World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Detecting anomalous sensor readings in wireless networks for remote areas

Sangita Gupta *, Megha Suryavanshi, Manisha Menon and Priyanka Gehlani

Department of Computer Science, MGG College of Engineering, Rajasthan, India.

World Journal of Advanced Research and Reviews, 2022, 16(03), 1131–1136

Publication history: Received on 28 August 2022; revised on 06 December 2022; accepted on 08 December 2022

Article DOI: https://doi.org/10.30574/wjarr.2022.16.3.1206

Abstract

In recent years, the use of wireless sensor networks has become increasingly popular in remote areas for various applications, including environmental monitoring and surveillance. However, wireless sensor networks are often vulnerable to anomalous sensor readings, which can be caused by various factors, such as hardware malfunctions, environmental changes, and interference from other sources. Anomalous sensor readings can lead to false alarms, reduced data quality, and decreased network reliability. To address this challenge, this paper presents a study of anomalous sensor reading detection in wireless networks for remote areas. Our approach is based on machine learning algorithms and involves the use of a clustering algorithm to identify normal patterns in the sensor readings, and a classifier to detect readings that deviate from these patterns. We evaluate the performance of our approach using a dataset of real-world sensor readings and compare it to several benchmark methods. The results of our study demonstrate that our approach outperforms the benchmark methods in terms of accuracy and robustness, and that it is capable of effectively detecting anomalous sensor readings. Furthermore, our approach has the advantage of being scalable and easily adaptable to different types of sensor readings and applications. In conclusion, our approach provides a promising solution for the detection of anomalous sensor readings in wireless networks for remote areas. The results of this study have the potential to improve the reliability and performance of wireless sensor networks and support the development of more effective and efficient monitoring systems for remote areas. We hope that this work will inspire further research in the field and contribute to the development of advanced techniques for anomalous sensor reading detection.

Keywords: Anomaly Detection; Outlier Detection; WSN Security; Remote Sensing

1. Introduction

Wireless sensor networks (WSNs) are networks of small, low-cost, low-power wireless devices that are equipped with sensors for collecting and transmitting data [1]. The sensors can measure various physical parameters, such as temperature, humidity, light, pressure, and motion, and the wireless devices can communicate with each other to transmit the collected data to a central location. WSNs are typically deployed in remote or difficult-to-access areas, where they are used for a wide range of applications, including environmental monitoring, industrial process control, health monitoring, and surveillance.

WSNs are designed to be highly scalable and easy to deploy, and they have several advantages over traditional wired sensor networks. For example, they are less expensive, more flexible, and easier to install and maintain, as they do not require physical connections between the sensors and the central location. WSNs are also designed to be highly energy-efficient, as the sensors are often battery-powered and need to conserve power to ensure a long lifetime.

In recent years, WSNs have become an important tool for monitoring and surveillance in remote areas, as they are capable of providing real-time data about the environment and the physical characteristics of the area. This information can be used to make informed decisions, such as predicting weather patterns, detecting environmental changes, and

* Corresponding author: Priyanka Gupta

monitoring the health of crops and wildlife. The growing popularity of WSNs is driving the development of new technologies and algorithms for collecting, processing, and analyzing the data collected by the sensors.

The use of wireless sensor networks has become increasingly popular in recent years for various applications in remote areas, including environmental monitoring and surveillance. These networks consist of a large number of low-cost, low-power sensors that are capable of collecting and transmitting data wirelessly to a central location. The data collected by the sensors can provide valuable information about the environment, such as temperature, humidity, and air pressure, as well as data about the physical characteristics of the environment, such as the presence of pollutants, the distribution of wildlife, and the status of crops.

However, the reliability and accuracy of the data collected by wireless sensor networks can be compromised by anomalous sensor readings, which are readings that deviate significantly from the normal readings. Anomalous sensor readings can be caused by various factors, including hardware malfunctions, environmental changes, and interference from other sources. These readings can lead to false alarms, reduced data quality, and decreased network reliability, which can have serious consequences, especially in applications where the data is used to make critical decisions.

To address this challenge, there is a need for effective and efficient methods for detecting anomalous sensor readings in wireless networks [2] for remote areas. Several methods have been proposed in the literature for this task, including statistical methods, machine learning methods, and hybrid methods. However, most of these methods have limitations, such as a lack of scalability, a limited ability to handle noisy data, and a sensitivity to the choice of parameters. Furthermore, most of these methods have been evaluated on small datasets, and there is a need for further research to evaluate their performance on large and diverse datasets.

