Position Paper: Research Challenges in the Design and Composition of Surrogate Models for Robust CPS

Yogesh Barve, Pranav Karve, Aniruddha Gokhale and Sankaran Mahadevan
Vanderbilt University
Nashville, TN, USA
{yogesh.d.barve,pranav.m.karve,a.gokhale,sankaran.mahadevan}@vanderbilt.edu

ABSTRACT

Data-driven Artificial Intelligence (AI)/Machine Learning (ML)-based models of CPS (also called *Digital Twins*) are becoming important in the design and control of modern CPS. CPS are a unique class of intelligent system, where the governing process models for the system (or its parts) are available to inform decision making. The governing physics or chemistry models provide an additional source of information, potentially improving the accuracy of CPS modeling, but also increasing the complexity of the information fusion process. The resource constraints of CPS, however, makes it hard to perform the information fusion within the CPS itself. Moreover, building a large, monolithic model of the complete CPS is also infeasible due to model size and training complexity. The hierarchical and compositional structure of CPS calls for distributed model training and composition. Resource constraints and real-time control needs call for rapid and low-cost training of models, which makes surrogate modeling a promising approach. However, when surrogate models are composed, their approximate nature may lead to larger errors. Further, data-driven models are prone to adversarial attacks, whose impacts may be even more pronounced in a federated model. This position paper highlights the key research challenges in this realm and need for new research.

CCS CONCEPTS

• **Computer systems organization → Embedded and cyber-physical systems**; • **Computing methodologies → Modeling and simulation**; **Machine learning**; • **Mathematics of computing → Probability and statistics**.

KEYWORDS

CPS, Surrogate Models, Composition, Accuracy, Adversarial.

ACM Reference Format:
Yogesh Barve, Pranav Karve, Aniruddha Gokhale and Sankaran Mahadevan. 2021. Position Paper: Research Challenges in the Design and Composition of Surrogate Models for Robust CPS. In *Design Automation for CPS and IoT (Destion '21), May 18, 2021, Nashville, TN, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3445034.3460505

1 INTRODUCTION

Cyber Physical Systems (CPS) are large-scale, networked systems with interconnected subsystems each of which may operate at different time scales and have different real-timeliness and dependability requirements. As the complexity and scale of CPS continue to grow and their deployment environments become increasingly challenging for humans to intervene for maintenance and upgrades, these CPS must be designed with self healing and self management attributes.

The first generation CPS were predominantly networked embedded systems, where traditional real-time scheduling and feedback control techniques were extended to address the needs of these CPS. However, with the adoption of Internet of Things (IoT) in CPS and the ensuing challenges stemming from the presence of Big Data, heterogeneity in resources and their constraints, the increased complexity of the subsystems, and fluctuating dynamics of these CPS in terms of workloads and resource availabilities, it is no longer feasible to design such CPS with *a priori*-defined control strategies. Any offline designed solution may at most serve to initially seed the system when it is deployed, but the strategies must adapt autonomously based on changing dynamics.

Data-driven Artificial Intelligence (AI)/Machine Learning (ML)-based models of CPS (also called *Digital Twins* [6, 8, 13, 15, 22]) are therefore becoming increasingly important in designing and controlling modern CPS as evident from emerging research literature [12, 17]. CPS are a unique class of intelligent systems, where, in contrast to natural language or image or video processing systems, the governing process models for the system (or its parts) are available to inform decision making. The governing physics or chemistry models provide an additional source of information, potentially improving the accuracy of building models of CPS, but also increasing the complexity of the information fusion process.

However, since CPS systems are often resource-constrained due to their embedded artifacts and demand real-time guarantees in their control loops, it is hard to perform the information fusion within the CPS itself. Moreover, building a large, monolithic model of the complete CPS is also infeasible for a variety of reasons. First, the size and complexity of such large models may require a cloud-based storage and training/execution on cloud-based servers both of which may be expensive and incur long and variable delays, which is detrimental to the real-time needs of the control loops. Second, since CPS are formed as a composition of many subsystems, each subsystem may need to be controlled at different time scales and may use different protocols and communication infrastructure making it infeasible for a single, large model to serve as the controller of many different subsystems all at once. Third, CPS are highly dynamic systems which may cause *concept drift* [24], i.e.,

Destion '21, May 18, 2021, Nashville, TN, USA

ACM ISBN 978-1-4503-8316-5/21/05…$15.00
https://doi.org/10.1145/3445034.3460505

degradation in accuracy, of the learned model due to significant deviations in the type of data used to train the models thereby requiring model re-learning or continual learning [23] . However, a continual learning process for the entirety of the CPS may be extremely expensive and incur extremely long time-scales, which may not be acceptable for the safe and correct operation of the deployed CPS. Finally, the loss in accuracy may propagate across the models when they are composed, or one or more models may be subject to adversarial attacks [10, 26] both of which may pose serious consequences for the effective control of the CPS.

