International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 6 | ISSN : 2456-3307

A Robust Architecture for Detecting Outliers in IoT Data using STCPOD Model

Priya Stella Mary, Dr. L. Arockiam

Department of Computer Science, St Joseph College(Autonomous). Trichy, Tiruchirappalli, Tamil Nadu, India

ABSTRACT

Internet of Things (IoT) is an ecosystem of interconnected physical devices that are accessible through the internet so that these devices can collect and exchange data. Outliers in IoT are generated either due to system malfunctions or because of unexpected transformation in the observed phenomenon. A novel outlier detection mechanism is crucial for IoT so as to achieve high detection rate and low false alarm rate by taking into consideration all the characteristics of IoT data while spotting outliers. In this paper a robust Architecture is proposed to efficiently detect outliers in IoT data using STCPOD (a novel STCPOD (Spatially and temporally correlated proximate Outlier Detection) model.

Keywords: IoT, sensors, outliers, outlier detection.

I. INTRODUCTION

Sensors are the essential units in the Internet of Things (IoT). But they are prone to several problems for instance limited energy, limited memory, computational abilities and communication bandwidth. These problems make sensor nodes susceptible to generate inappropriate data [3] and to have less chance of producing steady sensor readings. Thus the probability of generating incorrect sensor readings will increase sharply. Weird sensor readings are produced not only by poor environmental conditions but also by actual events. Indeed, outliers may also be converted into important information. Since in IoT, the sensed data will be supplied as an input for the data mining process so as to gain meaningful insights on observed phenomenon, unreliable data will lead to unsound decisions.

An outlier in IoT may be either an error caused because of system malfunction or an event which has been caused due to unforeseen transformation in the observed phenomenon or a point anomaly representing a single reading largely differs from the rest of the sensor readings or a context anomaly representing a sensor reading that could be considered as an outlier depending upon the context or a collective anomaly representing a group of sensor readings largely differ from the rest of the sensor readings in which case, a group of sensor readings is considered as outliers and not essentially the individual values [2]. An efficient outlier detection technique needs to be employed so as to ensure high detection rate and low false alarm rate. Owing to huge shortfalls in the current outlier detection techniques for general data which are not suitable to perform outlier detection in IoT, it becomes essential to develop new outlier detection techniques appropriate for IoT by taking into consideration all the characteristics of IoT data while spotting outliers.

Generally sensor readings are spatially and temporally correlated. Outliers in IoT data are abnormal data that significantly diverge from the preceding readings of the current sensor or diverge from the readings of the proximate sensor nodes. In IoT, An outlier may be a sign of an error or an event. Outlier detection in IoT is represented in the following figure 1.

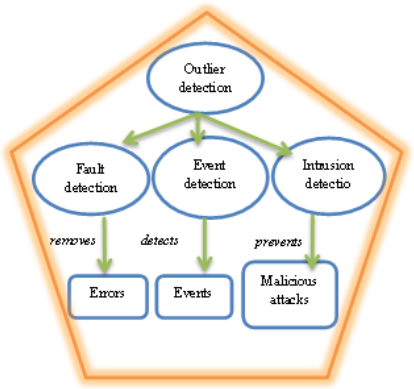

Figure 1. Outlier detection in IoT

The rest of this paper is systematized as follows. In Section 2, an overview of related works is presented. Section 3 presents the outlier detection techniques for sensor networks, in Section 4 the The architecture of the proposed STCPOD model is presented and Section 5 concludes the paper.

II. RELATED WORKS

Barakkath et al. [4] have proposed the relative correlation based clustering to detect outliers in the wireless sensor networks. During the first level, the spatial, temporal and attribute correlation were found. In the second level, optimum clusters were formed and outliers were classified based on correlation. The efficacy of the proposed model was appraised based on the computational complexity and communication overhead. The proposed method outperformed the existing outlier detection techniques taken for assessment and achieved high detection rate and low false alarm rate.

