Navid Hashemi, Bardh Hoxha, Tomoya Yamaguchi, Danil V. Prokhorov, Georgios Fainekos, Jyotirmoy Deshmukh, Xin Qin, Lars Lindemann
Main reference Navid Hashemi, Bardh Hoxha, Tomoya Yamaguchi, Danil V. Prokhorov, Georgios Fainekos, Jyotirmoy Deshmukh: "A Neurosymbolic Approach to the Verification of Temporal Logic Properties of Learning-enabled Control Systems", in Proc. of the ACM/IEEE 14th International Conference on Cyber-Physical Systems, ICCPS 2023, (with CPS-IoT Week 2023), San Antonio, TX, USA, May 9-12, 2023, pp. 98–109, ACM, 2023.
URL https://doi.org//10.1145/3576841.3585928

Synthesizing control policies for high-dimensional, highly nonlinear/hybrid systems that guarantee satisfaction of safety and performance properties of the system is a significant challenge problem. In this talk, we will review some recent work on neuro-symbolic techniques, i.e., techniques that combine learned neural models of the system dynamics with symbolic techniques to synthesize control policies. We assume that safety/performance properties of the system are specified using bounded horizon Signal Temporal Logic (STL) formulas, and provide algorithms to automatically synthesize controllers that are guaranteed to satisfy these STL specifications. We also show how the learned neural surrogate models can be used for both deterministic and probabilistic verification of the closed-loop system. Our verification results on surrogate models can be mapped to probabilistic correctness guarantees on the original system, which allows a mechanism to address real-world issues such as the sim2real gap and distribution shifts between the design and deployment.

References

1 N. Hashemi, B. Hoxha, T. Yamaguchi, D. Prokhorov, G. E. Fainekos, J. V. Deshmukh. A Neurosymbolic Approach to the Verification of Temporal Logic Properties of Learning-enabled Control Systems. In Proceedings of the ACM/IEEE 14th International Conference on Cyber-Physical Systems, 2023.
2 N. Hashemi, X. Qin, J. V. Deshmukh, G. E. Fainekos, B. Hoxha, D. Prokhorov, and T. Yamaguchi, Risk-Awareness in Learning Neural Controllers for Temporal Logic Objectives. In Proc. of American Control Conference (ACC), 2023.
3 N. Hashemi, X. Qin, L. Lindemann, and J. V. Deshmukh, Data-Driven Reachability Analysis of Stochastic Dynamical Systems with Conformal Inference. To appear in the Proc. of IEEE Conf. on Decision and Control (CDC), 2023.

3.4 Learning Featured Transition Systems

Sophie Fortz (University of Namur, BE)

Main reference Sophie Fortz: "Variability-aware Behavioural Learning", SPLC (B) 2023: 11-15.
URL https://researchportal.unamur.be/en/studentTheses/learning-featured-transition-systems

I recently obtained my PhD from the University of Namur (Belgium). During these last four years, I worked on learning the behaviour of Variability-intensive Systems (VISs). In my thesis, entitled "Learning Featured Transition Systems", I combined classical learning algorithms (L^*) and deep learning techniques (recurrent neural networks) to learn behavioural models of VISs.

Variability-intensive Systems (VISs) are software-based systems whose characteristics and behaviour can be modified by the activation or deactivation of some options. Addressing variability proactively during software engineering (SE) activities means shifting from reasoning on individual systems to *reasoning on families of systems*. Adopting appropriate variability management techniques can yield important *economies of scale* and quality improvements. Conversely, variability can also be a curse, especially for Quality Assurance (QA), i.e. *verification and testing* of such systems, due to the combinatorial explosion of the number of software variants. Indeed, by combining only 33 Boolean options, we can define more variants of a system than the number of people on Earth. Verifying or testing each variant individually is thus impossible in most practical cases.

About a decade ago, *Featured Transition Systems (FTSs)* were introduced as a formalism to represent, and reason on, *the behaviour of VISs*. Instead of representing each variant by a (classical) transition system, an FTS bears annotations that relate transitions to options through *feature expressions*. FTSs thus make it possible to reason at the family level by modelling all the variants of a system in a single behavioural model. FTSs have been shown to significantly improve the possibilities and execution time of automated QA activities such as model-checking and model-based testing. They have also shown their usefulness to guide design exploration activities. Yet, as most model-based approaches, FTS modelling requires both *strong human expertise and significant effort* that would be unaffordable in many cases, in particular for large legacy systems with outdated specifications and/or systems that evolve continuously.

Therefore, in this thesis we aim *to automatically learn FTSs from existing artefacts*, to ease the burden of modelling FTS and support continuous QA activities. To answer this research challenge, we propose a two-phase approach. First, we rely on deep learning techniques to locate variability from execution traces. For this purpose, we implemented a tool called *VaryMinions*. Then, we use these annotated traces to learn an FTS. In this second part, we adapt the seminal L^* algorithm to *learn behavioural variability*. Both frameworks are open-source and we evaluated them separately on several datasets of different sizes and origins (e.g., software product lines and configurable business processes).

3.5 Trust in Conversational Agents

Effie Lai-Chong Law (Durham University, GB)

My long-term research focus centres on usability and elevating user experience (UX) methodologies, pivotal aspects within the realm of Human-Computer Interaction (HCI). My current research foci cover three areas: Conversational agents (CAs), Multisensory emotion recognition, and eXtended Reality (XR). Specifically, my investigations delve into the intricate dynamics of trust within customer service Conversational Agents (CAs). Through a series of empirical studies, I study the nuanced impact of various factors – ranging from CA conversational performance and humanlikeness to the task's nature – on the perceived trust in these AI-infused applications [1, 2]. Within an ongoing project, I am examining the role of inclusive design in augmenting the trust of older adults in online banking applications seamlessly integrated with ChatGPT. Simultaneously, I endeavour to examine the application of multisensory emotion recognition to provide support for the mental health and well-being of young adults. In the expansive domain of eXtended Reality (XR), which encompasses augmented, virtual, and mixed reality, I am exploring their educational potential across diverse sectors [3].

