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

A research article on anomaly deduction techniques in IoT data

s.subha, Dr. J.G.R.Sathiaseelan

Research scholar, Associate Professor

Bishop Heber College/Bharathidasan University

Abstract

Anomaly detection is a significant problem in a number of applications. Due to the advancement of IoT technology, low-cost solutions, and significant influence in several application fields, anomaly detection has attracted significant attention from the research community in recent years. The discovery of innovative or unpredicted observations or sequences within the generating data is what it is concerned with. The majority of the available anomaly detection approaches are highly specialised to the unique use-case and need professional knowledge of both the methodology and the context. Numerous opportunities exist to employ this kind of data analysis in the IoT, a rapidly developing industry. This analysis offers a thorough grasp of the many anomaly detection methods used with IoT data. Finally, a summary of the present difficulties encountered in the IoT area while identifying anomalies has been provided to highlight prospective future research prospects.

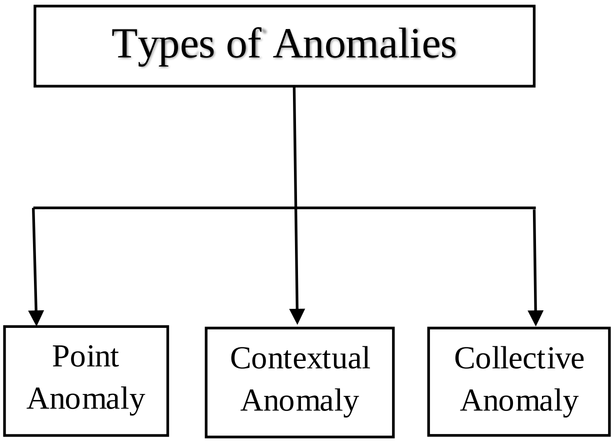
Keywords —IoT, Anomaly detection, analysis

I. Introduction

The Internet of Things (IoT) is a developing system of physical objects, or "things," that are equipped with sensors, software, and other technologies to connect to and exchange data with other objects and systems through the internet without the need for a human intermediary [1]. These gadgets might be smart gadgets, sensors, or actuators that can perceive or communicate with their internal and external environments. The development of several low-cost sensor and computing systems capable of operating in settings where it would have previously been impossible has facilitated the growth of IoT. According to projections, the worldwide economic effect of IoT will reach $11.1 trillion annually by 2025 [2]. The detection of novel or unusual states in a system being monitored by the sensors located in that system's surrounds is frequently a crucial need when data analysis is carried out on IoT data. This technique is frequently described as anomaly, outlier, or event detection [12].

The first stage in any anomaly detection system is to determine the type of data stream that has been gathered, such as time series data, geographical data, or graph data [13] [14]. This makes it easier to select the best anomaly detection method. Finding the kind of anomaly from a pre-set set—

point anomaly, contextual anomaly, and collective anomaly—is the second stage.



A. Point Anomaly

A point anomaly is a single observation that is separated from the rest of the data in the data stream. It's sometimes referred to as a "outlier" [4]. Prior to processing or conducting additional analysis on the data, point anomaly must be found.

B. Contextual Anomaly

An observation that is normal in one setting but odd in another is referred to as a contextual anomaly. These anomalies, usually referred to as conditional anomalies, need context comprehension [4]. In time series data streams, this kind of abnormalities are frequent.

C. Collective Anomaly

In order to understand how the data stream behaves collectively, a series of observations is investigated. Any variation from the typical pattern may cause a general anomaly affecting all data patterns across successive time spans [4].

Finding training data that can be used to build a supervised anomaly detection system, an unsupervised anomaly detection system, or a semi-supervised anomaly detection system is the third phase. The research community has already developed a number of methods to identify anomalies in historical data, perform real-time analysis, and forecast unusual behaviours in an Internet of Things context. This review focuses on analysing the various real-time anomaly detection methods used in IoT data streams. The majority of anomaly detection methods now in use require significant human input in order to extract and understand the generated data. Although it is relatively simple to observe a small subset of data using existing anomaly detection techniques to identify the trends and patterns that are of interest, as the number of interconnected devices increases, it becomes increasingly necessary to develop automated and efficient anomaly detection techniques to identify the trends and patterns of the most significant 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

In order to effectively manage the uncommon occurrences, a strong anomaly detection approach becomes crucial in any intelligent environment [1]. The following literature has been discovered and is organised in accordance with the stated research questions.

Yi Liu et al. [1] set up a federated learning system to let distributed edge devices train an anomaly detection model collectively. To properly find anomalies, an Attention Mechanism based Convolutional Neural Network-Long Short-Term Memory (AMCNN-LSTM) model was suggested. The AMCNN-LSTM model used CNNs with attention mechanism-based features to extract important fine-grained features and solve the gradient dispersion and memory loss problems. The advantages of the LSTM unit for forecasting time series data were kept in the suggested model. To improve communication effectiveness, a gradient compression strategy based on Top-k selection was suggested. Comprehensive experiment studies on four real datasets, including Power Demand, Space Shuttle, and ECG Engine, demonstrated that the proposed framework had accurately identified the anomalies and reduced the communication overhead by 50% when compared to the currently available federated learning frameworks and methods viz. Methods without a gradient compression technique include CNN-LSTM, LSTM, Gradient Recurrent Unit (GRU), Stacked Auto Encoder (SAE), and Support Machine Vector (SVM).