Min Wang et al. [5] presented an outlier detection algorithm to detect outliers based on spatial and temporal correlations between sensor nodes. The proposed algorithm comprises of three algorithms namely outlier self-detection algorithm, neighbour-voting algorithm and outlier confirming algorithm. Experiments were conducted using the OMNet++ platform. Experimental results proved the efficiency of the proposed algorithm over the existing majority voting algorithm.Fangfang Li et al. [6] proposed an event detection method depending upon the spatial and temporal correlations of the sensed data. The proposed method removed errors based on temporal linear comparison and spatial vertical comparison. Experimental results proved that the proposed event detection method has significantly improved the precision of event detection.

Kun Niu et al. [7] presented an innovative WSN clustering based outlier detection algorithm. First the proposed algorithm stated the time slot for sampling data. Then rational number of clusters was obtained. After this process maximum and minimum clusters were obtained. Finally latent clusters were identified through cluster labels of time slots. Experimental results showed the robustness of the presented algorithm on real WSN datasets.

Aymen Abid et al. [8] proposed an outlier detection method depending upon the distance between the current sensor readings and the neighbouring sensors readings. The size of the learning window was increased to observe the performance of the proposed method when there were more outliers. The study demonstrated that the proposed outlier detection method achieved high detection rate with low false alarm rate.

Farid Lalem et al. [9] proposed a faulty data detection approach in wireless sensor networks based on Copula theory to ensure data reliability. Experimentations performed on real world datasets proved that the proposed method detected significant percentage of faulty data. Andrade et al. [10] proposed an outlier detection technique to detect outliers in wireless sensor networks using clustering and other light weight statistical techniques. A case study was also done with temperature sensors in wireless sensor networks to evaluate the performance of the proposed technique. The experimental results proved that the proposed outlier detection technique has given accurate information even in the existence of outliers.

Wenjie Li et al. [11] proposed two distributed algorithms namely single-decision and an iterative algorithm and evaluated the performance of these algorithms in detecting faulty sensors. The sensor node first collected and processed data from the neighbouring node and found the existence of outliers using standard local outlier detection test. Consequently, the trade-off between false alarm rate and detection rate was found theoretically as well as by simulation. Mahsa Salehi et al. [12] proposed an outlier detection technique to detect local outliers on the streaming data. Experimental results on a range of datasets demonstrated the efficiency of the proposed

technique in detecting outliers accurately. It was also proved that the proposed incremental based outlier detection technique is also applicable to various application environments with restricted memory.

Hugo Martins et al. [13] presented a multi agent architecture to detect outliers in wireless sensor networks. Online detection of outliers in non-stationary time-series data was done using a machine learning technique. Empirical study proved the feasibility of the proposed architecture in WSNs.

III. OUTLIER DETECTION TECHNIQUES FOR SENSOR NETWORKS

Due to several shortcomings, traditional outlier detection techniques for general data cannot be straightaway applicable to the Internet of Things because of the nature of IoT data and also the specific needs of the IoT data. Though the techniques for wireless sensor networks cannot be applied directly to the IoT, but with slight modifications they can be used. This section provides a detailed view of existing outlier detection techniques specially developed for the sensor networks [14].

(i) Statistical based outlier detection techniques
(ii) Nearest Neighbor-based outlier detection techniques
(iii) Clustering-based outlier detection techniques
(iv) Classification-based outlier detection techniques
(v) Spectral-decomposition based outlier detection techniques

A. Statistical based outlier detection techniques

Statistical based outlier detection techniques are the basic techniques to tackle the outlier problem. These techniques are model-based techniques. A statistical model is developed to capture the data distribution and then data instances are assessed depending upon how well the data instances fit the model. When a data instance does not fit the model, it is declared as an outlier. Statistical based outlier detection techniques can also be unsupervised in which a statistical model can be built if it fits majority of the sensor readings while small number of outliers present in the data. Statistical-based approaches can be classified as follows

(i) Parametric-Based Approaches
(ii) Non-Parametric-Based Approaches

(i) Parametric-Based Approaches

These approaches assume the availability of data distribution and then parameters are assessed from the chosen data. Depending upon the type of distribution these parametric-based approaches are further classified into two models.

