IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Survey On Real Time Anomaly Detection Techniques In IoT Data Streams

Mrs. S.Subha
Assistant Professor
Department of computer applications
Bishop Heber College(Autonomous), Tiruchirappalli

Abstract

Anomaly detection is one of the biggest problems in a wide range of applications. Anomaly detection has enticed substantial attention from the research community in the past few years owing to the development of IoT technology, low-cost solutions, and great impact in different application domains. It is concerned with the detection of new or unforeseen observations or sequences within the data being generated. Most of the existing anomaly detection techniques are extremely specific to the individual use-case, demanding expert knowledge of the techniques as well as the situation to which it is being applied. The IoT as a rapidly emerging field provides several opportunities for this type of data analysis to be implemented. This review provides a deep understanding on the various anomaly detection techniques being applied to IoT data. Finally the current challenges being faced in the IoT domain while detecting anomalies have been summarized to showcase the potential research opportunities for the future.

Keywords —IoT, Anomaly detection, review

I. Introduction

The Internet of Things (IoT) is a growing network of physical devices—"things"—which are embedded with sensors, software, and other technologies for connecting and exchanging data with other devices and systems over the internet without human intermediation [1]. These devices may include sensors, actuators, or smart devices which are able to sense or communicate with their internal and external surroundings. The progression of IoT has been empowered by the development of various cost effective sensing and computing solutions able to work in environments which would have previously been unachievable. IoT is presently experiencing great expansion with estimates of global economic impact equals to $11.1 trillion per year by 2025 [2]. When data analysis is carried out in the IoT data, there is often an indispensible necessity to detect new or strange states within a system being monitored by the sensors installed in the surroundings of that system. This method is often stated as anomaly detection outlier detection or event detection[12].

In any anomaly detection system, the primary step is to identify the nature of the collected data stream viz. time series data, spatial data, or graph data [13] [14]. This facilitates in choosing the right anomaly detection technique. The second step is finding out the kind of anomaly from a predefined set viz. point anomaly, contextual anomaly, and collective anomaly.

A. Point Anomaly

Point anomaly is an individual observation in the data stream that is far from the rest of the data. It is also known as an ''outlier'' [4]. Point anomaly

has to be detected before processing or doing further analysis on the data.

B. Contextual Anomaly

The contextual anomaly is an observation that is normal in one situation, while abnormal in another situation. Such type of anomaly needs understanding of context and also known as conditional anomalies [4]. This type of anomalies is common in time series data streams.

C. Collective Anomaly

A sequence of observations is explored to comprehend the collective behaviour of the data stream. Any deviation from the usual pattern may result in collective anomaly pertaining to entire data patterns over consecutive time intervals [4].

The third step is identifying the availability of training data to construct an anomaly detection system viz. supervised anomaly detection system, unsupervised anomaly detection system or semi-supervised anomaly detection system.The research community has already established certain techniques to detect anomalies present in the historical data, real-time analysis and prediction of rare behaviours in IoT environment. The focus of this review is to study the various real time anomaly detection techniques in IoT data streams. At present, most of the anomaly detection techniques need substantial human intervention to extract and interpret the data being generated. Though it is comparatively easier to observe a small subset of data by deploying existing anomaly detection techniques to detect the trends and patterns which are of interest, but when the number of interconnected devices upsurges, there is a need for developing automated and efficient anomaly detection techniques to identify the trends and patterns of the most important events observed. In Section 2, an overview of related works is presented. Section 3 presents the research gaps. Finally Section 4 concludes the paper.

II. Review of literature

The robust anomaly detection technique becomes essential in any intelligent environment [1], to handle the unusual situations appropriately. The following literature is found and categorized according to defined research questions.

Mohsin Munir et al. [5] proposed a novel deep learning-based anomaly detection approach called DeepAnT for time series data. Various anomalies such as point anomalies, contextual anomalies in the time-series data were detected using this approach. The proposed approach deployed unlabelled data to understand the nature of the distribution which was employed to predict the normal behaviour of the time-series data. Times series predictor and anomaly detector were the two modules used. The first module used deep convolutional neural network(CNN)to forecast the next time stamp on the defined horizon whereas the second module took a window of time series and estimated the next time stamp. The forecast value was supplied to the anomaly detector module, which in turn tagged the matching time stamp either as usual or unusual. In case of DeepAnT, a model could be trained on relatively tiny data sets but with great generalization capabilities because of the effective parameter sharing of the CNN. Since the proposed DeepAnT was an unsupervised anomaly detection approach, it did not need anomaly labels while generating models. It was claimed that the proposed DeepAnT could be applied to real situations.

