Modeling Emergent Behaviour for Enhanced Autonomy in Cyber-Physical Systems

Matthias Schaffeld(✉) and Torben Weis

University of Duisburg-Essen, 47057 Duisburg, Germany
{matthias.schaffeld,torben.weis}@uni-due.de

Abstract. The convergence of Cyber-Physical Systems (CPS) and autonomous systems presents numerous application areas and challenges. Many CPS use cases require autonomy, which necessitates robust monitoring, maintenance, and longevity of these systems. However, handling unmodeled data, environmental uncertainties, and emergent behaviour poses significant challenges. This paper proposes the use of a novel formalism, Extended Hidden Markov Models with ϵ-emissions (ϵ-HMMT), to estimate the correctness of an autonomous system's emergent behaviour in complex and changing environments. We discuss the advantages and disadvantages of different modeling approaches, emphasizing the need for a more adaptable and robust modeling technique.

Keywords: behavioural modeling · monitoring · cyber-physical systems

1 Introduction

As cyber-physical systems (CPS) become more ubiquitous and pervasive, the number of applications that also require them to be more autonomous increases. To support this autonomy, we want to focus on enhancing the monitoring, analysis, verification, and longevity of these autonomous CPS. We argue, that all of these aspects can be enhanced by using robust modeling. There are however many challenges to model CPS. First, when creating and modeling complex CPS models, one has to deal with, incomplete technical data, possibly wrong system assumptions, hidden (physical) processes and interactions, or physical processes that are too intricate to model accurately. Therefore, in modeling, we must rely on simplifying assumptions and develop coarse models that only partially represent the real system's behaviour [7]. Furthermore, these models replicating the real-world engineered CPS, have to account for uncertainties within the physical environment, interactions, and communication systems. The magnitude of uncertainty's influence on a system often hinges on the degree of control attainable over its environment. In the case of autonomous systems, such as CPS interacting with the physical world, controllability typically remains quite constrained.

A. Zaslavsky et al. (Eds.): MobiQuitous 2023, LNICST 594, pp. 520–528, 2024.
https://doi.org/10.1007/978-3-031-63992-0_35

Consequently, these systems must possess the capability to accommodate data that was not accessible during their initial modeling, development, and testing phases. Additionally, it's essential to acknowledge that these systems are susceptible to parameter drift as a result of alterations and wear and tear. A model that has once been suitable does not necessarily have to remain so.

Secondly, CPS inherently require a combination of models and methods of various engineering disciplines with those of computer engineering, which do not combine well [5]. Even if they did, the system components are inherently interdependent and the emergent behaviour can not easily be determined by the heterogeneous models the individual system components are based on.

In the field of CPS, Model Driven Design is a widely used approach in which different categories of models are used, each of which has its own advantages and disadvantages. Evaluating system behaviour using these models is an important objective. However, due to the intricate nature of CPS, relying solely on any of these approaches or their combinations falls short of accurately predicting the emergent system behaviour during deployment. Hence, our proposed modeling formalism seeks to fill this gap by facilitating the learning and prediction of the emergent behaviour of the comprehensive system. Its purpose is not to encapsulate the entire system with all its components, dependencies, and physical laws. Instead, it aims to seamlessly integrate into existing systems as a supplementary tool, functioning as a heuristic to assess the correctness of emergent system behaviour by effectively capturing it.

2 Key Concepts and Desired Properties

This section presents key concepts that will be addressed in our research, as well as an overview of the desired properties of our formalism.

First, we share the definition of autonomous systems from Müller et al. [6] according to which an autonomous system is a defined technical system capable of autonomously and systematically achieving its predetermined objectives in uncertain environmental conditions. It exhibits self-governance, adapts to its environment, executes processes systematically, and is self-contained while allowing variation in instrumental goals to achieve its primary set objectives.

