Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Parv Venkitasubramaniam

Jeff Heflin

Paolo Bocchini

Jessecae Marsh

Ted Ralphs

Rick Blum

Brian Davison

Response to Networking and Information Technology Research and Development Request for Information on a National Plan for Cyber-Physical Systems Resilience (89 FR 78915)

Author Names: Parv Venkitasubramaniam, Jeff Heflin, Paolo Bocchini, Jessecae Marsh, Ted Ralphs, Rick Blum, Brian Davison.
Institution: Lehigh University
Statement: This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Cyber-Physical Systems Resilience RD Strategic Plan and associated documents without attribution

Explainable AI for Disaster Response

Despite vast amounts of data available, real-time data-driven decision making during extreme events such as natural disasters or large-scale cyber-attacks continues to be challenging, for multiple reasons. First, extreme events are infrequent and each event is unique. Available data overwhelmingly pertain to normal conditions, or at best to other real or simulated events; computational models trained using such data are not reliable under disaster conditions that are previously unseen by the models [8, 12]. Second, as a disaster unfolds, the state of infrastructure and underlying user behavior evolve rapidly [11]. Decision-making tools need to continually adapt to changing physical conditions and data patterns, without requiring high real-time computation. Third, emergency response actions during disasters are taken by human operators who bear professional liability for their decisions [6]. Any data-driven tool and outcome should be explainable to the operators, to gain their trust. Without such tools, emergency responders during a disaster often rely on experience or on pre-engineered solutions with limited usage of real-time data. Faster emergency response can not only lower the human costs of disasters [1, 10] but also the financial costs [2]. Addressing the aforementioned challenges requires fundamental research advances that cross-cuts multiple disciplines that include infrastructure modeling and disaster management, data science and machine learning, knowledge representation, efficient and explainable computation and cognitive psychology.

In this regard, there are four major areas of research that we propose:

Reliable prediction in unseen and unpredictable conditions: A major challenge in ML, referred to as out-of-distribution generalization, is to develop models that provide reliable outcomes when test data patterns are previously unseen during the training phase. Despite various approaches designed for ML in research areas such as vision [4, 13, 15] and text [14], addressing the challenge in dynamical and cyber-physical systems remains open. Most existing models for time-series prediction assume data are independent and identically distributed [3,5,9]. Newer machine learning approaches and computational architectures that adhere to underlying physical laws are required to improve prediction performance under conditions significantly different from those encountered in training data. Approaches that can deal with rapidly changing conditions and missing or false data as is common under extreme scenarios are crucial to enable robust and reliable prediction under these scenarios.

Adaptive Computation under rapidly evolving conditions: Computational algorithms for decision making in cyber-physical systems are typically designed based on static or predictable system states, and are not trained to include information such as impending risks of damage from an evolving disaster or a cyberattack. Modern catastrophe modeling approaches offer the opportunity to assess in near-real-time the probability of loss of functionality of system components and processes, and these measures along with the predicted system states must be used in performing the needed computations and optimized decision-making in an evolving disaster. Fast optimization and adaptive computational methods need to be developed to apply in disaster scenarios which can quickly adapt to changing conditions of the system and environment.

Explanations for computational outcomes: From a cognitive psychological perspective, very little is known about what kind of explanations for data-driven models can gain trust of actual users of the methods and outcomes. Explainable AI methodologies that have been developed, particularly for machine learning solutions, are focused on "simplifying" the computational process or outcomes rather than considering a holistic view of the user in the context of the application [7]. Different kinds of explanations have been proposed for computational models such as feature-importance, model interpretation, example-based, etc., are limited to generic ML models, and not tailored to domain-specific conditions and constraints, and additionally, the approaches are not adaptable to changing system or environmental conditions. Rather than a computation-only approach, what is required is a holistic framework to derive explanations that are computationally accurate, contextually interpretable, and guided by a use-inspired framework of human perception and decision-making. This requires bringing together concepts and tools from knowledge representation, cognitive psychology, data science, optimization and machine learning, to build explainable representations of domain knowledge, derive rules to contextualize the outcomes of data-driven decisions, and use cognitive psychological studies to formally understand the perception and use of explanations during extreme events to help formulate explanations that will be adopted and used.

