Safety Assurance and Managing Risk in Automated Driving

Workshop at SCSSS 2018
Stockholm, Sweden

Dealing with Risk and Uncertainty

Masoumeh Parseh/Fredrik Asplund/Sofia Cassel

Dealing with Risk and Uncertainty
(Introduction to the questions)

- Safety-critical applications are currently designed to, with certainty, lower the risk (probability of harm) below an acceptable threshold, a strategy that depends on minimizing uncertainty during operations.
- New strategies focus on dealing with uncertainty, which makes it difficult to appreciate the risk associated with a system. One example is the use of machine learning to identify objects in an environment, which cannot be proven to always work as intended.

Pre-defined Questions

- Should we allow non-deterministic methods to be used in safety-critical applications?
- If so, how do we calculate the associated risk?
- If so, do we accept certain types of risk, i.e. related to certain parts of the environment or design assumptions? Or, vice versa, do we only accept these methods in certain use cases or certain industrial domains?

Should we allow non-deterministic methods to be used in safety-critical applications?

- Nondeterministic methods are going to be used anyway, because they are cost-efficient (multi-threading etc.).
- Modern computers are inherently non-deterministic (unless single-threaded on single chip etc.). Eventually, we will need to handel them.
- People working in the aerospace believe that they already reached the limits of using deterministic methods for current systems. This means such systems will definitely be used in the automotive industry since they will have more complicated systems.

Should we allow non-deterministic methods to be used in safety-critical applications?

- Some argue that we should not even use algorithms / methods until we know exactly how they work. Some consensus on using monitoring methods (supervisory system) to help with uncertainty.
- Some argue that we have to use the nondeterministic methods, but the available data today is uncertain itself. So, maybe in 10 years with better data we are in a better place.
- Maybe not, and degrade performance for safety. But, we might have cars that drive very cautiously. So we have to make a trade-off.
- We have to use these methods sooner or later but the data is uncertain itself, so maybe we are not quite there yet.

Should we allow non-deterministic methods to be used in safety-critical applications?

- It is not a binary answer of yes or no. We start small and go step by step. We use what we understand today and not use what we do not understand.
- Predicting the error probability in safety critical software. Automating the okay/not okay decisions if you find a certain threshold. The assumption is that the code and the system has been deterministic.

“How can you supervise a complex system with a simple system?”
(asked by the participants)

-Supervising a nondeterministic system with a simple deterministic system-You put deterministic rules (clear rules) on your deterministic systems.

-There is a suggestion to use hard rules to regulate the boundaries of the nondeterminism, e.g., "if you see something unknown, you must stop" will overrule any underlying decision by a neural network. We can also always trade performance for increased safety --> reducing uncertainty.

If so, how do we calculate the associated risk?

- The questions led to other questions:
- The data that we use to train the module with it, how to make sure that the data is correct?
- Do we really need to quantify? Maybe, training the neural network in proper ways with proper and sufficient data. We need to build procedures around how to build it and how to train it.
- There was a discussion about neural networks and the quality of training data: can you certify your self-driving car for Beijing? How do you train it for that? How do we train AI algorithms so they can adapt and handle unforeseeable situations? We discussed corner cases and how to handle them.
- The training set would be different in different countries?

If so, how do we calculate the associated risk?

- Not with the traditional methods.
- The methods to calculate risk are non-deterministic themselves.
- New methods on how to simulate uncertainty e.g. Markov Chain Monte Carlo(MCMC). It could provide a hint of the risk.
- Be e.g. ten times better than the best drivers performance in dangerous situations.
- Making sure that the data used for training is high quality data.
- Using statistical data but it is more like building confidence than calculating risk.

If so, do we accept certain types of risk, i.e. related to certain parts of the environment or design assumptions? Or, vice versa, do we only accept these methods in certain use cases or certain industrial domains?

- Nondeterministic methods will mostly be used in high volume industries, e.g. cars and medical devices, but not in nuclear power plants.
- This also depends on what is the acceptable level of risk. For example, the fatality in the automotive industry is huge today. So, if we can just do a little better than that, we can accept high risk from the systems. However, in the nuclear industry, we have a very low risk in the current systems. If we want to introduce new systems/methods, we have to provide a lot of assurance. But, that is not a necessary case in the car industry since the drivers are bad anyway.

If so, do we accept certain types of risk, i.e. related to certain parts of the environment or design assumptions? Or, vice versa, do we only accept these methods in certain use cases or certain industrial domains?

- We accept a certain level of risk, and there is never zero risk. And all technical systems are associated with some level of it, but what this level is, changes over time (peoples' perception change.)
- Swedish defence wanted to implement a method where everything is factor of "the risk of normal everyday life". You accept a certain level of level risk and there is always a level of risk associated with the technical systems but what that level
- Societal acceptance is crucial -- if we are afraid of pilotless planes and do not trust them, then people will not go in them. The same goes for autonomous vehicles. It might be easier to accept nondeterministic methods in certain limited settings (for example, where people are not present).