JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Artificial Intelligence Techniques for Anomaly Detection: A Review

[1]Deepak Kumar, [2]Sanjay Kumar Pal
[1]M.Tech Scholar, [2]Assistant Professor
Department of Computer Science Engineering
Oriental Institute of Science and Technology, Bhopal, India

Abstract : Anomaly detection is a critical task in various domains, including cybersecurity, healthcare, finance, and industrial systems, where identifying deviations from normal behavior can prevent significant losses and enhance decision-making processes. The advent of Artificial Intelligence (AI) has revolutionized anomaly detection, offering sophisticated methods that can automatically learn complex patterns and detect anomalies with high accuracy. This paper presents a comprehensive review of AI techniques for anomaly detection, covering both traditional methods and modern approaches, such as machine learning and deep learning. The review examines various algorithms, including supervised, unsupervised, and semi-supervised learning techniques, and explores their applications across different fields.

Index Terms **– Anomaly, AI, Model, Big data, Artificial Intelligence.**

I. INTRODUCTION

Anomaly detection, also known as outlier detection, is a process aimed at identifying patterns in data that do not conform to expected behavior. These anomalies, or outliers, often signify critical events, such as fraud, cyber-attacks, equipment failures, or rare medical conditions, making their timely detection essential for mitigating risks and making informed decisions. Traditional methods for anomaly detection have relied on statistical models and rule-based systems, which, while effective in some scenarios, often struggle with the complexities of modern data environments, such as large-scale datasets, high-dimensional spaces, and non-linear relationships.

The emergence of Artificial Intelligence (AI) has introduced a new paradigm in anomaly detection, offering techniques that can automatically learn from data, adapt to changing patterns, and detect anomalies with unprecedented accuracy. AI-based approaches leverage machine learning, deep learning, and other advanced techniques to analyze data, uncover hidden patterns, and identify deviations that might be imperceptible to traditional methods. These capabilities have made AI indispensable in various fields, from cybersecurity, where it helps detect malicious activities in real-time, to healthcare, where it aids in the early diagnosis of diseases by identifying abnormal physiological signals.

One of the primary advantages of AI techniques in anomaly detection is their ability to handle the complexity and diversity of modern datasets. Unlike traditional methods, which often require extensive domain knowledge and manual feature engineering, AI models can automatically extract relevant features and learn the underlying structure of the data. This ability is particularly valuable in environments where the data is high-dimensional, unstructured, or continuously evolving, such as in IoT networks, financial transactions, and industrial control systems.

AI techniques for anomaly detection can be broadly categorized into supervised, unsupervised, and semi-supervised learning methods. Supervised learning approaches, which require labeled data, are typically used in scenarios where historical data with known anomalies is available. These methods, including decision trees, support vector machines, and neural networks, have shown great promise in detecting known types of anomalies but may struggle to identify new or rare types. On the other hand, unsupervised learning methods do not require labeled data, making them more flexible and applicable in situations where labeled data is scarce or unavailable. Techniques such as clustering, density estimation, and autoencoders are commonly used in unsupervised anomaly detection. Semi-supervised learning, which combines elements of both supervised and unsupervised learning, offers a middle ground by using a small amount of labeled data to guide the learning process while leveraging large amounts of unlabeled data.

II. BACKGROUND

X. Zhou et al.,[1] shows the public IBD dataset named UNSW-NB15 demonstrate that the proposed VLSTM model can efficiently cope with imbalance and high-dimensional issues, and significantly improve the accuracy and reduce the false rate in anomaly detection for IBD according to F1, area under curve (AUC), and false alarm rate (FAR).

L. Huang et al.,[2] In order to model the mutual influence between nodes and motif instances, the learning procedures of the node representation and the motif instance representation are integrated into a unified graph attention network with a novel hybrid-

order self-attention mechanism. After learning the node representation and the motif instance representation, two decoders are respectively designed to reconstruct the attribute information of the nodes and motif instances.

W. Liang et al.,[3] The abnormal characteristics in a Blockchain dataset are identified, a weighted combination is carried out, and the weighted coefficients among several nodes are obtained after multiple rounds of mutual competition among clustering nodes.

S. Han et al.,[4] propose Robust Online Evolving Anomaly Detection (ROEAD) framework which adopts Robust Feature Extractor (RFE) to remove the effects of noise and Online Evolving Anomaly Detection (OEAD) to dynamic update parameters. We propose Online Evolving SVM (OES) algorithm as the example of online anomaly detection methods.

