iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 8 **Issue:** VI **Month of publication:** June 2020

**DOI:** http://doi.org/10.22214/ijraset.2020.6155

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429*
*Volume 8 Issue VI June 2020- Available at www.ijraset.com*

# A Study of Anomaly Detection Techniques

Sanketh Harnoorkar[1], Rekha B.S.[2]
[1]*Student*, [2]*Asst. Professor, Dept. of ISE, RVCE*

***Abstract:*** ***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. Two assumptions have to hold off for the anomaly detection to be effective - number of normal patterns should be higher than the no of anomalies and anomalies should be distinguishable from normal patterns. Anomalies or outliers were detected as a part of cleaning the data. However, it was later found in 2000 that detection of anomalies can help in solving real world problems. Anomaly detection can help solving problems such as intrusion finding, fraud detection in credit card transactions, system health monitoring and industrial damage detection. The anomalies are classified into point, contextual and collective anomalies. If it's found that only a single data point differs significantly from other data points in terms of attributes, then it is called a point anomaly.***

***The main aim of anomaly detection is to detect cases which are not usually found within data that is of a similar kind.. Anomalies in the data can occur for different reasons.. Data analysts find these anomalies intriguing and interesting. . The bottom line is increase up of the time and the reduction of any downtime.***

***The paper makes an attempt to analyse the various anomaly detection techniques by bringing out a survey of the existing research which exists in this domain.***

## I. ANOMALY DETECTION ON TIME SERIES DATA

Data from the time series are data which are collected at different times. This is against cross-sectional data which at a single point in time observes individuals , companies, etc. Due to the fact that data points are collected in time series at adjacent time periods, correlation between observations is potential. This is one of the features that sets time series data apart from cross-sectional. Time series data is often seen in a variety of domains in today's world. They are seen in the field of Economics, Epidemiology, Social Sciences, ,Medicine, Physical sciences. [1] explores the isolation forest method to detect anomalies and measure how severe they are. It emphasises how the present anomaly detection strategies have their shortcomings in which many of them set an arbitrary threshold for detecting anomalies and the presence or absence or anomalies is determined based on how close the present feature is close to the threshold. The paper talks about the shortcomings of this approach and talks of a better alternative to this in the form of random forests. This is useful whenever there are many anomaly points and there is confusion on where to start the search from. The paper also says that the research is backed by experimental evidence. It deals with the main problem on finding out the threshold of anomaly. It proposes a k means clustering to decide on an adaptive algorithm to divide between outliers and normal values. It initially proposes a normal k means algorithm on two clusters and then proposes some modifications and enhancements to it and gives an anomaly score to each of the outlier detected. The methodology includes using an adaptive algorithm based on feature points, pattern feature extraction and mapping to a feature space. After this the isolation forest part comes in where with different tree construction algorithms a random forest is constructed. After this, anomaly score is calculated and anomaly sets are clustered based on K-means. After this we calculate the K nearest points of each point. According to the paper, experiments prove the Improved iForest algorithm is better suited for use in Big hydrological time series data and high characteristics precision in detecting anomalies.. [2] This paper deals with seasonal time series data and proposes an unsupervised algorithm. This proposed a look back window and takes a subset and takes a median as expected value. The actual value of new data is considered and compared with the median, it also proposes a percentile window and also a 3sigma detector. They've also suggested a validation procedure. They've named it as MULDER. According to the results they've obtained results better than amazon and twitter's algorithm. They've proposed the algorithm mainly for streaming systems. They've also laid out the challenges presently with streaming services. Mainly the problem with not having test/validation sets, not having a set normal defined, difficulties in accurately naming anomalies mainly due to their volume, velocity and variety of data and also manual parameter tuning. [4] deals with anomaly detection on time series data using markov chain. It first details about feature extraction, which involves transforming a pressure data segment into markov chain, one step probability, amplitude information of time series and combined feature vector. The dataset is obtained from Supervisory Control And Data Acquisition (SCADA) system from Tinjin, China.In the paper, a new feature extraction method is introduced and validated with the dataset.

## II. ANOMALY DETECTION ON TOPOLOGY DATA

[5] This paper deals with anomalies in Wireless Sensor Networks. This paper brings out their feature and compares it with outlier factors based on connectivity which is used as the data de Scoring function instances on the distribution of Next Data Instances. Wireless sensor networks ( WSNs) were widely used as well as deployed in different applications , for monitoring of agricultural related items and that of industrial items are few cases to quote of, to ensure their ease of use. Since Wireless Sensor Networks are low cost they are quite vulnerable to modifications due to external sources, that is to say, the environment or intrinsic changes Factors, i.e. faults with hardware or software. The issue can be, oftentimes, uncover by detecting unforeseen behaviors (Anomalies) of apparatuses. But detection of anomalies in WSNs is

Subject to the following objective: (i) Calculation limits Connectivity, (ii) Environment Dynamics and Network topology and (iii) the need to take measures in real time back to anomalies. This paper also details how the proposed network is better than the existing ones such as:

External Connectivity Factor (COF),it uses the metric of connectivity to define the scoring function for the next set of instances.

