IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 4, April 2020

Anomaly Detection: Different Machine Learning Techniques, A Review

Dhanush P.M. Naik[1], I. Rohit Satya[2], Chaitra B.H.[3], Vishalakshi Prabhu H[4]

Student, Dept. of Computer Science and Engineering, R V College of Engineering, Bengaluru, India[1,2]

Assistant Professor, Dept. of Computer Science and Engineering, R V College of Engineering, Bengaluru, India[3,4]

Abstract: Anomaly is something that deviates from normal, standard or unexpected. Anomaly detection in different applications has its own importance. The most common reason is that the unexpected behaviour always results in some kind of loss - it can either be theft of important data or damage to the system itself. Many anomaly detection techniques have been developed specific to application or to data. In this paper we have compiled a few machine learning algorithms that can be used for anomaly detection which can help researchers to select a particular algorithm for anomaly detection

Keywords: anomaly, anomaly detection, intrusion detection, outlier, supervised, unsupervised

I. INTRODUCTION

Anomaly detection refers to identification of data items, points or events that are rare, differ significantly from other data items, points or events or that have unexpected behaviour. These rare items are called anomalies, outliers, exceptions, defects or contaminants. Anomaly and outliers are the 2 commonly used words. Anomaly detection is used widely in various fields of application which include Intrusion detection system, fraud detection in credit card transactions, System (software and hardware) health monitoring systems, Industrial damage detection, texture surface defect, sensor networks.

Anomaly is defined as something that deviates from what is standard, normal or expected. So, in machine learning we can say anomalies are data items or points that are different from normal data or may cause unexpected behaviour. The anomalies are classified into point, contextual and collective anomalies

The reason for the importance of anomaly detection is that the anomalies might cause unexpected behaviour in the system or these anomalies arise due to unexpected behaviour in the system. This uncertainty in the behaviour is not safe. This unexpected behaviour can cause theft of important data, blocking of usual services or damage to the system. Hence, it is important to detect such anomalies and take necessary actions based on the anomaly detected

II. BASIC METHODOLOGY OF ANOMALY DETECTION

There are many ways of anomaly detection in machine learning - it is a four-step process as given in figure 1.

A. *Feature extraction stage and data pre-processing*

In this stage the required features are selected for detection and data is stored with the extracted features. The data is also pre-processed to handle missing or unknown data, normalization, scaling and other processes.

B. *Training*

In this stage the model would be selected and the model would be trained to learn normal (and/or abnormal) data / behaviour of the system.

C. *Detection*

Once the model is trained it is deployed to detect any anomaly in the real-world input. If the deviation is found to cross a predefined threshold it will be considered as an anomaly and an alert will be raised.

DOI 10.17148/IJARCCE.2020.9411

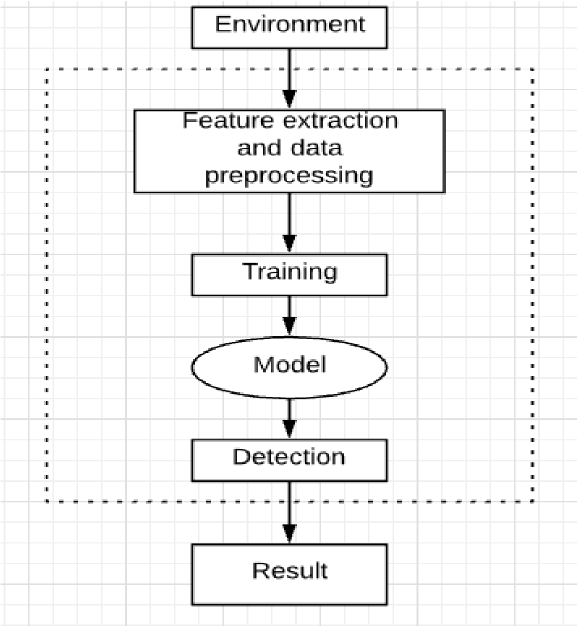

Fig 1. Basic Methodology of Anomaly Detection

III. Application Of Anomaly Detection

Important applications of anomaly detection include:

A. *Intrusion Detection System*

Anomalies computer-based system either in the computer networks or host. Anomaly detection plays an important role in intrusion detection systems by finding anomalies in the network traffic or in a computer which can be a call to the OS which might cause unusual behaviour in the computer.

B. *Industrial damage detection*

Industrial equipment or machines undergo damages due to continuous use of machines and the wear and tear. So, detection of such damages at an earlier stage can be helpful in preventing huge loss. Hence, anomaly detection plays an important role in detection of damages. Industrial damage can be divided into system health monitoring and by operation. In case of system health monitoring defect occurs in some part of the machine. In case of by operation the defect is in the structure of the machine.

