International Journal of Recent Trends in Engineering & Research

ISSN (ONLINE) : 2455 – 1457

IMPACT FACTOR : 4.101

A Comparative Review of Outlier Detection Techniques for Wireless Sensor Networks

Shivani Garg[1],Mukul Varshney[2],Abha Kiran Rajpoot[3],Jyotsna[4]

[1,2,3,4]CSE Department, Sharda University

Abstract—Outliers in the field of wireless sensor networks are the measured values that significantly differ from the normal pattern of data sensed by the sensors. Noise and errors, events, and malicious attacks on the network are the potential sources of outliers. Because of the nature of sensor data and specific requirements and limitations of the wireless sensor networks, traditional outlier detection techniques could not be applicable directly to wireless sensor networks. The survey presented in this paper provides a comprehensive overview of existing outlier detection techniques specifically developed for the wireless sensor networks. Additionally, it presents a technique-based classification and comparison to be used as a guideline to select a technique suitable for a particular application based on characteristics such as data type, outlier type, outlier identity, and outlier degree.

Index Terms—Outlier, outlier detection, wireless sensor networks, classification.

I. INTRODUCTION

A Wireless sensor network (WSN) is composed of many small, low-cost sensor nodes distributed over a large area with one or possibly more powerful sink nodes gathering readings of sensor nodes. The sensor nodes are equipped with capabilities like sensing, processing and wireless communication. Each node consists of a wireless radio transceiver, a small microcontroller, a power source and various sensors such as temperature, humidity, light, heat, pressure, sound, vibration, etc. The WSN is used to provide fine-grained real-time data about the physical world and to detect events which are time-critical. WSNs have widely diverse applications including those related to personal, industrial, business, and military domains, such as environmental and habitat monitoring, object and inventory tracking, health and medical monitoring, battlefield observation, industrial safety and control, to name but a few. In many of these applications, real-time data mining of sensor data to promptly make intelligent decisions is essential [1].

Several factors make wireless sensor networks (WSNs) especially prone to outliers.

- First, they collect their data from the real world using sensing devices which are imperfect.
- Second, they have limited battery power and hence their performance tends to deteriorate as power is exhausted.
- Third, since these networks may include several sensors, the chance of error increases.
- Finally, in their usage for security and military purposes, sensors are especially prone to manipulation by adversaries.

Hence, outlier detection should be an inseparable part of any data processing routine that takes place in WSNs.

The above internal and external factors lead to unreliability of sensor data, which further impact the quality of raw data and aggregated results. Since actual events occurring in the physical world, such as, forest fire, earthquake or chemical spill, cannot be accurately detected using inaccurate and incomplete data [4], it is highly important to be sure about the reliability and accuracy of sensor data before the decision-making process.

DOI : 10.23883/IJRTER.2017.3302.NEBJD

Since, outliers are one of the sources to greatly influence data quality, in this survey we provide a comprehensive overview of the research done in the field of outlier detection in WSNs, evaluate and compare existing outlier detection techniques specifically developed for WSNs, and identify potential areas for further research.

The rest of the paper is organized as follows: Section 2 describes the fundamental concepts of outlier detection in WSNs. Section 3 presents techniques for identifying important criteria associated with the classification of outlier detection. Section 4 provides a technique-based classification to categorize existing outlier detection techniques developed for WSNs. Section 5 presents a brief description of current outlier detection techniques.

II. FUNDAMENTAL CONCEPTS OF OUTLIER DETECTION IN WIRELESS SENSOR NETWORKS

This section throws a light on the definitions of outliers, various causes of outliers, motivation of outlier detection, and challenges of outlier detection in WSNs.