References

1 Law, E. L-C., Følstad, A., & Van As, N. (2022, October). Effects of Humanlikeness and Conversational Breakdown on Trust in Chatbots for Customer Service. In Nordic Human-Computer Interaction Conference (pp. 1-13).
2 Hobert, S., Følstad, A., & Law, E. L-C. (2023). Chatbots for active learning: A case of phishing email identification. International Journal of Human-Computer Studies, 179, 103108.
3 Thanyadit, S., Heintz, M., & Law, E. L-C. (2023, April). Tutor In-sight: Guiding and Visualizing Students' Attention with Mixed Reality Avatar Presentation Tools. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-20).

3.6 Reliable Learning for Safe Autonomy

Nicola Paoletti (King's College London, GB)

Main reference Francesca Cairoli, Nicola Paoletti, Luca Bortolussi: "Conformal Quantitative Predictive Monitoring of STL Requirements for Stochastic Processes", in Proc. of the 26th ACM International Conference on Hybrid Systems: Computation and Control, HSCC 2023, San Antonio, TX, USA, May 9-12, 2023, pp. 1:1–1:11, ACM, 2023.
URL https://doi.org//10.1145/3575870.3587113

With learning models increasingly being deployed in safety-critical systems, ensuring that predictions are reliable and providing guarantees for such models is paramount. In this talk, I will introduce conformal prediction, an assumption-free technique to obtain rigorous probabilistic guarantees on the predictions of any machine learning model. I will review recent work where my collaborators and I have applied this technique for data-driven verification of stochastic processes.

3.7 Robust and Reliable Reinforcement Learning

Marnix Suilen (Radboud University Nijmegen, NL)

My research focuses on planning and learning in probabilistic environments, typically modeled by Markov decision processes (MDPs) or partially observable MDPs (POMDPs). These MDPs are the standard models for decision-making problems where the outcome of an agent's choice is governed by a probability distribution. In the partially observable setting of POMDPs, the agent can only base its decisions on the given observations instead of the full state information. Recent results are in a model-based reinforcement learning (RL) setting, where the goal is to find an optimal decision-making policy by inferring a model from data and applying planning algorithms to find such a policy. Our results in this area can be split into two streams.

The first stream is the online RL setting, where we have sample access to the underlying model. By using robust dynamic programming as planning method, we can learn strategies that are probably approximately correct (PAC) optimal for the unknown true environment. Alternatively, linearly updating intervals (LUI) can find similarly performing strategies without any formal PAC guarantees. The advantage of LUI is that it is faster and requires less data than PAC learning, and it can easily adapt to new data that is inconsistent with previously encountered data, for example, due to a change in the probability distributions of the underlying true model. LUI was presented at NeurIPS 2022 [1].

The second stream is the offline RL setting, where only a fixed data set is given, and no further data can be collected. In the safe policy improvement (SPI) problem, we are given such a data set and the policy that collected it, known as the behavior policy. The goal is to, with a PAC guarantee, compute a new policy that outperforms this behavior policy with high confidence. We extended a standard method for SPI to the partially observable (POMDP) setting and devised alternative methods to compute the PAC guarantee such that it requires significantly less data. These results were published at AAAI 2023 [2] and IJCAI 2023 [3], respectively.

References

1 Marnix Suilen, Thiago Simão, David Parker, Nils Jansen. *Robust Anytime Learning of Markov Decision Processess*. NeurIPS 2022.
2 Thiago Simão, Marnix Suilen, Nils Jansen. *Safe Policy Improvement for POMDPs via Finite-State Controllers*. AAAI 2023.
3 Patrick Wienhöft, Marnix Suilen, Thiago Simão, Clemens Dubslaff, Christel Baier, Nils Jansen. *More for Less: Safe Policy Improvement with Stronger Performance Guarantees*.

3.8 Learning to avoid finding bugs

Thomas Arts (QuviQ AB – Gothenburg, SE)

Joint work of Thomas Arts, John Hughes, Ulf Norell, Nicholas Smallbone
Main reference John Hughes, Ulf Norell, Nicholas Smallbone, Thomas Arts: "Find more bugs with QuickCheck!", in Proc. of the 11th International Workshop on Automation of Software Test, AST@ICSE 2016, Austin, Texas, USA, May 14-15, 2016, pp. 71–77, ACM, 2016.
URL https://doi.org//10.1145/2896921.2896928

When testing industrial size software systems, one may be confronted with more than one fault. Clearly you want to report all faults found in a concise, understandable, to the software development team. The more faults they get at once, the better they can plan and fix errors in parallel.

With a QuickCheck state machine model, test cases are automatically generated and automatically shrunk to minimal test cases. The latter is important for size and understandability of the reported failure. However, there is an unexpected drawback when there are faults than can be demonstrated with a test sequence of only two or three commands. Such short sequences are very likely to appear as a subsequence of any larger generated test case. When shrinking the test case, it is very likely that just this sequence is reported, over and over again.

The reason to generate larger test cases is that in a realistic piece of software you may have 80 to 200 different API calls. To cover reasonably interesting sequences, one may need to cover quite a few to get to interesting states in the system. It turns out to be more effectiove to generate a long sequence and shrink when a failure is detected, then to explore subsets of the API in a more systematic way, by selecting a subset of the API and test with only that, after which one selects another subset... there are very many subsets.

Hughes et al [1] came up with some heuristics to avoid finding the same error twice by avoiding to generate it and avoiding to shrink to it. This resulted in being able to report more faults at once.

There is some clear open area for improvement in those heuristics as can be shown in a demo.

