European Journal of AI, Computing & Informatics

Vol. 1 No.1 2025

Article **Open Access**

Application of Anomaly Detection Mechanism in Large-Scale Data Processing

Yixin Zhou [1,*]

[1] Amazon, Ads API Infra, New York, 10001, USA
* Correspondence: Yixin Zhou, Amazon, Ads API Infra, New York, 10001, USA

Received: 26 March 2025
Revised: 01 April 2025
Accepted: 18 April 2025
Published: 23 April 2025

Abstract: Anomaly detection technology plays a crucial role in large-scale data processing and is widely used in multiple industries such as finance, the industrial Internet of Things, information security, and intelligent transportation systems such as finance, industrial Internet of Things, information security, and intelligent transportation systems. This technology is dedicated to discovering abnormal behaviors or patterns in large and complex datasets, with the aim of enhancing the accuracy and reliability of the data processing process. This article explores the specific applications of anomaly detection in abnormal transaction detection in the financial industry, device failure prediction in industrial Internet of Things, intrusion detection in network security, and abnormal traffic monitoring in smart transportation. It demonstrates the important role of this mechanism in optimizing business processes, strengthening security, and enhancing risk management capabilities. The trend of intelligent data processing is driven by anomaly detection technology, which significantly improves the ability to process large amounts of data and provides solid technological support for data-driven decision-making and management in various industries.

Keywords: anomaly detection mechanism; large scale data processing; financial monitoring; equipment failure prediction; network security

1. Introduction

In the face of the challenge of accurately identifying subtle abnormal behaviors in large amounts of data, anomaly monitoring systems play a key role. For example, industries such as finance, industrial Internet of Things, information security, and intelligent transportation management use anomaly monitoring systems to detect potential risks and unconventional patterns, greatly improving the speed and reliability of data processing. This article will discuss the definition, classification, and specific application scenarios of anomaly monitoring systems, and explore their important role in improving business processes, enhancing security management, and promoting intelligent processes.

2. Overview of Anomaly Detection Mechanism

2.1. Definition of Anomaly Detection

In data processing, anomaly detection refers to identifying data items or behavioral patterns in a dataset that are inconsistent with regular patterns. The goal of this process is to discover problems or special events hidden in complex data. These anomalies typically exhibit unexpected or irregular characteristics that differ significantly from the norm in the dataset, which are significantly different from the regular data in the dataset [1]. This

type of technology is widely used in various fields such as financial transaction regulation, network security intrusion detection, and machine equipment failure warning. The means of anomaly monitoring are based on in-depth analysis of data characteristics, including statistical analysis, machine learning, and rule-based techniques. In the specific implementation process, the challenges faced mainly stem from the diversity, high dimensionality, and noise interference of data. Therefore, it is particularly crucial to study effective monitoring strategies that adapt to different environments and data characteristics. As a core component of data processing, anomaly detection enhances the credibility of data processing and provides technical support for risk control and decision-making [2].

2.2. Classification of Anomaly Detection

According to the differences in algorithm principles and application backgrounds, anomaly detection techniques can be subdivided into numerous categories. At the algorithmic level, there are three common types of methods based on differences in learning methods: supervised, unsupervised, and semi supervised. Supervised learning relies on labeled datasets for training and uses classification algorithms to distinguish between regular and abnormal data. Unsupervised learning does not require labeled data and typically identifies data points with significant deviations in group behavior by implementing clustering analysis or calculating data density. Semi supervised learning uses a large amount of unlabeled data to enhance detection performance when labeled data is limited. From an application perspective, anomaly detection can be divided into detection types such as time series, data flow, and graphical structure according to the scenario. For example, time series anomaly detection focuses on discovering anomalous activities in time related patterns, data stream anomaly detection focuses on real-time processing of data, and graph data anomaly detection focuses on anomalous nodes within the network structure. These different classifications have their own challenges and goals, providing customized solutions for dealing with complex and ever-changing scenarios [3].

3. The Specific Application of Anomaly Detection Mechanism in Large-Scale Data Processing

3.1. Abnormal Transaction Detection in the Financial Industry

In the financial industry, abnormal transaction detection is a key strategy for preventing fraud and maintaining system stability. This strategy has been widely used in credit card fraud prevention, securities market supervision, and risk prevention of payment systems. In the specific implementation process, the system will track and analyze information such as transaction amount, transaction amount, transaction frequency, consumption location, and device parameters, and compare this information with the user's past transaction patterns to detect abnormal transaction behavior [4]. For example, users usually make purchases within a fixed area or a certain amount limit. If there are large cross-regional purchases or frequent transactions in a short period of time, the system will flag them as abnormal transactions. During the detection process, statistical analysis techniques such as normal distribution models can be used to calculate the probability density function of transaction data, in order to measure the normal probability of trading behavior. The formula is:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (1)$$

In this formula, μ represents the mean transaction amount, and σ represents the standard deviation. If the probability density value of the transaction is lower than the set threshold, it is judged as abnormal. Deep neural network technology, such as long short-term memory networks (LSTM) and autoencoders, is widely used in the analysis of transaction data in the financial field to capture the unique properties of its time series, thereby improving detection accuracy and response speed.