C. *Texture surface defect detection*

Texture surface consists of continuous repeated patterns. The defect can be in the texture or in the pattern. Anomaly detection helps in the detection of such defects.

D. *Credit card fraud detection*

Anomaly detection is mainly used to detect fraud applications for credit card and also for detection of fraud usage of credit card. Fraud is determined by high cost purchase, high rate purchase, purchase of items that has never been bought by a user before. Etc. It is a challenge to detect the fraud as soon as the transaction is done.

IV. Different Machine Learning Techniques For Anomaly Detection

There are many techniques in Machine learning that can be used for anomaly detection which can be mainly classified into supervised and unsupervised learning techniques.

A. *Supervised machine learning*

It is one of the main classes of machine learning where both input and output are available. The main task of this learning is to determine a function that is able to find a relation between input and output. In other words, find a mapping between input and output. Most commonly used machine learning techniques - algorithms used for anomaly detection are - Support vector machine, Decision tree, Bayesian networks and K nearest neighbors

DOI 10.17148/IJARCCE.2020.9411

1)Support Vector Machine (SVM): SVM is a popular supervised machine learning algorithm. It's most commonly used to solve classification problems, but regression problems can also be solved. The main task of SVM is to determine a hyperplane that is able to distinguish one class of data from another. i.e. data belonging to one class will be on one side of the hyperplane and rest on the other side. So, for multiclass classification SVM needs to determine multiple hyperplanes. SVM can be used for anomaly detection by labeling the training data as normal and abnormal or anomaly, there can be classes within the abnormal data to get more specific classes.

Authors of [2] used SVM for runtime anomaly detection using Hardware Performance Counters (HPC). They defined a SVM model for anomaly detection in applications. They took the CHStone benchmark applications suite for training. SVM was trained to detect anomalies based on features from HPC. They used a RBF (radial basis function) to define the hyperplane. The model could successfully differentiate between 2 similar applications. They were able to detect anomalies with high accuracy and not one of the anomalies was identified as normal data.

Authors of [3] used SVM for network anomaly traffic detection. They made a classification between legitimate and attack traffic. Four different attack patterns were considered and the model was able to achieve higher accuracy than incrementalLOF and HPStream.

2)K-nearest neighbor (KNN): KNN is mainly used for solving classification problems and just like SVM, it can also be used for regression as well. It makes the assumption that similar things are closer to each other. In this algorithm the data is stored rather than using for training, it is also called as lazy learning where the similarity metric between new data item and the available data is calculated and is sorted and the first k most similar instances or data are considered and a majority rule is applied to determine the class or label of the new data item. So, determining the appropriate K value to use is an important task.

Authors of [4] used KNN for anomaly detection in ELV DC Pico grids - status of different appliances in a circuit. They used KNN with Euclidean distance as similarity metrics and considering only data from a window block rather than whole data. They used a device to convert signals into features that is used by KNN. They tested their model in 3 different circuits and achieved a high accuracy of true positive and False positive.

3)Bayesian Networks: Bayesian networks is a popular supervised machine learning algorithm which is a probabilistic graph model. This algorithm contains two parts of learning - structure and parameter. The structure refers to a directed acyclic graph (DAG) where each node is associated with a random variable and the edges are associated with relationships. The DAG expresses the conditional dependencies and indepencies between the random variables. In parameter learning is to find conditional probability for each random variable in the network. So to build and use the Bayesian networks one must know what are the random variables of the problem, relationships between random variables and conditional probability of each random variable. This algorithm can be used for anomaly detection where there is class attribute and with conditional probability known for each feature, the task would be to find conditional probability given a particular class.

The authors of [5] used Bayesian networks to classify anomalies in network traffic along with Markov Chain monte carlo and Maximum Likelihood Estimation for structure and parameter learning respectively. They used UNB ISCX IDS 2012 and UAN W32.Worms 2008 datasets with 8 different features for training and testing. They achieved a high accuracy for classification of normal traffic and different type of attacks which was more efficient than other methods which used classical datasets

4)Decision Tree: Decision tree is a popular supervised machine learning which is most commonly used for classification than regression. It is also a non-parametric learning algorithm. This algorithm is based on a popular structure - tree, where each internal node represents a feature and each leaf node represents a class. So, the path from root to leaf node is the rule for classification of that class. The main objective of this algorithm is to construct the decision tree which is based on information gain at each node - how much information can the feature give about a class and hence, make decisions. This algorithm can be used for anomaly detection by selecting appropriate features for classification between normal and abnormal data.

