

Fusing Causality, Reasoning, and Learning for Fault Management and Diagnosis

Alessandro Cimatti*¹, Ingo Pill*², and Alexander Diedrich*³

1 Bruno Kessler Foundation – Trento, IT. cimatti@fbk.eu

2 Silicon Austria Labs – Graz, AT. ingo.pill@gmail.com

3 Helmut-Schmidt-Universität – Hamburg, DE. diedrica@hsu-hh.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar “Fusing Causality, Reasoning, and Learning for Fault Management and Diagnosis” (24031). The goal of this Dagstuhl Seminar was to provide an interdisciplinary forum to discuss the fundamental principles of fault management and diagnosis, bringing together international researchers and practitioners from the fields of symbolic reasoning, machine learning, and control engineering.

Seminar January 14–19, 2024 – <https://www.dagstuhl.de/24031>

2012 ACM Subject Classification Computing methodologies → Artificial intelligence; Theory of computation → Semantics and reasoning; Theory of computation → Theory and algorithms for application domains

Keywords and phrases cyber-physical systems, diagnosis, fault detection and management, integrative ai, model-based reasoning

Digital Object Identifier 10.4230/DagRep.14.1.25

1 Executive Summary

Alexander Diedrich (Helmut-Schmidt-Universität – Hamburg, DE)

Alessandro Cimatti (Bruno Kessler Foundation – Trento, IT)

Ingo Pill (Silicon Austria Labs – Graz, AT)

License © Creative Commons BY 4.0 International license
© Alexander Diedrich, Alessandro Cimatti, and Ingo Pill

Our goal for this Dagstuhl Seminar was to find approaches that leverage fault diagnosis to build resilient cyber-physical systems through combinations of symbolic, sub-symbolic, and control theoretic approaches.

Cyber-Physical Systems (CPSs), i.e. systems in which mechanical and electrical parts are controlled by computational algorithms, are not only continuously increasing in size and complexity, but they are also required to operate in evolving and uncertain environments, subject to frequent changes and faults. Detecting and correcting faulty behavior is a highly complex task that needs the help of computational algorithms. The constant advances in sensing technology and computational power, as well as the increase in data recording options, enables and also requires us to rely more and more on methods from Artificial Intelligence (AI) for these tasks, i.e. symbolic AI such as planning and reasoning engines, as well as subsymbolic AI like Machine Learning (ML). Sub-symbolic approaches are primarily used to detect symptoms; symbolic reasoning on the other hand provides diagnosis algorithms to identify root causes (from symptoms or observations) or reason about repairs. Furthermore,

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Fusing Causality, Reasoning, and Learning for Fault Management and Diagnosis, *Dagstuhl Reports*, Vol. 14, Issue 1, pp. 25–48

Editors: Alessandro Cimatti, Ingo Pill, and Alexander Diedrich



DAGSTUHL Dagstuhl Reports

REPORTS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

control engineering methods guide the system back to normal operation (based on the identified root cause). Since these methods come from different fields, they do not always work together in practice.

The research challenge at hand is to combine symbolic a-priori knowledge and learned data, as well as to develop an integrated concept taking both symbolic and sub-symbolic approaches into account. The leading research questions of this seminar are summarised as follows:

- How can a-priori knowledge be combined with data-centric, machine learning-based algorithms?
- Can we integrate a-priori knowledge such as background knowledge about functions, interfaces and operation modes into ML-algorithms to improve model performance?
- Can we use data to learn parts of the symbolic models?
- And can we develop new algorithms which are a synthesis of both worlds, symbolic and subsymbolic?

All of these research questions must be addressed to practical and resilient cyber-physical systems. To tackle these questions, we invited researchers from symbolic AI, sub-symbolic AI, and control engineering to develop a common notion of fault detection and fault handling tasks that takes also the practical needs from industry-scale problems into account. In this regard the seminar also had a secondary function: Traditional symbolic AI diagnosis is located within the Diagnostics community (DX), while sub-symbolic fault diagnosis was traditionally associated with the fault-detection and isolation (FDI) community within the control theory research field. More recently, also the research field of machine learning has created advances with regard to fault diagnosis. Since this seminar brought together researchers from all of these fields, we hope that the seminar created fertile ground for some cross-domain research initiatives.

Besides the individual contributions to the seminar, we used four breakout sessions to brainstorm ideas and next steps following from this seminar:

1) Breakout Session on Coupling Symbolic and Sub-symbolic Methods for Model Acquisition: Fusing symbolic methods with sub-symbolic methods in both directions is essential for the creation of resilient systems. The research gap that has been identified is that so far most approaches integrate some symbolic knowledge into the majority of sub-symbolic knowledge, or a small part of sub-symbolic knowledge into a large symbolic knowledge base. But both of these directions have drawbacks and do not automatically lead to models that are well-suited for resilient systems that can be used in practice. One takeaway is the idea to organise a competition that incentivises researchers to develop novel modelling formalisms and diagnosis algorithms that mitigate some of the current drawbacks.

2) Breakout Session on Causality – How to Generate Knowledge from Data: The breakout session detailed the importance of high-level causal models in capturing causal relationships within systems. It was discussed where the difficulties in manually crafting these models lie due to their complexity and the even greater challenge of learning causal models directly from data. Crafting causal models manually, one needs a deep understanding of the dependencies. For learning causal models, a large amount of data even for situations which barely occur is needed.

3) Breakout Session on LLMs for DX – Integrating Large Language Models for Root Cause Diagnosis: The breakout session featured a comprehensive exploration and discussion on the potential and challenges of using Large Language Models in the topics of the “DX” community. The central aspects that were discussed, revolved around (i) the models themselves and their current and potential future capabilities, (ii) the training data

for training and refining LLMs for diagnosis tasks, (iii) potential application areas, as well as (iv) current, and (v) future trends and topics that should be monitored or covered by the DX community. As a result, the attendees agreed on writing a position paper, which will capture the current potential and drawbacks of LLMs within DX domains.

4) Breakout Session on Resilient Systems: For resilient systems we saw that the application and scenario play a significant role when aiming to assess what would be “good” and “bad” behavior for some system. The same goes for the question of whether we would assess the performance of a system in a local or a global context. To this end we identified a set of such relevant scenarios ranking from an energy management scenario at a local home, via the operation of an electric grid, via agents/robots in a collaborative disaster or military scenario, to supply chain management. We also discussed and converged to a definition of resilience that would tailor to all the expressed needs.