3.2. Equipment Failure Prediction in Industrial Internet of Things

In the field of industrial Internet of Things, the real-time collection of sensor information is used to monitor unusual fluctuations in machine operation, thereby achieving early warning of mechanical failures. This technology is widely used in intelligent production and equipment maintenance. When observing wind turbines, vibration frequency, bearing temperature, and current fluctuations are often used as key monitoring indicators. Once any of these indicators show significant abnormalities, the warning system will be activated to promptly prevent further equipment failure. Power spectral density (PSD) is an important method for analyzing the frequency characteristics of vibration signals, and its formula is:

$$P(f) = \frac{1}{T}\left|\int_0^T x(t)e^{-j2\pi ft}dt\right| \quad (2)$$

Among them, $x(t)$ represents the vibration signal, T is the sampling time, and f is the frequency. By extracting the peak positions and amplitude changes from the power spectrum, to assess device operational status. If the energy of a specific frequency exceeds the normal range, it may indicate that may indicate wear or latent faults in the internal components of the equipment. By integrating sensor data from multiple angles, including temperature, pressure, vibration, and other factors, the system has the ability to analyze fault modes in depth, thereby enhancing the accuracy of predictions.

3.3. Intrusion Detection in Network Security

In the field of computer network security, intrusion detection technology is crucial, and its core task is to continuously monitor network data flows and system behavior in order to detect illegal intrusions or malicious operations in a timely manner. Intrusion detection systems (IDS) are responsible for real-time analysis of the attributes and behavior trajectories of network packets to detect abnormal traffic or suspicious behavior. Enterprise networks will use IDS to track user access frequency, data transmission volume, and protocol usage in order to identify security threats such as distributed denial of service attacks (DDoS) or information leaks. Once the traffic speed or access behavior exceeds the normal range, the system will immediately trigger an alert. The intrusion detection technology based on log files is equally critical, as it can detect illegal login attempts or abuse of permissions by collecting and analyzing the server or firewall logs. For example, consecutive login failures followed by successful remote login may indicate a password cracking attack. Deep learning techniques (convolutional neural networks) and clustering algorithms (K-Means) are also widely used to mine deep correlations in network data in order to discover more covert security threats. The application of these technologies significantly enhances the ability of enterprises and organizations to identify threats, ensuring the security and stable operation of information systems [5].

3.4. Abnormal Traffic Monitoring in Smart Transportation

Abnormal traffic monitoring in smart transportation identifies traffic congestion, accidents, or other abnormal situations through real-time analysis of road traffic data, providing efficient decision support for traffic management departments. In specific applications, various sensing devices, video surveillance, and vehicle networking technologies are used to collect real-time traffic information, including data on vehicle numbers, vehicle speeds, and road usage conditions. The initial cleaning and standardization of these data is to ensure the accuracy of subsequent analysis. In the feature analysis stage, by deeply analyzing the dynamic changes in traffic flow, core indicators such as lane utilization rate, vehicle speed, and traffic flow changes are extracted. The system adopts anomaly detection methods and relies on established rules or intelligent algorithms to identify abnormal traffic conditions. If a sudden decrease in vehicle speed and abnormal increase in traffic flow is detected in a certain area, the system will determine that congestion or accidents may occur in that area and automatically generate warning messages.

This monitoring information will be fed back in real-time to the traffic command platform, allowing relevant departments to respond in a timely manner, such as adjusting traffic signal cycles or dispatching personnel for traffic diversion. This technology has significant effects on alleviating traffic congestion during peak hours, responding to emergencies, and optimizing traffic flow distribution, greatly promoting the improvement of intelligent urban traffic management. Figure 1 shows the main process of monitoring abnormal traffic flow in smart transportation.