2.1. Outlier Definition

The term *outlier*, also referred to as *anomaly*, has its origin from the field of *statistics* [5]. The two classical definitions of outliers are:
(Hawkins [6]): *"an outlier is an observation, which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism".*
(Barnett and Lewis [7]): *"an outlier is an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data".*
In WSNs, outliers can be defined as, *"those measurements that significantly deviate from the normal pattern of sensed data"* [9]. This definition is based on the fact that in WSN sensor nodes are assigned to monitor the physical world and thus a pattern representing the normal behaviour of sensed data may exist. Potential sources of outliers in data collected by WSNs include *noise* and *errors*, *actual events*, and *malicious attacks*. Noisy data as well as erroneous data should be eliminated or corrected if possible as noise is a random error without any real significance that dramatically affects the data analysis [10]. Outliers caused by other sources need to be
identified as they may contain valuable information about events that are of great interest to the researchers.

2.2. Motivation of Outlier Detection in WSNs

One of the fundamental tasks of data mining is outlier detection, also known as *anomaly detection* or *deviation detection* [10]. Outlier detection has been widely researched in various disciplines such as statistics, data mining, machine leaning, information theory, and spectral decomposition [9]. Also, it finds its applications in domains such as fraud detection, network intrusion, performance analysis, weather prediction, etc [9].
Recently, the topic of outlier detection in WSNs has attracted much attention as it provides reliable and secure functioning of the network. It controls the quality of measured data, improves robustness of the data analysis under the presence of noise and faulty sensors so that the communication overhead of erroneous data is reduced and the aggregated results are prevented to be affected. Outlier detection also provides an efficient way to search for values that do not follow the normal pattern of sensor data in the network. The detected values consequently are treated as events indicating change of phenomenon that are of interest. Furthermore, outlier detection helps in identifying malicious sensors that always generate outlier values, detects potential network attacks by adversaries, and thus ensures the security of the network. Following are some real-life examples where outlier detection can be used to its advantage:

- *Environmental monitoring*, where sensors for sensing temperature and humidity are deployed in harsh and unattended regions to monitor the natural environment. Outlier detection triggers an alarm upon detection of occurrence of an event.
- *Habitat monitoring*, where endangered species can be equipped with small non-intrusive sensors for monitoring their behaviour. Outlier detection indicates abnormal behaviours of the species and provide a closer observation about behaviour of individuals and groups.
- *Health and medical monitoring*, where patients are equipped with small sensors on multiple parts of their body to monitor their well-being. Outlier detection indicates whether the patient has potential diseases and allow doctors to take effective medical care.
- *Industrial monitoring*, where machines are equipped with temperature, pressure, or vibration amplitude sensors to monitor their operation. Outlier detection quickly identifies anomalous readings to indicate possible abnormality in the machines and allow for their corrections.
- *Target tracking*, where sensors are embedded in moving targets to track them in real-time. Outlier detection can filter erroneous information to improve the estimation of the location of targets and to make tracking more efficiently and accurately.
- *Surveillance monitoring*, where multiple sensitive and unobtrusive sensors are deployed in restricted areas. Outlier detection identifying the position of the source of the anomaly can prevent unauthorized access and potential attacks by adversaries in order to enhance the security of these areas.

Several research topics have been developed for identifying specific sources of outliers occurred in WSNs. As illustrated in Figure 1, these topics include fault detection ([12], [13]), event detection ([4], [14], [15]) and intrusion detection ([16], [17]).

2.3. Outlier Detection in Event Detection Domain

The event-based applications require sensor nodes to report event to the sink node within specified time interval once an event is detected. A complex event, combing two or more atomic events, requires multiple types of sensors collaborating to detect an event [18]. The differences between event detection and outlier detection are summarised as:

- outlier detection techniques have no a priori knowledge of trigger condition of any event, while event detection techniques hold the trigger condition of certain event issued by the sink node.
- outlier detection identifies abnormal readings by comparing sensor measurements with each other, while event detection specifies a occurrence of a certain event by comparing sensor measurements with the trigger condition.
- outlier detection techniques need to lower the false alarm rate due to normal data being classified as outlier, while event detection techniques need to prevent erroneous data which conform to the event condition to influence reliability of the detection.

2.4. Challenges of Outlier Detection in WSNs

Extracting useful knowledge from raw sensor data is not a simple task [19]. The context of sensor networks and the nature of sensor data make design of an appropriate outlier detection technique more challenging. Due to following reasons, conventional outlier detection techniques might not be suitable for handing sensor data in WSNs.

