Towards Verification of Changes in Dynamic Machine Learning Models using Deep Ensemble Anomaly Detection

Tim Katzke[1], Bin Li[1], Simon Klüttermann[1], and Emmanuel Müller[1]

TU Dortmund, Dortmund, Germany
{tim.katzke,bin.li,simon.kluettermann,emmanuel.mueller}@cs.tu-dortmund.de

Abstract. Formal verification of machine learning models is of major importance for safety critical systems. However, in case of dynamically changing data distributions, such verification has to consider the temporal representation of state changes as the context of a dynamic machine learning model. For instance, one has to verify a time-dependent machine learning model w.r.t. distinct states of an evolving time series. As an illustration, environmental time series encompassing variables such as temperature, humidity, illumination are used to train such models w.r.t. seasonal states like spring, summer, autumn and winter.
In general, such machine learning models are trained under so-called concept drift and have been developed to capture changing data distributions with internal representation learning. For formal verification, however, this implicit representation lacks explicit delineation of states, descriptions of transitions, and their human-understandable interpretability. Hence, we aim to extend powerful neural network verification to dynamic machine learning models trained under dynamic data changes. We consider the formal representation of a minimal state transition model as the first open challenge for an interpretable as well as efficient verification of such time-dependent machine learning models.
This presentation presents work in progress on defining a novel co-training procedure that uses our deep ensemble anomaly detection (DEAN-TS) algorithm to detect state changes in an unsupervised fashion and learn a finite-state automaton in parallel to the training of any state-of-the-art machine learning model under concept drift. In a naïve setup, one could simply create one new state for each newly detected concept in the data. However, as we aim at an interpretable as well as efficient verification, we optimize for a minimal automaton by incrementally merging similar concepts/states. In particular, our novel DEAN-TS algorithm provides us with an un-likelihood of erroneous states by exploiting the unique characteristics of different anomaly types vs. normal data. Using feature bagging, each ensemble submodel considers a subset of time series components to identify characteristic feature importance distributions. This unique approach allows us to describe each state in our automaton with the most important features and consequently provide human users with the ability to add corresponding semantics.

Keywords: Learning · Finite-State Automaton · Verification · Time Series · Change Detection · Deep Ensemble · Anomaly Detection