2 Table of Contents

Executive Summary

<i>Alexander Diedrich, Alessandro Cimatti, and Ingo Pill</i>	25
--	----

Overview of Talks

Keynote on Reinforcement Learning for Control of Cyber Physical Systems <i>Gautam Biswas</i>	30
Diagnosability of Fair Transition Systems <i>Marco Bozzano</i>	31
Tree-based diagnosis enhanced with meta knowledge <i>Elodie Chanthery and Louise Travé-Massuyès</i>	32
Tutorial on Runtime verification and monitor synthesis <i>Alessandro Cimatti</i>	32
AI for predictive maintenance: domain adaptation, MLOps, and Edge computing. A case study <i>Marco Cristoforetti</i>	32
Keynote on Analogy for Diagnosis by and within Cognitive Architectures <i>Kenneth D. Forbus</i>	33
Keynote on Understanding Resilience <i>Johan de Kleer</i>	34
Data-driven diagnosis from an FDI practitioner’s perspective <i>Daniel Jung</i>	34
The Rayleigh-Ritz Autoencoder architecture for Machine Learning with hard Physical Constraints <i>Manfred Mücke</i>	35
Learning what to monitor: pairing monitoring and learning <i>Angelo Montanari</i>	35
Tutorial on Diagnosing Cyber-Physical Systems <i>Oliver Niggemann</i>	36
Tutorial on Basics of Model-Based Diagnosis <i>Ingo Pill</i>	36
Designing Fault-Tolerant Control Systems using Topological Systems Theory <i>Gregory Provan</i>	37
Root Cause Analysis via Anomaly Detection and Causal Graphs <i>Josephine Rehak</i>	38
Hybrid Model Learning for System Health Monitoring <i>Pauline Ribot, and Elodie Chanthery</i>	38
Tutorial on Bridge DX / FDI <i>Louise Travé-Massuyès</i>	39
Fault Detection, Diagnosis, and Mitigation for Space Propulsion Systems <i>Günther Waxenegger-Wilfing, Kai Dresia, and Ingo Pill</i>	39

Quality Assurance Methodologies for Resilient (Model-based) Systems <i>Franz Wotawa</i>	40
Working groups	
Breakout Session on coupling symbolic and sub-symbolic methods for model acquisition <i>Rene Heesch</i>	40
Breakout Session on coupling symbolic and sub-symbolic methods in both directions <i>Rene Heesch</i>	41
Breakout Session on Causality – How to generate knowledge from data? <i>Lukas Moddemann and Kaja Balzereit</i>	42
Breakout Session on LLMs for DX – Integrating Large Language Models for Root Cause Diagnosis <i>Lukas Moddemann and Jonas Ehrhardt</i>	43
Breakout Sessions on Resilience <i>Ingo Pill</i>	44
Panel discussions	
Panel on Current and Future Challenges in Resilient System Design <i>Ingo Pill</i>	44
Open problems	
LiU-ICE Industrial Fault Diagnosis Benchmark – Anomaly Detection and Fault Isolation with Incomplete Data <i>Daniel Jung</i>	47
Participants	48

3 Overview of Talks

3.1 Keynote on Reinforcement Learning for Control of Cyber Physical Systems

Gautam Biswas (Vanderbilt University – Nashville, US)

License © Creative Commons BY 4.0 International license
© Gautam Biswas

Joint work of Gautam Biswas, Marcos Quinones-Grueiro, Austion Coursey, Avisek Naug

Main reference Avisek Naug, Marcos Quinones-Grueiro, Gautam Biswas: “Deep reinforcement learning control for non-stationary building energy management”, *Energy and Buildings*, Vol. 277, p. 112584, 2022.

URL <https://doi.org/10.1016/j.enbuild.2022.112584>

The resilience of complex cyber-physical systems (CPS) or systems of systems that combine cyber and physical components and often include humans in the loop is critical for the safe and cost-effective operations of these systems. Moreover, these systems often operate in environments whose parameters are often not known in advance, therefore, these systems have to be robust to disturbances and changes that occur in their operating environment. In other words, these systems often operate in non-stationary environments, making traditional control methods less effective for operating these systems safely and reliably.

In my presentation, I discussed the use of Reinforcement Learning (RL) methods to design controllers for complex CPS that operate in non-stationary environments. In the first third of this talk, I presented a quick introduction to RL, covering basic topics, such as Markov Decision Processes (MDPs), reward signals, value and policy functions, and the Bellman optimality criterion. I will also cover very briefly the basic RL methods of value and policy iteration, Monte Carlo and TD-learning methods, Q-learning, and Policy Gradient approaches.

In the rest of the presentation, I detailed two studies we have conducted in developing RL controllers for real-world non-stationary problems. The first is Building energy management, where we used RL to develop a supervisory controller that has been deployed on a real building for the heating, ventilation, and air conditioning (HVAC) system. Given the non-stationarities in the operating environment (e.g., sudden weather changes), we monitored for performance degradation by tracking an aggregate metric that was derived from the overall accumulated reward. Degradation in performance triggered a relearning loop. Then, a set of data-driven models of the building behavior was updated with the latest data on the building operations. Subsequently, we returned the deployed controller by letting it interact with the model and was then redeployed on the system. The approach has resulted in significant energy savings for the deployed building.

As a second case study, we developed a hybrid control framework that combines a well-established cascade control architecture and data-driven methods to accommodate varying wind conditions and payloads for unmanned aerial vehicles (UAVs). We reframed the role of the data-driven methods to compensate for the limited adaptability of the traditional control approaches by dynamically modifying the reference velocities to account for disturbances that manifest as adverse wind and payload changes. We demonstrated the advantage of the proposed framework using a Tarot T18 octocopter simulation (validated with real data) under aggressive wind field changes and payload changes mid-flight. We also showed that our learned disturbance rejection controller generalized to a different octocopter, the DJI-S1000.

The talk concluded, by discussing future work in developing continual and safe RL schemes to further enhance RL-based control and make it applicable to real-world control problems.

3.2 Diagnosability of Fair Transition Systems

Marco Bozzano (*Bruno Kessler Foundation – Trento, IT*)

License © Creative Commons BY 4.0 International license
© Marco Bozzano

Joint work of Benjamin Bittner, Marco Bozzano, Alessandro Cimatti, Marco Gario, Stefano Tonetta, Viktória Vozárová

Main reference Benjamin Bittner, Marco Bozzano, Alessandro Cimatti, Marco Gario, Stefano Tonetta, Viktória Vozárová: “Diagnosability of fair transition systems”, *Artif. Intell.*, Vol. 309, p. 103725, 2022.

URL <https://doi.org/10.1016/J.ARTINT.2022.103725>

The integrity of complex dynamic systems often relies on the ability to detect, during operation, the occurrence of faults, or, in other words, to diagnose the system. The feasibility of this task, also known as diagnosability, depends on the nature of the system dynamics, the impact of faults, and the availability of a suitable set of sensors. Standard techniques for analyzing the diagnosability problem rely on a model of the system and on proving the absence of a faulty trace that cannot be distinguished by a non-faulty one (this pair of traces is called critical pair).

In this talk, we tackled the problem of verifying diagnosability under the presence of fairness conditions. These extend the expressiveness of the system models enabling the specification of assumptions on the system behavior such as the infinite occurrence of observations and/or faults.

We adopt a comprehensive framework that encompasses fair transition systems, temporally extended fault models, delays between the occurrence of a fault and its detection, and rich operational contexts. We show that in presence of fairness the definition of diagnosability has several interesting variants, and discuss the relative strengths and the mutual relationships. We proved that the existence of critical pairs is not always sufficient to analyze diagnosability, and needs to be generalized to critical sets. We defined new notions of critical pairs, called ribbon-shape, with special looping conditions to represent the critical sets.

Based on these findings, we provide algorithms to prove the diagnosability under fairness. The approach is built on top of the classical twin plant construction, and generalizes it to cover the various forms of diagnosability and find sufficient delays.

The proposed algorithms are implemented within the xSAP platform for safety analysis, leveraging efficient symbolic model checking primitives. An experimental evaluation on a heterogeneous set of realistic benchmarks from various application domains demonstrates the effectiveness of the approach.