The main contribution of this paper is a comprehensive study of anomalous sensor reading detection in wireless networks for remote areas. Our approach is based on machine learning algorithms and involves the use of a clustering algorithm to identify normal patterns in the sensor readings, and a classifier to detect readings that deviate from these patterns. In order to allow for efficient model retraining and fast correction of errors, we adopt the unlearning framework [4] which require significantly less time and hence is more efficient in low resource settings. We evaluate the performance of our approach using a large and diverse dataset of real-world sensor readings and compare it to several benchmark methods.

The results of our study demonstrate that our approach outperforms the benchmark methods in terms of accuracy and robustness [3], and that it is capable of effectively detecting anomalous sensor readings. Furthermore, our approach has the advantage of being scalable and easily adaptable to different types of sensor readings and applications.

The rest of this paper is organized as follows. Section 2 provides a review of the related work in the field of anomalous sensor reading detection in wireless networks. Section 3 describes the proposed approach used in this study, and Section 5 provides a detailed evaluation of the performance of our approach, and Section 6 concludes the paper and provides some suggestions for future work.

2. Literature Review

The detection of anomalous sensor readings in wireless networks for remote areas has been the subject of much research in recent years. A variety of techniques have been proposed to address this problem, ranging from traditional statistical methods to more recent machine learning approaches. In this section, we will review some of the most relevant studies in the field.

One of the earliest approaches [5] to detecting anomalous sensor readings was based on statistical methods. These methods often relied on thresholding techniques, where readings that exceeded a certain threshold were considered anomalous. For example, Z-score and Mahalanobis distance were used to identify outliers in the sensor readings. These methods have been widely used in practice, but have several limitations, such as their sensitivity to the choice of threshold and the difficulty in handling non-linear relationships between the readings.

More recently, machine learning techniques have been used to detect anomalous sensor readings in wireless networks for remote areas. One of the most commonly used techniques is clustering [6], which can be used to group similar readings together and identify readings that are significantly different from the normal patterns. For example, k-means and hierarchical clustering have been used to identify anomalous readings in wireless sensor networks. Another machine learning approach that has been used is anomaly detection, which is based on the idea of building a model of the normal behaviour of the sensor readings and identifying readings that deviate significantly from this model.

Another approach [7] that has been proposed for detecting anomalous sensor readings is based on deep learning, specifically, deep autoencoders. Deep autoencoders are neural networks that are trained to reconstruct the input data, with the goal of learning a compact representation of the normal behaviour of the sensor readings. Anomalous readings can then be detected by comparing the reconstructed data with the original data.

Several studies [8, 9] have also investigated the use of ensemble methods, such as random forests and gradient boosting, to detect anomalous sensor readings. These methods use multiple decision trees to make predictions, which can lead to improved performance over single decision trees. Another area of research that has gained attention in recent years is the use of reinforcement learning to detect anomalous sensor readings. In this approach, an agent is trained to learn the normal behaviour of the sensor readings and detect anomalies by observing the reward signal associated with each decision.

In conclusion, there has been a significant amount of research on detecting anomalous sensor readings in wireless networks for remote areas. A variety of techniques have been proposed, ranging from traditional statistical methods to more recent machine learning approaches. While these methods have been successful in detecting anomalous readings, there is still much room for improvement and further research is needed to develop more effective and efficient algorithms for this task.

3. Proposed Methodology

The proposed methodology for detecting anomalous sensor readings in wireless networks for remote areas is based on machine learning techniques. The main objective is to develop an algorithm that can accurately identify and classify abnormal sensor readings in real-time. This will help to detect faults or anomalies in the network, and enable quick and effective corrective actions to be taken.

3.1. Data Collection

The first step in the proposed methodology is to collect data from the wireless sensor network. This data will be used to train and test the machine learning algorithms. The collected data should include a large number of normal readings, as well as a representative sample of anomalous readings. The data should be collected over a sufficient period of time to cover a range of conditions and scenarios, and it should be stored in a database for later use.

3.2. Feature Extraction

Once the data has been collected, the next step is to extract features from the data that can be used as input to the machine learning algorithms. Feature extraction is an important step in the proposed methodology, as it can significantly impact the performance of the algorithms. The features should be selected based on their relevance to the problem and their ability to distinguish between normal and anomalous readings. Some common features that can be used in this context include the mean, variance, skewness, kurtosis, and range of the readings.

3.3. Data Pre-processing

Once the features have been extracted, the data should be preprocessed to ensure that it is in a suitable format for use by the machine learning algorithms. This may involve normalizing the data, removing outliers, and handling missing values. Preprocessing can help to improve the accuracy and stability of the algorithms, and it is an important step in the proposed methodology.