References

1 John Hughes, Ulf Norell, Nicholas Smallbone, and Thomas Arts. 2016. Find more bugs with QuickCheck! In Proceedings of the 11th International Workshop on Automation of Software Test (AST '16). Association for Computing Machinery, New York, NY, USA, 71–77.

3.9 Safe and Efficient Multi-agent Learning for Human–AI Collaboration

Mustafa Celikok

My work focuses on developing multi-agent learning methods that can learn to collaborate with both humans and AI agents with minimal prior coordination. The ideal outcome of this line of research is a learning agent who can be dropped in any team of agents which it has never seen before, and quickly learn how to collaborate with them towards a common goal. This capability is referred to as _ad hoc teamwork._ During my PhD, I have focused on

investigating model-based multi-agent reinforcement learning methods where the learning agent tries to infer an explicit model of its human partners. The model space in this work is derived from cognitive science research. My current work has three different paths.

(I) Bayesian Multi-agent Reinforcement Learning for Human–AI Collaboration
In this path, I address the following research questions.

RQ1. What assumptions can we make about the human partner in human–AI collaboration, and from where should they come?
RQ2. How can the AI take advantage of the assumptions and models of human behaviour?
RQ3. How can we guarantee safe and reasonable human–AI collaboration when our assumptions about the human partner are violated?

(II) Learning Dynamics of Continual Multi-agent Learning Agents for Long-Term Safety
There will come a point in time, where multi-agent learning systems are deployed in the real world to interact with humans and other learning systems, and learn continuously. Such systems might behave safely now, but since they are ever learning, their behaviour now does not guarantee future safety. They also create a complex system that is difficult to analyse. Dynamical systems theory allows us to study and _verify_ how the long-term learning dynamics of such agents will evolve. In our recent work, we stepped into this domain and studied what type of sets would learning agents in different ad hoc settings converge to.

(III) Sample Complexity and Non-asymptotic Results for Multi-agent Learning in Ad Hoc Collaboration
Even though there is a lot of work in terms of sample complexity when it comes to learning equilibria, most of these results either rely on all agents using similar learning algorithms, or that they are internally coupled. By relaxing these assumptions, I aim to complement the asymptotic results of Path II with finite-time results.

3.10 Chances and Challenges of AI-enhanced Automation for Automotive Health

Monique Dittrich (CARIAD – Berlin, DE)

Main reference Addam Mustapha, David Matusiewicz: "Automotive Health-A Systematic Overview of opportunities & boundaries" in Proceedings of the 12th edition of Numerical Analysis and Optimization Days.
URL https://hal.science/hal-02891566/document
Main reference Monique Dittrich: "Persuasive Technology to Mitigate Aggressive Driving : A Human-centered Design Approach", 2021.
URL https://doi.org//10.25972/OPUS-23022
Main reference Monique Dittrich: "The Car as a Personal Space to Improve your Health", in Proc. of the 18th International Conference on Persuasive Technology, Adjunct Proceedings co-located with PERSUASIVE 2023, Eindhoven University of Technology, Eindhoven, The Netherlands, April 19th – 21st, 2023, CEUR Workshop Proceedings, Vol. 3474, CEUR-WS.org, 2023.
URL https://ceur-ws.org/Vol-3474/paper23.pdf

Holding a PhD in Human-Computer-Interaction (Dittrich, 2020), I have been working in automotive research and development for around a decade, devoting myself to the human-centered design of Human-Machine-Interfaces of all kinds. By this, I dealt with overarching topics such as user experience and usability, emotion recognition, or behavior change.

Currently, my research concentrates on in-vehicle applications that are intended to improve the health and well-being of people the car. This area of research is scientifically manifested under the term "Automotive Health", whereby health (in the broader sense) is viewed through the lens of the automobile. Automotive Health deals with the questions of how health data can be sensed and interpreted in the vehicle environment, and how this information can be used to enable (or enrich) measures taken to maintain, restore, or strengthen occupants' well-being, (medical) health, and safety (Mustapha, & Matusiewicz, 2018). With a special focus on automatization through the use of Artificial Intelligence (AI) in this context, my research focusses on the following topics:

Data-driven decision making in the context of Automotive Health. Addressing the first pillar of Automotive Health, the sensing and interpretation of health data, inspiration can be found in the area of wearable technology. Wearables capture a wide range of vital signs (e.g., heart rate, respiratory rate, blood oxygen), and combine them into an meaningful information (e.g., stress, mood, sleeping quality), giving the user a quick and understandable indication of her state of health. For example, Garmin calculates the users performance condition based on her pace, heart rate, heart rate viability (HRV), and the maximum amount of oxygen that the body can take in (V02 max). However, compared to "lifestyle" use cases, such as sports or sleep, Automotive Health also addresses factors that are safety and medically relevant, such as fatigue, distraction, or sudden medical emergencies, and therefore require a higher level of accuracy and confidence. In this regard, the question of trustworthiness is primarily technical in nature. While the sensing and interpretation of health data is well established in the area of wearable technology and medicine, the topic is still in its infancy in the automotive domain.

Generative AI for mental health inventions. Generative AI, i.e., AI that generates texts, images, or other media content using generative models, existing knowledge, and prompts, is on everyone's lips. In the context of Automotive Health, I am primarily concerned with the generation of health-related content, more precise audio-guided mindfulness exercises (Dittrich, 2023). Mindfulness can be described as the process of bringing greater awareness to the present moment. Most practices to achieve this state are similar in their basic procedure, combining variations of breathing techniques, exercises for the awareness of body, mind, sensations and emotions, or mental images and thoughts. Mindfulness exercises can be also done while driving a car, since they are supposed to increase attention, lower stress, and reduce fatigue in order to promote driving safety. Generative AI would allow to adapt the exercise to the context in real time, i.e. the traffic situation or the person's current state of health, and to create all time new experiences for the user without the need to produce content through human speakers. However, in the context health one of the biggest problem is the generation of harmful content and the question of responsibility, when mental health deteriorates due to the content. Moreover, also trustworthiness plays an important role. When it comes to mental health practices, trust between the client and the counselors (even if embodied through audio content) is essential for the effectiveness of the intervention. So far, little is known about whether AI-generated content can overcome this hurdle.