References

- 1 B. Bittner, M. Bozzano, A. Cimatti, M. Gario, S. Tonetta, V. Vozarova, Diagnosability of fair transition systems, *Artificial Intelligence* 309.

3.3 Tree-based diagnosis enhanced with meta knowledge

Elodie Chanthery (LAAS – Toulouse, FR) and Louise Travé-Massuyès (LAAS – Toulouse, FR)

License © Creative Commons BY 4.0 International license
© Elodie Chanthery and Louise Travé-Massuyès

Joint work of Louis Goupil, Elodie Chanthery, Louise Travé-Massuyès, Sébastien Delautier

Main reference Louis Goupil, Elodie Chanthery, Louise Travé-Massuyès, Sébastien Delautier: “Tree based diagnosis enhanced with meta knowledge”, in Proc. of the 34th International Workshop on Principles of Diagnosis (DX’23), 2023.

URL <https://hal.science/hal-04186400>

This talk presents an online data and knowledge based diagnosis method. It leverages decision trees in which decisions are made based on diagnosis meta knowledge, namely knowledge about the properties of diagnosis indicators. This knowledge is used at the level of each node to set a symbolic classification problem that brings out discriminating functions. This results in a multivariate decision tree that produces a compact model for diagnosis. The use of decision trees increases the explicability of the results found, all the more so as one discovers the explicit formal expressions of diagnosis indicators in the process. The method has been tested on static systems. On the well-known polybox, the three diagnosis indicators known as analytical redundancy relations, that are generally computed from the model, are found.

3.4 Tutorial on Runtime verification and monitor synthesis

Alessandro Cimatti (Bruno Kessler Foundation – Trento, IT)

License © Creative Commons BY 4.0 International license
© Alessandro Cimatti

Runtime Verification (RV) is a lightweight verification technique that aims at checking whether a run of a system under scrutiny (SUS) satisfies or violates a given correctness specification. The tutorial first gave an overview about the general framework of RV, and the techniques to synthesize run-time monitors that can be efficiently executed in combination with the SUS. Then, we will cover the relationship between RV and the field of Fault Detection and Isolation (FDI). In FDI, runtime monitors are built taking into account models of the SUS, in order to monitor the occurrence of internal (faulty) conditions that are not directly observable.

3.5 AI for predictive maintenance: domain adaptation, MLOps, and Edge computing. A case study

Marco Cristoforetti (Bruno Kessler Foundation – Trento, IT)

License © Creative Commons BY 4.0 International license
© Marco Cristoforetti

Joint work of Andrea Gobbi, Mario Pujatti, Diego Calzà, Piergiorgio Svaizer, Marco Cristoforetti

In recent years, data-driven artificial intelligence (AI) has acquired relevance in diagnostics. Traditional methodologies are being complemented and, in some cases, supplanted by AI-powered solutions, suggesting a possible paradigm shift. AI algorithms have demonstrated remarkable capabilities in analyzing vast amounts of data with speed and precision, enabling early detection of anomalies and predictive insights into potential faults.

The necessity of monitoring the condition of devices with diverse characteristics, each of which may exhibit unique features, behaviors, and failure modes, makes it challenging to develop a universally applicable monitoring solution based on AI that usually necessitates extensive training data. This is even more true when considering classical predictive maintenance tasks such as Remaining Useful Life (RUL) estimation, with the typical scarcity of data covering the life of the monitored system until failure. Consequently, a critical need arises for methodologies that enable these algorithms to effectively perform on unseen cases, ensuring their reliability and accuracy in practical scenarios.

Domain adaptation techniques try to solve this problem by bridging the gap between the source domain (where the model is trained) and the target domain (where it is deployed), allowing for effective knowledge transfer and adaptation to specific contexts.

This contribution presents a comprehensive, modular, and scalable solution for data-driven diagnosis and prognosis that integrates deep learning algorithms and adversarial domain adaptation to permit transfer learning and increase generalization. The pipeline starts with the data acquisition and preprocessing steps, with a configurable feature extraction phase that produces a compressed latent representation of the input samples, computed using feature extraction techniques from multiple channels of raw signals. Additional features are calculated from the latent space of deep autoencoders. This latent space is deliberately composed of only a few variables, enforcing the compression of the information contained in the input. Domain adaptation based on adversarial learning uses the features extracted from all the samples available for the source and only the first few samples from the target system. This is to align the Deep Learning regressor responsible for estimating the health index of the target necessary to compute the RUL.

In our solution, the setup includes a communication system via MQTT, enabling an online data stream for real-time monitoring and maintenance. The overall infrastructure was deployed in an industrial setting and tested in a real-time experiment, demonstrating the validity of the proposed approach.

3.6 Keynote on Analogy for Diagnosis by and within Cognitive Architectures

Kenneth D. Forbus (Northwestern University – Evanston, US)

License © Creative Commons BY 4.0 International license
© Kenneth D. Forbus

This talk described two big ideas:

1. Analogy plays key roles in human diagnosis It provides reasoning from experience and detection of novel situations Analogical generalization constructs probabilistic relational schema Provides a source of priors for model-based diagnosis Analogical learning is incremental, inspectable, and data/training efficient
2. Cognitive architectures need diagnosis
Goal: Software social organisms instead of tools Systems need to manage their own learning Achieving agency needs internal diagnostic capabilities How to build software that never blue-screens?

3.7 Keynote on Understanding Resilience

Johan de Kleer (c-infinity – Mountain View, US)

License © Creative Commons BY 4.0 International license
© Johan de Kleer

Joint work of Johan de Kleer, Alex Feldman, Ion Matei, Saigopal Nelaturi, Morad Behandesh, Ingo Pill, Jan VanderBrande, Shiwali Mohan, Wiktor Piotrowski, Sachin Grover, Sookyung Kim, Jacob Le, Roni Stern

Main reference Ion Matei, Wiktor Piotrowski, Alexandre Perez, Johan de Kleer, Jorge Tierno, Wendy Mungovan, Vance Turnewitsch: “System Resilience through Health Monitoring and Reconfiguration”, ACM Trans. Cyber Phys. Syst., Vol. 8(1), pp. 7:1–7:27, 2024.

URL <https://doi.org/10.1145/3631612>

The real world is made up of cyber-physical systems – we want them not to fail, to be invisible. How can we improve the resilience of our CPSs? Recently, a space craft designer said to me “all failures are failures of imagination.” By that he meant that it’s the designers’ responsibility to imagine all the things that could possibly go wrong with the space craft and design around or compensate for them. With the advances in AI including planning and ML we can now put AIs inside of our systems which can address the designer’s unknown unknowns. To make significant advances we need to define what resilience is and how to measure it. I will describe a number of definitions of resilience. In this talk I will describe a comprehensive approach to achieving certain types of resilience with examples ranging from printers to unmanned ships.

3.8 Data-driven diagnosis from an FDI practitioner’s perspective

Daniel Jung (Linköping University, SE)

License © Creative Commons BY 4.0 International license
© Daniel Jung

Joint work of Daniel Jung, Arman Mohammadi, Matthias Krysander

Main reference Daniel Jung: “Automated Design of Grey-Box Recurrent Neural Networks For Fault Diagnosis using Structural Models and Causal Information”, in Proc. of the Learning for Dynamics and Control Conference, L4DC 2022, 23-24 June 2022, Stanford University, Stanford, CA, USA, Proceedings of Machine Learning Research, Vol. 168, pp. 8–20, PMLR, 2022.

