Seminar Talk: "Towards Safe and Trustworthy Cyber-Physical Systems" (Speaker: Dr. Lu Feng)

Matthew Whitesides

Abstract

In today's presentation Dr. Lu Feng, an Assistant Professor of Computer Science at the University of Virginia discuesses the implementation and saftey concers when implementing Cyber Physical Systems. Her research focuses on assuring the safety and trustworthiness of cyber-physical systems, with applications ranging from medical devices, to autonomous robots, to smart cities.

I. Introduction and Background

CYBER-PHYSICAL Systems (CPS) are becoming an ever-increasing part of modern society (CPS). A CPS is any computer system in which a mechanism is controlled or monitored by computer-based algorithms that interact with the natural world (as opposed to purely virtual systems). These systems can range from simple Internet of Things (IoT) smart home devices such as thermostat controls to fully autonomous self-driving cars. Security in these systems is paramount as attackers gain access to sensitive information. They potentially can use these systems to inflict real-world damage. One prominent example includes the landmark case where researchers showed that certain connected vehicles could be remotely hacked and the power breaks disabled. Another implication in securing a CPS is that the attacker can physically and virtually intercept the system, adding another layer of consideration when designing these systems.

Three significant security factors to consider in CPS systems include human interaction, AI-enabled decision making, and the large scale of the systems. These all have specific safety concerns, failure points, and aspects of required trust.

One example includes modern medical treatment in both the home and hospital settings which require the involvement of multiple medical devices helping the doctor achieve the treatment goals. These devices have become increasingly connected IoT devices that communicate information from sensors to the healthcare's IT infrastructure. Security concerns can occur when these devices are not interoperable among different vendors or IT systems. The data transferred relies on the security of the weakest individual system.

Traditional reinforcement learning has an agent supplied with some action in an environment that will give you observation and the rewards based upon those actions. On the other hand, multi-agent to reinforcement learning is similar to reinforcement learning, but now you have multiple agents who have to cooperate to achieve optimal results. This complexity makes learning with a profile necessary for many safety-critical applications. Essentially you do not want the learning phase of these CPS systems to cause any real-world disruptions, so reinforcement learning must be done with safety parameters in place.

II. Research Contributions and Results

A. Atonomous Vehicle Handover

One of the first studies performed was to estimate the trust dynamics in autonomous vehicles in scenarios where the user may have to intervene. The example included when a vehicle approaches an incident in the road, and the vehicle needs to decide if it can handle the situation or give control over to the user. Various factors include is a pedestrian involved, what vehicles are involved in the incident, and how to handle it after the incident has been passed. They then formed a partially observable Markov decision process (POMDP) model for this scenario that involves the system's trust, capability, incident, and location among various incidents in a route. When gathering the model data, a simulator was set up to safely put users in this scenario and track how they and the system interact.

These factors are vital in deciding if the system is in a scenario to hand over the control to the user, but the models must also predict what the user will do if they receive control. This scenario is another full prediction model to consider, from the vehicle's physical situation to the user's biometric patterns, reaction time, and ability to take over. These are all valuable data points to ensure the user is able and willing to take over the situation or if the vehicle would have a better chance of success staying in control. This situation is a fascinating one to consider. Most autonomous vehicle proposals think of safety only considering the AI or human driver, not the interactions between the two.

Using labels for the various data labels from the human participants and autonomous simulations, they collected, pre-processed, and then built a deep neural network model. This model predicts the takeover intention time and the quality of driving after the takeover. This data is based on a deep neural network they call Deep Tech Framework Prediction. They compared the prediction accuracy and F1 scores with six other classical machine learning models. Their Deep Tech Framework showed much better prediction accuracy than the six different models alone. Ultimately, this allows the vehicles to make optimal decisions based on the predictive driver's takeover.

B. Shield Synthesis

How can we guarantee the safety of training models in the training and testing phase, so not only the final model but the learning models ensure safety and accuracy to keep the CPS and outside environment safe? One solution involves Sheild Synthesis essentially is a safety game applied in the training phase which introduces rewards and punishments based upon predefined safety concerns that may come up in the environment. To simulate these safe and unsafe states, they took inspiration from path tracing models that establish unsafe locations and scenarios. When the training gets closer to an unsafe location, the heuristic applies a punishment and vice versa for safe situations.

C. Preditive Modeling Uncertinty

The following research revolves around predictive monitoring with logic-calibrated uncertainty. Like in the study done above, traditional predictive monitoring can have promising results on decision-making for a CPS. However, when making subsequent decisions or far-reaching ones, more significant uncertainty and difficulty are created. Dr. Lu Feng and the team then created a new approach to monitoring long-form sequential prediction data. They used data from air quality sensors in a large area in China with a lot of inherent uncertainty due to human interaction, differing sensor quality, and vast scale that causes a lot of variations. Using a typical model based upon historical data can lead to incorrect decision-making, so instead of predicting the air quality in an area precisely, they are predicting the uncertainty of the prediction and can give an accurate confidence level to the estimated air quality.

III. Lessons Learned

As the world becomes increasingly reliant on smart systems, AI, and machine-learned systems, we must consider the safety risks involved in handing over control to these systems. Even if, statistically, a situation is safer in the hands of an autonomous system, it seems worse when things go wrong. Not only in cases that are apparent dangers like autonomous vehicles but ones that have far-reaching but unseen impacts like smart infrastructure in cities.

The work Dr. Lu Feng and the team, have done is very interesting. They are looking one step beyond the apparent building CPS ML models but looking at how multiple agents interact, how these systems can be safe while in the learning phases, and how to predict beyond the obvious next step and into subsequent actions. It was fascinating the work done on how an autonomous vehicle can make better decisions on if the human driver or the vehicle can safely handle a given situation. This consideration is something I've never thought of and will be a crucial factor in autonomous vehicle safety. So far, most autonomous accidents have been due to humans not paying attention or the vehicle misreading the situation. Minimizing those situations and making better choices per driver when they need to intervene will help the adoption and trust of driverless vehicles.

IV. Conclusion

Overall the work, Dr. Lu Feng and the team are researching is inspiring. There is so much to consider when dealing with CPS and intelligent systems. Dr. Lu Feng is thinking beyond the obvious and tackling some of the more complex CPS decision-making and safety challenges of implementing these systems. Just the idea that you need to consider safety in the training model and how to deal with uncertainty in subsequent decisions is stepping beyond what I would think of when dealing with implementing or researching CPS. It's good to know research like the ones done by Dr. Lu Feng and the team is being done, especially as CPS and predictive models collide more and more with our daily lives.

Acknowledgment

The author would like to thank Professor Sajal Das with the Department of Computer Science, Missouri University of Science and Technology and Dr. Lu Feng with the University of Virginia.

Matthew Whitesides Master's Student at Missouri University of Science and Technology.