J. Zhang et al.,[5] The proposed model outperforms binary classification models on the clinical X-VIRAL dataset that contains 5,977 viral pneumonia (no COVID-19) cases, 37,393 non-viral pneumonia or healthy cases. Moreover, when directly testing on the X-COVID dataset that contains 106 COVID-19 cases and 107 normal controls without any fine-tuning, our model achieves an AUC of 83.61% and sensitivity of 71.70%, which is comparable to the performance of radiologists reported in the literature.

P. Rathore et al.,[6] propose an incremental siVAT algorithm, called inc-siVAT, which deals with the streaming data in chunks. It first extracts a small size smart sample using an intelligent sampling scheme, called maximin random sampling (MMRS), then incrementally updates the smart sample points on the fly, using our novel incremental MMRS (inc-MMRS) algorithm, to reflect changes in the data stream after each chunk is processed.

O. Abdelrahman et al.,[7] present techniques gave the highest performance are KNN, ABOD for both product series datasets with 0.95 and 0.99 AUROC respectively. Finally, we applied a statistical root cause analysis on the detected anomalies with the use of Pareto chart to visualize the frequency of the possible causes and its cumulative occurrence.

A. Alnafessah et al.,[8] To address this challenge, we introduce TRACK-Plus a black-box training methodology for performance anomaly detection. TRACK-Plus has been extensively validated using a real Apache Spark Streaming system and achieve a high F-score while simultaneously reducing training time by 80% compared to efficiently detect anomalies.

A. Sonkar et al.,[9] pays a close attention to pattern representation, and proposes three sets of numeric features for representing road conditions. Also, three deep learning approaches, i.e. Deep Feedforward Network (DFN), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN), are considered to tackle the classification problem. The detectors, with respect to the three deep learning approaches, are trained and evaluated through data collected from a test vehicle driven on various road anomaly conditions.

Y. Lu et al.,[10] propose a KQIs-based QoE anomaly detection framework using semi-supervised machine learning algorithm, i.e., iterative positive sample aided one-class support vector machine (IPS-OCSVM).

A. Libri et al.,[11] report on a novel lightweight and scalable approach to increase the security of DCs/SCs, which involves AI-powered edge computing on high-resolution power consumption.

T. Sui et al.,[12] presented blockchain innovation has as of late drawn in a great deal of interest as a potential major advantage for various applications. By the by, there are as yet significant snags with blockchain organizations' presentation and versatility.

III. CHALLENGES

1. Imbalanced Data

Anomaly detection typically involves highly imbalanced datasets, where normal instances vastly outnumber anomalies. This imbalance poses a significant challenge as most AI models are designed to optimize accuracy, which may lead to a bias toward the majority class (normal behavior). As a result, the model might fail to detect rare but critical anomalies, leading to a high rate of false negatives. Addressing this imbalance requires specialized techniques, such as resampling, synthetic data generation, or cost-sensitive learning, but these methods can introduce their own complexities and potential biases.

2. High Dimensionality

Many modern datasets used in anomaly detection are high-dimensional, meaning they contain a large number of features. High dimensionality can lead to the "curse of dimensionality," where the distance between data points becomes less meaningful, and the model's performance degrades. This issue complicates the learning process and can result in overfitting, where the model becomes too tailored to the training data and fails to generalize well to new data. Dimensionality reduction techniques, such as Principal Component Analysis (PCA) or t-SNE, are often employed, but these can also result in the loss of important information.

3. Interpretability of Models

One of the significant challenges with AI-based anomaly detection, especially when using deep learning models, is the interpretability of the results. Many AI models, particularly neural networks, are often considered "black boxes" because they do

not provide clear explanations for their decisions. In anomaly detection, understanding why a particular instance is classified as an anomaly is crucial for validating the results and ensuring that the model is making accurate and reasonable predictions. Techniques like explainable AI (XAI) are being developed to address this challenge, but achieving a balance between model performance and interpretability remains difficult.

4. Scalability

Anomaly detection systems need to be scalable to handle large volumes of data, especially in real-time applications like cybersecurity or IoT networks. AI models, particularly those based on deep learning, can be computationally intensive and may struggle to scale efficiently when applied to big data. This scalability issue can lead to delays in detection, reducing the effectiveness of the system in real-time scenarios. Optimizing these models for scalability often requires significant computational resources or the development of more efficient algorithms.

