Introduction to Accountability, Evaluation and Obscurity of AI Algorithms Minitrack

Radmila Juric
University of South East Norway
Kongsberg, Norway
Radmila.Juric@usn.no

Robert Steele
Capitol Technology University
Laurel, MD 20708, USA
rjsteele@captechu.edu

This Minitrack has four interesting papers which look at the AI algorithms from different perspectives. We read about the problem of ignoring small data sets for algorithms, which dominate within predictive and learning technologies; we gain insights into the characteristics and rate of machine learning-based malware detection performance deterioration and training set selections; we follow the development of an adversarial training based ML approach for malware classification, under adversarial conditions, and read about undisputed bias of Google Maps, which systematically underestimates necessary car driving time, and thus has an impact on users' choice of transportation.

The paper entitled "How Useful are Hand-crafted Data? Making Cases for Anomaly Detection Methods" focuses on the importance of small data sets, which have not been widely adopted as a necessity in current machine learning or data mining research. The authors state that our current perception, that the more complex data we use in machine learning, the better its performance might be, is detrimental to interpretability, explainability, and the sustained development of the AI field. Therefore they run several classical anomaly detection methods against small, mindfully crafted cases on which the results are examined in detail. To their surprise, this has led to the discovery of some novel uses of classical anomaly detection methods.

The paper entitled "The Effect of Training Set Timeframe on Future Performance of Machine Learning-based Malware Detection Models" looks at the relative rate of performance deterioration of various machine learning-based malware detection models after training. In their study, a range of machine learning models were applied to the features extracted from a large collection of software executables in Portable Executable format, ordered by the date the binary was first encountered, consisting of both malware and benign examples, whilst considering different training set configurations and timeframes. The paper offers the analysis and quantification of the relative performance deterioration of these machine learning models on future test sets of these features, and discusses insights into the characteristics and rate of machine learning-based malware detection performance deterioration and training set selection.

The paper entitled "An Adversarial Training Based Machine Learning Approach to Malware Classification under Adversarial Conditions" focuses on the role of ML in the classification of malware in cybersecurity, in order to deal with adversaries, which target the underlying data and/or models responsible for the functionality of malware classification and map its behavior or corrupt its functionality. The authors developed an adversarial training based ML approach for malware classification, under adversarial conditions, that leverages a stacking ensemble method, which compares the performance of 10 base ML models when adversarially trained on three data sets of varying data perturbation schemes. Their work reveals that a malware classifier can be developed to account for potential forms of training data perturbation with minimal effect on performance.

The piper entitled "Bias in Geographic Information Systems: The Case of Google Maps" examines predictions generated by Google maps, when providing relevant travel information, on user's perception and travel choices. The authors found out that Google Maps systematically underestimated car driving time, which has an impact on users' choice of transportation. Therefore this is a useful case study on the limitation of geographic information systems in urban environments, which in turn affects user behavior when it comes to choosing driving over other forms of transport. The authors added that, due to the habitual nature of choices of transportation, the effects of their research on decision-making are likely to be particularly strong on multi-modal transport users, who are able to choose between multiple different forms of transport.

URI: https://hdl.handle.net/10125/70713
978-0-9981331-4-0