URL <https://proceedings.mlr.press/v168/jung22a.html>

A diagnosis system can be described as a function that uses observations from the monitored system to compute diagnoses. Because of its industrial and scientific relevance, the fault diagnosis problem has been approached in many different communities. A popular approach is data-driven fault diagnosis which refers to methods that use historical data from different fault scenarios to learn the relation between observations and diagnoses. Compared to model-based diagnosis, which uses physically based models that have a long theoretical foundation, data-driven fault diagnosis is often treated as a general classification problem. This presentation has looked at data-driven fault diagnosis from a model-based diagnosis perspective. It is shown that central model-based concepts, like redundancy, can be interpreted in a data-driven framework and it is also illustrated how these ideas can be used to develop new data-driven fault diagnosis methods.

3.9 The Rayleigh-Ritz Autoencoder architecture for Machine Learning with hard Physical Constraints

Manfred Mücke (Material Center Leoben, AT)

License © Creative Commons BY 4.0 International license
© Manfred Mücke

Joint work of Anika Terbuch, Paul O’Leary, Dimitar Ninevski, Elias Hagendorfer, Elke Schlager, Andreas Windisch, Christoph Schweimer, Matthew Harker, Manfred Mücke

Main reference Anika Terbuch, Paul O’Leary, Dimitar Ninevski, Elias Jan Hagendorfer, Elke Schlager, Andreas Windisch, Christoph Schweimer: “A Rayleigh-Ritz Autoencoder”, in Proc. of the IEEE International Instrumentation and Measurement Technology Conference, I2MTC 2023, Kuala Lumpur, Malaysia, May 22-25, 2023, pp. 1–6, IEEE, 2023.

URL <https://doi.org/10.1109/I2MTC53148.2023.10176014>

I present the Rayleigh-Ritz Autoencoder (RRAE) architecture [1] for unsupervised hybrid machine learning. It is suitable for applications where the system being observed by multiple sensors is well modeled as a boundary value problem. The embedding of the admissible functions in the decoder implements a truly physics-informed machine learning architecture. The RRAE provides an exact fulfillment of Neumann, Cauchy, Dirichlet or periodic constraints. Only the encoder needs to be trained; the RRAE is numerically more efficient during training than traditional autoencoders.

We extended the RRAE architecture [2] to distribution-free statistics to achieve stability with respect to non-Gaussian data. This provides consistent results for sensor data with both Gaussian and non-Gaussian perturbations. The necessity for handling non-Gaussian data in sensor applications is documented by the behavior of inclinometer sensors where the perturbations are characterized by Cauchy-Lorentz distribution. In such cases variance does not provide a reliable measure for uncertainty; consequently, 1-norm error measures are investigated thoroughly.

References

- 1 Anika Terbuch, Paul O’Leary, Dimitar Ninevski, Elias Hagendorfer, Elke Schlager, Andreas Windisch, and Christoph Schweimer, *A Rayleigh-Ritz Autoencoder* in 2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2023.
- 2 Anika Terbuch, Dimitar Ninevski, Paul O’Leary, Matthew Harker and Manfred Mücke. *Extended Rayleigh-Ritz Autoencoder with Distribution-Free Statistics* in 2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2024.

3.10 Learning what to monitor: pairing monitoring and learning

Angelo Montanari (University of Udine, IT)

License © Creative Commons BY 4.0 International license
© Angelo Montanari


Joint work of Angelo Montanari, Andrea Brunello, Dario Della Monica, Luca Geatti, Nicola Saccomanno

Monitoring is a runtime verification technique that can be used to check whether an execution of a system (trace) satisfies or not a given set of properties. Compared to other formal verification techniques, e.g., model checking, one needs to specify the properties to be monitored, but a complete model of the system is no longer necessary. In the talk (uploaded slides), we first introduce the notion of monitoring and display a simple architecture of a monitoring system. Then, we provide a characterisation of positively and negatively monitorable properties, and we define the safety and cosafety fragments of Linear Temporal Logic (LTL). We complete the picture by showing that monitorability goes behind safety and

cosafety LTL fragments, and that there are natural properties which are not monitorable. Next, we proceed by pointing out that monitoring suffers from some significant limitations. In particular, modern systems have such a level of complexity that it is impossible for a system engineer to specify in advance all properties to be monitored, and even minor changes to the system to be monitored can introduce unforeseen bugs. To overcome these limitations, we provide a multi-objective genetic programming algorithm to automatically extend the set of properties to monitor on the basis of the history of failure traces collected over time. The monitor and the learning algorithm are then integrated in a unifying framework, whose distinguishing features are (i) interpretability (the machine learning methods manipulate and produce only formulae, that can be easily inspected by a system engineer), (ii) formal guarantees on monitorability (every formula produced during the learning phase is guaranteed to be monitorable), and (iii) generality (different monitoring and machine learning backends). The framework has been experimentally validated on various public datasets, and the outcomes of the experimentation confirm the effectiveness of the proposed solution.

3.11 Tutorial on Diagnosing Cyber-Physical Systems

Oliver Niggemann (Helmut-Schmidt-Universität – Hamburg, DE)

License  Creative Commons BY 4.0 International license
© Oliver Niggemann

The transition from sub-symbolic representation in artificial intelligence such as time series to symbolic representations such as expressions in formal logic or in language is crucial to creating resilient cyber physical systems. The first step for this is often the discretization, i.e. the identification of symbolic concepts that come true at certain points in time. The tutorial presents typical discretization algorithms for time series, especially with a focus on engineering and scientific applications. In the last step, a brief overview of possibilities to further develop the identified concepts into causalities is given.

3.12 Tutorial on Basics of Model-Based Diagnosis

Ingo Pill (Silicon Austria Labs – Graz, AT)

License  Creative Commons BY 4.0 International license
© Ingo Pill

For Seminar 24031, we aimed to invite a diverse audience from a variety of fields connected to the seminar's focus. For us organizers it was thus important to provide the attendees with some basic knowledge, common grounds concerning well-established techniques in the field, and also a basic context and terminology. In this introductory talk, I focused on providing some brief basics about model-based diagnosis (MBD), also known as consistency-based diagnosis or DX approach. Due to its attractive features, MBD is such a central technology in the scope of this seminar, and I covered the following aspects in my presentation:

- the basic underlying approach of reasoning from first principles
- the standard scenario of explaining some unexpected behavior like a failed test case
- connections to verification tasks and techniques
- the very basic definitions of diagnoses, conflicts and minimal hitting sets
- the impact of using a weak fault model or strong fault models

- two basic algorithmic concepts for MBD: conflict-driven and direct
- MBD's flexibility in terms of application and deployed algorithm
- diagnosing multiple scenarios at the same time (e.g. like results from a test suite) and the resulting opportunity to characterize a system (when using a representative test suite, e.g., obtained with combinatorial testing)
- diagnosing static scenarios and sequential behavior
- improving the basic algorithmic concepts via algorithmic optimizations (example RC-Tree) and structural, diagnosis problem-specific information (like exploiting its parse tree when diagnosing some LTL description)
- completeness and soundness of MBD in relation to the model and the scenario(s)
- information about which authors of the covered papers participated in the seminar.