(i) Gaussian-based models
(ii) non-Gaussian-based models

(ii) Non-Parametric-Based Approaches

Non-parametric techniques do not assume the availability of data distribution. These techniques specify a distance measure between a new instance and the statistical model and deploy thresholds to determine whether the sensor reading is an outlier or not. These non-parametric-based approaches are further classified into two models.

(i) Histograms
(ii) Kernel density estimator.

Statistical based techniques are mathematically proved. These techniques can efficiently detect outliers once the proper model has been built. But in actual real- life scenarios, a priori knowledge of sensor readings is not available to construct the model. So the parametric techniques are not suitable when the sensor readings do not follow the pre-defined distribution of the model. Since non-parametric techniques do not assume any distribution, they are considered better than parametric techniques.

B. Nearest Neighbour-based outlier detection techniques

As per the nearest neighbour-based approach, a data object is considered as an outlier if it is far away from its neighbours. Euclidean distance is the common distance measure used to compute the distance between two data objects. Nearest neighbour outlier detection techniques are unsupervised in nature and do not make any assumptions about the data distribution. Several distance based outlier detection techniques were proposed to detect outliers in sensor networks. In one such technique [15], sensor node deploys distance similarity to detect outliers. After detection, the outliers are sent to nearby nodes for confirmation.

But this technique does not assume any network structure so that each sensor node communicates with other nearby nodes through broadcasting. As a result, it is not suitable for use in larger networks. When the distance-based outlier detection techniques adopt

network structure [14], communication overhead can be greatly reduced and the need for broadcasting is eliminated. The spatial-temporal correlation based outlier detection techniques identify outliers efficiently but defining appropriate threshold is not easy.

C. Clustering-based outlier detection techniques

As per the clustering-based outlier detection techniques, data objects are considered as outliers if they do not fit into any clusters. Euclidean distance is used as the dissimilarity measure between two data objects. Clustering-based outlier detection techniques are unsupervised in nature and do not make any assumptions about the data distribution. But calculating the distance between data objects in multivariate datasets is expensive.

D. Classification-based outlier detection techniques

Classification based outlier detection techniques operate in two phases. The first phase namely the training phase learns a classifier using the available labelled training data. The second phase namely the testing phase classifies a test instance as normal or an outlier using the classifier.

Prevailing classification based outlier detection techniques for sensor networks are classified into support vector machines based outlier detection techniques and Bayesian network-based approaches outlier detection techniques.

E. Spectral-decomposition based outlier detection techniques

Spectral decomposition-based approaches determine normal behaviour of data by deploying principal components. Principal component analysis (PCA) is used to shrink the dimensionality prior to the detection of outliers and to find new subset of dimensions to capture the behaviour of data. Several Spectral decomposition-based outlier detection techniques were proposed. In one such technique [16], PCA is used to detect outliers by modelling the spatial-temporal data correlations in a distributed manner so that outliers spanning through neighbouring nodes are easily found. However, Using PCA to exactly assess the correlation matrix of normal patterns is highly expensive.

IV. ARCHITECTURE OF THE PROPOSED MODEL

The proposed model as shown in figure 2 is proximity based outlier detection model. This model is implemented online to detect outliers at faster rate owing to the low power, low memory and poor computing capability of sensor nodes. Air quality sensors

Smart air quality monitoring sensors fixed on the 449 observation points near the roadside measure, observe and send the concentration of various gases such as carbon-di-oxide, carbon monoxide etc. in the ambient air at fixed intervals.

IoT gateway

IoT gateways sit between the intersection of IoT devices, controllers and sensors and the cloud. The internet of things will not only comprise of new IoT compatibility devices but also the devices that are already in place. So the already existing non IP-based devices can be connected to the internet via IoT gateways.

Cloud Storage

The air quality sensors generated data is in the form of time-series data. Usually, the measurements are taken at regular time intervals. A trustworthy storage location like cloud is needed to store time-series data.

