Abstract

A Novel Outlier Detection Method Exploiting the Incomplete Data Handling Capability of Artificial Neural Networks

Hamdi Fatma

This thesis introduces a novel algorithm for finding outliers in datasets, the IDOD: Incomplete Data based Outlier Detection algorithm. The core idea of the proposed solution is to find outliers through overwriting them as missing data values, using HIDD a neural network based algorithm created for handling missing data.

During its creation it was found to be significantly different from other novel outlier detection methods.

The IDOD algorithm, unlike other state-of-the-art methods is **assignment dependent** and furthermore while the overwhelming majority of algorithms consider an outlier as a complete data vector, the proposed method takes on a more refined approach and considers it instead as a component.

This can be seen as an advantage, in the fact that it can give a deeper insight into the dataset and the origin of the outlier. Or it can be considered a drawback, since it is significantly different from most other currently available outlier detection methods, which makes evaluating and comparing the novel results to other current techniques quite difficult. So usual benchmarking methods had to be scrapped and instead of utilizing the most commonly used outlier benchmark datasets, different methods had to be implemented in order to start evaluating the IDOD algorithm.

This novel method introduces a new approach to outlier detection and moves in the direction of a new outlier formulation as well. It can be considered as the first trial of this kind.

The goal of this thesis is to present this new algorithm and the ideas it's based on, while also introducing the classic and state-of-the-art outlier detection methods mentioned in the recent literature. Additionally, the first test results about the behaviour of the algorithm is also described, leading to multiple conclusions and ideas for its future development.