

A comparative study on anomaly detection algorithms in non-stationary time series

N.A.M.R. Senaviratna^{1*}, D.D.M. Jayasundara², and A.P. Hewarachchi²

¹Department of Mathematics, Faculty of Natural Sciences, The Open University of Sri Lanka

²Department of Statistics and Computer Science, Faculty of Science, University of Kelaniya

Anomaly detection in time series data has been a research topic for many years. Numerous algorithms have been proposed for various applications, including fraud detection, fault detection, network intrusion detection, and healthcare. However, it remains challenging, particularly when dealing with non-stationary time series that exhibit trends, seasonality, and change points, as the system behavior monitored by time series is likely to change over time, either smoothly or abruptly. There is a lack of comprehensive performance evaluation of the state-of-the-art algorithms in non-stationary time series data. This makes it difficult for users to choose an appropriate method for real-world applications. In general, anomaly detection methods are classified into different taxonomies. One taxonomy is proximity-based, clustering-based, density-based, and isolation-based methods. For this study, one method was selected from each category based on their public availability. The chosen methods are Isolation Forest (IF) for isolation-based detection, Local Outlier Factor (LOF) for proximity-based detection, Support Vector Machine (SVM) for density-based detection, and K-means clustering for clustering-based detection. These methods were tested using the Yahoo Webscope S5 dataset. The dataset was divided into four data classes: *A1*, *A2*, *A3*, and *A4*. Each data class has a set of real or synthetic web traffic metrics annotated with anomalies. Data sets with stationary time series and non-stationary time series that exhibit trends, seasonality, and change points were selected among these four classes. The performance of these methods was compared using precision, recall, f1 score, receiver operating curve (ROC), and precision-recall curve (PR-AUC). For stationary time series, the isolation forest algorithm performed well in identifying anomalies, with precision=1.00, recall= 0.94, f1 score=0.97, ROC=1.00, and PR-AUC=1.00. In non-stationary time series with seasonality, both isolation forest and local outlier factor methods outperformed other methods, with precision=1.00, recall=0.94, f1 score=0.97, ROC=1.00, and PR-AUC=1.00. The K-means clustering showed better time series results with trend, seasonality, and change points.

Keywords: Anomaly detection, Non-stationary, Time series.

*nasen@ou.ac.lk

ORCID ID: <https://orcid.org/0000-0001-6267-7832>