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

False Positive Error Reduction in Anomaly Detection

Abdulkarim Shaikh
Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering Pune, India

Ankit Jadhav
Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering Pune, India

Imran Shaikh
Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering Pune, India

Jyoti Dupare
Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering Pune, India

Prof. Yogesh Pawar
Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering Pune, India

Abstract—**Anomaly detection systems play a pivotal role in identifying deviations from expected patterns in various domains, such as cybersecurity, industrial processes, and healthcare. However, these systems often suffer from the issue of generating false positives, which can lead to unnecessary alerts, operational disruptions, and resource wastage. This paper presents a comprehensive approach to address the challenge of reducing false positives in anomaly detection systems.**

Our proposed methodology combines advanced machine learning techniques, feature engineering, and domain-specific knowledge to improve the accuracy of anomaly detection while minimizing false positives. We investigate the importance of feature selection and extraction methods, model selection, and hyperparameter tuning in achieving a more reliable anomaly detection system. Additionally, we explore the incorporation of contextual information and feedback mechanisms to enhance the system's adaptability to evolving data patterns.

Through extensive experimentation and evaluation on real-world datasets, we demonstrate the effectiveness of our approach in significantly reducing false positives without compromising true anomaly detection rates. The results highlight the practical applicability of our methodology in diverse domains and underscore its potential to enhance system reliability, reduce operational costs, and improve overall security and efficiency.

This research contributes valuable insights and techniques to the ongoing efforts in anomaly detection, offering a promising avenue for organizations and industries seeking to bolster their security and operational resilience by mitigating the impact of false positives in their anomaly detection systems.

Keywords - Anomaly Detection, False Positive error reduction, Ensemble Model

1. INTRODUCTION

The proposed method for reducing false positive errors in anomaly detection improves upon the prior art by using an ensemble of anomaly detection models that are trained using a variety of different anomaly detection algorithms. This approach makes the proposed method more robust to noise and outliers in the data.

In addition to using an ensemble of anomaly detection models, the proposed method also uses a variety of other techniques to reduce false positive errors. For example, the proposed method uses a threshold on the anomaly scores to identify anomalies. This threshold is set based on the desired balance between false positives and false negatives.

The proposed method also uses a variety of post-processing techniques to further reduce false positive errors. For example, the proposed method can use a technique called clustering to identify groups of similar data points. Data points in a cluster are less likely to be anomalies than data points that are not in a cluster.

Anomaly detection is a well-known problem in machine learning. It is the task of identifying data points that are different from the rest of the data. Anomaly detection is used in a variety of applications, such as fraud detection, network intrusion detection, and medical anomaly detection.

However, all anomaly detection algorithms are susceptible to false positive errors. A false positive error occurs when an anomaly detection algorithm identifies a data point as an anomaly when it is actually not an anomaly. False positive errors can lead to significant costs and penalties, such as lost revenue, damaged reputation, and even loss of life.

The proposed method for reducing false positive errors in anomaly detection improves upon the prior art by using an ensemble of anomaly detection models that are trained using a variety of different anomaly detection algorithms. This approach makes the proposed method more robust to noise and outliers in the data.

I. LITERATURE REVIEW

In Paper [1]:Ri-Xian L , Ming-Hai Y , Xian-Bao

Defect detection is an important step in the field of industrial production. Through the study of deep learning and transfer learning, this paper proposes a method of defect detection based on deep learning and transfer learning. Our method firstly establishes Deep Belief Networks and trains it according to the source domain sample feature, in order to obtain the weights of the network according to source domain samples. Then, the structure and parameters of the source domain DBN is transferred to the target domain and target domain samples are used to fine-tune the network parameters to get the mapping relationship between the target domain training sample and defect-free template. Finally, the defects of testing samples will be detected by compared with the reconstruction image. This method not only can make full use of the advantages of DBN, also can solve over-fitting in DBN network training through parameters transfer learning. These experiments show that DBN is a successful approach in the high-dimensional-feature-space information extraction task, which can perfectly establishes the mapping relationship, and can quickly detect defects with a high accuracy.

