OUTLIER DETECTION USING IOT AND MACHINE LEARNING

Deore Priyanka[1], Devhadrao Gayatri[2], Kadam Monali[3], Bornare Apoorva[4], Prof.Chandgude A.S[5]

Department of Computer Engineering,
SND College of Engineering and Research Center, Yeola, Nashik 423401.

***Abstract*: Outliers once upon a time regarded as noisy data in statistics, has turned out to be an important problem which is being researched in diverse fields of research and application domains. Many outlier detection techniques have been developed specific to certain application domains, while some techniques are more generic. Some application domains are being researched in strict confidentiality such as research on crime and terrorist activities. The techniques and results of such techniques are not readily forthcoming. A number of surveys, research and review articles and books cover outlier detection techniques in machine learning and statistical domains individually in great details. In this paper we make an attempt to bring together various outlier detection techniques, in a structured and generic description. With this exercise, we hope to attain a better understanding of the different directions of research on outlier analysis for ourselves as well as for beginners in this research field who could then pick up the links to different areas of applications in details.**

***Keywords*: Machine Learning, Processing, Dataset, Support Vector Machine, Database, IOT.**

INTRODUCTION

Outliers once upon a time regarded as noisy data in statistics, has turned out to be an important problem which is being researched in diverse fields of research and application domains. Many outlier detection techniques have been developed specific to certain application domains, while some techniques are more generic. An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. Examination of the data for unusual observations that are far removed from the mass of data. These points are often referred to as outliers. An outlier is an observation of data that does not fit the rest of the data. It is sometimes called an extreme value. When you graph an outlier, it will appear not to fit the pattern of the graph.

MOTIVATION

The term exception, otherwise called irregularity is initially taken from the field of insights. Exceptions can be raised on account of human blunder, machine mistake, mechanical blames and changes in the conduct of framework or might be because of regular abnormality in the climate. Overcome the this all causes data entry. An experiment measurement errors, sampling problems, and natural variation. An error can occur while experimenting/entering data. During data entry, a typo can type the wrong value by mistake

PROBLEM DEFINATION

The term outlier, also known as anomaly is originally taken from the field of statistics. Outliers can be raised because of human error, machine error, mechanical faults and changes in the behavior of system or may be due to natural deviance in the environment.

LITERATURE SURVEY ON

An Improved LOF Outlier Detection Algorithm LOF (Local Outliers Factor) algorithm is a very classic anomaly detection algorithm. In order to detect the outliers more accurately, avoid too much testing error, and ensure the detection can be implemented relatively accurately in the data set without professional knowledge, on the basis of traditional LOF algorithm, an improved detection algorithm LOF Outliers is proposed. According to the different distribution densities of logarithmic data points, all the data point sets A1 that are most likely to become outliers are found out. Then, the information entropy weighted LOF algorithm is used to detect the data set to get the result A2. The point set A1 is intersected with A2 to get the final point set A which is the final result. The experimental results show that the algorithm is feasible, and it is more accurate and contains fewer false detection points[1].

A k-Nearest Neighbor Medoid-Based Outlier Detection Algorithm Outlier detection techniques are well known for identifying a small amount of data objects named outliers that are far away from clusters and exist in sparse regions of data space. However, most outlier detection algorithms based on k nearest neighbors are sensitive to parameter k. The outlier detection algorithms based on clustering rely on specific clustering algorithms, and outliers are by-products. To partially circumvent these problems, motivated by the medoid concept of Kmedoids clustering algorithms, in this paper, we propose a k-nearest neighbor medoid-based outlier detection method that is easy to implement and can provide competing performances with existing solutions. At the same time, a method to determine the parameter k is proposed in combination with the outlier detection algorithm proposed in this paper. Experiments performed on real datasets demonstrate the efficacy of our method [2].

UN-AVOIDS: Unsupervised and Nonparametric Approach for Visualizing Outliers and Invariant Detection Scoring he visualization and detection of anomalies (outliers) are of crucial importance to many fields, particularly cyber security. Several approaches have been proposed in these fields, yet to the best of our knowledge, none of them has fulfilled both objectives,

simultaneously or cooperatively, in one coherent framework. Moreover, the visualization methods of these approaches were introduced for explaining the output of a detection algorithm, not for data exploration that facilitates a standalone visual detection. This is our point of departure in introducing UNAVOIDS, an unsupervised and nonparametric approach for both visualization (a human process) and detection (an algorithmic process) of outliers, that assigns invariant anomalous scores (normalized to [0, 1]), rather than hard binary-decision [3].