A CPS is an integration of physical entities and associated processes equipped with computational capabilities that enable a seamless connection between the physical world in which sensors and actuators operate and the virtual world of information processing. It combines elements from different domains, such as embedded computing, physical systems, and communication technologies, to monitor, control, and coordinate operations in a variety of applications, focusing on the interaction between computational components and physical processes. Given that CPS are required to make decisions, adapt, and execute actions to achieve their designated objectives, all while contending with incomplete sensory information, approximate modeling, uncertainty within the physical system, and limitations in environmental control, it becomes evident that there is a significant overlap between autonomous systems and CPS.

Within this system, the cyber elements exhibit discrete dynamics, while the physical aspects demonstrate continuous behaviour. This necessitates a pivotal challenge: discretizing the continuum [7] effectively to facilitate accurate interpretation, prediction, and interaction between the cyber and physical components. However, our focus lies not in exhaustive detailing of every individual component but rather in comprehending the emergent behaviour of the entire system and its constituent parts. Consequently, we adopt a phenomenological perspective, viewing the behaviour of the CPS as a sequence of discrete events.

In the context of CPS design and deployment, we distinguish three types of instances of the same system. The first is the *conceptual instance*, which encapsulates the developers' mental model of the CPS. This model encompasses all knowledge and assumptions related to the desired system behaviour, addressing questions such as the environment's characteristics, the system's objectives, the requisite parameters for cyber components, decision-making processes, action execution methods, and the impact of actions on the environment. Given the inherent complexity and uncertainties of CPS, conceptual instances inevitably incorporate assumptions, approximations, and the potential for errors. Unfortunately, since these are only mental models, their comprehension remains elusive.

Nonetheless, the subsequent form of instance we refer to as *specification instance*, is constructed upon these mental constructs. Specification instances formally articulate the entirety of system components, encompassing the environment, physical processes, and physical and cyber components, as well as their interdependencies, behavioural patterns, constraints, and objectives. Ideally, these specification instances encompass all the information found in the conceptual models providing detailed representations, such as differential equations for physical processes. However, this means that the inherent approximations, assumptions, and errors found in the conceptual instance are also adopted. Furthermore, while these concretize and enhance the tangibility of conceptual instances, they often also sacrifice a degree of precision inherent in their conceptual counterparts. Notably, not every conceptual instance readily translates into a suitable specification instance as a fitting modeling formalism may be missing, and not all types of models seamlessly combine, for example, to represent their interactions. Moreover, existing implicit assumptions remain implicit in the transition to specification instances, and dynamics that are too complex or intricate to capture in modeling require further simplifications, resulting in further loss of information. After all components of the system considered in the conceptual instance have been defined, it becomes possible to conduct a thorough verification process in alignment with best practices. At this point, the intended system behaviour can be verified through the use of simulations and test data. Nevertheless, the uncertainty persists regarding the extent to which the test data accurately mirrors the real-world environment and whether the simulations adequately capture the genuine behaviour of the implemented system.

The third type of instance, which we want to focus on, is the *implementation instance*. These are the realization of their specification counterparts, both in the physical and cyber domains, thereby inheriting the same set of assumptions, errors, and the potential for information loss. However, this transition from specification to implementation introduces its own set of challenges. Issues may arise due to resource constraints, misinterpretations of the specifications, implicit assumptions, and makeshift solutions, which often stem from the system's inherent uncertainties and the inherent incompleteness of specification instances. This transformation from conceptual to implementation instance can, therefore, be likened to a game of "Telephone", where information is passed along and becomes distorted in the process. In light of these complexities, we contend that our primary focus should be on capturing, observing, and verifying the emergent behaviour of the implementation instance. This perspective is further reinforced by the fact that even when models and software undergo verification, manufacturers are compelled to validate the actual component behaviour in conjunction with their hardware, recognizing that behaviour may diverge depending on the specific combination of software and hardware elements [5].

Based on this reflection on autonomous systems, CPS, and their instances, we want to develop a modeling formalism that satisfies the following properties:

Capture of Emergent System Behaviour: Emergent system behaviour refers to the system's behaviour that arises from the interactions of its components, which are often not explicitly specified. We aim to develop a modeling formalism that has the ability to capture and represent such emergent behaviours. This is crucial for understanding complex systems where global behaviour cannot be easily deduced from the properties of individual components alone.

