RNN-Time-series-Anomaly-Detection paper

MengLin Han

Tongji university  
JunZe Zhu

Tongji university

INTRODUCTION:Time-series anomaly detection has been an active area of research in recent years, with applications spanning diverse domains such as finance, healthcare, transportation, and industrial monitoring. With the proliferation of data-driven systems, the need for effective anomaly detection techniques has become increasingly important.

# 1 Situation

## *1.1 others' invention*

Previous studies have explored various machine learning approaches for time-series anomaly detection,including **traditional statistical methods**[1],**distance-based algorithms**, **density-based models**[2].

these methods have many disadvantage.

*1.1.1:the-traditional-statistical methods*

assumpt the data is stationnary:Traditional statistical methods often assume that data is stationary, meaning that its statistical properties (such as mean and variance) do not change over time. However, time series data is often non-stationary, with its statistical properties varying over time. This non-stationarity poses difficulties for the application of traditional statistical methods, as they are unable to effectively handle the dynamic changes in the data.

*1.1.2the distance-based algorithms*

High algorithmic complexity:Distance-based algorithms typically require calculating the distances between data points, which can become extremely

time-consuming when dealing with large datasets, leading to high algorithmic complexity.Especially in time-series data, where data points possess temporal order and logical relationships, calculating distances may necessitate considering additional factors such as the selection of time windows and the optimization of distance metrics, further increasing the complexity of the algorithm.

*1.1.3the density-based models*

the density-based models have difficult in Threshold Setting:**Density-based models** typically require setting an anomaly factor threshold based on the density of data points to distinguish between normal and abnormal points. However, there is often no unified standard for setting this threshold, and researchers mostly need to manually set it based on the characteristics of the dataset and domain knowledge, which **increases the uncertainty** of the algorithm.

and more recently, deep learning techniques. Among these, **Recurrent Neural Networks** (RNNs) have shown promising results due to their ability to capture temporal dependencies in time-series data.

## *1.2start of our research*

Several researchers have investigated the use of RNNs for anomaly detection. Malhotra et al. (2015, 2016) demonstrated the effectiveness of **Long Short-Term Memory (LSTM)**[3] networks, a type of RNN, in detecting anomalies in time-series data. Their work focused on using LSTM networks for multi-step prediction and anomaly scoring, where anomalies were identified as deviations from the predicted values.

# 2.Background

In addition to LSTM networks, other variants of RNNs have also been explored for anomaly detection. For instance, the proposed RNN-based model in this paper (Park, 2018) utilizes a stacked RNN architecture to recursively predict future values of time-series data. This model then calculates anomaly scores based on the deviation between the predicted and actual values, using a multivariate Gaussian distribution model.

more exactly,The model employs a two-stage strategy comprising time-series prediction and anomaly score calculation.

1:In the first stage, a stacked RNN model is trained on a clean dataset (without anomalies) to recursively predict future values of the time series. This model is then utilized to detect anomalies in a test dataset where anomalies are present.

2:The second stage involves fitting a multivariate Gaussian distribution to the predicted time-series data andcalculating anomaly scores based on the deviation from the distribution.

### *2.1Dataset*

The model is evaluated using multiple datasets

*2.1.1NYC taxi passenger count*

Source: Provided by the New York City Transportation Authority, preprocessed by Cui, Yuwei et al., aggregated every 30 minutes.

Purpose: To test the model's ability in predicting and detecting urban traffic flow abnormalities, such as passenger volume sudden changes during peak hours or due to special events.

#### *2.2.2ECGs*

Content: An ECG dataset containing a single abnormal point corresponding to a premature contraction.

Purpose: Used to detect abnormalities in cardiac activity, such as arrhythmias.

Characteristics: The data exhibits high frequency and periodicity, while potentially containing multiple types of noise and artifacts.

#### *2.2.3 2D gesture data*

Content: Time series data of X and Y coordinates representing gestures in videos.

Purpose: To detect abnormal gestures or behaviors in video surveillance, such as sudden changes in movement or unusual paths.

Characteristics: The data may encompass multiple dimensions (X and Y coordinates) and potentially exhibit complex movement patterns.

#### *2.2.4respiration measurements*

Content: Respiratory data of patients, measured through thoracic expansion with a sampling rate of 10Hz.

Purpose: To detect respiratory abnormalities, such as apnea and abnormal respiratory rates.

Characteristics: The data is typically periodic but may be influenced by various physiological and pathological factors.

#### *2.2.5space shuttle time-series*

Content: Time series data from the Marotta valves of a space shuttle.

Purpose: To detect potential faults or abnormal operations in spacecraft systems, such as valves.

Characteristics: The data may encompass multiple types of failure modes, and requires high real-time and accurate anomaly detection capabilities.

#### *2.2.6power demand data.*

Content: One year's worth of electricity demand data from a research institution in the Netherlands.

Purpose: Used for predicting and detecting abnormal fluctuations in electricity demand, assisting grid operators in scheduling and planning.

Characteristics: The data may contain periodic patterns (such as daily and weekly cycles), seasonal patterns, as well as random fluctuations.

### *2.2impovement*

While these approaches have shown promising results, they also have limitations. For instance, traditional RNNs can suffer from vanishing or exploding gradient problems, which can impact their performance over long time-series data. To address this issue, more recent studies have explored the use of advanced RNN variants, such as**Gated Recurrent Units (GRUs)**, which have been found to be more robust to these issues

# 3.Summary

Overall, the literature on time-series anomaly detection using RNNs is rich and diverse, with many promising approaches and techniques being proposed. However, there is still a need for further research to develop more accurate and efficient models that can effectively detect anomalies in real-world time-series data. The proposed RNN-based model in this paper aims to contribute to this field by offering a two-stage strategycomprising time-series prediction and anomaly score calculation.

## 4.REFERENCE

[1]Cao Chenxi, Tian Youlin, Zhang Yukun, et al. Application of Anomaly Detection Using Statistical Methods in Time Series Data. Journal of Hefei University of Technology (Natural Science).

[2]Sun Meiyu. Research on Time Series Anomaly Detection Method Based on Distance and Density. Computer Engineering and Applications.

[3]W. Xie et al., "PCA-LSTM Anomaly Detection and Prediction Method Based on Time Series Power Data," 2022 China Automation Congress (CAC), Xiamen, China, 2022, pp. 5537-5542, doi: 10.1109/CAC57257.2022.10054757.