3.13 Designing Fault-Tolerant Control Systems using Topological Systems Theory

Gregory Provan (University College Cork, IE)

License © Creative Commons BY 4.0 International license
© Gregory Provan

Joint work of Gregory M. Provan, Marcos Quiñones-Grueiro, Yves Sohege

Main reference Gregory M. Provan, Marcos Quiñones-Grueiro, Yves Sohege: “Generating Minimal Controller Sets for Mixing MMAC”, in Proc. of the 61st IEEE Conference on Decision and Control, CDC 2022, Cancun, Mexico, December 6-9, 2022, pp. 3009–3014, IEEE, 2022.

URL <https://doi.org/10.1109/CDC51059.2022.9993251>

Given information about (a) the desired operating conditions for a system and (b) the tasks the system must carry out, the fault-tolerant control design (FD) task is to design a set of controllers to ensure that conditions (a) and (b) are guaranteed, even when faults and/or external disturbances occur.

Designing control systems from requirements and model specifications is a challenging task, and has been addressed from many perspectives, most notably design optimization and multi-controller tuning. Our approach extends both design optimization and multi-controller tuning. Multi-controller tuning adopts a “divide and conquer” approach, decomposing the system’s operating range into smaller local sub-spaces, each associated with a “local” model. These local models are then combined to create the global system response. The primary advantage of the (MM) approach lies in its simplification of complex modeling through the use of these local models.

We develop an optimisation-based approach to designing systems with fault-tolerance and resilience capabilities, i.e., it enables multi-mode operation. Standard design optimisation approaches assume a single nominal mode; in contrast, we explicitly define an approach that generalises this framework to enable the design of systems that operate in multiple modes. Multi-controller tuning methods typically use methods to compute the set of possible modes, and then tune a controller for each mode. In contrast to multi-controller tuning methods, this new approach does not optimize just performance of individual controllers, but task-centric performance, based on topological transformations from plant spaces to control spaces.

3.14 Root Cause Analysis via Anomaly Detection and Causal Graphs

Josephine Rehak (KIT – Karlsruher Institut für Technologie, DE)

License © Creative Commons BY 4.0 International license
© Josephine Rehak

Main reference Josephine Rehak, Anouk Sommer, Maximilian Becker, Julius Pfrommer, Jürgen Beyerer: “Counterfactual Root Cause Analysis via Anomaly Detection and Causal Graphs”, in Proc. of the 21st IEEE International Conference on Industrial Informatics, INDIN 2023, Lemgo, Germany, July 18-20, 2023, pp. 1–7, IEEE, 2023.

URL <https://doi.org/10.1109/INDIN51400.2023.10218245>

In industrial processes, anomalies in the production equipment may lead to expensive failures. To avoid and avert them, the identification of the right root cause is crucial. Ideally, the search for a root cause is backed by causal information like causal graphs. We presented an extension of a framework that fuses causal graphs with anomaly detection to infer likely root causes in a process setup. The causal graph is required to contain measurement and root cause variables and causal relations with annotations for process steps. The framework uses this graph to compute for each root cause variable which measurement variable it might affect in which process step. Thereby, it considers that a cause must always precede its effect. Independently, an anomaly detection algorithm is performed on given sensor measurements to provide information about anomalies and the corresponding process step. Finally, the framework computes the likelihood of each potential root cause by comparing the results from the graph preprocessing and the anomaly detection using the Jaccard similarity to identify the most likely root cause. We demonstrated the use of this framework on a simulated robotic gripping process. Future research might investigate how to learn the causal graph from the provided data using causal discovery methods and how to apply the framework in an online fashion on given process data.

3.15 Hybrid Model Learning for System Health Monitoring

Pauline Ribot (LAAS – Toulouse, FR) and Elodie Chanthery (LAAS – Toulouse, FR)

License © Creative Commons BY 4.0 International license
© Pauline Ribot, and Elodie Chanthery

Joint work of Amaury Vignolles, Elodie Chanthery, Pauline Ribot

Main reference Amaury Vignolles, Elodie Chanthery, Pauline Ribot: “Hybrid Model Learning for System Health Monitoring”, IFAC-PapersOnLine, Vol. 55(6), pp. 7–14, 2022.

URL <https://doi.org/10.1016/j.ifacol.2022.07.098>

Health monitoring approaches are usually either model-based or data-based. This work aims at using available data to learn a hybrid model to profit from both the data-based and model-based advantages. The hybrid model is represented under the Heterogeneous Petri Net formalism. The learning method is composed of two steps: the learning of the Discrete Event System (DES) structure using a clustering algorithm (DyClee) and the learning of the continuous system dynamics using two regression algorithms (Support Vector Regression or Random Forest Regression). The method is illustrated with an academic example.

3.16 Tutorial on Bridge DX / FDI

Louise Travé-Massuyès (LAAS – Toulouse, FR)

License © Creative Commons BY 4.0 International license
© Louise Travé-Massuyès

Joint work of Marie-Odile Cordier, Philippe Dague, François Lévy, Jacky Montmain, Marcel Staroswiecki, Louise Travé-Massuyès

Main reference Marie-Odile Cordier, Philippe Dague, François Lévy, Jacky Montmain, Marcel Staroswiecki, Louise Travé-Massuyès: “Conflicts versus analytical redundancy relations: a comparative analysis of the model based diagnosis approach from the artificial intelligence and automatic control perspectives”, IEEE Trans. Syst. Man Cybern. Part B, Vol. 34(5), pp. 2163–2177, 2004.

URL <https://doi.org/10.1109/TSMCB.2004.835010>

Two distinct and parallel research communities have been working along the lines of the Model-Based Diagnosis approach: the FDI community and the DX community that have evolved in the fields of Automatic Control and Artificial Intelligence, respectively. This talk clarifies and links the concepts and assumptions that underlie the FDI analytical redundancy approach and the DX consistency-based logical approach. A formal framework is proposed to compare the two approaches. It is shown that by adopting the same assumptions regarding fault exoneration, they produce the same diagnostic results.

3.17 Fault Detection, Diagnosis, and Mitigation for Space Propulsion Systems

Günther Waxenegger-Wilfing (Universität Würzburg, DE), Kai Dresia (DLR – Hardthausen, DE), and Ingo Pill (Silicon Austria Labs – Graz, AT)

License © Creative Commons BY 4.0 International license
© Günther Waxenegger-Wilfing, Kai Dresia, and Ingo Pill

Joint work of Günther Waxenegger-Wilfing, Kai Dresia, Ingo Pill, Chiara Gei, Radchenko Gleb, Andrea Urgolo, Federico Pinto, Manuel Freiburger, Heike Neumann


Space propulsion systems continue to be a significant source of faults in space activities, necessitating dedicated fault management strategies to meet stringent safety requirements. The operational nature of these systems, pushed to the limits of technical feasibility to minimize weight, makes them susceptible to a diverse set of faults, with abnormal behavior having potentially catastrophic consequences. The substantial costs associated with the loss of a launch vehicle or spacecraft underscore the critical need for effective fault detection, diagnosis, and mitigation functionalities. Real-time detection and assessment of faults based on available sensor data are imperative to initiate emergency shutdowns or reconfigurations, further compounded by the constraint of limited computing resources.

The German Aerospace Center (DLR), in collaboration with various partners, has long been engaged in exploring the application of AI methods for the operation of space propulsion systems. While past efforts primarily focused on intelligent control in the absence of faults, recent initiatives, such as the collaboration with Silicon Austria Labs (SAL) within the SUNRISE project, mark the initial strides towards fault management. The SUNRISE project is dedicated to researching dependable sensor concepts for resilient control.