Learnability from Data: Our focus is the emergent behaviour of CPS during deployment, which involves handling uncertainties and previously unseen data. However, we lack precise models that fully capture this behaviour. Consequently, a modeling formalism must be equipped with the capacity to derive crucial model parameters and relationships from observed data.

Determination of Deviations: The ability to detect and quantify deviations in system behaviour is an essential requirement for our formalism. It is especially relevant in contexts like emergency management, as highlighted by the example of aging infrastructures increasing the risk of system unreliability [2].

Smoothing of Observed System Behaviour: In many practical applications, observed system behaviour may contain noise or erratic fluctuations. A modeling formalism should incorporate algorithms or mechanics for smoothing this behaviour to enhance the accuracy and analysis of system dynamics.

Confidence Measure in Prediction: To instill trust and reliability in the predictions generated by a suitable modeling formalism, it has to incorporate some form of a confidence measure. This measure provides an indication of the degree of certainty associated with the model's predictions.

Adaptability: As systems evolve a modeling formalism must be able to accommodate changes in system behaviour. This ensures that our modeling approach remains relevant and effective as the system behaviour changes.

Non-disruptive Modeling: The formalism has to be non-invasive, meaning that the system under consideration is not compelled to undergo substantial modifications to conform to the modeling formalism. This principle aligns with the concept of non-disruptive modeling, where the modeling process should not disrupt the normal operation of the system. This also enables the capability of retrofitting, allowing the formalism to be applied to existing systems without necessitating extensive changes.

3 Approach and Challenges

In the realm of modeling emergent behaviour in Cyber-Physical Systems (CPS), a range of modeling techniques presents itself as viable options. The survey conducted by Rai et al. [7] has categorized these techniques into four distinct categories, namely physics-based equations, state machine-based modeling, rule and agent-based modeling, and data-driven modeling.

Physics-based equations excel in capturing the physical dynamics of CPS, while state machines are particularly adept at modeling discrete system dynamics. Rule and agent-based modeling, on the other hand, excel at representing intricate dependencies between cyber and physical components, effectively capturing system semantics. All these categories offer the advantage of traceability in modeling but face challenges when dealing with the complexity of CPS. Moreover, they share a common limitation-they can only represent known system dynamics, leaving out the nuances of real-world system behaviour during deployment.

To address the need for capturing actual system behaviour in deployment scenarios, these models must integrate data-driven learning methodologies. However, the fourth category, data-driven modeling, stands out by virtue of its capability to encompass both continuous and discrete dynamics. Nevertheless, it comes with the drawback of reduced explainability and the requirement for substantial data to achieve accuracy [7].

Furthermore, when modeling CPS behaviour, the choice between deterministic and non-deterministic modeling is a key consideration. In his survey [5], Lee discusses the pros and cons of both approaches, which we briefly outline below.

Deterministic models offer the advantage of clarity and precision, making them valuable for analysis and the development of CPS. They establish a definitive system behaviour providing an unambiguous definition of what "correct" behaviour is. However, deterministic models have inherent limitations, such as potential incompleteness in capturing uncertainties and complexities, particularly in intricate CPS scenarios, where they can grow unwieldy.

In contrast, non-deterministic models excel in accommodating the inherent uncertainties of the real world, offering the ability to represent entire families of possible behaviours. This quality allows them to deal with sources of non-determinism, including physical noise, component failures, imperfect actuation, and unpredictable delays. Furthermore, non-deterministic models are well-suited for situations where unknown or unknowable properties play a central role. Building on this foundation, Lee asserts that deterministic modeling, when coupled with faithful implementations, remains relevant in the context of CPS and leverages the established advantages of deterministic models. However, our main focus is on capturing the emergent behaviour of CPS that serves as a basis for analysis, verification, and maintenance. Given the inherent complexity and non-deterministic nature of CPS, we argue for the extension of a non-deterministic probabilistic formalism to represent behaviour - Hidden Markov Models (HMM).