- *Resource constraints*. The low cost and low-quality sensor nodes have rigorous constraints in resources. They have limited energy, memory, computational capacity and communication bandwidth. They are usually computationally expensive and require memory for data analysis and storage. Thus, minimizing energy consumption while using a reasonable amount of memory for storage and computational tasks is a challenge for outlier detection.

- *High communication cost*. In WSNs, a lot of energy is consumed for radio communication rather than computation. Most of traditional outlier detection techniques using centralized approach for data analysis cause too much energy consumption and communication overhead. Thus, another challenge for outlier detection in WSNs is how to minimize the communication overhead in order to relieve the network traffic and prolong the lifetime of the network.
- *Distributed streaming data*. Distributed sensor data coming from many different streams may dynamically change. Moreover, the underlying distribution of streaming data may not be known a priori. Most of traditional outlier detection techniques that analyze data in an offline manner do not meet the requirement of handling distributed stream data. Thus, a challenge for outlier detection in WSNs is how to process distributed streaming data online.
- *Dynamic network topology, frequent communication failures, mobility and heterogeneity of nodes*. A sensor network deployed in unattended environments over extended period is susceptible to dynamic network topology and frequent communication failures. Moreover, sensor nodes may move among distinct locations at any point in time, and may have different sensing and processing capacities. Each sensor node may even be equipped with different number and types of sensors. Such dynamicity and heterogeneity increase the complexity of designing an appropriate outlier detection technique for WSNs.
- *Large-scale deployment*. Deployed sensor networks can have massive size (up to hundreds or even thousands of sensor nodes). The key challenge of traditional outlier detection techniques is to maintain a high detection rate while keeping the false alarm rate low. This is a very difficult task for large-scale sensor network applications.
- *Identifying outlier sources*. The sensor network is expected to provide the raw data sensed from the physical world and detect events occurred in the network. However, it is difficult to identify what has caused an outlier in sensor data due to the resource constraints and dynamic nature of WSNs. Thus, a challenge of outlier detection in WSNs is identifying outlier sources and making distinction between errors, events and malicious attacks.

III. IMPORTANT ASPECTS OF OUTLIER DETECTION TECHNIQUES FOR WSNS

This section identifies and discusses several important aspects of outlier detection techniques specially developed for WSNs. These aspects will be used as metrics to compare characteristics of different outlier detection techniques in Section VI.

3.1. Input Sensor Data

Sensor data can be viewed as *data streams*, i.e., a large volume of real-valued data that is continuously collected by sensor nodes [21]. The type of input data determines which outlier detection techniques can be used to analyse the data. Outlier detection techniques usually consider the two following aspects of sensor data.

1. *Attributes:* A data measurement can be identified as outlier when its attributes have anomalous values [10]. Outlier detection techniques for WSNs should be able to analyse *multivariate data* and identify whether the attributes together display anomaly.
2. *Correlations:* Attributes of multivariate sensor data may induce certain correlation, e.g., the readings of humidity and barometric pressure sensors are related to the readings of the temperature sensors. Capturing the attribute correlations helps to improve the mining accuracy and computational efficiency.

3.2. Type of Outliers

Depending on the scope of data used for outlier detection, outlier may be either *local* or *global*.

1. *Local Outliers:* Techniques for detecting local outliers save communication overhead and enhance the scalability as they are identified at individual sensor nodes. Two variations exist for local outlier identification in WSNs. In the first one, each node identifies the abnormal values only depending

on its historical values. In an alternative technique, each sensor node collects readings of its neighbouring nodes in addition to its own historical readings to collaboratively identify the abnormal values. The second approach is more accurate and robust in outlier detection as compared to the first one.

2. *Global Outliers:* Global outliers are identified in a more global perspective. They are particularly of interest since analysts would like to have a better understanding of overall data characteristics in WSNs. In a centralized architecture, all data is transmitted to the sink node for identifying outliers. This mechanism consumes much communication overhead and delays the response time. In aggregate/clustering-based architecture, the aggregator/clusterhead collects the data from nodes within its controlling range and then identifies outliers.

