Introduction to Accountability and Evaluation of AI Algorithms Minitrack

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

Robert Steel
Florida Polytechnic University
USA
rsteele@floridapoly.edu

The minitrack on the Accountability and Evaluation of AI Algorithms is a new within the track on Decision Analytics and Service Science. This is a rather novel but important topic which concerns many of us, but it did not attract too many papers, for many reasons. One of them is that the problem of AI accountability still does not appear as the mainstream research topic across our academic institutions, and consequently, we could not find academically viable research results which could crack this quite difficult problem.

There are only two papers accepted in our session. They are different, they do not directly resolve the problem of AI evaluation and accountability, but they address issues which can have an impact on the quality and reliability of algorithms we run behind the spectrum of AI computational models.

The paper entitled "A comparison between a two-feedback control loop and a reinforcement learning algorithm for compliant low-cost series elastic actuators" proposes a novel type of a controller for an elastic actuator which uses an artificial neural network, trained with reinforcement learning. The authors claim that there is significance in using knowledge generated through machine learning procedures and instantly create properties of the elastic actuator to be controlled. This in turn reduces the need of manually defining fuzzy rues for controlling actuators. Trustworthy and reliable results from such AI algorithms, in safety critical applications, could be obtain by building more robust ML systems, particularly in the presence of hardware level faults. A non-biased comparison between AI and classical two-loop control systems does give indications that with reinforcement learning, ANN algorithm might excel, particularly if the model of the system is difficult or impossible to obtain.

The paper entitled "Reliability of Training Data Sets for ML Classifiers: a Lesson Learned from Mechanical Engineering" gives an insight into the role and responsibility of the semantic of the data set in the functioning of an ML classifier. It shows an example from mechanical engineering. The calculation of adhesion and friction, which appear between hard surfaces, should be sufficient for building an ML classifier to predict slippery conditions in rail transport. However, the paper outlines that the ML classifier definition is not directly responsible for achieving reliable results in predictions. It is that our attitude, in the name of data science practices, towards changing the semantic of the training data set, in order to fit prediction algorithms, represent a real danger.

HICSS