Local outlying distance factor (LDOF) uses metric of local distance for defining relationships Data instances and average distance ratio calculated from one particular instance of data to an average of it's distance from neighbours.

[6] This paper gives out an interesting perspective on developing solutions for anomalies in large system with the concept of E-alarms. This paper mainly deals with the process of detecting anomalies in case of a large attack on network such as DDoS attack and worm attack. The paper proposes anomaly detection technique based on clustering and correlativity of random variables. The paper also details on another method with reference to a scalable and a low network topology overhead measurement. This model mainly relies on bayesian network to detect outliers and anomalies. A drawback includes gradual training, but that can be overcome by finding out features and determining major correlations between the features.

[7]This survey Uses GHSOM 's advantage in network analysis Traffic data and display pattern distribution with The relationship here is hierarchical. The geometric stage of mining Distances between the patterns and their descriptive data Are disclosed in the topological space. [8]The sample density and Size of each node can help detect anomalous traffic on the network. In this study extends the detection stage, the traditional GHSOM And uses the vector support machine (SVM)[6] for classification Data on network traffic into predefined categories. Suggested Approaches (1) help people understand the behaviors of Anomalous data concerning network traffic (2) provide effective classifications. The private dataset and the public dataset are Used for assessment of the approach proposed. The expected outcome is to confirm that the approach proposed can help us understand Network traffic data and an effective detection mechanism for identifying abnormal behavior.

[9]An anomaly detection system is proposed in this work proposes three different levels to monitor the network. First, to detect behavioral changes, Simple Network Management Protocol(SNMP) data is collected and then they are compared with normal traffic data. In the second, they make a graph of dependency representing the relationships between the objects. It is used as a means to study in detail about the behaviors at first level, which can help us be sure about the device level anomalies. The alerts at second level are grouped according to information on network topology in the third level of analysis, information on the context in which anomaly occurred is conveyed to the network administrators. Testing was carried out in an actual environment, and the results achieved were good

## III. TELEMETRY ANOMALY DETECTION

[10]The operations engineers monitor spacecraft housekeeping telemetry at flight control centers using simple trend analysis tools and limit checking tools. The latest developments in anomaly detection techniques helps us to implement more sophisticated systems aimed at increasing the current proprietary tools to give required support for decisions.The work covered in this paper describes the project's first phase which includes performing anomaly detection on the individual telemetry channels and also providing the current high-criticality anomaly report to the operator.

Detection of anomalies is an essential task that allows for condition-based maintenance (CBM), fault diagnostics and system prognostics of complex satellite systems. Some prognostic methods were used to detect outliers among satellite telemetry series. The advantage was that it required no labels and also gave a very strong outlier detection performance. In particular, the models for the probability prediction. [11]This paper proposed a prediction anomaly for LUBE intervals Model detection based on multi-object optimisation.A new concept of knee point was introduced to be incorporated into the forecast interval. Compared to the original model of interval forecast, Combining optimized range width and interval Officially, it was documented that this method performed better than the original method. Additionally, some work needs to be done in the future for better results. This paper is using the data from mostly repetitive and obvious information, so in the future work, further application of the algorithm to aperiodic dataPerformance assessment. Secondly, model training is performed offline.

## REFERENCES

[1] Y. Qin and Y. Lou, "Hydrological Time Series Anomaly Pattern Detection based on Isolation Forest," 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 2019, pp. 1706-1710, doi: 10.1109/ITNEC.2019.8729405.

[2] L. von Werra, L. Tunstall and S. Hofer, "Unsupervised Anomaly Detection for Seasonal Time Series," 2019 6th Swiss Conference on Data Science (SDS), Bern, Switzerland, 2019, pp. 136-137, doi: 10.1109/SDS.2019.00036.

[3] D. Zang, J. Liu and H. Wang, "Markov chain-based feature extraction for anomaly detection in time series and its industrial application," 2018 Chinese Control And Decision Conference (CCDC), Shenyang, 2018, pp. 1059-1063, doi: 10.1109/CCDC.2018.8407286.

[4] Y. Tsou, H. Chu, C. Li and S. Yang, "Robust Distributed Anomaly Detection Using Optimal Weighted One-Class Random Forests," 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 2018, pp. 1272-1277, doi: 10.1109/ICDM.2018.00171.

[5] M. Sun, Y. Wang and Y. Luo, "E-Alarm: An Anomaly Detection System on Large Network," 2009 International Joint Conference on Artificial Intelligence, Hainan Island, 2009, pp. 555-558, doi: 10.1109/JCAI.2009.197.