Given that state machines are inherent to the computational components of CPS that determine the executed system behaviour, we aim to address existing limitations, such as the lack of robustness and adaptability, associated with traditional approaches. To achieve this, we propose leveraging the ϵ-Hidden Markov Model with Transition Emissions (ϵ-HMMT) which we introduced in previous work [9]. It extends HMMs by incorporating ϵ-emissions, that allow for unobservable transitions, which enable the handling of missing or erroneous observations. It also moves the emissions from states to transitions, facilitating more fine-grained modeling of emissions, leading to an alignment to the more accustomed formalism of the discrete automaton. Furthermore, in our previous work (Bernemann et al. [1]) we extended and verified the Baum-Welch and Viterbi algorithms for ϵ-HMMT. Consequently, the Baum-Welch algorithm can learn system behaviour by observing event sequences while accounting for ϵ-emissions. The Viterbi algorithm on the other hand considers both past and future emissions when determining the best explanation for a sequence of observations.

However, the use of HMM comes with inherent limitations. This approach assumes that the system's behaviour is observable through discrete events. Furthermore, it assumes the Markov property, meaning that the next event depends solely on the previous one. Deviations from this assumption necessitate an exponential increase in the number of states. Currently, the model also does not account for temporal dependencies. Nonetheless, we plan to extend the formalism with techniques such as Good Turing and Back-off Estimation.

In the survey done by Gunes et al. [3], which describes 21 challenges in the field of CPS, we identified seven challenges that we perceive as pertinent to our research objectives focused on enhancing CPS autonomy through the modeling of emergent behaviour. These challenges are succinctly summarized below and elaborated on within the context of our proposed methodology:

Predictability: Predictability denotes the extent to which one can predict the system's state, behaviour, or functionality, through qualitative or quantitative means. Problems such as the inherent uncertainty of the environment and non-deterministic processes must be dealt with to achieve predictability.

Considering the inherent utilization of probabilities and the incorporation of ϵ-emissions, the ϵ-HMMT represents a significant step towards addressing the challenges posed by non-determinism and uncertainty in CPS. By virtue of the Viterbi algorithm's ability to consider past and future observations, including potentially erroneous ones, when determining the optimal explanation for an observed sequence, we can harness this capability to introduce confidence measures for various explanations, current states, and predictions.

Accuracy denotes the extent to which the observed/measured behaviour of the system matches its calculated/actual behaviour. Accuracy assessment using the ϵ-HMMT presents a straightforward process. It involves a comparison between the observed system behaviour and the behaviour computed using the Viterbi algorithm. This assessment can be done through various approaches, ranging from a comprehensive method employing a window-based strategy in conjunction with system-specific knowledge to simpler techniques such as measuring the edit distance between the calculated and observed event sequences.

Dependability denotes the system's ability to consistently deliver required functionalities, maintain performance, and produce expected outcomes without significant degradation throughout its operation. It represents the extent of trust that can be placed in the entire system. Dynamic interdependencies between cyber and physical components as well as issues like timing uncertainties in sensor data and actuation make achieving dependability a challenging task.

After applying the extended Baum-Welch algorithm to learn the emergent system behaviour, one can subsequently assess the model's accuracy and employ it as a benchmark for monitoring the system. If there is a decline in the accuracy of the system or its components, targeted maintenance can be scheduled to ensure dependability. By incorporating a confidence metric rooted in the likelihood of observation explanations as determined by the Viterbi algorithm, we can establish a quantifiable confidence indicator of the system behaviour's correctness.

Maintainability denotes the system's ability to perform repairs and recoveries in the event of failures. This includes continuous monitoring and testing of the infrastructure, which enables the identification and remediation of components that are prone to recurring failures. While monitoring the system against a learned benchmark, one can schedule targeted maintenance if the accuracy of the system or specific components drops. The extent to which erroneous behaviours and components can be identified and isolated remains to be evaluated.

Robustness denotes the system's capacity to maintain stability and functionality, even in the presence of failures, disturbances, and unforeseen events. A robust CPS effectively handles failures and disruptions, including sensor noise, actuator inaccuracies, communication issues, and hardware errors. Robustness is an essential CPS challenge due to the incomplete knowledge about the system and its environment during design, leading to unavoidable disturbances [8].