3.11 Trust in and acceptance of technology

Ellen Enkel (Universität Duisburg-Essen, DE)

Joint work of Ellen Enkel, Magnus Liebherr, Monika Hengstler, Sabrina Dueli
Main reference Ellen Enkel: "To get consumers to trust AI, show them its benefits", Harvard Business Review Blog, 17 April 2017.
URL https://hbr.org/2017/04/to-get-consumers-to-trust-ai-show-them-its-benefits
Main reference Monika Hengstler, Ellen Enkel, Selina Duelli: "Applied artificial intelligence and trust–The case of autonomous vehicles and medical assistance devices", Technological Forecasting and Social Change, Vol. 105, pp. 105–120, 2016.
URL https://doi.org//10.1016/j.techfore.2015.12.014

Automation with inherent artificial intelligence (AI) is increasingly emerging in diverse applications, for instance, autonomous vehicles and medical assistance devices. However, despite their growing use, there is still noticeable skepticism in society regarding these applications. Drawing an analogy from human social interaction, the concept of trust provides a valid foundation for describing the relationship between humans and automation. Accordingly, we explore how firms systematically foster trust regarding applied AI. In the paper presented, the empirical analysis using nine case studies in the transportation and medical technology industries, illustrates the dichotomous constitution of trust in applied AI. Concretely, it emphasizes the symbiosis of trust in the technology as well as in the innovating firm and its communication about the technology. In doing so, it provides tangible approaches to increase trust in the technology and illustrate the necessity of a democratic development process for applied AI and provides a basis for our joined paper at this Dagstuhl Seminar.

3.12 Extracting Behavioral Models of Users/Applications by Combining Automata and Machine Learning

Fatemeh Ghassemi (University of Tehran, IR)

Joint work of Fatemeh Ghassemi, Zeynab Sabahi-Kaviani
Main reference Zeynab Sabahi-Kaviani, Fatemeh Ghassemi: "An Enhanced Encrypted Traffic Classifier via Combination of Deep Learning and Automata Learning", 2023.
URL https://doi.org//10.21203/rs.3.rs-3290610/v1
Main reference Fatemeh Marzani, Fatemeh Ghassemi, Zeynab Sabahi-Kaviani, Thijs van Ede, Maarten van Steen: "Mobile App Fingerprinting through Automata Learning and Machine Learning", in Proc. of the IFIP Networking Conference, IFIP Networking 2023, Barcelona, Spain, June 12-15, 2023, pp. 1–9, IEEE, 2023.
URL https://doi.org//10.23919/IFIPNETWORKING57963.2023.10186420
Main reference Zeynab Sabahi-Kaviani, Fatemeh Ghassemi, Zahra Alimadadi: "Combining Machine and Automata Learning for Network Traffic Classification", in Proc. of the Topics in Theoretical Computer Science – Third IFIP WG 1.8 International Conference, TTCS 2020, Tehran, Iran, July 1-2, 2020, Proceedings, Lecture Notes in Computer Science, Vol. 12281, pp. 17–31, Springer, 2020.
URL https://doi.org//10.1007/978-3-030-57852-7_2

My research focuses on the verification of real-world systems using model-based techniques such as model checking. I am concerned about approaches to make this technique feasible through reduction techniques or runtime monitoring. In recent years, I became interested in using the automata-based models as the behavioral fingerprint of applications for classification by combining automata learning and machine deep learning.

Machine/deep learning approaches are the main used technique to extract the fingerprint of applications from a set of sample data. The false positive rates of these methods are not negligible if they do not consider the temporal relations among events although they

are fast to generate and predict. On the other hand, learned models through automata learning are very precise with high true positives but too sensitive to the order of the events which leads to overfitting and not tolerable to noise and events loss in the distributed setting. So, the false negative rates of these methods are also high while they are not fast to either generate or use. By combining these approaches, we can take advantage of both; we use the FSM-learned models of applications to train our machine learning classifier.

We are using our approach to extract the behavioral models of users in the social network domain. We can interpret the events generated by a group of users with similar age, interest, and gender, ... identified as a specific user cluster in a social network, and generate a predictive behavioral model for those users. The following are possible directions for my research:

- Extending such combination for learned register automata
- Deriving the behavioral model of applications run within the cloud for predictive resource management
- Extending and using such behavioral models as digital twins

3.13 Automation challenges to human perception, emotion, and well-being

Heiko Hecht (Universität Mainz, DE)

Joint work of Heiko Hecht, Carina Röckel, Elisabeth Wögerbauer, Robin Welsch, Marlene Wessels, John Stins
Main reference Carina Röckel, Heiko Hecht: "Regular looks out the window do not maintain situation awareness in highly automated driving", Transportation Research Part F: Traffic Psychology and Behaviour, Vol. 98, pp. 368–381, 2023.
URL https://doi.org//10.1016/j.trf.2023.09.015

My field is human-machine interaction in the broadest sense. I am a cognitive experimental psychologist by training, with current specialization in the field of human factors and engineering psychology. In this context, the following current research might be of interest.