The first segment of this presentation unveils preliminary findings, showcasing how trained virtual sensors can effectively estimate critical quantities, such as combustion mixture ratios, and detect faults with high accuracy. In the second part, we introduce the LUMEN demonstrator engine, currently in development, serving as an ideal test bed for diverse control and diagnosis approaches. This includes intentionally injecting faults into the system and utilizing the platform for generating training data for machine learning algorithms.

3.18 Quality Assurance Methodologies for Resilient (Model-based) Systems

Franz Wotawa (TU Graz, AT)


License  Creative Commons BY 4.0 International license
© Franz Wotawa

Deploying systems requires to show that the system complies with its requirements and specifications. Hence, quality assurance is an essential part of any system development, which also holds for systems with attached resilience functionality utilizing model-based reasoning. In my talk, I discuss the general challenge of quality assurance applied to systems with monitoring and diagnosis functionality and the importance of residual risks. In particular, I mention which faults that come when introducing monitoring and diagnosis have to be considered and their effects on residual risks. Afterward, I present early work on using testing for quality assurance for models used for diagnosis and challenges. In particular, I discuss the challenge of coming up with methods for checking the quality of the tests and its consequences. Finally, I summarize the findings and present open challenges.

4 Working groups

4.1 Breakout Session on coupling symbolic and sub-symbolic methods for model acquisition

Rene Heesch (Helmut-Schmidt-Universität – Hamburg, DE)

License  Creative Commons BY 4.0 International license
© Rene Heesch

The discussion within this breakout session built upon the outcomes of the previous day, focusing on integrating prior knowledge with data to enhance model development and refinement. A critical insight from the session was the importance of creating a unified language or framework to facilitate the integration of diverse knowledge types. Additionally, it was proposed that utilizing two distinct languages or foundational models, along with a mapping between them, could also bridge the divide between symbolic and sub-symbolic methods. The session explored model checking for continuous systems and discussed algorithms for learning hybrid automata, leading to discussions on the generalizability of results from specific applications, such as ECG diagnoses, and the feasibility of learning hybrid automata directly from data. This process entails aligning a known hybrid automaton with new data and modifying the automaton based on discrepancies between the model predictions and the actual data. This method is currently being investigated by a PhD student at Laas. Furthermore, Signal Temporal Logic (STL) was discussed as a tool for applying logical systems to continuous signals or variables, demonstrating the adaptability of symbolic methods to continuous data streams. The session finally revisited the DX competition held in 2010, proposing a new competition for 2024 focusing on the development of new diagnostic models. Unlike the previous competition, which focused on algorithms, the proposed DX competition aims to challenge participants to integrate data with parts of knowledge to discover new models. Emphasis was placed on using real data from technical processes, with a focus on incorporating both nominal and faulty data into the dataset. The faulty data labels would be part of the knowledge provided to participants, allowing for a nuanced approach to model

learning and improvement. The idea is to first learn a model based on data and subsequently refine this model through the integration of additional knowledge. The session suggested not to frame this as a conventional competition but rather as a kind of special session where results could be compared and discussed, potentially leveraging past PHM (Prognostics and Health Management) challenges to minimize the preparatory work required.

4.2 Breakout Session on coupling symbolic and sub-symbolic methods in both directions


Rene Heesch (Helmut-Schmidt-Universität – Hamburg, DE)

License  Creative Commons BY 4.0 International license
© Rene Heesch

The discussion within this session was focused on exploring two main pathways for integration: a predominantly symbolic combination approach and a mainly sub-symbolic combination approach. Additionally, the concept of neuro-symbolic integration was discussed. It was clarified that sub-symbolic AI encompasses more than just Machine Learning (ML), although ML approaches were predominantly considered within the session as examples of sub-symbolic AI methods. Regarding the first pathway, the predominantly symbolic approach, which primarily utilizes sub-symbolic AI to generate models for the symbolic components, the discussion was brief. It was concluded that this topic would not be the focus due to potential overlap with another breakout session titled “Model Acquisition (Suitable for Diagnosis) from Real-World Observation”. Consequently, the discussion within in this breakout session concentrated on a primarily ML-based combination approach, outlining different key strategies. One strategy was the use of outputs from symbolic systems as inputs for ML models, providing context-rich information that is not available in raw data. This not only addresses the lack of data but also reduces reliance on large datasets by supplementing them with symbolic insights. Furthermore, the session explored incorporating logical formalisms to describe model phenomena, which enhances the interpretability and transparency of ML models in terms of their explainability. The integration of background knowledge into ML models was discussed as a crucial strategy for influencing model architecture and improving performance, particularly in data-limited scenarios or those requiring a nuanced understanding. This approach uses existing knowledge to guide the design process and boost model effectiveness. Lastly, the potential of neuro-symbolic approaches that combine neural networks with the reasoning capabilities of symbolic AI was discussed. These approaches aim to create AI systems that are not only powerful in analysis but also capable of human-like reasoning. No further steps have been defined so far, as this session merged into the session “Coupling symbolic and sub-symbolic Methods for model acquisition” later.

4.3 Breakout Session on Causality – How to generate knowledge from data?

Lukas Moddemann (Universität der Bundeswehr – Hamburg, DE) and Kaja Balzereit (Hochschule Bielefeld, DE)

License  Creative Commons BY 4.0 International license
© Lukas Moddemann and Kaja Balzereit

The breakout session on causality, held during the Dagstuhl Seminar 24031, provided a deep dive into the complexities and methodologies surrounding the identification and analysis of causal relationships within systems, especially cyber-physical systems. Several approaches to represent causality such as fault propagation graphs [Trave], causal orderings of equations [Bozzano], or causal graphs were discussed. The discussion focussed especially on fault propagation graphs, as a foundational formalism used to understand the sequential dependencies among components. These graphs are crucial for determining which components must be operational for subsequent components to function correctly, illustrating the direct causal links within a system.

A significant portion of the discussion focused on the challenges presented by cycles within fault propagation graphs. These cycles can complicate the analysis by introducing feedback loops where components influence each other in a cyclic manner, making the isolation of causal paths more complex. The session also highlighted the application of Hidden Markov Models (HMMs) as a method to model similar structures causing responses in other structures, offering a statistical approach to understanding how components influence one another even when not directly observable. The relevance of causal models for diagnosis tasks and recent work in this area [Rehak] have also been discussed.

A key takeaway from the session was the importance of high-level causal models in capturing the overarching causal relationships within systems. However, the attendees were reminded of the difficulties in manually crafting these models due to their complexity and the even greater challenge of learning causal models directly from data. When manually crafting causal models, one needs a deep understanding of the dependencies between a – usually high – number of system variables which is often not available. For learning causal models, a large amount of data even for situations which barely occur is needed. Furthermore, the distinction between correlation and causality cannot be done purely data-driven.


In conclusion, the causality breakout session provided valuable insights into the current state of causal analysis, emphasizing both the potential and the limitations of existing methodologies. The discussions underscored the need for continued research and innovation in the field to overcome the challenges of acquiring data and constructing models that can effectively capture the intricate web of causality in complex systems.