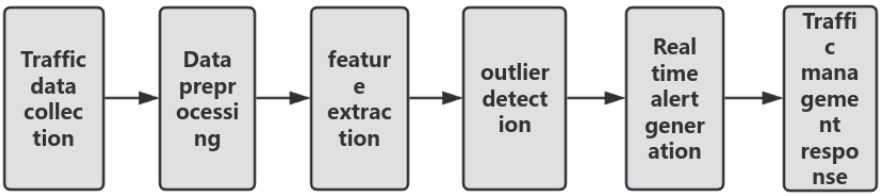

Figure 1. Flow Chart of Intelligent Traffic Abnormal Flow Monitoring.

4. The Application Effectiveness of 3 Anomaly Detection Mechanisms in Large-Scale Data Processing

4.1. Improve the Accuracy and Reliability of Data Processing

The anomaly detection mechanism effectively improves the accuracy and reliability of large-scale data processing by filtering and processing outliers and noisy data. In the industrial Internet of Things, due to the performance degradation of device sensors or interference from external environments, the collected data often contains abnormal components. By utilizing anomaly detection technology, these abnormal data can be effectively removed, thereby improving the precise monitoring of equipment operation status. When processing vibration signals of equipment, the assessment of equipment health status after excluding abnormal data points will be more in line with the actual situation, preventing misjudgment caused by data abnormalities. In the financial industry, the application of anomaly detection can eliminate erroneous transaction records and ensure the accuracy of analysis results. The system will comprehensively analyze multiple dimensions such as transaction amount, occurrence time, and transaction frequency to identify behaviors with large abnormal transactions or abnormal transaction frequency, and prevent these data from causing adverse effects on the prediction model. Anomaly detection also plays an important role in preprocessing model input data, ensuring that the algorithm performs more robustly during training and prediction phases. One commonly used method is the anomaly detection formula based on Euclidean distance:

$$D_E = \sqrt{\sum_{i=1}^{n}(x_i - \mu_i)^2} \quad (3)$$

Among them, x_i is a single data point, and μ_i is the mean of the corresponding dimension. When the Euclidean distance (D_E) between a data point and the mean exceeds the set threshold, the point is marked as abnormal. Euclidean distance is highly efficient in low dimensional data detection and can quickly locate deviation points.

4.2. Optimizing Business Processes and Decision Support

Anomaly detection systems play a crucial role in optimizing business processes and enhancing decision support capabilities, especially in real-time information flow and data-driven decision-making processes. By quickly detecting and processing abnormal data, enterprises can adjust their business processes in a timely manner. In the logistics and inventory management process, anomaly detection can track real-time operations such as stock levels and delivery schedules, timely reveal inventory shortages or logistics delays, and help enterprises quickly optimize supply chain strategies and reduce opera-

tional risks. In the manufacturing industry, detecting abnormal behavior on the production line, such as machine overload or production errors, can effectively improve the production process, reduce resource consumption, and enhance production efficiency. In the decision support process, anomaly detection ensures high-quality data analysis and prediction. In the field of marketing, the system can accurately identify problematic customers or adjust marketing strategies by identifying abnormal purchasing habits or complaint volumes in customer behavior. Enterprises utilize anomaly monitoring systems to ensure the stability and credibility of analyzed data, thereby enhancing the accuracy of decision-making. Anomaly detection often uses density-based Local Outlier Factor (LOF) method, whose calculation formula is:

$$LOF(A) = \frac{\sum_{B \in N(A)} \frac{lrd(B)}{lrd(A)}}{|N(A)|} \tag{4}$$

Among them, A is the target point, B is its neighbor point, $lrd(A)$ is the local reachable density of point A, and $N(A)$ is the set of neighbors. When the value of $LOF(A)$ is large, point A is considered an outlier. This method can effectively identify outliers in complex business scenarios, providing important basis for optimizing decisions.

4.3. Strengthen Security and Risk Management Capabilities

In the process of enhancing security and risk management, anomaly detection systems play a crucial role by detecting abnormal data or activities, providing accurate risk warnings for organizations and individuals. In the financial industry, this system helps identify fraudulent transactions and abnormal fund transfers, preventing potential economic losses. For example, real-time tracking of credit card transactions, identifying abnormal changes in transaction frequency and amount, can effectively prevent fraudulent behavior and ensure that user funds are not compromised. In the field of information security, anomaly detection technology is widely used in intrusion detection systems. By monitoring abnormal patterns in network traffic, it can effectively identify threats such as distributed denial of service attacks (DDoS) or malicious packet injection, providing robust protection for the system. In the industrial Internet of Things, the anomaly detection system is responsible for tracking the operation status of machines. It can analyze abnormal changes in sensor information to timely detect possible faults and hidden dangers, in order to prevent safety accidents. In advanced power grid systems, by continuously detecting current and voltage data, the system can identify circuit abnormalities in advance and take timely measures to reduce operational risks. Table 1 shows the specific effectiveness of anomaly detection mechanisms in enhancing security and risk management in different fields.