Jianwu Wang et al. [2] suggested the System-Level Anomaly Prediction in Manufacturing (SAPIM) framework to accurately forecast system anomalies from sensor data to prevent future damage and manufacturing maintenance expenses. The recommended system-level anomaly prediction method discovered simultaneous occurrences across several sensors and their temporal connections. In order to make collective predictions, the suggested system-level anomaly prediction framework mined anomaly dependency network from sensor data. The proposed method was tested against the dataset of actual power plants, and its viability was then assessed. By comparing the mined sub sequences with the sequences in the anomaly report, system-level anomalies in the dependency graph were quantified.

ZhipengLiu et al. [2] used various machine learning techniques to quickly identify anomalies in the IoT Network Intrusion Dataset. The dataset used in this study was created expressly for use with smart home IoT devices. On the loT Network Intrusion Dataset, experiments were conducted using a variety of machine learning algorithms. KNN was determined to have the second-highest accuracy, at 99%, with an average runtime of two minutes. With just 10.8 seconds of average run time, XGBoost displayed the second-highest accuracy of 97%.Fl scores obtained using various machine learning methods were reliable. The experiment's results showed how effective each machine learning method used in this study was at finding anomalies in the loT Network Intrusion Dataset.

Vikram Patil et al. [3] put out a technology termed GeoSClean that, while maintaining user privacy, used a cutting-edge anomaly detection technique to clean GPS trajectory data. By examining the GPS trajectory data's distance, acceleration, and velocity characteristics, anomaly spots were found. To find anomalous spots with high confidence, the hypothesis testing-based anomaly detection method was validated. Extensive tests performed on real-world datasets showed the effectiveness of the suggested technique in identifying anomalous spots.

Mohsin Munir et al. [4] proposed DeepAnT, a unique deep learning-based anomaly detection method for time series data. By employing this method, various anomalies in the time-series data, including point anomalies and contextual anomalies, were found. The suggested method made use of unlabelled data to comprehend the nature of the distribution and forecast the typical behaviour of the time-series data. The two modules employed were the times series predictor and the anomaly detector. The first module used a deep convolutional neural network (CNN) to predict the subsequent time stamp on the specified horizon, whereas the second module predicted the subsequent time stamp using a window of time series. The anomaly detecting module received the forecast value and labelled the corresponding time stamp as either ordinary or exceptional. With DeepAnT, a model might be trained on very small data sets while yet having excellent generalisation abilities because to the CNN's efficient parameter sharing. As an unsupervised anomaly detection method, the suggested DeepAnT does not require anomaly labels for creating models. The potential use of the proposed DeepAnT in actual circumstances was asserted.

Nusaybah Alghanmi et al. [5] suggested a hybrid learning model that used classification and clustering to automate labelling and find anomalies in IoT data. The two phases of the proposed model were automatic labelling and anomaly detection. The data were grouped into normal clusters and anomalous clusters using Hierarchical Affinity Propagation. The Decision Trees (DTs) were trained using the labelling data obtained during the clustering phase, and classification for upcoming unobserved data was completed. In terms of assessment metrics, it was discovered that the suggested hybrid learning model based on clustering and classification (HLMCC) performed better than decision trees on the initially labelled datasets. False Positive Rate (FPR), recall, precision, and Area Under the Precision-Recall Curve (AUCPR).

Haotian Chang et al. [6] proposed the HADIoT framework, a three-hierarchy joint local and global anomaly detection system. In the suggested framework, pre-processing approaches for sensory data from IoT devices included re-framing, normalisation, complexity reduction through

Principal Component Analysis, and symbol mapping. The data was pre-processed before being delivered to the local edge servers to perform local anomaly detection. Later, the edge server forwarded this processed data to the cloud server for widespread anomaly detection. Achieving high detection accuracy was made possible by the local and global anomaly detection processes. The Gated Recurrent Unit was used during the local anomaly detection phase to verify that each device's data pattern was consistent, while Conditional Random Fields were used during the global anomaly detection process to analyse the correlations between different IoT devices' data patterns. Using the Information Security Centre of Excellence (ISCX)2012 real world dataset, simulations were used to experimentally evaluate the performance of the proposed architecture. In comparison to the other three benchmark systems, simulation results demonstrated the suggested framework's effectiveness in terms of True Positive Rate, False Positive Rate, Precision, Accuracy, and F score.

III. Research Gaps

Numerous obstacles must be overcome for research to successfully identify abnormal behaviour, including

- There is a significant research gap in standardising how to collect data logs and sensory data streams in order to create models and test their effectiveness in practical contexts.
- To deal with the complicated real-world events, even more exact and reliable models are urgently required.
- In order to extract useful information and knowledge from datasets, well-made procedures or approaches are required. The exponential growth of streaming data produced by smart devices makes it impossible for the available statistical methods to keep up.
- There is a major difference in the evaluation and use of innovative models for identifying abnormalities in data produced by smart things.
- A gap in displaying the anomalies for analysis has been discovered.
- The research community is still working very hard to provide effective methods to find abnormalities in streaming IoT data.

IV. Conclusion

In the last several years, the relevance of anomaly detection in IoT data streams has grown. The functional hazards, unforeseen problems, and system downtime may all be avoided with an anomaly detection. A thorough literature evaluation has been conducted in this study to assess the effectiveness of the available anomaly detection methods. The created methodologies, their traits, and performance metrics are thoroughly covered in this study. It has been noted that different data streams generated by a wide variety of IoT devices need to be processed by utilising new methods and techniques, so this information can help the research community acquire the detailed information on current techniques, approaches, and methodologies in the anomaly detection domain.

References

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

[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.