Addressing these challenges calls for handling most of the model (re)learning and execution activities closer to or within the CPS – a paradigm referred to as *edge computing* [19]. Models that can still be afforded to be learned over longer time scales and which are based on aggregate information can be trained using traditional cloud resources. For those models that must be trained at the edge, due to the constraints on resources in the CPS and contention for these resources by the different subsystems, learning a model of the CPS in its entirety and that too using exhaustive set of training data samples is not feasible.

Accordingly, two solution approaches used in combination hold the most promise. First, surrogate modeling [5] as an alternative to expensive and exhaustively trained models can rapidly train models of the systems at lower cost. The surrogate models for CPS can be trained using the process models. Second, distributed (re)learning [3] of surrogate models at the individual subsystem level and composing them to form models of the larger CPS can scale the modeling activity, and promote reuse via transfer learning [18].

Although distributed (re)learning of surrogate models of CPS and their composition is an attractive approach, it is fraught with many technical challenges. This position paper provides an exposition of such research challenges and the research needed to realize the vision of edge-based, distributed surrogate model design and composition for CPS. The rest of the paper focuses on detailing these research challenges and solution needs.

2 RESEARCH NEEDS FOR COMPOSABLE, DATA-DRIVEN CPS

We present details of the research challenges. To better appreciate these challenges, we first present a motivating case study.

2.1 Motivating Case Study

Consider a large manufacturer, such as an auto maker, who must ensure the efficiency and reliability of its CPS enterprise comprising a large number of assets including the many different geographically distributed manufacturing and assembly plants, its showrooms and service centers, its deployed fleet, suppliers of parts, and the overall supply chain. These assets, which range from closed to open systems, may be owned and managed by the enterprise itself (e.g., the assembly plants), some independently owned (e.g., value added resellers and part manufacturers), some operated by franchises/family-owned businesses (e.g., showrooms and service centers), and some operated by public/private operators (e.g., state/federal transportation systems that manage roads, trucking companies that operate car transporters or carry parts to assembly plants or service centers, weather agencies that inform potential disruptions on routes, etc).

Each such asset produces significant amounts of data in different volumes and velocities. The enterprise must exploit this data in managing and controling its assets by developing models or digital twins, which means that **data analytics** is key to advancing the state of art. However, building a single, large enterprise-wide model or digital twin is hard and infeasible due to the scale of the CPS, the different functionalities and responsibilities of the individual assets, the different time scales at which these assets operate and must be controlled and managed, difficulty in ensuring a temporally consistent snapshot of the overall system operation, the ineffectiveness of controlling the dispersed, heterogeneous and often independently owned/managed assets from a centralized location, and the substantial scale of compute resources needed to train and execute such a digital twin.

These attributes call for the composition of models/digital twins. However, even at the level of each asset, the amount of data may be significant and the real-time control needs of these assets will require faster training of these models using novel machine learning-based approaches, which could be surrogate models. Developing such composable surrogate models will require distributed infrastructure ranging from in-house resources within each asset, which could even be Internet of Things (IoT) resources that we collectively refer to as edge computing resources all the way to traditional cloud computing resources.

Since the trained models cannot permanently remain accurate due to emergent properties of the CPS, different models may need retraining, and hence detecting the drift in models and retraining them must be done autonomously. Finally, since models are composed hierarchically across the different assets to define the enterprise model, and since these assets range from being closed to open, each asset can be vulnerable to different kinds of attacks, notably adversarial attacks, which are security issues concerning the correctness of the trained models, and hence ensuring appropriate defense mechanisms to guard against such attacks is critical.

2.2 Research Challenges

We now present the key research challenges that we have identified and situate these in a generic architecture shown in Figure 1.

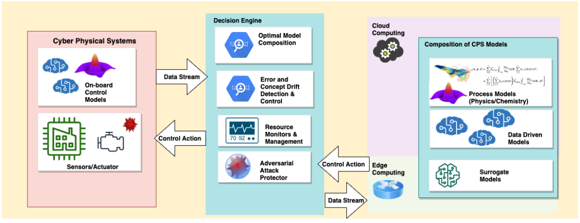

Figure 1: Architecture Overview and Research Challenges

2.2.1 *Learning surrogate models across the edge and cloud.*

High-fidelity models for physical or chemical phenomena of interest are often multi-disciplinary and multi-scale, and are computationally expensive. Using such models for various types of analyses that require many repeated runs (such as "what-if" analysis, sensitivity analysis, uncertainty quantification, reliability analysis, model calibration, system health diagnosis and prognosis, and design optimization) quickly becomes unaffordable. Computational speed is even more crucial when real-time or on-line decision making to enable optimal performance or operations of the system of interest is the main goal. Therefore, construction of inexpensive, accurate, and up-to-date surrogate models is central to improving the performance of cyber physical systems.