The authors of [10] used decision trees to detect anomalies in real time servers - Short Message Service Center (SMSC). They wrote a script to collect about 500 attributes of data from the server and ran the script for 30 days and collected 8600 records. They divided the recorded data into normal and abnormal classes. They achieved a very high accuracy in classification of anomalies and normal data.

DOI 10.17148/IJARCCE.2020.9411

B. Unsupervised machine learning

It is a machine learning technique where only input data is available for training. The main objective in this type of machine learning algorithms is to find common recurring patterns in the data and also to find relations, similarity between data. This type of machine learning algorithm also helps in summarizing and also explaining the features of the data. Different unsupervised machine learning algorithms for anomaly detection are Self Organising Map, K-means, C-means and Adaptive Resonance theory.

1)Self Organising Map (SOM): SOM is an unsupervised machine learning algorithm which is an artificial neural network where it reduces the dimensionality of the input space which is also discrete. This method hence can be used for dimensionality reduction. The main objective of SOM is to make changes in the network such that some parts of the network respond similarly for certain types of inputs. Neurons are initialized with weights which are randomly selected. This algorithm applies competitive learning where distance is calculated between the input and the weight of the neurons and the neurons with weights most similar to the weights of the input are defined as the best matching unit (BMU). The BMU neurons that are close to input in the network of SOM are changed to input weights. The input is fed to the network for a large number of cycles. Hence, the network adjusts itself giving output nodes to mapping the input to groups or patterns in the data. SOM can be used for anomaly detection by giving inputs to find patterns of how normal data are and once the network responds differently, we can infer as anomaly.

Authors of [11] used SOM for anomaly detection on warp knit fabric surfaces was used at 2 levels. On 1st level it was used for detection of type of fabric and at second level it was used to detect defects in the texture. Dataset are not readily available of texture images hence were taken directly from industries and are unpredictable. Since, the defects of texture surfaces are not predictable, unsupervised learning is favorable. 8 different types of defects needed to be detected and SOM was able to achieve decent accuracy of 80%.

2)K-means: K-means is a popular unsupervised machine learning algorithm. The main objective is to find similarity between the data, group similar data as clusters and find common patterns in the data. So K-means tries to find a fixed number of clusters defined by K. K is defined by the user. In case of machine learning the data can be clustered as normal and abnormal data and hence, once the new data - input is applied it can choose one of the clusters by the similarity metric.

Authors of [12] used K-means clustering algorithm to detect anomaly in network traffic. They made modifications on the amount initially determined of clusters. They used the concept of a sliding window, which limits the amount of data transmitted or processed. They used the KDDCup99 dataset for training. They achieved a high accuracy in classification of normal and abnormal traffic

3)Fuzzy C-means (FCM): Fuzzy clustering is different from hard clustering in the sense that in fuzzy clustering, we can have data points belonging to more than one cluster at a given instance, whereas in hard clustering, data points are assigned only to one cluster. Data points belonging to different clusters are as different as possible and those belonging to the same cluster are as alike as possible.

For each data point lying in any cluster initially, the fuzzy membership value is obtained using fuzzy logic. In FCM, we are first required to find out the centroid of the various data points/items. For each of these items, the distance is calculated from the centroid, each cluster has its own centroid. On doing cluster analysis if we find that the data point is “closer” to another cluster, then it is considered a part of that cluster thereafter. This process continues until the clusters so formed reach a constant state.

Authors of [13] presented a way of using fuzzy c means clustering in an unsupervised manner for anomaly detection for intrusion detection and the results were compared with other known methods such as Expectation Maximisation, K-medoids, and it was recorded that FCM obtained fairly better results.

4)Adaptive Resonance Theory (ART): Adaptive resonance theory is a theory on aspects of how the brain processes information. It describes a number of neural network models which use supervised and unsupervised learning methods, and address problems such as pattern recognition and prediction.

This procedure comprises two steps. The first one being the learning step. In this step, a data set with numerous samples which are recorded under normal conditions, according to their similarity, is categorised. The collection of process variables acquired at an instance is called a sample. The result of the learning step, i.e., the categories acquired, are hereafter referred to as normal categories. The second step is the outlier detection step. Each sample is taken and its category is decided by first evaluating the normality using ART. If the normality so obtained can be put into one

DOI 10.17148/IJARCCE.2020.9411

of the normal categories, then the sample does not show any behavior of an anomaly/outlier. Otherwise, a new category is to be formed which can be hereafter called an anomaly for future samples.

