Evaluation Metrics for Anomaly Detection Algorithms in Time Series

György Kovács, Gheorghe Sebestyen, Anca Hangan
Department of Computer Science
Technical University of Cluj-Napoca
Romania

July 17, 2019

Abstract

Anomaly detection is the process of identifying erroneous data in big data sets, in order to improve the quality of further data processing. An anomaly detection method classifies data into normal and abnormal values. The selection of the best detection method greatly depends on the data set characteristics. Therefore, we need metrics to evaluate the performance of different methods, on a given data set.

Traditionally, in order to evaluate the quality of a classification, the confusion matrix, or one of its derived metrics is used. These metrics work well when the data set does not have a temporal dimension.

The anomaly detection task has certain particularities when it comes to time-series data. The temporal dimension that may be lacking in other types of data sets can be taken into account in order to improve the evaluation of these methods.

In this paper we propose some evaluation metrics which are more appropriate for time series. The basic idea of the new metrics take into consideration the temporal distance between the true and predicted anomaly points. This way, a small time shift between the true and the detected anomaly is considered a good result as opposed to the traditional metric that will consider it an erroneous detection.

Through a number of experiments, we demonstrate that our proposed metrics are closer to the intuition of a human expert.