Nusaybah Alghanmi et al. [6] proposed a hybrid learning model that deployed both clustering and classification to automate the labelling process and detected anomalies in IoT data. The proposed model consisted of two phases namely automatic labelling and detecting anomalies. Hierarchical Affinity Propagation was employed to assemble the data into normal clusters and anomaly clusters. After receiving the labelling data from the clustering phase, the Decision Trees (DTs) were trained using this data and also classification was done for future unseen data. It was found that the proposed hybrid learning model based on clustering and classification (HLMCC) outperformed the decision trees on the originally labelled datasets in terms of evaluation metrics viz. False Positive Rate (FPR), recall, precision and the Area Under the Precision-Recall curve (AUCPR).

Haotian Chang et al. [7] proposed a 3-hierarchy joint local and global anomaly detection framework named as HADIoT. In the proposed framework, IoT devices generated sensory data were pre-processed using the re-framing, normalization, complexity reduction via Principal Component Analysis, and symbol mapping

techniques. After pre-processing, the data was sent to the local edge servers for carrying out local anomaly detection. This processed data was later sent from the edge server to the cloud server for global anomaly detection. Due to the local anomaly detection and global anomaly detection processes, high detection accuracy was accomplished. During the local anomaly detection process, the Gated Recurrent Unit was employed to ensure the data pattern consistency of individual devices whereas in case of global anomaly detection process, Conditional Random Fields was deployed to analyse the data pattern correlations among various IoT devices. The performance of the proposed framework was empirically appraised through simulations, using the Information Security Center of Excellence (ISCX)2012 real world dataset. Simulation outcomes proved the efficiency of the proposed framework in terms of True Positive Rate, False Positive Rate, Precision, Accuracy and F_score, in compared with other three benchmark schemes. Hongyu Sun et al. [8] proposed an effective anomaly detection approach in multiple multi-dimensional data streams. The drawbacks such as failure to handle streaming data, more time consumption while handling streaming applications in the existing tree isolation based detection techniques were overcome by the proposed approach. The proposed approach was based on the stream pre-processing, locality sensitive hashing and dynamic isolation forest. Extensive experiment studies on four real-world datasets demonstrated that the proposed approach attained high accuracy in compared with other contemporary approaches which were taken for comparative analysis.

Yi Liu et al. [9] deployed a federated learning framework to permit decentralized edge devices to collaboratively train an anomaly detection model. An Attention Mechanism based Convolutional Neural Network-Long Short Term Memory (AMCNN-LSTM) model was proposed to precisely discover anomalies. The AMCNN-LSTM model employed attention mechanism-based CNN units to extract significant fine-grained features and eliminated the memory loss and gradient dispersion issues. The proposed model retained the benefits of LSTM unit to forecast time series data. A gradient compression mechanism based on Top-k selection was proposed to enhance communication efficiency. Comprehensive experiment studies on four real datasets viz. Power Demand, Space Shuttle, ECG Engine proved that the proposed framework had identified the anomalies precisely and reduced the communication overhead by fifty per cent in compared to the existing federated learning frameworks and methods viz. CNN-LSTM, LSTM, Gradient Recurrent Unit (GRU), Stacked Auto Encoder (SAE) and Support Machine Vector (SVM) methods which did not deploy gradient compression mechanism.

Vikram Patil et al. [10] proposed a method called GeoSClean which cleansed the GPS trajectory data by deploying an innovative anomaly detection technique at the same time maintained the location of the users confidential. Anomaly points were detected by taking into account the properties of the GPS trajectory data viz. distance, acceleration, and velocity. The hypothesis testing based anomaly detection method was validated to discover anomalous points with high confidence. Comprehensive experimentations conducted on the real-life datasets proved the efficiency of the proposed method in detecting anomalous points over the existing methods. ZhipengLiu et al. [11] deployed different machine learning algorithms to efficiently spot anomalies on the IoT Network Intrusion Dataset. The dataset which was taken for this research was designed specifically using the smart home IoT devices. Multiple machine learning algorithms were used to conduct experiment on the loT Network Intrusion Dataset. It was found that the second highest accuracy of 99% was attained using KNN while the average runtime was two minutes. XGBoost showed the next highest accuracy of 97% with just 10.8 seconds of average run time. Fl scores acquired through diverse machine learning algorithms were reliable. The outcomes of the experiment demonstrated the efficiency of each machine learning algorithm deployed in this research for detecting anomalies in the loT Network Intrusion Dataset.