(1) Cyborg perception: On the road to fully autonomous vehicles, driver assistance systems enter a new phase in which human perception can be augmented. We explore how camera-monitor systems can replace traditional mirrors and how they can be used to compensate perceptual deficiencies by enhancing the digital mirror image with traffic-relevant information. Here the reduction of cognitive load is essential.
(2) Situation awareness maintenance: As more and more daily action become partially automated, the user is likely to become engrossed in secondary tasks, which makes it difficult if not impossible to switch back to manual control in critical cases where the automation fails. We explore how situation awareness can be best maintained or regained.
(3) Social robots: We have investigated how people regulate interpersonal distance among each other. The needs for trust, personal space, feelings of crowding, etc. likewise apply to robots. We explore how they may change as a function of robot appearance, situational factors, and habituation.
(4) Cybersickness: Car sickness is an unresolved problem in autonomous driving. We investigate the causes and potential remedies of car sickness and the related phenomena of simulator sickness, sea sickness, etc. Among others we found that even with see-through displays, moving virtual scenes can induces strong cybersickness.

References

1 Wögerbauer, E. M., Hecht, H., Wessels, M. (2023). Camera–monitor systems as an opportunity to compensate for perceptual errors in time-to-contact estimations. Vision, 7, 65. doi.org/10.3390/ vision7040065
2 Röckel, C., & Hecht, H. (2023). Regular looks out the window do not maintain situation awareness in highly automated driving. Transportation Research Part F: Traffic Psychology and Behaviour, 98, 368–381. doi.org/10.1016/j.trf.2023.09.015
3 Welsch, R., Hecht, H., & Stins, J. (2023). Task-relevant social cues affect whole-body approach-avoidance behavior. Scientific Reports 13:8568. doi.org/10.1038/s41598-023-35033-7
4 Kaufeld, M., Mundt, M., Forst, S., & Hecht, H. (2022). Optical see-through augmented reality can induce severe motion sickness. Displays, 74, doi.org/10.1016/j.displa.2022.102283

3.14 Learning and Testing of Real Systems

Leo Henry (University College London, GB)

After a PhD in the SUMO team in Rennes (France) I have been working as a Research Fellow at UCL.

I am mainly interested by how to use complex models to advance the verification (and general understanding) of real systems, which includes tackling questions relating to the availability of models, the complexity to interact with a real system or to control it.

My research stands on two legs: (1) verification of complex systems, with an emphasis on timed systems and timed automata stemming from my PhD. My main work in this area concerns the use of advanced game-strategies to create test case for systems modelled as input-out timed automata; (2) active learning, which I have mostly focused on at UCL. I have worked on practical questions, such as the handling of noise and changes to the target system during learning, or learning of a network of synchronizing models.

Ultimately, I would love to see off-the-shelf tools for learning, testing and verifying systems that developers and engineers could use without a great investment in theoretical matters.

3.15 Enhancing Human Interaction with Automated Technologies: The Role of Adaptability, Trust, and Cognition

Magnus Liebherr (Universität Duisburg-Essen, DE)

I hold a Ph.D. in psychology and currently lead the 'Human Factors' working group at the Chair of Mechatronics at the University of Duisburg-Essen. This position reflects my deep interest in exploring the impact of human factors within the realms of digital technologies, highly automated systems, and their interaction with our ever-evolving environment. Within the scope of my work, I concentrate on three fundamental areas: adaptability, trust, and cognition. Additionally, I delve into associated constructs such as mental workload. In the pursuit of my interdisciplinary research initiatives, I employ these themes across a diverse spectrum of technologies. This encompasses automation within the automotive and shipping sectors, collaborative robot systems, electromobility, m-health, and digital media.

Adapting to new technologies empowers individuals to effectively navigate changing environments or conditions, resulting in heightened performance, reduced mental workload, as well as increased comfort. Drawing from evolutionary biology, we distinguish between three facets of adaptability: Trait Adaptability, which represents a stable characteristic; State Adaptability, signifying the extent of immediate adjustment; and Adaptation, encapsulating the ongoing process of adaptation. Our previous research indicates that individuals' trait adaptability significantly influences the level of trust placed in a technology. Furthermore, we identified significant differences in state adaptability between subjectively perceived and objectively measured. Intriguingly, neither age nor prior experience with the technology accounted for this discrepancy. However, individuals exhibiting limited adaptation to a simulated environment spent considerably less time in the simulator, primarily due to the onset of simulator sickness.

Trust in technology encompasses individuals' confidence, belief, and reliance on the integrity, reliability, and security of a technological system. Findings on the interaction between humans and automation highlight the dependency of trust on perceived control over automated operations. Our contention is that heightened perceived control fosters trust in technology by diminishing users' uncertainties about transaction outcomes. Additionally, our findings reveal that familiarity and experience with automated technologies play a pivotal role in increasing trust.

The term cognition basically means the use of the brain but includes various complex activities. Within my work I focus on working memory, cognitive flexibility, and attentional processes. In the use of new technologies, working memory is crucial for grasping and remembering system features, navigating complex interfaces, and executing multitasking operations. Its role in problem-solving, decision-making, and managing cognitive load is essential for users to effectively engage with and adapt to evolving technological interfaces and functionalities. Cognitive flexibility enables rapid adaptation to new technologies and therefore improved performances, decreased error rate, and less mental workload. In terms of attentional processes, our findings highlight the greater impact of environmental complexity on attention during cognitive task performance compared to the difficulty of simultaneously executed motor tasks.

References

1 Vasile, L., Seitz, B., Staab, V., Liebherr, M., Däsch, C., & Schramm, D. (2023). Influences of Personal Driving Styles and Experienced System Characteristics on Driving Style Preferences in Automated Driving. Applied Sciences, 13(15), 8855.

2 Liebherr, M., Kohler, M., Zerr, M., Brand, M., & Antons, S. (2022). Effects of digital media use on attention subdomains in children aged 6 to 10 years. Children, 9(9), 1393.

3 Sauce, B., Liebherr, M., Judd, N., & Klingberg, T. (2022). The impact of digital media on children's intelligence while controlling for genetic differences in cognition and socioeconomic background. Scientific Reports, 12(1), 1-14.