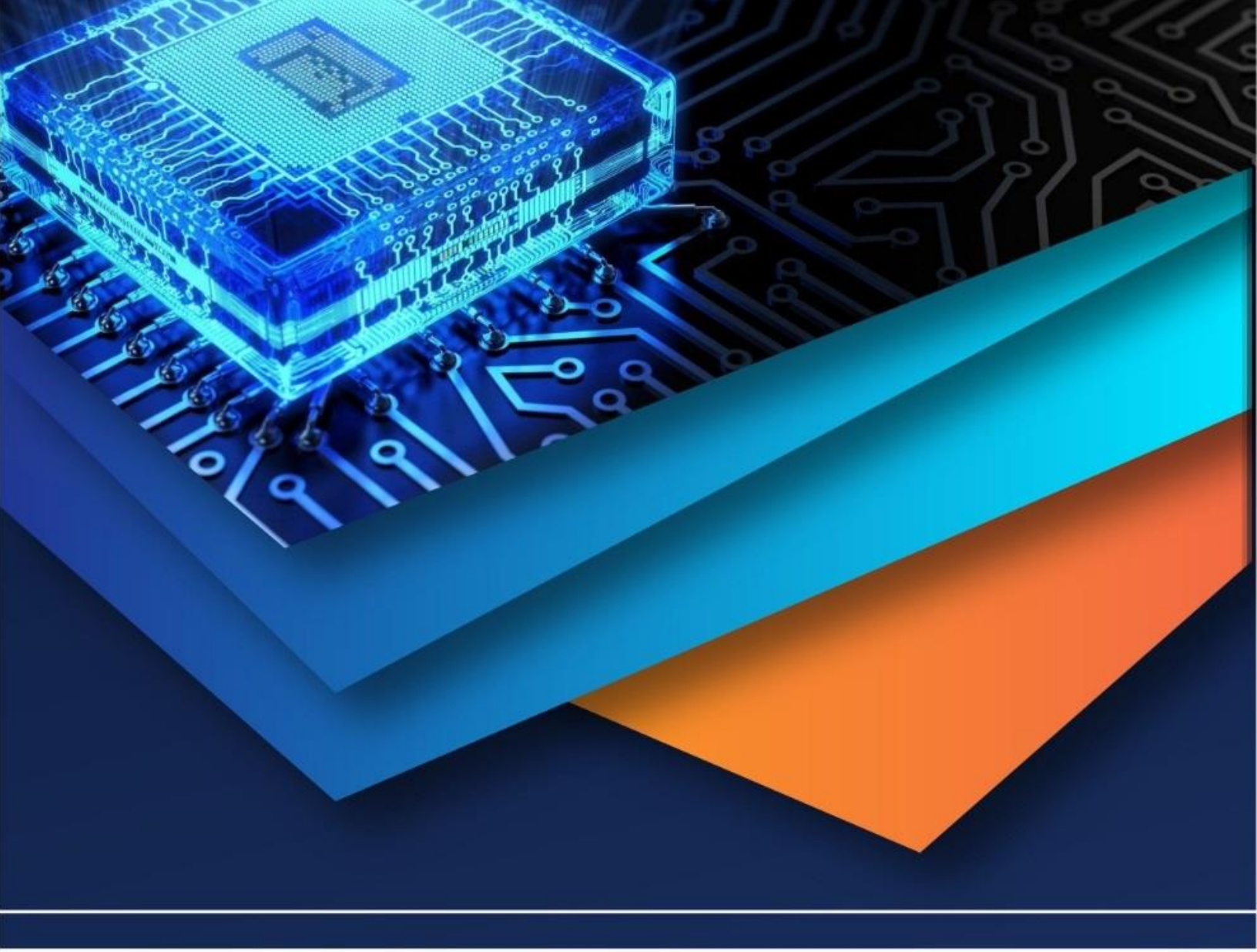
[6] S. Huang and Y. Huang, "Network traffic anomaly detection based on growing hierarchical SOM," 2013 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Budapest, 2013, pp. 1-2, doi: 10.1109/DSN.2013.6575338.

[7] B. B. Zarpelao, L. S. Mendes, M. L. Proenca and J. J. P. C. Rodrigues, "Three levels network analysis for anomaly detection," SoftCOM 2009 - 17th International Conference on Software, Telecommunications & Computer Networks, Hvar, 2009, pp. 281-285.

[8] M. M. Fernández, Y. Yue and R. Weber, "Telemetry Anomaly Detection System Using Machine Learning to Streamline Mission Operations," 2017 6th International Conference on Space Mission Challenges for Information Technology (SMC-IT), Alcala de Henares, 2017, pp. 70-75, doi: 10.1109/SMC-IT.2017.19.

[9] J. Pang, D. Liu, Y. Peng and X. Peng, "Anomaly Detection for Satellite Telemetry Series with Prediction Interval Optimization," 2018 International Conference on Sensing,Diagnostics, Prognostics, and Control (SDPC), Xi'an, China, 2018, pp. 408-414, doi: 10.1109/SDPC.2018.8664879.

[10] X. Li, T. Zhang, K. Li and Y. Liu, "Spacecraft Telemetry Data Anomaly Detection Based On Multi-objective Optimization Interval Prediction," 2019 Prognostics and System Health Management Conference (PHM-Qingdao), Qingdao, China, 2019, pp. 1-8, doi: 10.1109/PHM-Qingdao46334.2019.8942998.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY