International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2019): 7.583

Anomaly Detection Techniques in Time Series Forecasting: Identifying Outliers

Sowmya Ramesh Kumar

Seattle, WA
Email: *rsowmyash[at]gmail.com*

Abstract: ***Time series forecasting, a linchpin in data science, faces challenges due to the inherent fluctuations, trends, and patterns in time series data, making it susceptible to anomalies. Detecting outliers becomes crucial to preserve the accuracy of forecasting models, especially in domains like finance, supply chain, healthcare, manufacturing, and cybersecurity. This paper explores the significance of anomaly detection in time series forecasting, emphasizing both statistical and machine learning - based approaches. Statistical methods like the Z - Score, Grubbs' Test, and Modified Z - Score provide foundational techniques, while machine learning algorithms like Isolation Forests, One - Class SVM, and Autoencoders offer advanced anomaly identification. Additionally, time series - specific methods such as Seasonal Hybrid ESD, the Prophet Algorithm, and Dynamic Time Warping address challenges unique to time series data. These techniques cater to seasonal patterns, varying speeds, and the dynamic nature of evolving patterns. The paper concludes by addressing challenges in anomaly detection, emphasizing real - time detection, balancing false positives and negatives, and managing imbalanced datasets. As technology advances, integrating sophisticated anomaly detection remains critical for resilient and effective time series forecasting in the evolving landscape of data science.***

Keywords: anomaly detection, time - series forecasting, z - score, isolation forests, one - class SVM, dynamic time warping, real - time detection

1. Introduction

Anomaly detection in time - series forecasting plays a pivotal role in enhancing the robustness and reliability of predictive models across diverse industries. Leveraging statistical methods, machine learning algorithms, and specialized time - series techniques empowers organizations to identify outliers, mitigate risks, and make informed decisions based on accurate forecasts. As technology advances, the integration of sophisticated anomaly detection techniques will continue to be a critical aspect of ensuring the resilience and effectiveness of time series forecasting systems.

Importance of Anomaly Detection in Time Series Forecasting

Time series data [1] inherently contains fluctuations, trends, and patterns, making it susceptible to anomalies that can disrupt forecasting models. Anomalies can be indicative of critical events, errors in data collection, or even emerging trends that may impact future predictions. Detecting outliers in time series data is essential for maintaining the accuracy and reliability of forecasting models across diverse domains such as finance, healthcare, manufacturing, and cybersecurity.

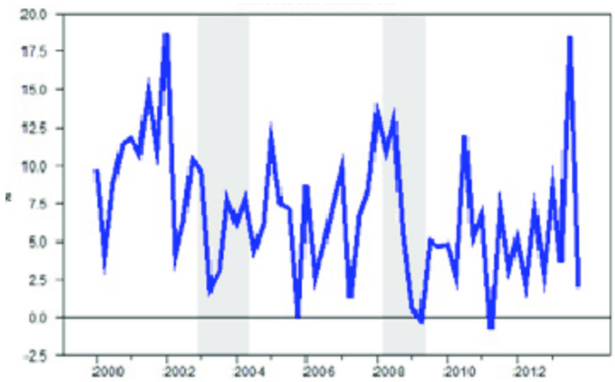

Figure 1: A time - series with trends, anomalies, fluctuations [1]

There are different types of outliers. Types of outliers that are usually dealt with are additive outliers (AO), innovational outliers (IO), temporary change and level shift.

Additive Outlier: The Additive Outlier which is also known as Type I outlier [4]. An AO only affects a single observation, which is either smaller or larger in value compared to the expected values in the data. After this disturbance, the series returns to its normal path as if nothing has happened.

The effect of an additive outlier is independent of the ARIMA model and is bounded.

If an AO outlier occurs at time t=T, the observed series can be represented as

$$Y(t) = U(t) + \omega IT(t)$$

Where, $I_T(t) = \begin{cases} 0, & yt \neq T \\ 1, & yt = T \end{cases}$ is a pulse function and ω is the deviation from the true U(T) caused by the outlier. The true rule suggests that the shock caused by an AO affects the original observation at t = T only with the magnitude of ω and the rest remained unaffected.

In monthly sales data for a retail store, an additional outlier might occur when a one - time marketing promotion leads to an unexpectedly high number of sales for a particular month. This sudden and isolated increase in sales is not part of the typical sales pattern.

Innovational Outlier: In contrast to the AO, innovational outlier is known as Type II outlier [4] that affects several observations. An AO affects only one residual, at the date of the outlier. The effect of the IO on an observed series consists of an initial shock that propagates in the subsequent observations with the weights of the moving average (MA) representation of the ARIMA model.

If an IO outlier occurs at time t = T, then

$$Y(t) = \mu(t) + \frac{\theta(B)}{\Delta\phi(B)}\left(a(t) + \omega I_T(t)\right)$$

A company adopts revolutionary production technology, causing a sudden and sustained increase in productivity, leading to an innovational outlier in the production output time series.

The third category of outliers is *temporary change*, also known as a temporary shock or impulse, refers to a sudden and short - term deviation from the regular pattern or trend in a time series. It is characterized by a brief and abrupt increase or decrease in the values of the time series. This deviation is temporary, and the time series tends to return to its usual pattern after a short duration. For example, In financial markets, a temporary change might be observed as a sudden spike or drop in stock prices due to unexpected news or events. This deviation is expected to be short - lived.