4 Liebherr, M.*, Corcoran, A.W.*, Alday, P.M., Coussens, S., Bellan, V., Howlett, C.A., Immink, M.A., Kohler, M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2021). EEG and behavioral correlates of attentional processing while walking and navigating naturalistic environments. Scientific Reports, 11, 22325. *Both authors contributed equally to this work.

5 Liebherr, M., Müller, M. S., Schweig, S., Maas, N., Schramm, D., & Brand, M. (2021). Stress and simulated environments – Insights from physiological marker. Frontiers in Virtual Reality, 2(18), 1-10.

6 Liebherr, M., Schweig, S., Brandtner, A., Averbeck, H., Maas, N., Schramm, D., & Brand, M. (2020). When virtuality becomes real: Relevance of mental abilities and age in system adaptation and the occurrence of simulator sickness. Ergonomics, 63(10), 1271-1280.

7 Brandtner, A., Liebherr, M., Schweig, S., Maas, N., Schramm, D., & Brand, M. (2019). Subjectively estimated vs. objectively measured adaptation to driving simulators – Effects of age, driving experience, and previous simulator adaptation. Transportation Research Part F: Psychology and Behaviour, 64, 440-446.

3.16 Compositional Learning of Synchronous Systems

Thomas Neele (TU Eindhoven, NL)

Joint work of Thomas Neele, Matteo Sammartino
Main reference Thomas Neele, Matteo Sammartino: "Compositional Automata Learning of Synchronous Systems", in Proc. of the Fundamental Approaches to Software Engineering – 26th International Conference, FASE 2023, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2023, Paris, France, April 22-27, 2023, Proceedings, Lecture Notes in Computer Science, Vol. 13991, pp. 47–66, Springer, 2023.
URL https://doi.org//10.1007/978-3-031-30826-0_3

My main research in the field of automata learning is efficient learning of concurrent systems, in the shape of synchronously communicating automata [1]. The large amount of interleaving behaviour of these automata can cause significant scalability issues when applying traditional automata learning techniques in a monolithic fashion. Our approach is to instantiate one learner per automaton inside the composite SUL. These learners are able to independently learn significant parts of their respective automata, and only synchronise when proposing a hypothesis (which is the parallel composition of their individual hyptheses).

Furthermore, I have an extensive background in the areas of model checking, parity games, bisimulation and partial-order reduction (see e.g. [2, 3, 4]).

References

1 Thomas Neele and Matteo Sammartino. *Compositional Automata Learning of Synchronous Systems*. In *FASE 2023*, volume 13991, pages 47–66. 2023. doi:10.1007/978-3-031-30826-0_3.

2 Thomas Neele, Antti Valmari, Wieger Wesselink, and Tim A. C. Willemse. *Partial-Order Reduction for Parity Games and Parameterised Boolean Equation Systems. Software Tools for Technology Transfer*, 24:735–756, 2022. doi:10.1007/s10009-022-00672-0.

3 Thomas Neele, Antti Valmari, and Tim A. C. Willemse. *A Detailed Account of The Inconsistent Labelling Problem of Stutter-Preserving Partial-Order Reduction. Logical Methods in Computer Science*, 17:8:1–8:27, 2021. doi:10.46298/lmcs-17(3:8)2021.

4 Thomas Neele, Tim A. C. Willemse, and Jan Friso Groote. *Finding Compact Proofs for Infinite-Data Parameterised Boolean Equation Systems. Science of Computer Programming, FACS 2018 special issue*, 188:102389, 2020. doi:10.1016/j.scico.2019.102389.

3.17 Approximation, abstraction and apartness in automata learning

Jurriaan Rot (Radboud University Nijmegen, NL)

URL http://jurriaan.creativecode.org/approximation-abstraction-and-apartness-in-automata-learning-apple/

My research focuses on the analysis and specification of models of computation, with the use of algebraic and coalgebraic techniques as a central theme. In recent years I have become interested in automata learning, and have worked on categorical generalisations of learning, extensions to various types of models, and algorithmic aspects. I currently lead the NWO VIDI project "Approximation, Abstraction and Apartness in Automata Learning", where the overall aim is to enhance the scope and scalability of automata learning through the development of learning algorithms that feature approximation and abstraction.

3.18 Automata learning algorithms for concurrent, infinite-state, and hardware systems

Matteo Sammartino (Royal Holloway, University of London, GB)

Joint work of Thomas Neele, Joshua Moerman, Alexandra Silva, Bartek Klin, Michal Szynwelski, Loris D'Antoni, Thiago Ferreira, Amir Naseredini, Martin Berger, Shale Xiong
Main reference Thomas Neele, Matteo Sammartino: "Compositional Automata Learning of Synchronous Systems", in Proc. of the Fundamental Approaches to Software Engineering – 26th International Conference, FASE 2023, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2023, Paris, France, April 22-27, 2023, Proceedings, Lecture Notes in Computer Science, Vol. 13991, pp. 47–66, Springer, 2023.
URL https://doi.org//10.1007/978-3-031-30826-0_3

My research is at the intersection of formal methods and AI. I am interested in both theoretical aspects, such as algebraic/coalgebraic approaches, formals semantics, and automata models with specific expressivity and decidability properties; and practical aspects, such as the automated verification of those systems.