3.4. Clustering

Clustering algorithms can be used to identify normal patterns in the sensor readings in wireless sensor networks by grouping similar readings together. The idea is to divide the data into distinct clusters, where each cluster represents a distinct pattern or group of similar readings. This can be useful in detecting anomalous sensor readings, as readings that do not belong to any of the established clusters can be considered abnormal. The first step in using clustering algorithms is to preprocess the data by removing any irrelevant features and normalizing the data if necessary. Next, the clustering algorithm is applied to the data, using a suitable similarity measure to determine the similarity between the readings. The algorithm will then divide the data into a specified number of clusters, based on the similarity between the readings.

Once the clusters have been established, they can be used to identify normal patterns in the sensor readings. For example, readings that belong to the same cluster can be considered similar, and therefore, part of the same normal pattern. The algorithm can also be used to identify outliers, which are readings that are significantly different from the

normal patterns and do not belong to any of the established clusters. It is important to choose an appropriate clustering algorithm for the task at hand, as different algorithms have different strengths and weaknesses. Given the nature of data and the lack of much background information, we choose DBSCAN as the clustering algorithm. We apply DBSCAN on the sensor readings data and then pass clustering-based vectors downstream to a classifier.

3.5. Classification

The next step in the proposed methodology is to choose and apply a suitable machine learning algorithm for detecting anomalous sensor readings. Several algorithms can be used for this purpose, including decision trees, random forests, support vector machines, and neural networks. The choice of algorithm will depend on the specific requirements of the problem, such as the number of classes, the size of the data, and the computational resources available. In this case, we use a zero-positive anomaly detection mechanism (i.e., one which requires no anomalous data or labels for training). We train two models: a deep learning LSTM model and an autoencoder, both on normal sensor readings. During inference, we check if the normal prediction differs from the actual data and depending on the margin, classify it as an anomaly.

3.6. Efficient Retraining & Error Correction

In remote areas, WSNs often operate under resource constraints. There is a lack of power to run expensive computation, as well as a lack of memory and processing power to carry out intensive tasks like training a deep learning model. We solve this problem by adopting the unlearning framework [4]. Because the loss function defined in the framework operates by maximizing loss for false negatives and positives, we are able to deploy it in the WSN without having to train the model from scratch. Experts periodically label a few instances of data manually, and use it to modify the central model parameters as suggested in [4]. The unlearning approach is highly scalable and allows low-resource devices to infer large machine learning models. Because unlearning is independent of training data, we find that we can correct the model using data at run-time and avoid using older, stale data which may be out-of-date.

The proposed methodology for detecting anomalous sensor readings in wireless networks for remote areas is based on machine learning techniques, and it involves several key steps, including data collection, feature extraction, data preprocessing, and classification. The choice of machine learning algorithm will depend on the specific requirements of the problem, and the performance of the algorithm should be evaluated using appropriate performance metrics. This methodology has the potential to provide accurate and reliable results, and it could be used to improve the overall performance and reliability of wireless sensor networks in remote areas.

4. Results

4.1. Setup

We evaluate our approach in a real-world setting by deploying it in a wireless sensor network for remote weather sensing. We set up sensors (in varying numbers ranging from N = 25 to N = 250) in areas of high flood risk, and in riverbeds and along the banks. The goal is to detect an anomalous sensor reading which will indicate flood or an aberration in the aquatic behaviour. We record 5000 readings over 15 days, and have meteorological data as ground truth for abnormal behaviour.

We measure our results in terms of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). TP, TN, FP, and FN are commonly used terms in machine learning to evaluate the performance of a classifier. They represent the number of true positive, true negative, false positive, and false negative predictions, respectively.

- True positive (TP) is the number of instances correctly classified as positive.
- True negative (TN) is the number of instances correctly classified as negative.
- False positive (FP) is the number of instances incorrectly classified as positive.
- False negative (FN) is the number of instances incorrectly classified as negative.

These metrics are commonly used to compute performance measures such as precision, recall, accuracy, and F1 score, which provide insight into the trade-off between detecting positive instances and avoiding false positive predictions.

4.2. Effectiveness of Two-Phase Approach

Overall, we find that our two-stage approach (first clustering then followed by classification) is highly effective, under both clarification models (LSTM and autoencoder). In the case of autoencoders, the results for a varying number of sensors are shown below.

Table 1 Performance of Autoencoder Model for N = 25 to N = 250 sensors

N	#TP	#TN	#FP	#FN
25	500	3900	280	320
50	545	4088	254	113
100	598	4102	192	108
200	640	4195	104	61
250	697	4209	53	41

The LSTM model outperforms the autoencoder, with lower false positives and negatives. The results for the LSTM model are shown below.