PROPOSED SYSTEM

While outlier detection is the method of finding patterns from a given set of data that significantly differs or deviates drastically from the normal or average of the data set. Outlier detection is defined as finding patterns in data that behaves unexpectedly. Objective of outlier detection is to find devices by their behavior that differs from the expected and previously observed in the field of IoT. The evaluation of used machine learning method provides high accuracy of 97.8 percent with proposed outlier detection methods and almost 2 percent improvement in the accuracy of localization process in indoor environment after eliminating outliers.

SYSTEM ARCHITECTURE

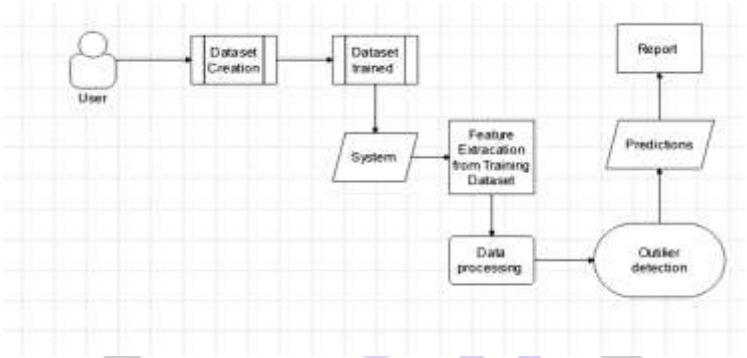

Fig -1: System Architecture Diagram

ADVANTAGES

- To recognize patterns in dataset
- To reduce the detection time
- Training will provide proper accuracy
- Detecting a Oulier automatically
- Easy to system
- Provide better solution in Low Cost

APPLICATIONS

- Personal
- Forensic Department
- Police Department
- Investigation department

CONCLUSION

Hence we are overcoming the drawback of existing system , we are providing the better solution than existing system in affordable cost. We proposed a system which is use to identify the Outlier detection using CNN algorithm, which is based deep learning.

REFERENCES

1] K. K. Almuzaini and A. Gulliver, “Range-based localization in wireless networks using density based outlier detection,” Wireless Sensor Network, vol. 2, no. 11, p. 807, Nov. 2010.

[2] H. Aly, A. Basalamah, and M. Youssef, “Accurate and energy-efficient GPS-less outdoor localization,” ACM Trans. Spatial Algorithms Syst., vol, 3, no. 2, pp.
1–31, July 2017.

[3] Y. Zhang, N. Meratnia, and P. Havinga, “Outlier detection techniques for wireless sensor networks: A survey,” IEEE Common. Surveys Tuts., vol, 12. no. 2,
pp. 159–170, 2010.

[4] M. Ahmed, A. N. Mahmood, and J. Hu, “A survey of network anomaly detection techniques,” J. Network Computer Applicat., vol. 60, pp. 19–31, 2016.

[5] I. Cramer et al., “Detecting anomalies in device event data in the IoT,” in Proc. IoTBDS, 2018, pp. 52–62

[6] P. Yang and B. Huang, "A modified density based outlier mining algorithm for large dataset," in Proc. FITME, 2008.

[7] J. F. Liu and Z. Ning, "Localization anomaly detection for wireless sensor networks," in Proc. IEEE ICIS, 2010.
[8] Y. C. Chen and J. C. Juang, "Outlier-detection-based indoor localization system for wireless sensor networks," Int. J. Navigation Observation, vol. 2012, pp. 1687–5990, 2012.

[9] S. Capkun, S. Ganeriwal, F. Anjum, and M. Srivastava, "Secure RSSbased localization in sensor networks," Technical report/Swiss Federal Institute of Technology Zurich, Department of Computer Science, vol 529, 2011.

[10] Z. Yang et al., "Detecting outlier measurements based on graph rigidity for wireless sensor network localization," IEEE Trans. Veh. Technol., vol. 62, no. 1,
pp. 374–383, 2012