Jianwu Wang et al. [12] proposed a system-level anomaly prediction in manufacturing (SAPIM) framework to perfectly foretell system anomalies

from the sensor data so as to save manufacturing maintenance costs and thwart further damages. The proposed system-level anomaly prediction framework detected events across sensors collectively and their temporal dependencies. The proposed system-level anomaly prediction framework mined anomaly dependency graph from sensor data for collective prediction. The proposed approach was applied to the real-world power plant dataset and its feasibility was subsequently evaluated. System-level anomalies in the dependency graph were measured using the similarities between the mined sub sequences with sequences in anomaly report. The review work has been categorized in the following table 1 shown below.

Table 1. Anomaly detection techniques for IoT streaming data

S.No.	Authors	Nature of data	Type of anomaly	Proposed anomaly detection approach/method/technique /framework/Model/ Algorithm	Anomaly detection system	Dataset	Evaluation Metrics/accuracy indicators
1.	Mohsin Munir et al.[5]	Time series data	Point anomaly and contextual	DeepAnT (deep learning-based anomaly detection approach)	unsupervised	433 real and synthetic time series data sets	recall, precision ,F-score, true positive
2	Nusaybah Alghanmi et al.[6]	IoT Time series data	Point anomaly	a Hybrid Learning Model which uses both Clustering and Classification methods(HLMCC)	Supervised and unsupervised	The Labelled Wireless Sensor Network Data Repository (LWSNDR) dataset, LANDSAT SATELLITE dataset	False Positive Rate (FPR), recall, precision, Area Under the Precision-Recall curve (AUCPR)
3	Haotian Chang1 et al. [7]	sensory data	Contextual anomaly	HADIoT: A Hierarchical Anomaly Detection Framework for IoT	supervised	the Information Security CenterofExcellence(ISCX)2012dataset	True Positive Rate, False Positive Rate, Precision,accuracy and F_score
4	Hongyu Sun et al. [8]	Stream data	Point anomaly	effective anomaly detection approach	Unsupervised	Occupancy dataset, Buzz in social media dataset	the AUC (Area under Curve), F1-score
5.	Yi Liu et al. [9]	Time series data	Contextual anomaly	Attention Mechanism based Convolutional Neural Network-Long Short Term Memory (AMCNN-LSTM) model	Unsupervised	Power Demand, Space Shuttle, ECG Engine	Root Mean Square Error (RMSE)
6	Vikram Patil et al. [10]	GPS trajectory data	Point anomaly	GeoSClean method	supervised	Microsoft's Geolife dataset	Residual value
7.	Zhipeng Liu [11]	Smart home devices generate data	-	Intrusion Detection System (IDS)	Supervised, regression	loT Intrusion Network Dataset	Recall , FI score, CPU time
8	Jianwu Wang [12]	Sensor data	collective anomaly	system-level anomaly prediction in manufacturing (SAPIM) framework	supervised	real sensor event set from a coal power plant	Precision, recall

III. Research Gaps

Research on the detection of anomalous behaviour faces many challenges such as

- A big research gap exists to standardize the way to extract data logs and sensory data stream so as to build a model and validate the efficiency of the model in real-world settings
- Still more precise and robust models are in dire need to deal with the complex real-world scenarios
- Well-made methods or techniques are needed to pre-process the datasets in order to extract the meaningful information and knowledge
- The exponential increase in the streaming data generated smart devices cannot be handled by the existing statistical methods
- A big gap also exists in the application and assessment of new models in detecting anomalies in data generated by smart objects
- It has been found that a gap in visualizing the anomalies for analysis
- Lack of fusion techniques to fuse the sensory data streams and aid in the exploration of anomalous behaviour in the data streams
- Still the research community is striving hard to develop efficient techniques to detect anomalies in streaming IoT data

IV. Conclusion

Anomaly detection in IoT data streams is an active research area which has gained importance in recent times. An anomalous detection can eliminate the functional risks; sidestep unseen issues, and downtime of the systems. In this research work, a systematic literature review has been done to evaluate the existing anomaly detection techniques. This review provides a comprehensive overview of the developed techniques, their characteristics and performance measures. Consequently, this information can facilitate the research community to acquire the detailed information on current techniques, approaches and methodologies in the anomaly detection domain, It has been observed that diverse data streams generated by a wide variety of IoT devices need to processed by deploying the new methods and techniques.