The authors of [14] compared the performances of four anomaly detection systems based on the ART model. They were Single ART model, Multiple ART models, Distributed ART model, and Multiple and distributed ART models. A criterion was developed which further complemented the work done on distributed model systems by evaluating the occurrence of the anomaly as well as the quantity, in other words, the anomaly's level.

V. CONCLUSION

In this paper different machine learning algorithms are listed and given a brief description, which can be used for anomaly detection. This review will help researchers and others to get to know different machine learning techniques and their basics for anomaly detection. Every day new kinds of anomalies arise in every application and a technique is needed to adapt for this situation. Hence, it is better to use unsupervised machine learning techniques to handle new kinds of anomalies. But in situations where it is guaranteed that only a fixed number of anomalies arise, then supervised learning is a better choice yielding better accuracy results.

REFERENCES

[1]. Shikha Agrawal, Jitendra Agrawal, "Survey on Anomaly Detection using Data Mining Techniques", Published in : Procedia Computer Science,Volume 60, 2015, pp 708-713
[2]. Muhamed Fauzi Bin Abbas, Sai Praveen Kadiyala, Alok Prakash,Thambipillai Srikanthan,Yan Lin Aung, "Hardware performance counters based runtime anomaly detection using SVM", Published in: 2017 TRON Symposium (TRONSHOW), DOI: 10.23919/TRONSHOW.2017.8275073
[3]. Gao Yan "Network Anomaly Traffic Detection Method Based on Support Vector Machine", Published in: 2016 International Conference on Smart City and Systems Engineering (ICSCSE), DOI : 10.1109/ICSCSE.2016.0011
[4]. Yang Thee Quek, Wai Lok Woo, Logenthiran Thillainathan, "IoT Load Classification and Anomaly Warning in ELV DC Picogrids Using Hierarchical Extended k-Nearest Neighbors " ,Published in : IEEE Internet of Things Journal (Volume 7, Issue 2, Feb 2020), pp.863-873
[5]. M. J. Vargas-Muñoz, R. Martínez-Peláez, P. Velarde-Alvarado, E. Moreno-García, D. L. Torres-Roman, J. J. Ceballos-Mejía, "Classification of network anomalies in flow level network traffic using Bayesian networks", Published in: 2018 International Conference on Electronics, Communications and Computers (CONIELECOMP),DOI:10.1109/CONIELECOMP.2018.8327205
[6]. M. Panda, A. Abraham y M. R. Patra, "Hybrid intelligent systems for detecting network intrusions," Security Communication Networks, vol. 8, p. 2741–2749, 2015
[7]. B. Balasingam, P. Mannaru, D. Sidoti, K. Pattipati, P. Willett, "Online Anomaly Detection in Big Data: The First Line of Defense Against Intruders", Studies in Big Data, vol. 24, pp 83-107, 2017
[8]. V. Hodge, J. Austin, "A survey of outlier detection methodologies", Artif. Intell. Rev., 22(2), 2004
[9]. H. Jantan, "Human talent prediction in HRM using C4.5 classification algorithm", IJCSE, vol. 02, no. 08, p. 2529, 2010
[10]. Georges Chaaya, Hoda Maalouf, "Anomaly detection on a real-time server using decision trees step by step procedure", Conference: 2017 8th International Conference on Information Technology (ICIT), DOI:10.1109/ICITECH.2017.8079989
[11]. Dimuthu Wijesingha, Buddhika Jayasekara, "Detection of defects on Warp-knit Fabric surfaces Using Self Organising Map", Published in: 2018 Moratuwa Engineering Research Conference (MERCon), DOI : 10.1109/MERCon.2018.8421944
[12]. I Wayan Oka Krismawan Putra, Yudha Purwanto, Fiky Yosef Suratman, "Modified k-means algorithm using timestamp initialization sliding window to detect anomaly traffic", Published in: 2015 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC), DOI : 10.1109/ICCEREC.2015.7337042
[13]. B S Harish, S V Aruna kumar, "Anomaly based intrusion detection using modified Fuzzy clustering", Published in: International Journal of Interactive Multimedia and Artificial Intelligence · January 2017,pp 54-59.
[14]. Yoshinari Hori, Yoshiharu Hayashi, Takaaki Sekiai, Shinji Hasebe, "Evaluation of Performance of Anomaly Detection Systems Based on Adaptive Resonance Theory", Journal of Chemical Engineering of Japan, Vol. 52, No. 11, pp. 843–850, 2019

DOI 10.17148/IJARCCE.2020.9411