Many different techniques exist in building surrogate models, however, the challenge for CPS is to rapidly build surrogate models for different subsystems of CPS trading off granularity of the trained model versus its accuracy all while leveraging the limited resources the CPS subsystems (i.e, the edge) for shorter time-scale needs (i.e., lower fidelity models), and use the powerful cloud resources for decisions that can be taken over longer time-scales (higher fidelity models). A multi-fidelity approach, which aims to identify the optimal combination of low-fidelity and high-fidelity physics model runs to build the surrogate model, can be leveraged to achieve the balance between training effort and accuracy.

The problem of building such models involves two key challenges: a) development of improved surrogate model building algorithms (surrogate model form selection, training point selection to maximize information gain, analysis of surrogate model accuracy, adaptive improvement of the surrogate model, accounting for surrogate model error in prediction, etc.), and b) development of intelligent computational resource allocation algorithms for performing computations across the cloud and edge computing spectrum for activities such as physics model runs to collect training data, surrogate model training, and prediction with the surrogate model.

2.2.2 Composability of the surrogate models. Due to the multi-disciplinary nature of real-world CPS, composition of multiple process models or the corresponding surrogate models is inherent to these systems. Accordingly, individual, trained surrogate models of the different subsystems of CPS will need to be composed in some way to realize the model of the overall CPS. However, research on composition of data-driven AI/ML models is still in its infancy, which requires new ideas and formalisms to compose models to form larger models [11]. Most work that we found in the literature has focused on composition in the realm of human brain models[9] or in natural language processing (NLP) [11].

Yet, we believe that existing research in this space may help connect-the-dots between their findings and applicability to CPS, which provides new directions of research. For instance, although the five test criteria proposed in [11] are specific to linguistics, they appear generic enough to be researched in the context of CPS. These include (a) **systematicity,** i.e., whether the composition can recombine known parts and rules to form new sequences – for CPS, this could entail determining if combining subsystems using different network links and scheduling algorithms can give rise to a different but operational CPS; (b) **productivity,** i.e., whether the composition can predict beyond the length of the training data for individual models – for CPS this could mean testing the composed model for its ability to make predictions for emergent behaviors stemming from subsystem composition or workload changes, (c) **substitutivity,** i.e., whether the predictions of the composition is robust to synonym substitution – in CPS we could switch one scheduling algorithm with another and check for the composed model's robustness, (d) **localism,** i.e., whether the operations of the composition are local or global – for CPS this would indicate whether individual models execute on their own and their outcomes aggregated or whether the composed model makes the prediction at the global level, and (e) **overgeneralization,** i.e., whether the composition favors specific rules or exceptions – for CPS this could mean how the composed models handle *concept drifts* and what would it take to retrain based on newer evidence [23].

2.2.3 Handling error propagation and concept drift in composed models: Since surrogate models are approximate, errors can propagate and accumulate when the models are composed. Thus, it is critical to limit the error spread and improve the accuracy of the composed model. Moreover, since one or more of the trained models in the composition can incur different degrees of concept drift, it is important to be able to pinpoint the set of drifted models and determine which among them need to be re-learned without impacting existing control logic.

For any surrogate model, there are two main sources of error: a) the inability of the surrogate model ($\mathcal{S}_i$) to accurately approximate the governing process model ($\mathcal{P}_i$), and b) the inability of the process model ($\mathcal{P}_i$) to accurately describe the real-world phenomenon. The error introduced due to the first reason can be alleviated by adding training points from previously under-represented regions (also known as adversarial input) of the input space. For a cyber physical system, one approach to continual improvement could be by training the surrogate model in parallel with the system operations. That is, while the system is operating (and while the j-th version of a surrogate model ($\mathcal{S}_{i(j)}$) is being used to perform system optimization/control), execute the process model ($\mathcal{P}_i$) to obtain additional training data corresponding to adversarial inputs, re-train the surrogate model ($\mathcal{S}_i$), and update the version from j to $j + 1$ when the error has reduced sufficiently. The error introduced due to the second reason can possibly be alleviated by refining the process model ($\mathcal{P}_i$), or by incorporating model discrepancy in the model estimate.

