Budapesti Műszaki és Gazdaságtudományi Egyetem
Méréstechnika és Információs Rendszerek Tanszék

Evaluating tests on neural networks

Szántó Tamás I. évf., (MSc) mérnökinformatikus szakos hallgató
Konzulens: dr. Micskei Zoltán, MIT
Kritikus rendszerek főspecializáció
Önálló laboratórium 1. összefoglaló
2017/18. II. félév

As autonomous driving systems are evolving, partly and fully driverless vehicles are starting to appear on public roads and gaining more and more publicity. In my opinion, there is a great potential for self-driving cars to improve road safety and to change people lifestyle by enabling to utilize their time spent on driving. Of course, this technology comes with great responsibility too to ensure the safety of the passengers and environment of the vehicle.

Waymo is a self-driving car company, started by Google, which recently released a safety report about their testing and validation method. The first part is the base vehicle safety, the next part is the self-driving hardware testing, mostly it means that we want to ensure that the vehicle and the sensors are working perfectly. The interesting part is testing and verifying the self-driving software, because of the used machine learning techniques like deep neural networks we do not have a source code like direct access to the running mechanisms.

The main problem in the testing mechanism is that the core of the self-driving software is some kind of deep neural network, that mostly can be represented only as a black-box. For testing, the result is just a passed or failed, maybe some levels between them, but it is challenging to generate actual precise metrics.

What are the reasons behind the output? This is the main question about our results with most of the testing systems. Is it possible that we were just lucky, the system is passed the test, but based on the wrong reasons. Another problem is identifying the source of a bad decision because we have no information what happening on the inside of the system. In general, the question is how good the system's prediction is.

Recently appeared tools (LIME, SHAP) offer interpretation for the predictions of Deep Neural Networks. The goal is gaining confidence about the results by revealing how the input features contributed to the final prediction.

My goal was to find out how can a prediction explanation system improve the test evaluation process. In order to create an actual working system that can demonstrate this method and the possibilities in it, we need to simplify the problem. In summary, the first step is building the test images from parts then running an image classifier and then using the prediction explanation and the (also generated) masked image for the test evaluation.

My test case is that a taxi passes through the front of a bus in a crossroads and the goal is to examine the bus and taxi predictions of the classifier. Basically, with a given step interval the taxi goes through the scene, first without the bus (for reference) and then with a bus, in front of different backgrounds. I created several metrics for the evaluation based on the pixels that have a positive impact on the prediction. These metrics could successfully reveal more in-depth differences between the precision of the predictions.