Table 1. Specific Effectiveness of Anomaly Detection Mechanism in Security and Risk Management.

Application area	Effect	Example
Financial industry	Identify fraudulent transactions and protect user assets	Detecting abnormal credit card consumption behavior, preventing money laundering and fraudulent activities
Network security	Prevent network attacks and enhance network protection capabilities	Discover DDoS attacks, malware propagation, or abnormal logins
Industrial internet of things	Detecting equipment failures in advance and reducing safety accidents	Monitor equipment vibration or abnormal temperature to prevent equipment malfunction during operation

Supply Chain Management	Avoiding logistics delays and optimizing inventory management	Discovering transportation delays, insufficient inventory, or data errors
Traffic control	Enhance traffic flow monitoring capability and reduce accident occurrence	Real time detection of abnormal vehicle speed, traffic flow, or vehicle behavior

This table summarizes the application and effectiveness of anomaly detection mechanisms in different fields, from fraud transaction detection in the financial industry to abnormal traffic monitoring in smart transportation, fully demonstrating the diverse role and significant achievements of anomaly detection technology in improving system security, optimizing business processes, and risk management.

4.4. Promote the Development of Intelligence and Automation

In the development process of intelligence and automation, anomaly detection technology has played a crucial role, and its application in many industries has created significant economic benefits. On industrial production lines, this technology can monitor key operating parameters of equipment in real time, including temperature, vibration, and energy consumption, effectively detecting potential faults such as component wear or unbalanced operation, and automatically initiating maintenance programs or adjusting production strategies to prevent downtime and associated costs caused by equipment failures. The application of this technology greatly improves the stability of the production process and resource efficiency. In intelligent transportation systems, anomaly detection technology can quickly detect traffic congestion or accidents through real-time analysis of traffic flow and speed changes. The system can automatically adjust traffic signal timing and provide dynamic navigation information to help vehicles effectively avoid congested areas, thereby improving traffic flow and efficiency. In smart grid management, anomaly detection technology can timely detect voltage anomalies or load imbalances through the real-time monitoring of power system data, and automatically optimize power dispatch to ensure stable operation of the power grid and reduce the possibility of accidents.

5. Conclusion

In the field of large-scale data processing, accurately detecting data anomalies and quickly implementing corresponding strategies are the core technical means to improve efficiency and ensure security. This technology has achieved significant results in enhancing the accuracy and credibility of data processing, and has also shown extensive application value in improving business processes, strengthening risk management, and promoting the development of intelligence and automation. In industries such as finance, industrial Internet of Things, network security, and smart transportation systems, practical applications have greatly improved the operational efficiency and security of the system, providing solid support for the sustainable and efficient development of the industry. With the advancement of technology and the continuous expansion of application fields, this technology strategy is expected to play a greater role in more industries, providing a stronger foundation for building an intelligent society and achieving automated management.

References

1. W. Liu, L. Yan, N. Ma, G. Wang, X. Ma, P. Liu, and R. Tang, "Unsupervised deep anomaly detection for industrial multivariate time series data," *Appl. Sci.*, vol. 14, no. 2, p. 774, 2024, doi: 10.3390/app14020774.
2. Z. Wang, C. Pei, M. Ma, X. Wang, Z. Li, D. Pei, and G. Xie, "Revisiting VAE for unsupervised time series anomaly detection: A frequency perspective," in *Proc. ACM Web Conf. (WWW)*, 2024, pp. 3096–3105, doi: 10.1145/3589334.3645710.
3. D. K. Thakur, K. P. Sivaraj, M. Gulhane, J. Alahari, R. Jena, and P. Goel, "Efficient network anomaly detection and mitigation strategies for large-scale networks," in *Proc. 2025 Int. Conf. Pervasive Comput. Technol. (ICPCT)*, 2025, pp. 279–283, doi: 10.1109/ICPCT64145.2025.10940544.

4. H. Luo, Y. Zheng, K. Chen, and S. Zhao, "Probabilistic temporal fusion transformers for large-scale KPI anomaly detection," *IEEE Access*, vol. 12, pp. 9123–9137, 2024, doi: 10.1109/ACCESS.2024.3353201.
5. Z. Jia, Z. Wang, Z. Sun, X. Sun, P. Liu, and F. Ruzzenenti, "A multi-scenario data-driven approach for anomaly detection in electric vehicle battery systems," *eTransp.*, vol. 24, p. 100418, 2025, doi: 10.1016/j.etran.2025.100418.