References

- 1 Trave-Massuyes, Louise, and Renaud Pons. “Causal ordering for multiple mode systems.” Proceedings of the eleventh international workshop on qualitative reasoning. 1997.
- 2 Rehak, Josephine, et al. “Counterfactual Root Cause Analysis via Anomaly Detection and Causal Graphs.” 2023 IEEE 21st International Conference on Industrial Informatics (INDIN). IEEE, 2023.
- 3 Bozzano, Marco, et al. “SMT-based validation of timed failure propagation graphs.” Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 29. No. 1. 2015.

4.4 Breakout Session on LLMs for DX – Integrating Large Language Models for Root Cause Diagnosis

Lukas Moddemann (Universität der Bundeswehr – Hamburg, DE) and Jonas Ehrhardt (Universität der Bundeswehr – Hamburg, DE)

License  Creative Commons BY 4.0 International license
© Lukas Moddemann and Jonas Ehrhardt

The breakout session on “Large Language Models for Root Cause Diagnosis” featured a comprehensive exploration and discussion on the potential and challenges of using Large Language Models in the topics of the “DX – Principles of Diagnosis” community. The central aspects that were discussed, revolved around (i) the models themselves and their current and potential future capabilities, (ii) the training data for training and refining LLMs for diagnosis tasks, (iii) potential application areas, as well as (iv) current, and (v) future trends and topics that should be monitored or covered by the DX community. As a result, the attendees agreed on writing a position paper, which will capture the current potential and drawbacks of LLMs within DX domains.

The beginning of the session included a brief introduction into the principles of LLMs. Introducing the capability of state-of-the-art LLMs and their capability on simple diagnostic benchmark problems like the Polybox. Subsequently, the prerequisites for current LLMs were discussed to effectively perform diagnoses, including necessary data, semantics, and specialized training. The capabilities of current LLMs and LLM-ensembles were discussed, highlighting the capability of formulating programming code for simple, testable and traceable reasoning, as well as the capability of understanding image data, like circuits or technical drawings, for an easier and more precise recognition of concepts of diagnosable systems. Extending the capability of pre-trained LLMs for diagnosis by fine-tuning them on diagnosis problems, as well as ensemble approaches of different Expert-LLMs for different diagnostic tasks, were highlighted as low-hanging fruits. Lastly, the capability of continuous training of LLMs for their application on changing systems was identified as a challenge.

Regarding the training data for LLMs that perform diagnostic tasks, a broad field was identified, reaching from image data, to time-series data from system observations, natural language, technical documentations, or structured knowledge graphs. The discussion came to the consensus that for training from ground up general world-knowledge should be included, whereas for fine-tuning models only specific information would be needed.

Identifying causality with LLMs was considered as a fundamental aspect of LLMs for fostering applications in diagnosis, like root cause analysis or root cause identification. Additionally, the ability to capture causality and contradiction was identified as a major aspect that should not only be considered and researched on a phenomenological level which considers LLMs as black-boxes, but also by looking into the functioning of the models.

Current trends and topics that should be monitored by the DX community revolve around the short term perspective of current LLMs in the application field of diagnosis. This includes understanding the immediate capabilities and limitation of current models, as well as application scenarios in which LLMs could already perform diagnostic tasks or at least pose as a component within a diagnosis framework.

Future trends and topics include the long term perspective of LLMs in the DX context. These trends revolve around enhancing the accuracy of LLMs and should be driven by autonomy research for structuring the requirements for LLMs in diagnostic roles.

The session concluded in writing a statement paper toward the current state of LLMs in diagnostic applications, highlighting aspects that LLMs already can achieve, and limitation they occur. The claims will be proven with empirical evaluations, like testing LLMs for creating causal graphs or evaluating on standard diagnosis problems.

4.5 Breakout Sessions on Resilience

Ingo Pill (Silicon Austria Labs – Graz, AT)

License  Creative Commons BY 4.0 International license
© Ingo Pill

In respect of these specific questions we saw that the application and scenario play a significant role when aiming to assess what would be “good” and “bad” behavior. The same goes for the question of whether we would assess the performance of a system in a local or a global context, an example for the latter being a system of resilient systems context. To this end we identified a set of such relevant scenarios ranking from an energy management scenario at a local home, via the operation of an electric grid, via agents/robots in a collaborative disaster or military scenario, to supply chain management. We also discussed and converged to a definition of resilience that would tailor to all the expressed needs. Enabled by our discussions, we identified also some follow-up actions:

- Authoring a white paper on resilience by a group of the attendees of this seminar (in 2024)
- Submitting a proposal for a follow-up Dagstuhl Seminar proposal that focuses on resilience, and where we will invite scientists from more relevant fields as well as relevant stakeholders (agency, psychology and societal sciences, security & safety, law and public regulations, ...)

Please note that the discussions led in an interdisciplinary context at Dagstuhl will also contribute to the evolution of the Int. Workshop on Principles of Diagnosis to International Conference on Principles of Diagnosis and Resilient Systems that we are implementing in 2024.

5 Panel discussions

5.1 Panel on Current and Future Challenges in Resilient System Design

Ingo Pill (Silicon Austria Labs – Graz, AT)

License  Creative Commons BY 4.0 International license
© Ingo Pill

Joint work of Ingo Pill, Gautam Biswas, Alessandro Cimatti, Johan De Kleer, Ken Forbus, Oliver Niggemann, Franz Wotawa

Directly in succession to Johan de Kleer’s keynote on Resilience discussed in an additional report, we organized this panel discussion to which we invited panelists with diverse backgrounds such as to cover topics like software engineering, intelligent agent design, automation in production and manufacturing, AI-based control, rigorous system design, formal verification, run-time verification and monitoring, intelligent sensing, and other related topics that are related to the diverse challenges connected to designing resilient systems. Similar to the term “artificial intelligence”, there seems to be an intuitive understanding of the concept’s purpose and the meaning of resilience on an abstract level. As we saw in our discussions, there are, however, also differences in how to interpret the concept and what to expect from a resilient system. As an initial characterization, let us thus describe resilience as a system’s intrinsic ability of sustaining its operation also when impacted by anticipated and unexpected contingencies. In this context, we would like to distinguish between basic and extreme resilience as follows: Basic resilience would allow a system to cope with anticipated

issues, while extreme resilience would enable a system to deal also with challenges that were not anticipated when the system was designed. While resilience could relate also to resilient design concepts that would allow a designer or developer to react more easily to design/requirement changes (or that certain components are resilient to changes in other components). In contrast, we consider the major focus of resilience to be on maintaining expected operation during its operation, no matter the circumstances. It is important to note though that we have to design a system such as to add the capabilities of dealing with (unexpected) issues (faults, threats, environmental changes, ...) at design time – it is only the effects achieved that we are experiencing at run-time. To the end of discussing relevant technologies, we invited our panelists to give short lightning talks where we tasked them to provide some background information about resilience aspects in their individual expertise to the audience, and to comment on the most recent questions and thoughts covered by frontier research. Including the discussion among the panelists, with the moderator and also the entire audience, the lightning talks inspired the following discussions:

