Anomaly Detection in Time Series Data: Approaches, Results, and Future Scope

Karan Kumar

July 28, 2025

Abstract

This report presents an overview of anomaly detection methods applied to time series data, focusing on a statistical approach and the Vector Autoregression (VAR) model. Results highlight the detection capabilities of each method for different anomaly types. Finally, future scope emphasizes the potential benefits of implementing advanced models like Orion and Luminaire to improve detection accuracy.

1 Introduction

Anomaly detection in time series data is crucial for identifying unusual behavior that can indicate faults, fraud, or novel phenomena. This work investigates two approaches: a statistical method and a multivariate VAR model, analyzing their strengths and limitations in detecting two types of anomalies labeled here as *Type A* and *Type B*.

2 Methodology

2.1 Statistical Approach

Statistical methods form the backbone of anomaly detection by examining deviations in data properties such as mean, variance, and distribution. Techniques like Z-score analysis, moving averages, and Seasonal-Trend decomposition (STL) help isolate residuals that represent anomalies.

These methods are effective at detecting both point anomalies (isolated abnormal points) and contextual anomalies (deviations in local temporal context). However, they often rely on stationarity assumptions and parameter tuning for thresholds.

2.2 Vector Autoregression (VAR) Model

VAR models capture linear interdependencies among multiple time series variables, using past values of each variable and others to predict current values. Anomalies are signaled by large prediction errors or abnormal residuals.

This model is well-suited for detecting *Type B* anomalies, which often reflect contextual or collective anomalies disrupting the dynamic relationships between multiple series. However, it is less effective for *Type A* anomalies, which tend to be more persistent and isolated in nature.

3 Results and Analysis

The anomaly detection performance was evaluated on two major categories:

- **Type A Anomalies:** These anomalies tend to last for at least 4 hours. The statistical approach detected 24 such anomalies, indicating its capability to capture longer-lasting anomalous events.
- **Type B Anomalies:** These are typically instantaneous anomalies occurring at single timestamps. The statistical method detected 1347 such events, while the VAR model detected 1511 anomalies, confirming VAR's effectiveness in capturing instantaneous multivariate anomalies.

The VAR model's inability to detect Type A anomalies highlights its limitation in identifying persistent anomalies, unlike the statistical approach which covers both types adequately.

4 Scope and Future Work

Due to time constraints, implementation of advanced state-of-the-art models such as **Orion** and **Luminaire** could not be completed. Based on a survey of their methodologies and reported performance, these models promise substantial improvements:

- **Orion** is an end-to-end unsupervised anomaly detection framework utilizing sophisticated machine learning techniques. It is expected to excel at detecting *Type A* anomalies by characterizing anomalies with richer feature representations and incorporating human feedback for refinement.
- **Luminaire** specializes in detecting *Type B* anomalies, leveraging advanced statistical and pattern recognition models to identify contextual and collective anomalies effectively.

While integrating these models demands significant effort in design, tuning, and computational resources, their adoption could enhance detection accuracy and robustness for both anomaly types beyond what is achievable with the current statistical and VAR-based methods.

Exploring hybrid approaches that combine the strengths of statistical, VAR, and advanced machine learning frameworks represents a promising direction for future research.

5 Conclusion

This work provides a comparative analysis of two anomaly detection methods applied to time series data, clearly outlining their respective capabilities and limitations. The statistical approach offers comprehensive detection of both persistent and instantaneous anomalies, whereas the VAR model effectively detects multivariate instantaneous anomalies. Future implementation of sophisticated models like Orion and Luminaire holds promise for significantly improving anomaly detection performance, particularly when tailored to anomaly type.

Keywords: anomaly detection, time series, statistical methods, VAR model, Orion, Luminaire

6 Requirements

The following Python libraries are required to run the anomaly detection implementations discussed in this report:

- `pandas` — Data manipulation and analysis.
- `numpy` — Numerical computations.
- `matplotlib` — Data visualization and plotting.
- `seaborn` — Statistical data visualization built on matplotlib.
- `pykalman` — Kalman filtering for time series smoothing and inference.
- `scikit-learn` — Machine learning utilities including preprocessing and evaluation metrics.
- `statsmodels` — Statistical modeling, including time series models such as VAR and stationarity tests.

These libraries can be installed using `pip` with the following command:

```
pip install pandas numpy matplotlib seaborn pykalman scikit-learn statsmodels
```