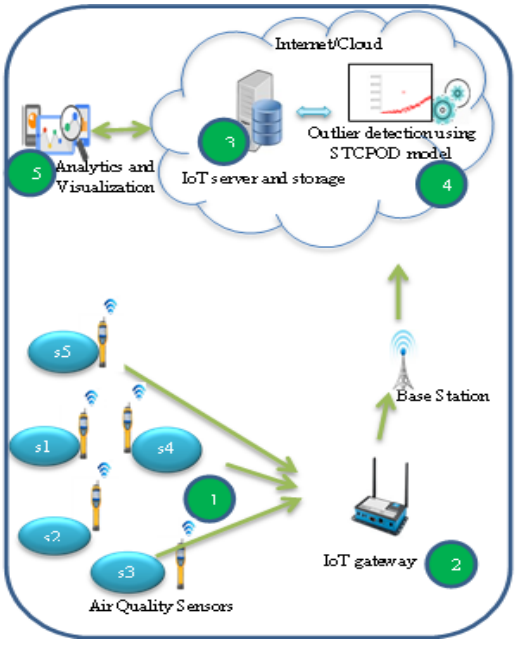

Figure 2. The architecture of the proposed STCPOD model

Outlier detection

Sensor nodes are usually placed in the harsh environment. While monitoring air quality, they are prone to generate faulty data and frequently tend to produce untrustworthy readings due to environmental interference and other quality problems [17]. To enhance the quality of sensor data, detection of outliers becomes the prime task to ensure reliability and accuracy and robustness of data. The proposed STCPOD model plays a vital role in detecting outliers at a faster rate to produce complete dataset.

In this paper, a robust architecture is built to detect outliers effectively using the proposed STCPOD model which attains high detection rate and low false alarm rate.

V. CONCLUSION

Outlier detection is indispensable to ensure quality of data and improved analytics. A robust Architecture proposed in this paper efficiently detects outliers in IoT data by deploying novel STCPOD (a novel STCPOD (Spatially and temporally correlated proximate Outlier Detection) model so as to accomplish high detection rate and low false alarm rate by considering the prominent characteristics of IoT data.

VI. REFERENCES

[1]. Shen, Q. , Zhao, Z. , Niu, W. , Liu, Y. and Tang, H. , "Tolerance-Based Adaptive Online Outlier Detection for Internet of Things", In Proceedings of the 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing, doi={10. 1109/GreenCom-CPSCom. 2010. 23}, 2010, pp. 560-565.

[2]. Karkouch A, Mousannif H, Al Moatassime H and Noel T, "Data quality in internet of things: A state-of-the-art survey", Journal of Network and Computer Applications, Vol. 73, 2016, DOI: https://doi. org/10. 1016/j. jnca. 2016. 08. 002, pp. 57-81.

[3]. Zhang, Y. , Meratnia, N. and Havinga, P. , "Why general outlier detection techniques do not suffice for wireless sensor networks", Intelligent Techniques for Warehousing and Mining Sensor Network Data, 2009, DOI: 10. 4018/978-1-60566-328-9. ch007, p. 136.

[4]. Nisha, U. B. , Maheswari, N. U. , Venkatesh, R. and Abdullah, R. Y. , 2014, December. " Robust estimation of incorrect data using relative correlation clustering technique in wireless sensor networks", IEEE International Conference on Communication and Network Technologies (ICCNT), 2014, ISBN: 978-1-4799-6266-2, pp. 314-318.

[5]. Wang, Min, and Zhongbo Wu. "Spatio-temporal correlation based outlier detection algorithm in sensor network. " In Second IEEE International Conference on Computer and Automation Engineering (ICCAE), Vol. 4, 2010, doi={10. 1109/ICCAE. 2010. 5451639}, pp. 424-427.

[6]. Li, Fangfang, and Zhibo Feng. "An efficient real-time event detection approach based on temporal-spatial correlations in wireless sensor networks. " In IEEE International Conference on Computer Science and Network Technology (ICCSNT), Vol. 2, 2011, doi={10. 1109/ICCSNT. 2011. 6182185}, pp. 1245-1249.