Despite the above-proposed ideas, surrogate models are often inadequate to describe the data or the process model. Several cases may arise in this context, e.g., the case where the process model is available but complex to run, or the case where the process model is represented extensionally by data only. Irrespective, this results in two different problems that need resolution. The first one is lack of data to train the surrogate model, the second is the form of the surrogate model which might be inadequate. In the latter case, adding data points is ineffective and a different hypothesis space should instead be required.

2.2.4 Handling adversarial attacks: Data-driven techniques, particularly deep learning, have made deep strides in addressing a variety of CPS issues like health management and prognostics, e.g., power disturbance classification [25] and remaining useful life prediction [16], or in smart power grids where data from smart power meters is used by the deep neural network to forecast load and thereby inform effective power distribution [20]. However, recent research [2] shows that current machine learning algorithms proposed for such use cases can be vulnerable to adversarial attacks, which are small but specially designed modifications to normal data inputs that can adversely affect the quality of the machine-learned predictor [4]. For instance, with increasing IoT sensors in many CPS applications, an attacker could intercept and maliciously modify sensor readings to conduct such adversarial attacks, leading to critical damage caused due to inaccurate control decisions by the data-driven model. Surrogate models are no exception to this rule; in fact given their approximate nature, we surmise that they may be even more susceptible to such attacks.

The impact of adversarial attacks in deep learning [4, 21] has given rise to many concerns [7]. Prior research [26] has shown the threats to current machine learning systems. Most prior work in the field of adversarial machine learning has focused on classification tasks [1]. As regression tasks, such as power load forecasting, temperature forecast, remaining useful life prediction, crime prediction, traffic prediction, accident prediction and many others, start playing an increasingly important role in CPS use cases, the topic of adversarial regression is attracting research attention and is in need for solutions to defend against adversarial attacks.

CPS systems can vary from being *mostly closed* (e.g., an aircraft) to *semi open* (e.g., connected vehicles or a data center) to *mostly open* (e.g., IoT-based systems or global supply chain). Consequently, each category of CPS may be vulnerable to different kinds of attacks. Here, we focus on data-driven models that can be prone to adversarial attacks [10, 26] and hence must be protected against such attacks otherwise the control algorithms will behave incorrectly.

2.2.5 *Dynamic resource management.* Crosscutting the above research thrusts is the need to dynamically allocate and share resources across the spectrum of resources comprising the edge to the cloud for all the model (re)learning, composition, adversarial defense, and model inference tasks.

Deep learning model training is a resource and time intensive activity. One approach to speeding up training is to modify the internals of the training process. Modular Networks [14] views a deep neural network as being made up of a set of modules, and instead of activating each module in every layer per training sample, only a subset of the modules per layer are chosen stochastically by a controller, and the resulting output from the subset is concatenated or summed up and passed on to the subsequent layer.

While such a technique can certainly speed up modeling of individual subsystems of a CPS, composition of such models and their effectiveness still remains an unresolved issue including knowing when and how many resources are needed. Moreover, the work has been applied to image classification and language modeling examples but not concretely to solve any CPS problem.

Inferencing for large deep neural networks is expensive too, however, the use of surrogate models can help alleviate this challenge.

3 CONCLUSIONS

This position papers laid out key challenges and possible research directions in realizing robust CPS that are composed of data-driven, surrogate models of individual subsystems of the CPS.

REFERENCES

[1] Battista Biggio and Fabio Roli. 2018. Wild Patterns: Ten Years After the Rise of Adversarial Machine Learning. *Pattern Recognition* 84 (2018), 317–331.

[2] Yize Chen, Yushi Tan, and Baosen Zhang. 2019. Exploiting Vulnerabilities of Load Forecasting Through Adversarial Attacks. In *Proceedings of the Tenth ACM International Conference on Future Energy Systems.* ACM, 1–11.

[3] Jeffrey Dean, Greg Corrado, Rajat Monga, Kai Chen, Matthieu Devin, Mark Mao, Marc'aurelio Ranzato, Andrew Senior, Paul Tucker, Ke Yang, et al. 2012. Large Scale Distributed Deep Networks. In *Advances in neural information processing systems.* 1223–1231.

[4] Javier Echauz, Keith Kenemer, Sarfaraz Hussein, Jay Dhaliwal, Saurabh Shintre, Slawomir Grzonkowski, and Andrew Gardner. 2019. Adversarial Campaign Mitigation via ROC-Centric Prognostics. In *Proceedings of the Annual Conference of the PHM Society*, Vol. 11.

[5] Alexander Forrester, Andras Sobester, and Andy Keane. 2008. *Engineering Design via Surrogate Modelling: A Practical Guide.* John Wiley & Sons.