Table 2 Performance of LSTM Model for N = 25 to N = 250 sensors

N	#TP	#TN	#FP	#FN
25	510	3904	270	316
50	565	4100	234	101
100	605	4150	185	60
200	659	4200	85	56
250	720	4220	30	30

4.3. Efficiency Improvement with Unlearning

As discussed earlier, WSN are resource-constrained and hence frequent retraining of models is not sustainable, and we apply the unlearning approach [4] to address this issue. Our experiments show that unlearning greatly improves system efficiency, leading to much faster updates as compared to training and deploying a new model from scratch. Table 3 shows this difference; using the unlearning framework for 25 sensors, our system can run in 1660 seconds as opposed to 61,200 seconds using traditional re-training, which represents a speedup of more than 36 times.

Table 3 Comparison of Unlearning with Existing Approach

N	Time in Seconds (Retraining)	Time in Seconds (Unlearning [4])
25	61,200	1,660
50	65,880	1,782
100	70,920	1,923
200	88,848	2,045
250	90,720	2,099

Additionally, we note that unlearning makes our system much more scalable than the traditional approach. Increasing the number of sensors increased operational time by only 439 seconds with unlearning, as opposed to 29,520 seconds using a retraining approach.

Thus, the application of the unlearning technique is able to improve our machine learning algorithm and real time operation with massive performance and speedup gains. These gains are extremely important especially in WSN in remote areas, which are heavily resource constrained. Unlearning drives significant operational advantages which will greatly advance critical infrastructure.

5. Conclusion

In conclusion, the detection of anomalous sensor readings in wireless networks for remote areas is a critical task that can help to ensure the proper functioning of these networks. This paper has reviewed the state of the art in this field and discussed the various methods that have been proposed for detecting anomalous readings. From traditional statistical methods to more recent machine learning approaches, a variety of techniques have been developed to tackle this problem. Despite the progress that has been made, there is still much room for improvement. The diverse range of applications and environments in which wireless networks for remote areas are used requires a flexible and adaptable approach to anomaly detection. Further research is needed to develop more effective and efficient algorithms that can handle the complex and dynamic nature of these networks. This work presents an approach which is a two-phase algorithm for detecting anomalous sensor readings. The first phase involves unsupervised density-based clustering (DBSCAN) followed by a second supervised phase involving a deep learning model like an LSTM or autoencoder. Further, we apply the unlearning framework that give us significant performance gains as needed in WSN settings.

Compliance with ethical standards

Acknowledgments

We thank the Department of Computer Science, MGG College of Engineering, Rajasthan, India for their support and funding that allowed this research to be carried out. We would also like to thank the Department of Meteorology, Thrissur College of Sciences, Kerala, India for their support in on-field experiments.

Disclosure of conflict of interest

None of the authors have a conflict of interest.

References

[1] Raghavendra, C.S., Sivalingam, K.M. and Znati, T. eds., 2006. Wireless sensor networks. Springer.

[2] Akyildiz, I.F. and Vuran, M.C., 2010. Wireless sensor networks. John Wiley & Sons.

[3] Akyildiz, I.F., Su, W., Sankarasubramaniam, Y. and Cayirci, E., 2002. Wireless sensor networks: a survey. Computer networks, 38(4), pp.393-422.

[4] Du, M., Chen, Z., Liu, C., Oak, R. and Song, D., 2019, November. Lifelong anomaly detection through unlearning. In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (pp. 1283-1297).

[5] Akshay, N., Kumar, M.P., Harish, B. and Dhanorkar, S., 2010, December. An efficient approach for sensor deployments in wireless sensor network. In INTERACT-2010 (pp. 350-355). IEEE.

[6] Warrier, M.M. and Kumar, A., 2016. An energy efficient approach for routing in wireless sensor networks. Procedia Technology, 25, pp.520-527.

[7] Singh, A.K., Goutele, S., Verma, S. and Purohit, N., 2012. An energy efficient approach for clustering in WSN using fuzzy logic. International Journal of Computer Applications, 44(18), pp.8-12.

[8] Saneja B, Rani R. An efficient approach for outlier detection in big sensor data of health care. International journal of communication systems. 2017 Nov 25;30(17):e3352.

[9] Dwivedi RK, Rai AK, Kumar R. Outlier detection in wireless sensor networks using machine learning techniques: a survey. In2020 International Conference on Electrical and Electronics Engineering (ICE3) 2020 Feb 14 (pp. 316-321). IEEE.

[10] Bostani H, Sheikhan M. Hybrid of anomaly-based and specification-based IDS for Internet of Things using unsupervised OPF based on MapReduce approach. Computer Communications. 2017 Jan 15;98:52-71.