In Paper [2]:I .Guyon and A.Elisseeff

Variable and feature selection have become the focus of much research in areas of application for which datasets with tens or hundreds of thousands of variables are available. These areas include text processing of internet documents, gene expression array analysis, and combination chemistry. The objective of variable selection is three-fold: improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and providing a better understanding of the underlying process that generated the data. The contributions of this special issue cover a wide range of aspects of such problems: providing a better definition of the objective function, feature construction, feature ranking, multivariate feature selection, efficient search .

In Paper [3] :T. Shon, and J. Moon

Cyber attacks such as worms and spy-ware are becoming increasingly widespread and dangerous. The existing signature-based intrusion detection mechanisms are often not sufficient in detecting these types of attacks...First, we create a profile of normal packets using Self-Organized Feature Map (SOFM), for SVM learning without pre-existing knowledge. Second, we use a packet filtering scheme based on Passive TCP/IP Fingerprinting (PTF), in order to reject incomplete network traffic that either violates the TCP/IP standard or generation policy inside of well-known platforms. Third, a feature selection technique using a Genetic Algorithm (GA) is used for extracting optimized information from raw internet packets. Fourth, we use the flow of packets based on temporal relationships during data preprocessing, for considering the temporal relationships among the inputs used in SVM learning. Lastly, we demonstrate the effectiveness of the Enhanced SVM approach using the above-mentioned techniques, such as SOFM, PTF, and GA on MIT Lincoln Lab datasets, and a live dataset captured from a real network. The experimental results are verified by m-fold cross validation, and the proposed approach is compared with real world Network Intrusion Detection Systems (NIDS).

In paper [4]: Sapna Malik and Kiran Khatter

The Android Mobiles constitute a large portion of mobile market which also attracts the malware developer for malicious gains. Every year hundreds of malware are detected in the Android market. Unofficial and Official Android marke tsuch as Google Play Store are infested with fake and malicious apps which is a warning alarm for naive user. Guided by this insight, this paper presents the malicious application detection and classification system using machine learning techniques by extracting and analyzing the Android Permission Feature of the Android applications. For the feature extraction, the authors of this work have developed the AndroData tool written in shell script and analyzed the extracted features of 1060 Android applications with machine learning algorithms. They have achieved the malicious application detection and classification accuracy of 98.2% and 87.3%, respectively with machine learning techniques.

In Paper [5]:C. Tsai, Y. Hsu, C. Lin, and W. Lin

The popularity of using Internet contains some risks of network attacks. Intrusion detection is one major research problem in network security, whose aim is to identify unusual access or attacks to secure internal networks. In literature, intrusion detection systems have been approached by various machine learning techniques. However, there is no a review paper to examine and understand the current status of using machine learning techniques to solve the intrusion detection problems. This chapter reviews 55 related studies in the period between 2000 and 2007 focusing on developing single, hybrid, and ensemble classifiers. Related studies are compared by their classifier design, datasets used, and other experimental setups. Current achievements and limitations in developing intrusion detection systems by machine learning are present and discussed. A number of future research directions are also provided.

In Paper [6]:Vincent P, Larochelle H, Bengio Y

Previous work has shown that the difficulties in learning deep generative or discriminative models can be overcome by an initial unsupervised learning step that maps inputs to useful intermediate representations. We introduce and motivate a new training principle for unsupervised learning of a representation based on the idea of making the learned representations robust to partial corruption of the input pattern. This approach can be used to train autoencoders, and these denoising autoencoders can be stacked to initialize deep architectures. The algorithm can be motivated from a manifold learning and information theoretic perspective or from a generative model perspective. Comparative experiments clearly show the surprising advantage of corrupting the input of autoencoders on a pattern classification benchmark suite.

