technische universität dortmund

computer science 12

Bachelor/Master Thesis

Framework for Evaluating the Trustworthiness of Anomaly Detection in Autonomous Vehicles

Dr. Ching-Chi Lin
Prof. Dr. Jian-Jia Chen

Otto-Hahn Str. 16
Technische Universität Dortmund
Email: chingchi.lin@tu-dortmund.de
September 29, 2022

Different machine learning applications have been widely applied and attached to our daily life. Most of these applications are based on complicated neural network models trained with previously existing datasets. The interpretability of these models and accuracy of the outputs cannot be guaranteed. As an user, we can only determine if we trust and rely on the output results.

It is important that the performance of a machine learning application aligns with users expectations. An application can be *misused* if the quality of the outputs is low but we use it anyway because we put too much trust on the application. On the other hand, an application with a nice output quality can be *underused* due to the lack of trust from the users.

Before aligning the actual performance of a machine learning application with trust from the users, a framework for estimating trust is required. Since trust is a very vague and user-dependent idea, currently we rely on the direct feedback from the user, using methods such as questionnaires. A framework that can systematically adjust different control factors and collect user feedback accordingly would help us to develop the relationship between performance and trust more efficiently.

Our target application in this thesis is the anomaly detector on an autonomous vehicle. Anomaly detection [3] is a technique that detects potential problems from the data streams generated by the sensors. We consider the scenario that once an anomaly on the vehicle is detected by the detector, the corresponding warning lights on the car dashboard are turned on, and the driver will be asked to manually takeover the car if necessary.

To evaluate users trust to an anomaly detector, users will first be asked to observe several simulations on a driving simulator. During the simulation, there might be some observable anomalies such as drifting from time to time, and the warning lights on the car dashboard may (or may not) light up. After all the simulations, the user will be asked to gives a score of how likely he/she trusts the anomaly detector based on these simulations.

In this thesis[1], the objective is design and build a framework for evaluating the trustworthiness of an anomaly detector based on CARLA [1, 2], an open-source simulator for autonomous driving. The student should first get familiar with the CARLA simulator. After that, the student is expected to generate different anomalies during the simulated driving, and implement the warning lights indicating the corresponding anomalies. Note that the warning lights **do not** have to be turned on every time an anomaly is detected, but following a certain probability distribution, leading to *false positives* and *false negatives*. Finally, a feedback sub-system for evaluating the trust from users, and the analysis on the relation between the trust and the performance, i.e., the accuracy of the warning lights, should also be carried out. Students should note that, the involved source code this thesis will be publicly released and should be fully documented to comply the rationale of open-source software development.

Required Skills:

- Knowledgeable of C and C++ programming
- Basic knowledge in statistic and probability

Acquired Skills after the thesis:

- Knowledge about the state-of-the-art anomaly detection techniques in autonomous vehicles.
- Knowledge about evaluating / quantizing human trust to an AI system.

References:

[1] Dosovitskiy, Alexey, et al. “CARLA: An open urban driving simulator.” Conference on robot learning. PMLR, 2017.

[2] Shah, Tarang, et al. ”A Simulation-Based Benchmark for Behavioral Anomaly Detection in Autonomous Vehicles.” 2021 IEEE International Intelligent Transportation Systems Conference (ITSC). IEEE, 2021.

[3] Chandola, Varun, Arindam Banerjee, and Vipin Kumar. ”Anomaly detection: A survey.” ACM computing surveys (CSUR) 41.3 (2009): 1-58.

[1] Other suggestions and related topics are also welcome. Please do not hesitate to make an appointment.