[7]. Niu, Kun, Fang Zhao, and Xiuquan Qiao. "An outlier detection algorithm in wireless sensor network based on clustering", In 15th IEEE International Conference on Communication Technology (ICCT), 2013, doi={10. 1109/ICCT. 2013. 6820415}, pp. 433-437.

[8]. Abid, Aymen, Abdennaceur Kachouri, and Adel Mahfoudhi. "Anomaly detection through outlier and neighborhood data in Wireless Sensor Networks", In 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 2016, doi={10. 1109/ATSIP. 2016. 7523045} pp. 26-30.

[9]. Lalem, Farid, Ahcene Bounceur, Rahim Kacimi, Reinhardt Euler, and Massinissa Saoudi, "Faulty Data Detection in Wireless Sensor Networks Based on Copula Theory", In Proceedings of the ACM International Conference on Big Data and Advanced Wireless Technologies, 2016, ISBN: 978-1-4503-4779-2 doi={10. 1145/3010089. 3010114}, p. 29.

[10]. Andrade, A. T. C. , C. Montez, R. Moraes, A. R. Pinto, Francisco Vasques, and G. L. da Silva. "Outlier detection using k-means clustering and lightweight methods for Wireless Sensor Networks", In 42nd Annual Conference of the IEEE on Industrial Electronics Society, 2016, doi={10. 1109/IECON. 2016. 7794093}, pp. 4683-4688.

[11]. Li, Wenjie, Francesca Bassi, Davide Dardari, Michel Kieffer, and Gianni Pasolini. "Defective sensor identification for WSNs involving generic local outlier detection tests. " IEEE transactions on Signal and Information Processing over Networks, Vol. 2, No. 1, 2016, doi={10. 1109/TSIPN. 2016. 2516821}, ISSN={2373-776X}, pp. 29-48.

[12]. Salehi, Mahsa, Christopher Leckie, James C. Bezdek, and Tharshan Vaithianathan. "Local outlier detection for data streams in sensor networks: Revisiting the utility problem invited paper", In IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2015, doi={10. 1109/ISSNIP. 2015. 7106978}, pp. 1-6.

[13]. Martins, Hugo, Fábio Januário, Luís Palma, Alberto Cardoso, and Paulo Gil. "A machine learning technique in a multi-agent framework for online outliers detection in Wireless Sensor Networks" , In 41st Annual Conference of the IEEE on Industrial Electronics Society, 2015, doi={10. 1109/IECON. 2015. 7392180}, pp. 000688-000693.

[14]. Zhang, Yang, Nirvana Meratnia, and Paul Havinga. "Outlier detection techniques for wireless sensor networks: A survey. " In IEEE Communications Surveys & Tutorials, Vol. 12, No. 2, 2010, doi={10. 1109/SURV. 2010. 021510. 00088}, ISSN={1553-877X}, pp. 159-170.

[15]. Branch, Joel W. , Chris Giannella, Boleslaw Szymanski, Ran Wolff, and Hillol Kargupta ", In-network outlier detection in wireless sensor networks", In Springer Journal of Knowledge and information systems, Vol. 34, No. 1, 2013, https://doi. org/10. 1007/s10115-011-0474-5, pp. 23-54.

[16]. Chatzigiannakis, Vasilis, Symeon Papavassiliou, Mary Grammatikou, and B. Maglaris. "Hierarchical anomaly detection in distributed large-scale sensor networks. " In 11th IEEE Symposium on Computers and Communications, 2006, doi={10. 1109/ISCC. 2006. 1691116}, ISSN={1530-1346} pp. 761-767.

[17]. Ghorbel, Oussama, Mohamed Wassim Jmal, Walid Ayedi, Hichem Snoussi, and Mohamed Abid, "An overview of outlier detection technique developed for wireless sensor networks", In 10th International Multi-Conference on Systems, Signals and Devices (SSD), 2013, doi={10. 1109/SSD. 2013. 6564165}, pp. 1-6.