3.3. Identity of Outliers

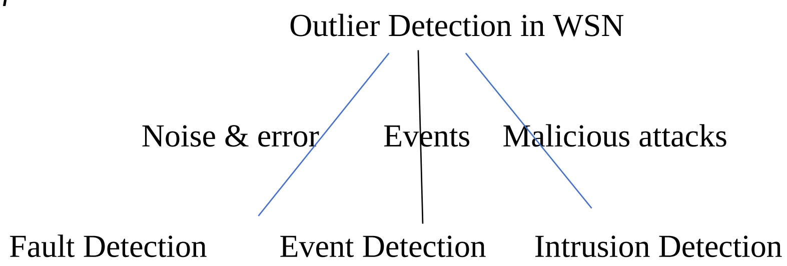

Fig 1. Three outlier sources in WSN and their detection techniques

There are three sources of outliers occurred in WSNs: (1) noise and errors, (2) events, and (3) malicious attacks. The sort of outliers caused by malicious attacks is concerned with the issue of network security and is out of the scope of this paper. For outliers resulted from various sources, outlier detection techniques are desired to specify the identity of these outliers and deal further with them.

1. *Errors:* An error refers to a noise-related measurement or data coming from a faulty sensor. Outliers caused by errors may occur frequently [4]. Erroneous data is normally represented as an arbitrary change and is extremely different from the rest of the data. Since such errors influence data quality, they need to be identified and corrected if possible as data after correction may still be usable for data analysis.
2. *Events:* An event is defined as a specific phenomenon that changes the real-world state, e.g., forest fire, chemical spill, air pollution, etc. This sort of outlier normally lasts for a relatively prolonged time and changes historical pattern of sensor data. However, faulty sensors may also generate similar long segmental outliers as events and therefore it is hard to distinguish the two different outlier sources only by examining one sensing series of a node itself [27].

IV. CLASSIFICATION FRAMEWORK FOR OUTLIER DETECTION TECHNIQUES DESIGNED FOR WSNS

Related work on classification framework for outlier detection techniques for general data has been addressed in various literature. Markou and Singh [28] and [29] present an extensive review of novelty detection techniques based in statistical and neural network fields. Hodge and Austin [5] address outlier detection methodologies from perspective of three fields of computing, i.e., statistics, neural networks and machine learning. Chandola et al. [9] classify anomaly detection techniques in terms of various application domains and several knowledge disciplines. Zhang et al. [8] provide a classification for outlier detection techniques with respect to multiple type of data sets. Although there may be some overlaps between these taxonomies and the one presented here, existing taxonomies are not directly applicable to WSNs due to the nature of sensor data and specific requirements and limitations of WSNs. Additionally, recently, many outlier detection techniques specifically developed for WSNs have emerged. This calls for a classification addressing techniques and requirements of WSNs specifically. In this section, we provide a technique-based t classification framework to categorize these techniques designed for WSNs.

As illustrated in Figure 2, outlier detection techniques for WSNs can be categorized into *statistical-based*, *nearest neighbour-based*, *clustering-based*, *classification-based*, and *spectral decomposition-based* approaches. Statistical-based approaches are further categorized into *parametric* and *nonparametric* approaches based on how the probability distribution model is built [28]. *Gaussian-based* and *non-Gaussian based* approaches belong to parametric approaches, and *kernel-based* and *histogram-based* approaches belong to nonparametric approaches. Classification-based approaches are categorized as *Bayesian network-based* and *support vector*

machine-based approaches based on type of classification model that they use. Bayesian network-based approaches are further categorized into *naive Bayesian network*, *Bayesian belief network*, and *dynamic Bayesian network* based on the degree of probabilistic independencies among variables. Spectral decomposition-based approaches use *principle component analysis* for outlier detection.