Given that the ϵ-HMMT offers ϵ-emissions and introduces more intricate modeling through transition emissions, it holds the potential to increase robustness. This enhancement is achieved by enabling established HMM algorithms to proficiently manage imperfect or erroneous data while supporting intricate modeling.

Adaptability denotes the system's ability to (autonomously) modify its configuration to deal with the continuous change of environmental conditions.

The behavioural model embedded within a trained ϵ-HMMT aims to validate the anticipated system state, refine potentially inaccurate observation data, and assess system accuracy, enabling crucial adjustments. Reiteratively training the model with data from a modified CPS while leveraging the prior model as input offers valuable insights into the extent of CPS modifications. Additionally, this approach introduces adaptability to a state machine-based modeling paradigm.

Reliability denotes the system's ability to consistently perform its intended function in uncertain, evolving environments. The system has to be able to manage uncertainties such as incomplete knowledge and parameter variations. The limits of modeling uncertainty, component accuracy, network connections, and software errors are all factors that decrease reliability [4].

Selecting a modeling formalism that embraces uncertainty as a CPS tool and enhancing its robustness holds significant potential for evaluating and bolstering CPS reliability. By striving to address at least some, if not all, of the aforementioned challenges, we can anticipate an improvement in reliability as well.

4 Conclusion

In conclusion, we emphasized the need to model emergent behaviour in CPS to promote their autonomy and recognize the inherent challenges and uncertainties associated with modeling these systems. We defined key concepts, addressed challenges in transitioning from conceptual models to real-world implementation, and outlined the desired properties for our proposed modeling formalism. We also explored a range of modeling techniques and discussed the choice between deterministic and non-deterministic modeling, ultimately advocating for the use of an extended HMM formalism (ϵ-HMMs) to capture the dynamic emergent behaviour of CPS. In parallel, we identified CPS challenges including dependability, predictability, accuracy, reliability, robustness, adaptability, and maintainability, and discussed the application of our methodology. Our forthcoming work will focus on evaluating the limitations of our approach as a basis for further useful and needed extensions to the formalism.

References

1. Bernemann, R., König, B., Schaffeld, M., Weis, T.: Probabilistic systems with hidden state and unobservable transitions. arXiv:2205.13871 (2022)
2. F.E.M.A.: Crisis response and disaster resilience 2030: forging strategic action in an age of uncertainty. Tech. rep., The Strategic Foresight Initiativ (SFI) (2012)
3. Gunes, V., Peter, S., Givargis, T., Vahid, F.: A survey on concepts, applications, and challenges in cyber-physical systems. KSII Trans. Internet Inf. Syst. **8**, 4242–4268 (2014). https://doi.org/10.3837/tiis.2014.12.001

4. Sztipanovits, J., Ying, S., I.C.D.C.J.D.H.K.P.J.M.V.P.L.S.: Strategic r&d opportunities for 21st century cyber-physical systems. Tech. rep., Steering Committee for Foundations in Innovation for Cyber-Physical Systems. (2013)
5. Lee, E.A.: The past, present and future of cyber-physical systems: a focus on models. Sensors **15**(3), 4837–4869 (2015). https://doi.org/10.3390/s150304837
6. Müller, M., Müller, T., Talkhestani, B.A., Marks, P., Jazdi, N., Weyrich, M.: Industrial autonomous systems: a survey on definitions, characteristics and abilities. at - Automatisierungstechnik **69**(1), 3–13 (2021). https://doi.org/10.1515/auto-2020-0131
7. Rai, R., Sahu, C.K.: Driven by data or derived through physics? A review of hybrid physics guided machine learning techniques with cyber-physical system (CPS) focus. IEEE Access **8**, 71050–71073 (2020)
8. Rungger, M., Tabuada, P.: A notion of robustness for cyber-physical systems. IEEE Trans. Autom. Control **61**(8), 2108–2123 (2016)
9. Schaffeld, M., Bernemann, R., Weis, T., König, B., Matkovic, V.: Lifecycle-based view on cyber-physical system models using extended hidden Markov models. In: 2022 20th ACM-IEEE International Conference on Formal Methods and Models for System Design (MEMOCODE) (2022)