Human-in-the-loop Testing Platform Given the challenges with real-time disaster testing of developed methods, there is a need to build a human-in-the-loop experimental platform that can simulate disaster or cyberattack scenarios to generate data on the real-time impact on infrastructural systems, as a function of the unfolding disaster as well as the human decision making in its aftermath. Such an experimental platform should have the ability to provide risk analysis on the evolving disaster scenario and the time-varying state of the infrastructure. The testing platform should provide the necessary feedback to validate the methodological outcomes as well as the validity and usability of the explanations by human operators.

References

[1] Saman Ahmadi, Guido Tack, Daniel D Harabor, and Philip Kilby. A fast exact algorithm for the resource constrained shortest path problem. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12217–12224, 2021.

[2] Daniel Brent and Louis-Philippe Beland. Traffic congestion, transportation policies, and the performance of first responders. *Journal of Environmental Economics and Management*, 103:102339, 2020.

[3] Mark Granroth-Wilding and Stephen Clark. What happens next? event prediction using a compositional neural network model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

[4] Dan Hendrycks, Steven Basart, Norman Mu, Saurav Kadavath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli, Mike Guo, et al. The many faces of robustness: A critical analysis of out-of-distribution generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8340–8349, 2021.

[5] Balázs Hidasi and Alexandros Karatzoglou. Recurrent neural networks with top-k gains for session-based recommendations. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pages 843–852, 2018.

[6] Nan D Hunter. *The law of emergencies: public health and disaster management*. Butterworth-Heinemann, 2017.

[7] Uday Kamath and John Liu. *Explainable artificial intelligence: An introduction to interpretable machine learning*. Springer, 2021.

[8] Vaishnavh Nagarajan, Anders Andreassen, and Behnam Neyshabur. Understanding the failure modes of out-of-distribution generalization. In *International Conference on Learning Representations*, 2020.

[9] Zhi Qiao, Shiwan Zhao, Cao Xiao, Xiang Li, Yong Qin, and Fei Wang. Pairwise-ranking based collaborative recurrent neural networks for clinical event prediction. In *Proceedings of the twenty-seventh international joint conference on artificial intelligence*, 2018.

[10] Rocío Sánchez-Mangas, Antonio García-Ferrrer, Aranzazu De Juan, and Antonio Martín Arroyo. The probability of death in road traffic accidents. how important is a quick medical response? *Accident Analysis & Prevention*, 42(4):1048–1056, 2010.

[11] Wenjuan Sun, Paolo Bocchini, and Brian D Davison. Applications of artificial intelligence for disaster management. *Natural Hazards*, 103(3):2631–2689, 2020.

[12] Yue Sun, Chao Chen, Yuesheng Xu, Sihong Xie, Rick S Blum, and Parv Venkitasubramaniam. On the generalization discrepancy of spatiotemporal dynamics-informed graph convolutional networks. *Frontiers in Mechanical Engineering*, 10:1397131, 2024.

[13] Florian Wenzel, Andrea Dittadi, Peter Gehler, Carl-Johann Simon-Gabriel, Max Horn, Dominik Zietlow, David Kernert, Chris Russell, Thomas Brox, Bernt Schiele, et al. Assaying out-of-distribution generalization in transfer learning. *Advances in Neural Information Processing Systems*, 35:7181–7198, 2022.

[14] Linyi Yang, Yaoxiao Song, Xuan Ren, Chenyang Lyu, Yidong Wang, Lingqiao Liu, Jindong Wang, Jennifer Foster, and Yue Zhang. Out-of-distribution generalization in text classification: Past, present, and future, 2023.

[15] Chongzhi Zhang, Mingyuan Zhang, Shanghang Zhang, Daisheng Jin, Qiang Zhou, Zhongang Cai, Haiyu Zhao, Xianglong Liu, and Ziwei Liu. Delving deep into the generalization of vision transformers under distribution shifts. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 7277–7286, 2022.