References

[1]V. R. Jakkula and D. J. Cook, ‘‘Detecting anomalous sensor events in smarthomedataforenhancingthelivingexperience,’’ inProc.Artif.Intell. Smarter Living, 2011, pp. 1–2.

[2] J. Manyika, M. Chui, P. Bisson, J. Woetzel, R. Dobbs, J. Bughin, and D. Aharon, “The internet of things: Mapping the value beyond the hype,” McKinsey Global Institute, Tech. Rep., 2015.

[3] M. Hung, “Leading the IoT,” Gartner Research, Tech. Rep., 2017.

[4] Fahim, Muhammad, and Alberto Sillitti. "Anomaly detection, analysis and prediction techniques in iot environment: A systematic literature review." *IEEE Access* 7 (2019): 81664-81681.

[5] Munir, Mohsin, Shoaib Ahmed Siddiqui, Andreas Dengel, and Sheraz Ahmed, "DeepAnT: A deep learning approach for unsupervised anomaly detection in time series", IEEE Access 7 2018, pp.1991-2005.

[6] Alghanmi, Nusaybah, Reem Alotaibi, and Seyed M. Buhari. "HLMCC: a hybrid learning anomaly detection model for un labeled data in Internet of Things", IEEE Access 7, 2019, pp. 179492-179504.

[7] Chang, Haotian, Jing Feng, and Chaofan Duan. "HADIoT: A hierarchical anomaly detection framework for IoT." IEEE Access 8 (2020): 154530-154539.

[8] Sun, Hongyu, Qiang He, Kewen Liao, Timos Sellis, Longkun Guo, Xuyun Zhang, Jun Shen, and Feifei Chen. "Fast anomaly detection in multiple multi-dimensional data streams", IEEE International Conference on Big Data , 2019, pp. 1218-1223.

[9] Liu, Yi, Sahil Garg, Jiangtian Nie, Yang Zhang, Zehui Xiong, Jiawen Kang, and M. Shamim Hossain. "Deep anomaly detection for time-series data in industrial iot: A communication-efficient on-device federated learning approach.",IEEE Internet of Things Journal, 2020.

[10] Patil, Vikram, Priyanka Singh, Shivam Parikh, and Pradeep K. Atrey. "Geosclean: Secure cleaning of gps trajectory data using anomaly detection." , IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2018, pp. 166-169.

[11] Liu, Zhipeng, Niraj Thapa, Addison Shaver, Kaushik Roy, Xiaohong Yuan, and Sajad Khorsandroo. "Anomaly Detection on IoT Network Intrusion Using Machine Learning", In International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), 2020 pp. 1-5.

[12] Wang, Jianwu, Chen Liu, Meiling Zhu, Pei Guo, and Yapeng Hu. "Sensor data based system-level anomaly prediction for smart manufacturing" , IEEE International Congress on Big Data (BigData Congress), 2018, pp. 158-165.

[13] Widanage, Chathura, Jiayu Li, Sahil Tyagi, Ravi Teja, Bo Peng, Supun Kamburugamuve, Dan Baum, Dayle Smith, Judy Qiu, and Jon Koskey. "Anomaly detection over streaming data: Indy500 case study.", in IEEE 12th International Conference on Cloud Computing (CLOUD), 2019, pp. 9-16.

[14] Chenaghlou, Milad, Masud Moshtaghi, Christopher Leckie, and Mahsa Salehi. "An efficient method for anomaly detection in non-stationary data streams", In IEEE Global Communications Conference, 2017, pp. 1-6.

[15] Cook, Andrew A., Göksel Mısırlı, and Zhong Fan. "Anomaly detection for IoT time-series data: A survey", IEEE Internet of Things Journal, Vol. 7, Issue. 7, 2019, pp. 6481-6494.