Recent work focuses on learning algorithms for systems with "real-world" features, namely: concurrent/distributed systems, infinite-state systems, and physical hardware systems:

- *Compositional automata learning:* traditional model learning algorithms (eg, L*) are monolithic, and do not scale well when learning composite systems. In [1] we provide a compositional version of L*, where each component of a given synchronous system is learned independently, and independent learners cooperate in the presence of synchronisations. Experiments show that our approach may require up to six orders of magnitude fewer membership queries and up to ten times fewer equivalence queries than L* (applied to the monolithic system). Ongoing work focuses on relaxing assumption on the target system, for instance knowledge of the number of components.
- *Learning infinite-state models:* model learning algorithms exist for a wide range of automata models. In [2] we have introduced an extension for *nominal* automata, which are automata over infinite alphabets admitting a rich theory. Unlike the finite-alphabet case, non-deterministic nominal automata are strictly more expressive than deterministic ones (language equivalence is undecidable!), hence learning algorithm cannot deal with the full class. In [3] we have characterised and investigated learning of *residual* nominal

automata, which retain the required decidability properties for model learning to work. Future work will focus on richer classes of infinite-state automata, for instance [4], nominal ω-regular acceptors, and automata equipped with global freshness.

- *Learning from physical hardware:* due to their complexity, hardware components such as CPUs and DRAMs are hard to model and reason about, and often models are "idealised versions" that do not reflect the actual behaviour. Model learning can help in deriving models that are closer to the implementation. In [5] we have developed *ALARM*, a tool that uses learning to reverse-engineer undisclosed security features of DRAMs (= Dynamic Random-Access Memory). This was a proof-of-concept implementation targeting a software-simulated DRAM. Ongoing work is about enabling learning from physical hardware via a an FPGA-based testing harness.

References

1 Neele, T. & Sammartino, M. Compositional Automata Learning of Synchronous Systems. *Fundamental Approaches To Software Engineering – 26th International Conference, FASE*. **13991** pp. 47-66 (2023)

2 Moerman, J., Sammartino, M., Silva, A., Klin, B. & Szynwelski, M. Learning nominal automata. *Proceedings Of The 44th ACM SIGPLAN Symposium On Principles Of Programming Languages, POPL*. pp. 613-625 (2017)

3 Moerman, J. & Sammartino, M. Residuality and Learning for Nondeterministic Nominal Automata. *Log. Methods Comput. Sci.*. **18** (2022)

4 D'Antoni, L., Ferreira, T., Sammartino, M. & Silva, A. Symbolic Register Automata. *Computer Aided Verification – 31st International Conference, CAV*. **11561** pp. 3-21 (2019)

5 Naseredini, A., Berger, M., Sammartino, M. & Xiong, S. ALARM: Active LeArning of Rowhammer Mitigations. *Proceedings Of The 11th International Workshop On Hardware And Architectural Support For Security And Privacy, HASP*. pp. 1-9 (2022)

3.19 Safe, reliable and trustworthy AI

Philipp Sieberg (Schotte Automotive GmbH & Co KG – Hattingen, DE)

My research focuses on creating reliable, safe and trustworthy artificial intelligence.

During my doctoral research, I focused on the use of AI in safety-critical areas, where I believe reliability is the necessary basis for acceptance and trust. Consequently, I developed innovative hybrid approaches that integrate knowledge-based techniques with artificial intelligence methods. Through the implementation of hybrid methods, the benefits of incorporating artificial intelligence – including improved performance and accuracy – can be realized while maintaining a reliable process that instills a proven level of trust. In my doctoral research, hybrid methods were applied to the domain of automated driving, allowing for reliable implementation of virtual sensors and centralized predictive control.

As a principal investigator at the Chair of Mechatronics of the University of Duisburg-Essen, I am conducting further research on the subject of trustworthy and reliable artificial intelligence. For instance, an interdisciplinary project concentrates on utilizing AI to build a toolbox for automatically identifying wear mechanisms. The research initiative is organized into two distinct phases. During the first phase, the necessary technical methods will be developed to achieve reliable and comprehensible results. In the second stage of the research

project, the outcomes will be integrated into a toolbox in order to make the specialized materials engineering knowledge available to a wide range of users in industry and society. Interaction with users is critical in this context. The topics of acceptance and trust in this toolbox and in artificial intelligence are very important. Besides this project, I am also actively involved in the technical development of reliable and trustworthy AI methods in the field of automated driving. The focal point is the integration of physical knowledge to promote transparency and ensure the safety of AI.

As the General Manager of Schotte Automotive GmbH & Co. KG, one of my main goals is to drive the digitalization of the company. I concentrate on supporting and improving operational processes through the implementation of AI systems. In cooperation with Prof. Enkel and Dr. Liebherr from the University of Duisburg-Essen, a research project has been initiated that focuses on improving disposition with the help of artificial intelligence. In addition to predicting order quantities and times, achieving acceptance and trust in an AI-based tool by employees in the disposition department poses significant challenges.

References

1 Sieberg P.M. and Hanke S. (2023) Challenges and Potentials in the Classification of Wear Mechanisms by Artificial Intelligence. Wear, Vol. 522, 204725. https://doi.org/10.1016/j.wear.2023.204725

2 Stockem Novo, A., Hürten, C., Baumann, R., & Sieberg, P. (2023). Self-evaluation of automated vehicles based on physics, state-of-the-art motion prediction and user experience. Scientific Reports, 13(1), 12692.