- Gautam Biswas brought up in his statement the fact that designing resilience into a system can be thought about in two directions. That is, we can anticipate issues and design a system in a way that it would be “robust enough” against certain problems by design. The second concept would be to allow a system to assess and consider a situation at run-time, reason about an appropriate mitigation strategy and then take mitigation actions – all done at run-time. We can emphasize on the second option when using the term operational resilience. There are several stages that a system goes through in the context of such operational resilience, in that a system would suffer from degraded performance before the mitigation strategy’s effects manifest and the system’s desired performance is restored (to the degree to which this is possible).
- Alessandro Cimatti focused in his statement on the challenge of defining resilience, and he referred to multiple example domains for showcasing relevant questions. He brought up the implicit connection to fail-operational concepts, to the operation of adaptive systems, to planning in the context of non-determinism and uncertain duration, and he observed that such planning alone won’t go far on its own. That is, it is a combination of techniques that will be necessary to tackle the challenges faces when aiming for resilience in a system’s behavior (like the ability to extract models for evolving dynamic environments). A specific question of interest is that of enabling resilience from a short- and a long-term perspective
- Ken Forbus focused in his statement titled “Analogy and Cognitive Architecture as Sources of Software Resilience” on the importance of the concept of analogy, as well as the design of a cognitive architecture propelling the performance in resilient, intelligent systems. Especially in the context of extreme resilience, a system faces incomplete information (not only in terms of the environment, but also referring to domain knowledge) like when we initially did not know how to deal with Covid-19 as a society. Drawing on analogies and exploiting earlier expertise for analogous situations could be one fundamental technology to drive solutions for achieving resilience. This will require us to enable agency in a system, such that systems will elevate from being simple tools to becoming intelligent and evolving agents. The cognitive architectures that would allow us in implementing such agency (that then enables a system to efficiently come up and effectively execute appropriate mitigation strategies) are among the currently most relevant challenges.
- Oliver Niggemann discussed the necessity of considering backloading instead of frontloading when designing a resilient system, and that we need to adapt our design processes as well as the education of engineers accordingly. In particular we see that, while a well-engineered system is a prerequisite for a system to be trustworthy, the complexity and

dynamics of applications requires us to come up with trustworthy AI-based solutions for operating a system. Designers and engineers thus will need to think about the operation phase in more detail when developing future systems. Our education and engineering concepts have to be adapted in order to support and being able to leverage resilient system design in practical designs. This includes also addressing the fact that we have currently a set of technologies available that are promising in being able to address one or the other resilience aspect from a scientific perspective. We lack, however, integrated approaches and methodologies that we can then deploy in practice and transfer to an industrial context, so that a big challenge in this context of resilience is to develop those.

- Franz Wotawa discussed in his statement three fundamental questions, considering resilience not only from a design perspective but also from the perspective of evaluating a resilient system design: What is the best resilient system design? What are desired properties of resilient systems? How do we ensure the correctness of resilient systems? Addressing those questions requires us to think about architectural aspects but also about our development processes that now have to facilitate resilience in a design. This entails the need to establish not only a common understanding of the purpose and meaning of resilience, but also of the degrees of freedom a resilient system is allowed to operate within. This is crucial not only from a design perspective, but especially so in an evaluation and verification context. Enabling the latter, we need to come up with concrete evaluation metrics, and we need to define exact bounds of acceptable autonomy in resilience.

From the discussions we had in the panel and two break-out sessions, we can immediately conclude that achieving resilience is a very complex task. Aside apparent technical and technological questions, there are also legal ones, like who would be responsible if the required autonomy to achieve operational resilience causes harm, damage, or the loss of revenue. Psychological and societal questions are also relevant, like in the sense that they influence the definition of acceptable behavior or are of interest in a technological context when we think of human-machine collaboration or also systems of resilient systems where we could have humans in the loop or where we are (at the very least) operating in a shared environment. While some resilience can be achieved with anticipating challenges and making a design resilient by default to those, it is especially the extreme resilience where we equip a system with the intelligence to overcome unexpected issues at run-time that requires us to rethink our current design, development, evaluation and verification processes. Due to the complexity of the discussions, they were led not only in the time-limited context of the panel discussion, but specific aspects were discussed also in two break-out sessions:

- What is resilience and how do we measure it?
- Dealing with unknown unknowns in resilience?

6 Open problems

6.1 LiU-ICE Industrial Fault Diagnosis Benchmark – Anomaly Detection and Fault Isolation with Incomplete Data

Daniel Jung (Linköping University, SE)

License © Creative Commons BY 4.0 International license

© Daniel Jung

Joint work of Daniel Jung, Mattias Krysaner, Erik Frisk

URL https://vehsys.gitlab-pages.liu.se/diagnostic_competition/

A common challenge of designing diagnosis systems in industrial applications, is limited data availability from relevant fault scenarios and a lack of knowledge of model uncertainty. Development of fault diagnosis design techniques in this situation is the theme of the competition.

The case study is the air-flow of an internal combustion engine. The complexity of modeling the engine together with noisy measurements makes is a challenging system to diagnose because of its non-linear dynamic behavior and wide operating range.

Competition Objectives

- Design a diagnosis system that can detect and isolate faults.
- Handle that availability of representative data from all fault scenarios and fault sizes is limited.
- The diagnosis system should handle faults that are not represented in training data.

Participants

- Kaja Balzereit
Hochschule Bielefeld, DE
- Gautam Biswas
Vanderbilt University –
Nashville, US
- Marco Bozzano
Bruno Kessler Foundation –
Trento, IT
- Elodie Chanthery
LAAS – Toulouse, FR
- Alessandro Cimatti
Bruno Kessler Foundation –
Trento, IT
- Marco Cristoforetti
Bruno Kessler Foundation –
Trento, IT
- Philippe Dague
University Paris-Saclay –
Orsay, FR
- Johan de Kleer
c-infinity – Mountain View, US
- Alexander Diedrich
Helmut-Schmidt-Universität –
Hamburg, DE
- Kai Dresia
DLR – Hardthausen, DE
- Jonas Ehrhardt
Universität der Bundeswehr –
Hamburg, DE
- Alexander Feldman
NextFlex – San Jose, US
- Kenneth D. Forbus
Northwestern University –
Evanston, US
- Rene Heesch
Helmut-Schmidt-Universität –
Hamburg, DE
- Daniel Jung
Linköping University, SE
- Lukas Moddemann
Universität der Bundeswehr –
Hamburg, DE
- Angelo Montanari
University of Udine, IT
- Manfred Mücke
Material Center Leoben, AT
- Edi Muskardin
Silicon Austria Labs – Graz, AT
- Oliver Niggemann
Helmut-Schmidt-Universität –
Hamburg, DE
- Ingo Pill
Silicon Austria Labs – Graz, AT
- Gregory Provan
University College Cork, IE
- Belarmino Pulido
University of Valladolid, ES
- Josephine Rehak
KIT – Karlsruher Institut für
Technologie, DE
- Pauline Ribot
LAAS – Toulouse, FR
- Martin Sachenbacher
Universität Lübeck, DE
- Anika Schumann
IBM Research-Zurich, CH
- Gerald Steinbauer-Wagner
TU Graz, AT
- Markus Stumptner
University of South Australia –
Mawson Lakes, AU
- Anna Szyber
Warsaw University of
Technology, PL
- Louise Travé-Massuyès
LAAS – Toulouse, FR
- Günther Waxenegger-Wilfing
Universität Würzburg, DE
- Katinka Wolter
FU Berlin, DE
- Franz Wotawa
TU Graz, AT
- Marina Zanella
University of Brescia, IT
- Alois Zoitl
Johannes Kepler Universität
Linz, AT

