Comparison of Anomaly Detection Techniques in Univariate Time Series Analysis

Barış Kaan Alagöz, Berk Niyazi Aydın, İpek Aydın

Assoc. Prof. Ceylan Yozgatlıgil

December 22, 2021 - METU Culture and Convention Center

ABSTRACT

Anomaly detection is one of the most crucial steps in data analysis. It can be used in time series data analysis, as a quality control tool or for forecasting performance refinement. It is used in various applications such as realization of abnormal situations at entry, fraud detection, healthcare, homogeneity of climate variables, private account intrusion detection and surveillance systems. We want to reveal the best anomaly detection method(s) by comparing the performances of the methods on the labeled real data. In this way, the quality of the time series will increase, and we will be able to determine the predictions for the data correctly.

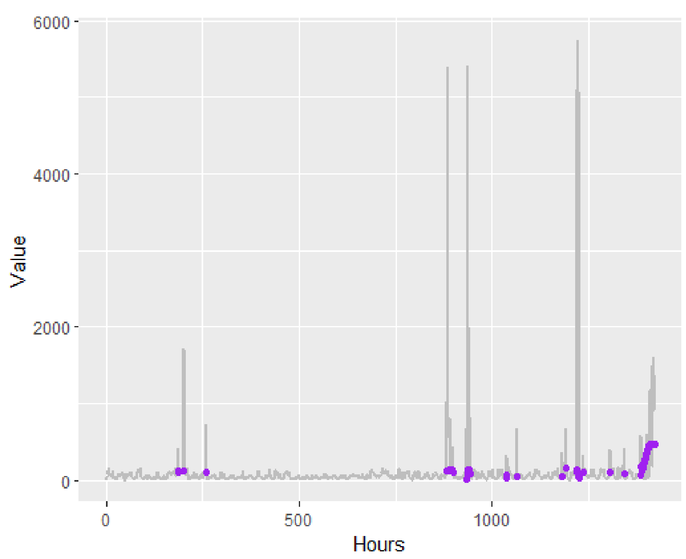

Figure 1: Output of Tsoutliers Method

EXPERIMENTAL PROCEDURE

The dataset consists of sixty-seven univariate data containing actual hourly production traffic to some Yahoo properties. The datasets contain labels that show the amount of traffic generated and whether there are anomalies. Seven unsupervised anomaly detection algorithms were selected for review to compare the algorithms and observe which might be best for the situation. Local Outlier Factor (LOF) algorithm compares the local readability density (lrd) of a point to the lrd of its neighbors. A LOF score of approximately 1 indicates that the lrd around the point is comparable to the lrd of its neighbors and the point is not an outlier. It was decided to use 3 of the nearest neighbors while defining the local neighborhoods. Output is quite successful to detect the outliers, indicating LOF scores greater than 1 for points that are at the end of our datasets. STL is the default method used for decomposition. It is a seasonal decomposition utilizing a Loess smoother. The seasonal values were removed, and the remainder smoothed to find the trend. The overall level was removed from the seasonal component and added to the trend component. The Twitter method is identical to STL for removing the seasonal component (RS,2019). The difference is in removing the trend is that it uses a piece-wise median of the data (one or several median split at specified intervals) rather than fitting a smoother. Detected outliers and algorithm metrics can be seen in the results and finding's part. Tsoutliers algorithm was designed to identify outliers and to suggest potential replacement values. Visualizing it with autopilot function indicated potential outliers.

It can be seen in Figure 1, the Output of Tsoutliers Method is successful at the right end but there are some mis guessed values throughout the data. It will be enhanced in the future. Self-organizing maps are decent ways to detect outliers visually. It calculates the distance from each data points' weights to the sample vector by operating each one through all weight vectors, then assigns a darker or lighter color according to it (Ralhan, 2018). The darker the color gets, the better. Then, k-Means clustering (it was chosen k=2) partitioned the points into 2 groups such that the sum of squares from points to assigned cluster centers is minimized. It assessed all the data according to that and classified. However, its outputs were not successful. It can be improved in further analysis.

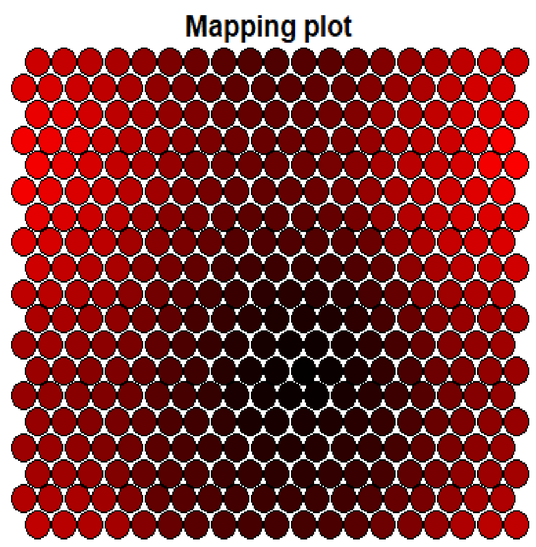

Figure 2: Mapping Plot

MOTIVATION

Anomaly detection methods are used to detect data points, events, and observations that deviate from the normal behavior of the dataset. Thanks to the anomaly detection methods, critical events can be seen in the analysis of different sectors, a new successful marketing campaign that increases potential customers can be found in businesses, or a promotional discount that increases sales, a price error that affects revenue can be detected. In addition to these, seasonality and cyclical behavior patterns within important datasets can also be observed. But there are millions of unique ways to gain insights from anomaly detection as there are so many different methods by which anomalies can be observed. The motivation of the project is to compare these methods and find the method that will give the most accurate result of anomaly detection and increase its applicability in different fields.