3 Sieberg, P.M. and Schramm, D. (2022) Ensuring the Reliability of Virtual Sensors Based on Artificial Intelligence within Vehicle Dynamics Control Systems. Sensors, Vol. 22, 3513. https://doi.org/10.3390/s22093513

4 Hürten C., Sieberg P.M. and Schramm D. (2022) Generating a Multi-fidelity Simulation Model Estimating the Models' Applicability with Machine Learning Algorithms. In Proceedings of the 12th International Conference on Simulation and Modeling Methodologies, Technologies and Applications 2022: SIMULTECH – Vol. 1, pp. 131-141. ISBN 978-989-758-578-4, https://doi.org/10.5220/0011318100003274

5 Sieberg P.M., Blume S., Reicherts S., Maas N. and Schramm D. (2021) Hybrid State Estimation – A Contribution Towards Reliability Enhancement of Artificial Neural Network Estimators. In IEEE Transactions on Intelligent Transportation Systems, Vol. 23, pp. 6337-6346. https://doi.org/10.1109/TITS.2021.3055800

6 Sieberg P.M., Hürten C. and Schramm D. (2020) Representation of an Integrated Non-Linear Model-Based Predictive Vehicle Dynamics Control System by a Co-Active Neuro-Fuzzy Inference System. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV) 2020, pp. 572-577, Las Vegas, USA – ISBN 978-1-7281-6673-5

4 Open problems

4.1 Sound Control Synthesis with Logics and Data

Alessandro Abate (University of Oxford, GB)

We are witnessing an inter-disciplinary convergence between scientific areas underpinned by model-based reasoning and by data-driven learning. Original technical work across these areas is justified by numerous applications, where access to information-rich data has to be traded off with a demand for safety criticality: cyber-physical systems are exemplar applications.

Within OXCAV, the Oxford Control and Verification group, I focus on control synthesis for complex objectives, and study how techniques from formal verification (logics and SAT, automata theory, abstractions) and from learning (sample-driven approaches and neural architectures) can be together leveraged to attain both sound and effective synthesis outcomes.

More broadly, I argue that, on the one hand, control theory and formal methods can provide certificates to learning algorithms and, on the other hand, that learning can bolster formal verification and strategy synthesis objectives.

4.2 Automated generation of efficient and systematic test suites

Robert M. Hierons (University of Sheffield, GB)

My research interests largely concern the automated generation of efficient, systematic test suites on the basis of models or specifications, although I have done some work on generating test cases based on the code. Within this, I have had a particular interest in testing based on state-based models, typically expressed either as a form of state-machine or using a process algebra such as CSP. Within testing, I have looked at how the nature of the interaction between the tester(s) and the system under test affect any notion of correctness (conformance) used, including work on distributed testing and asynchronous testing. In recent years, I have also become interested in the use of concepts regarding causality to support testing. Finally, I have longstanding interest in the relationship between testing and learning, with this including recent work on the use of reinforcement learning.

4.3 Testing causal properties, and reasoning about uncertainty

Neil Walkinshaw (University of Sheffield, GB)

My research focusses on model inference and statistical reasoning for the purpose of testing. My current research interests focus on two primary areas: (1) Testing for causal relationships between input and output variables, and (2) reasoning about uncertainty in reverse-engineered models.

Testing causality. I am the PI on the CITCOM project, which is concerned with the development of techniques to test for the presence or absence of causal relationships in complex systems. For this we have adapted and applied a family of statistical reasoning techniques called Causal Inference. Here, the developer provides a model (in the form of a DAG) setting out the causal relationships between the inputs and outputs they are interested in, and our testing technique can test for them in data.

We have shown how this form of testing can be framed as Metamorphic Testing. However, it also has the added benefit that we are able to test for causal relationships purely from observed data, without needing to control the inputs.

Reasoning about uncertainty. One strand of this research has focussed on the inference of models that explicitly incorporate second-order uncertainty. Second order uncertainty pertains our confidence (or lack thereof) in a particular model element or value. This is particularly important when dealing with inferred models, where some aspects of the inferred model will invariably be better supported by data than others.

Much of my research has used Subjective Logic – a way of reasoning about probabilities whilst explicitly incorporating second-order uncertainty. Most recently, we have developed a generalisation of probabilistic finite state machines, called Subjective Opinion State Machines. For this, we have also developed an inference approach that can be "layered over" any conventional state machine inference algorithm to produce these state machines with explicit second-order uncertainty.

References

1 `https://github.com/CITCOM-project/CausalTestingFramework`
2 Andrew G. Clark, Michael Foster, Benedikt Prifling, Neil Walkinshaw, Robert M. Hierons, Volker Schmidt, and Robert D. Turner. 2023. Testing Causality in Scientific Modelling Software. ACM Trans. Softw. Eng. Methodol. 33, 1, Article 10 (January 2024).Reference:
3 Walkinshaw, Neil, and Robert M. Hierons. "Modelling Second-Order Uncertainty in State Machines." IEEE Transactions on Software Engineering (2023).

Participants

- Alessandro Abate
University of Oxford, GB
- Wolfgang Ahrendt
Chalmers University of Technology – Göteborg, SE
- Bernhard K. Aichernig
TU Graz, AT
- Thomas Arts
QuviQ AB – Gothenburg, SE
- Thorsten Berger
Ruhr-Universität Bochum, DE
- Jyotirmoy Deshmukh
USC – Los Angeles, US
- Monique Dittrich
CARIAD – Berlin, DE
- Ellen Enkel
Universität Duisburg-Essen, DE
- Sophie Fortz
University of Namur, BE
- Fatemeh Ghassemi Esfahani
University of Tehran, IR
- Heiko Hecht
Universität Mainz, DE
- Leo Henry
University College London, GB
- Robert M. Hierons
University of Sheffield, GB
- Falk Howar
TU Dortmund, DE
- Nils Jansen
Ruhr-Universität Bochum, DE
- Effie Lai-Chong Law
Durham University, GB
- Magnus Liebherr
Universität Duisburg-Essen, DE
- Mohammad Reza Mousavi
King's College London, GB
- Thomas Neele
TU Eindhoven, NL
- Nicola Paoletti
King's College London, GB
- Jurriaan Rot
Radboud University Nijmegen, NL
- Kristin Yvonne Rozier
Iowa State University – Ames, US
- Matteo Sammartino
Royal Holloway, University of London, GB
- Philipp Sieberg
Schotte Automotive GmbH & Co KG – Hattingen, DE
- Bernhard Steffen
TU Dortmund, DE
- Marnix Suilen
Radboud University Nijmegen, NL
- Neil Walkinshaw
University of Sheffield, GB