The fourth and last category is *level shift*, also known as a structural change, occurs when there is a sustained and permanent change in the baseline or average level of the time series. Unlike a temporary change, a level shift represents a more persistent alteration in the overall magnitude of the time series values. Once a level shift occurs, the time series tends to follow the new baseline. Consider a manufacturing process where the introduction of new technology leads to a permanent increase in production efficiency. This could result in a level shift in the time series representing daily production levels.

Statistical Approaches for Anomaly Detection:

a) Z - Score Method [7]: The Z - Score is a statistical measure that quantifies how far a data point is from the mean of a dataset in terms of standard deviations. In anomaly detection, data points with Z - Scores beyond a certain threshold are considered outliers.
b) Grubbs' Test [7]: Grubbs' Test, or the Maximum Absolute Deviation test, is employed to detect a single outlier in univariate data. It calculates the Z - Score for the maximum deviation and compares it to a critical value to identify outliers.
c) Modified Z - Score [7]: The Modified Z - Score is a robust version of the traditional Z - Score that is less sensitive to outliers. It uses the median and median absolute deviation (MAD) instead of the mean and standard deviation.

Machine Learning - Based Anomaly Detection Techniques:

a) Isolation Forests: Isolation Forests utilize the principle that anomalies are often isolated and require fewer splits in a decision tree. This algorithm efficiently isolates anomalies in time series data, making it particularly suitable for large datasets.
b) One - Class SVM: Support Vector Machines (SVM) can be adapted for one - class classification, making them effective in identifying anomalies when trained on normal instances only. This approach is useful when most data points represent the normal behavior.
c) Autoencoders [6]: Autoencoders are neural network models trained to learn efficient representations of input data. Anomaly detection using autoencoders involves identifying instances where the reconstructed output deviates significantly from the input, signaling the presence of outliers.

Time Series - Specific Anomaly Detection Methods

a) Seasonal Hybrid ESD: Seasonal Hybrid Extreme Studentized Deviate (ESD) is designed for time series data with seasonal patterns. It combines classical decomposition methods with statistical tests to identify anomalies.
b) Prophet Algorithm: Developed by Facebook, the Prophet algorithm is designed for forecasting with daily observations that display patterns on different time scales. It includes anomaly detection capabilities to identify outliers in the time series.
c) Dynamic Time Warping: Dynamic Time Warping is a technique that measures the similarity between two sequences with varying speeds. It is effective in detecting anomalies in time series data with irregular patterns or variable speeds.

2. Challenges and Considerations in Anomaly Detection

Anomaly detection in time series forecasting poses challenges related to the dynamic nature of data, evolving patterns, and the need for real - time detection. Balancing the trade - off between false positives and false negatives, selecting appropriate features, and handling imbalanced datasets are crucial considerations in implementing effective anomaly detection techniques.

In navigating these challenges, organizations can enhance the resilience of their forecasting models, enabling them to make informed decisions, mitigate risks, and adapt to changing circumstances. As technology evolves, the continued integration of sophisticated anomaly detection techniques will be essential for maintaining the effectiveness and reliability of time series forecasting systems across various sectors. The future landscape of time series forecasting relies on a proactive approach to anomaly detection, ensuring that organizations stay ahead in a dynamic and ever - changing business environment.

References

[1] Jayasuriya, Sisira & Leu, Shawn. (2017). Volatile Capital Flows and Macroeconomic Performance in Indonesia: An SVAR Analysis. Economic Papers: A journal of applied economics and policy.36.10.1111/1759 - 3441.12168

[2] Kim, S. and Roubini, N. (2000), 'Exchange Rate Anomalies in the Industrial Countries: A Solution with a Struc - tural VAR Approach', Journal of Monetary Economics, 45, 561–86.

[3] Snedecor, G. and Cochran, W. (1989), Statistical Method. Iowa State University Press, Iowa, IA.

[4] Fox, A. J. (1972). Outliers in Time series. Journal of the Royal Statistical Society, 55, 559 - 567

[5] Deneshkumar, V., & Kannan, K. S. (2011). Outliers in time series data. *Int. J. Agricult. Stat. Sci*, *7* (2), 685 - 691.

[6] Bozbura M, Tunc H, Kusak M, Sakar CO. Detection of e - Commerce Anomalies using LSTM - recurrent Neural Networks; 2019. p.217–224.

[7] Zhang Y, Hamm NAS, Meratnia N, Stein A, van de Voort M, Havinga PJM. Statistics - based outlier detection for wireless sensor networks. International Journal of Geographical Information Science.2012; 26 (8): 1373–1392.

[8] Zhou Y, Qin R, Xu H, Sadiq S, Yu Y. A Data Quality Control Method for Seafloor Observatories: The Application of Observed Time Series Data in the East China Sea. Sensors.2018; 18: 2628. pmid: 30103440

[9] Kim, S. and Roubini, N. (2000), 'Exchange Rate Anomalies in the Industrial Countries: A Solution with a Struc -

[10] tural VAR Approach', Journal of Monetary Economics, 45, 561–86.

[11] Kim, S. and Roubini, N. (2000), 'Exchange Rate Anomalies in the Industrial Countries: A Solution with a Struc -

[12] tural VAR Approach', Journal of Monetary Economics, 45, 561–86.