5. Adaptability to Evolving Data

In many applications, the data distribution can change over time due to evolving user behavior, new types of attacks, or changes in the environment. AI models must be able to adapt to these changes to maintain their effectiveness in anomaly detection. However, adapting to evolving data, also known as concept drift, is challenging. Traditional models might need to be retrained frequently, which can be resource-intensive and time-consuming. Continuous learning techniques and adaptive algorithms are being explored to address this issue, but they come with their own set of challenges, such as maintaining stability and avoiding overfitting to recent data.

IV. CONCLUSION

Anomaly detection is a process of finding those rare items, data points, events, or observations that make suspicions by being different from the rest data points or observations. Anomaly detection is also known as outlier detection. Anomaly detection can effectively help in catching the fraud, discovering strange activity in large and complex Big Data sets. This paper studies about various previous works on the anomaly detection.

REFERENCES

1. X. Zhou, Y. Hu, W. Liang, J. Ma and Q. Jin, "Variational LSTM Enhanced Anomaly Detection for Industrial Big Data," in IEEE Transactions on Industrial Informatics, vol. 17, no. 5, pp. 3469-3477, May 2021, doi: 10.1109/TII.2020.3022432.
2. L. Huang et al., "Hybrid-Order Anomaly Detection on Attributed Networks," in IEEE Transactions on Knowledge and Data Engineering, doi: 10.1109/TKDE.2021.3117842.
3. W. Liang, L. Xiao, K. Zhang, M. Tang, D. He and K. -C. Li, "Data Fusion Approach for Collaborative Anomaly Intrusion Detection in Blockchain-based Systems," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2021.3053842.
4. S. Han et al., "Log-Based Anomaly Detection With Robust Feature Extraction and Online Learning," in IEEE Transactions on Information Forensics and Security, vol. 16, pp. 2300-2311, 2021, doi: 10.1109/TIFS.2021.3053371.
5. J. Zhang et al., "Viral Pneumonia Screening on Chest X-Rays Using Confidence-Aware Anomaly Detection," in IEEE Transactions on Medical Imaging, vol. 40, no. 3, pp. 879-890, March 2021, doi: 10.1109/TMI.2020.3040950.
6. P. Rathore, D. Kumar, J. C. Bezdek, S. Rajasegarar and M. Palaniswami, "Visual Structural Assessment and Anomaly Detection for High-Velocity Data Streams," in IEEE Transactions on Cybernetics, vol. 51, no. 12, pp. 5979-5992, Dec. 2021, doi: 10.1109/TCYB.2020.2973137.
7. O. Abdelrahman and P. Keikhosrokiani, "Assembly Line Anomaly Detection and Root Cause Analysis Using Machine Learning," in IEEE Access, vol. 8, pp. 189661-189672, 2020, doi: 10.1109/ACCESS.2020.3029826.
8. A. Alnafessah and G. Casale, "TRACK-Plus: Optimizing Artificial Neural Networks for Hybrid Anomaly Detection in Data Streaming Systems," in IEEE Access, vol. 8, pp. 146613-146626, 2020, doi: 10.1109/ACCESS.2020.3015346.
9. A. Sonkar, S. K. Sahu, A. Nayak, D. Sahu, P. Verma and R. Tiwari, "An Efficient Privacy-Preserving Machine Learning for Blockchain Network," 2024 4th International Conference on Intelligent Technologies (CONIT), Bangalore, India, 2024, pp. 1-6, doi: 10.1109/CONIT61985.2024.10627061.
10. Y. Lu et al., "Semi-Supervised Machine Learning Aided Anomaly Detection Method in Cellular Networks," in IEEE Transactions on Vehicular Technology, vol. 69, no. 8, pp. 8459-8467, Aug. 2020, doi: 10.1109/TVT.2020.2995160.
11. A. Libri, A. Bartolini and L. Benini, "pAElla: Edge AI-Based Real-Time Malware Detection in Data Centers," in IEEE Internet of Things Journal, vol. 7, no. 10, pp. 9589-9599, Oct. 2020, doi: 10.1109/JIOT.2020.2986702.
12. T. Sui et al., "A Real-Time Hidden Anomaly Detection of Correlated Data in Wireless Networks," in IEEE Access, vol. 8, pp. 60990-60999, 2020, doi: 10.1109/ACCESS.2020.2984276.