In Paper[7]:Mohammed S. Alsahli1 , Marwah M. Almasri2 ,etc

Technology has revolutionized into connecting “things” together with the rebirth of the global network called Internet of Things (IoT). This is achieved through Wireless Sensor Network (WSN) which introduces new security challenges for Information Technology (IT) scientists and researchers. This paper addresses the security issues in WSN by establishing potential automated solutions for identifying associated risks. It also evaluates the effectiveness of various machine learning algorithms on two types of datasets, mainly, KDD99 and WSN datasets. The aim is to analyze and protect WSN networks in combination with Firewalls, Deep Packet Inspection (DPI), and Intrusion Prevention Systems (IPS) all specialized for the overall protection of WSN networks. Multiple testing options were investigated such as cross validation and percentage split. Based on the finding, the most accurate algorithm and the least time processing were suggested for both datasets.

In Paper[8]:Mahmudul Hasan, Md. Milon Islam, Md Ishrak Islam Zarif, etc

Attack and anomaly detection in the Internet of Things (IoT) infrastructure is a rising concern in the domain of IoT. With the increased use of IoT infrastructure in every domain, threats and attacks in these infrastructures are also growing commensurately. Denial of Service, Data Type Probing, Malicious Control, Malicious Operation, Scan, Spying and Wrong Setup are such attacks and anomalies which can cause an IoT system failure.In this paper, performances of several machine learning models have been compared to predict attacks and anomalies on the IoT systems accurately. The machine learning (ML) algorithms that have been used here are Logistic Regression (LR), Support Vector Machine(SVM), Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN). The evaluation metrics used in the comparison of performance are accuracy, precision, recall,f1 score, and area under the Receiver Operating Characteristic Curve. The system obtained99.4% test accuracy for Decision Tree, Random Forest, and ANN. Though these technique shave the same accuracy, other metrics prove that Random Forest performs comparatively better.

I. CONCLUSION

Ensemble learning has emerged as a powerful and effective technique for anomaly detection, offering significant advantages over traditional single-model approaches. By combining the strengths of multiple base models, ensemble learning can achieve higher accuracy, reduce false positives, and enhance adaptability to changing data patterns. This makes ensemble learning well-suited for a wide range of applications, including fraud detection, network intrusion detection, industrial process monitoring, and medical diagnosis. As research and development in ensemble learning continue to advance, we can expect even more sophisticated and effective anomaly detection systems to emerge, further expanding the potential of this promising technique.

II. REFERENCES

[1] Jeong-Hyeon Moon a, Jun-Hyung Yu b, Kyung-Ah Sohn . An approach to anomaly detection using high- and low-variance principal components . Date of Publication: 15 February 2022

[2] Tin Lai, Farnaz Farid, Abubakar Bello, Fariza Sabrina . Anomaly Detection for IoT Cybersecurity via Bayesian Hyperparameters Sensitivity Analysis . Date of Publication: 20 Jul 2023

[3] Robin Tibor Schirrmeister, Yuxuan Zhou, Tonio Ball, Dan Zhang . Understanding Anomaly Detection with Deep Invertible Networks through Hierarchies of Distributions and Features . Published at NeurIPS 2020

[4] Yonghyeok Ji; Hyeongcheol Lee . Event-Based Anomaly Detection Using a One-Class SVM for a Hybrid Electric Vehicle . Date of Publication: 11 May 2022 . DOI: 10.1109/TVT.2022.3165526

[5] Lev V. Utkin, Andrey Y. Ageev, Andrei V. Konstantinov . Improved Anomaly Detection by Using the Attention-Based Isolation Forest . Date of Publication: 5 Oct 2022

[6] Fernando Arévalo ,M. Tahasanul Ibrahim ,M. P. Christian Alison ,Andreas Schwung . Anomaly Detection Using Classification and Evidence Theory Date of Publication: 25 May 2023 . DOI: 10.1109/ACCESS.2023.3280048

[7]Khloud Al Jallad, Mohamad Aljnidi, Mohammad Said Desouki . Anomaly detection optimization using big data and deep learning to reduce false-positive . Date of Publication: 28 Sep 2022