[6] Edward Glaessgen and David Stargel. 2012. The digital twin paradigm for future NASA and US Air Force vehicles. In *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference.*

[7] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. 2014. Explaining and Harnessing Adversarial Examples (2014). *arXiv preprint arXiv:1412.6572* (2014).

[8] Michael Grieves and John Vickers. 2017. Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In *Transdisciplinary perspectives on complex systems.* Springer, 85–113.

[9] Barbara Hammer. 2003. Compositionality in Neural Systems. *The handbook of brain theory and neural networks* (2003), 244–248.

[10] Ling Huang, Anthony D Joseph, Blaine Nelson, Benjamin IP Rubinstein, and J Doug Tygar. 2011. Adversarial Machine Learning. In *Proceedings of the 4th ACM workshop on Security and artificial intelligence.* 43–58.

[11] Dieuwke Hupkes, Verna Dankers, Mathijs Mul, and Elia Bruni. 2019. The compositionality of neural networks: integrating symbolism and connectionism. *arXiv preprint arXiv:1908.08351* (2019).

[12] Yuchen Jiang, Shen Yin, and Okyay Kaynak. 2018. Data-driven Monitoring and Safety Control of Industrial Cyber-Physical Systems: Basics and Beyond. *IEEE Access* 6 (2018), 47374–47384.

[13] Pranav M. Karve, Yulin Guo, Berkcan Kapusuzoglu, Sankaran Mahadevan, and Mulugeta A. Haile. 2020. Digital twin approach for damage-tolerant mission planning under uncertainty. *Engineering Fracture Mechanics* 225 (2020), 106766. https://doi.org/10.1016/j.engfracmech.2019.106766

[14] Louis Kirsch, Julius Kunze, and David Barber. 2018. Modular Networks: Learning to Decompose Neural Computation. In *Advances in Neural Information Processing Systems (NIPS).* 2408–2418.

[15] Jay Lee, Behrad Bagheri, and Hung-An Kao. 2015. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters* 3 (2015), 18 – 23. https://doi.org/10.1016/j.mfglet.2014.12.001

[16] Xiang Li, Qian Ding, and Jian-Qiao Sun. 2018. Remaining Useful Life Estimation in Prognostics Using Deep Convolution Neural Networks. *Reliability Engineering & System Safety* 172 (2018), 1–11.

[17] Oliver Niggemann, Gautam Biswas, John S Kinnebrew, Hamed Khorasgani, Sören Volgmann, and Andreas Bunte. 2015. Data-Driven Monitoring of Cyber-Physical Systems Leveraging on Big Data and the Internet-of-Things for Diagnosis and Control. In *DX@ Safeprocess.* 185–192.

[18] Sinno Jialin Pan and Qiang Yang. 2009. A Survey on Transfer Learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2009), 1345–1359.

[19] Mahadev Satyanarayanan. 2017. The Emergence of Edge Computing. *Computer* 50, 1 (2017), 30–39.

[20] Raffi Sevlian and Ram Rajagopal. 2018. A Scaling Law for Short Term Load Forecasting on Varying Levels of Aggregation. *International Journal of Electrical Power & Energy Systems* 98 (2018), 350–361.

[21] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. 2013. Intriguing Properties of Neural Networks. *arXiv preprint arXiv:1312.6199* (2013).

[22] Fei Tao, Jiangfeng Cheng, Qinglin Qi, Meng Zhang, He Zhang, and Fangyuan Sui. 2018. Digital Twin-driven Product Design, Manufacturing and Service with Big Data. *The International Journal of Advanced Manufacturing Technology* 94, 9-12 (2018), 3563–3576.

[23] Huangshi Tian, Minchen Yu, and Wei Wang. 2018. Continuum: A Platform for Cost-Aware, Low-Latency Continual Learning. In *Proceedings of the ACM Symposium on Cloud Computing (SoCC).* 26–40.

[24] Alexey Tsymbal. 2004. The Problem of Concept Drift: Definitions and Related Work. *Computer Science Department, Trinity College Dublin* 106, 2 (2004), 58.

[25] Martin Valtierra-Rodriguez, Rene de Jesus Romero-Troncoso, Roque Alfredo Osornio-Rios, and Arturo Garcia-Perez. 2013. Detection and Classification of Single and Combined Power Quality Disturbances Using Neural Networks. *IEEE Transactions on Industrial Electronics* 61, 5 (2013), 2473–2482.

[26] Yevgeniy Vorobeychik and Murat Kantarcioglu. 2018. Adversarial Machine Learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 12, 3 (2018), 1–169.