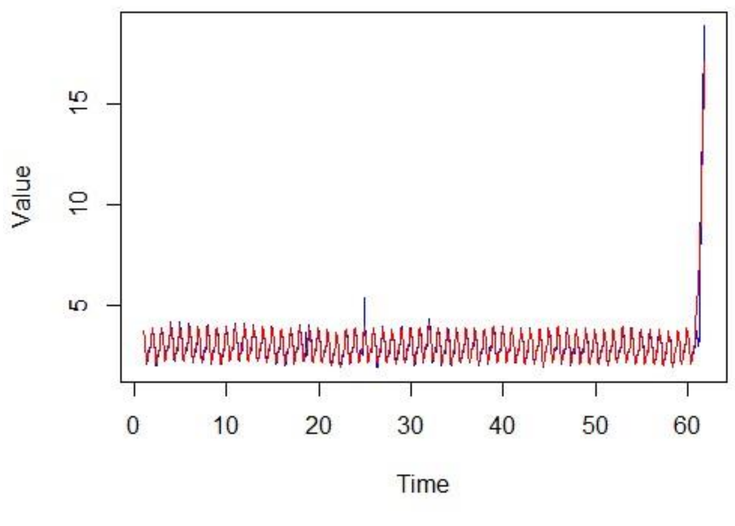

Figure 3: STL Method

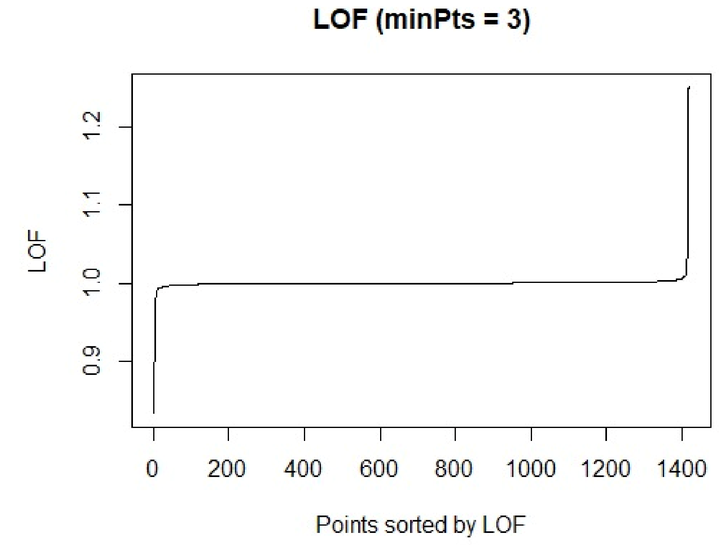

Figure 4: LOF Method

CONCLUSION

Seven anomaly detection methods were analyzed. Their efficiencies differ in terms of their accuracy, precision, specificity, and sensitivity. These values are calculated where the values detected with the anomaly detection methods and detected in our dataset are different. When they are compared, it can be seen from the productivity chart the best anomaly detection method is the twitter method. Also, the k-fold method is efficient but does not show the exact anomaly points. LOF and isolation forests are efficient, too. However, isolation forests detect only extreme points in the dataset. So, there become many differences in comparison. On the other hand, the LOF method detects the anomaly points more accurately. However, the number of points and their indexes cannot be seen from the output. Tsoutliers method is similar to the LOF method. It detects the anomaly points and shows them in the graph. Also, it shows the number of anomaly points and indexes, but it detects much more anomaly points than it has to find.

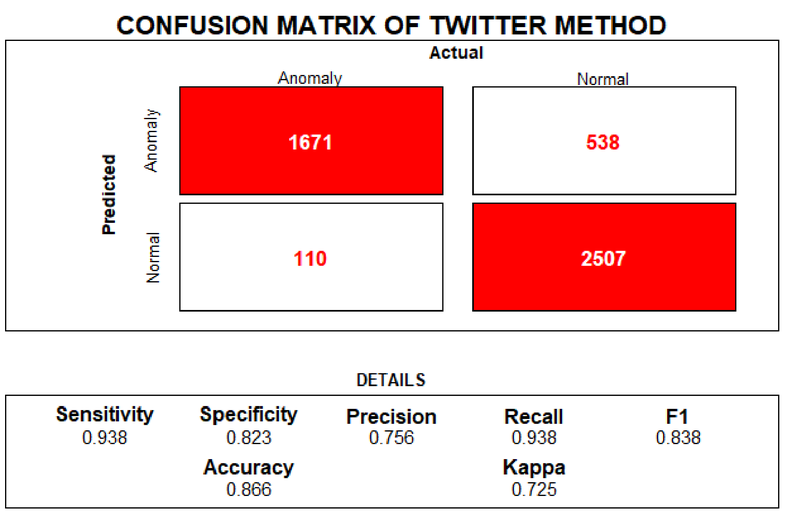

Figure 5: The Final Outcome To Analyze Anomaly Detection Methods

RESULTS AND DISCUSSIONS

Each of the seven anomaly detection algorithms tested offers different solutions. Although it has been decided that the Twitter anomaly detection method is the most useful in the preliminary analysis, further studies will develop findings and obtain clearer results.

REFERENCES

RS, A.M.R (2019, April 6). Tidy anomaly detection using R. Medium. Retrieved December 16, 2021, from https://towardsdatascience.com/tidy-anomaly-detection-using-r-82a0c776d523

Ralhan, A. (2018, September 17). Self organizing maps. Medium. Retrieved December 16, 2021, from https://medium.com/@abhinavr8/self-organizing-maps-ff5853a118d4