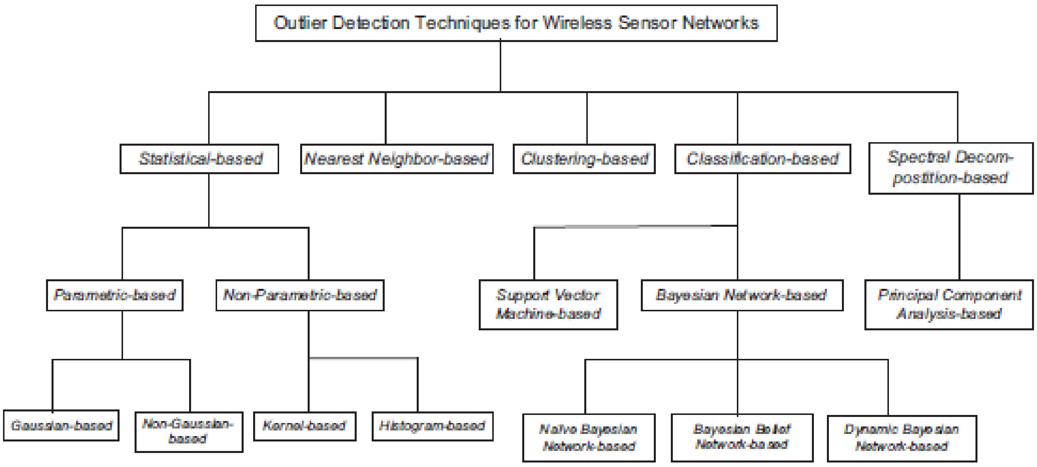
Outlier Detection Techniques for Wireless Sensor Networks
Statistical-based
Nearest Neighbor-based
Clustering-based
Classification-based
Spectral Decom-position-based
Parametric-based
Non-Parametric-based
Support Vector Machine-based
Bayesian Network-based
Principal Component Analysis-based
Gaussian-based
Non-Gaussian-based
Kernel-based
Histogram-based
Naive Bayesian Network-based
Bayesian Belief Network-based
Dynamic Bayesian Network-based

Fig 2. classification of outlier detection techniques in WSNs

V. OUTLIER DETECTION TECHNIQUES FOR WSNS

In this section, we classify outlier detection techniques designed for WSNs based on the discipline from which they adopt their ideas and address the key characteristics and performance analysis of each outlier detection technique using the classification framework presented in Section 4. Furthermore, we provide a brief evaluation for each of these disciplines.

5.1. Statistical-Based Approaches

Statistical-based approaches are the earliest approaches to deal with the problem of outlier detection. The statistical outlier detection techniques are essentially *model-based* techniques. The statistical-based approaches are categorized into parametric and non-parametric based on how the probability distribution model is built.

1. *Parametric-Based Approaches:* Parametric techniques assume availability of the knowledge about underlying data distribution. It then estimates the distribution parameters from the given data. Based on type of distribution assumed, these techniques are further categorized into Gaussian-based models and non-Gaussian-based models. In Gaussian models, the data is assumed to be normally distributed.

- *Gaussian-based models*. Wu et al. [30] present two local techniques for identification of outlying sensors as well as identification of event boundary in sensor networks. These techniques employ the spatial correlation of the readings existing among neighboring sensor nodes to distinguish between outlying sensors and event boundary..
- *Non-Gaussian-based models*. Jun et al. [33] present a statistical-based technique, which uses a symmetric α- stable (SαS) distribution to model outliers being in form of impulsive noise. The

technique utilizes the spatiotemporal correlations of sensor data to locally detect outliers. Each node in a cluster first detects and corrects temporal outliers by comparing the predicted data and the sensing data. However, the SαS distribution may not be suitable for real sensor data and the cluster-based structure may be susceptible to dynamic changes of network topology.

2. *Non-Parametric-Based Approaches:* Non-parametric techniques do not assume availability of data distribution. They typically define a distance measure between a new test instance and the statistical model and use some thresholds on this distance to determine whether the observation is an outlier. Two most widely used approaches in this category are histograms and kernel density estimator.

- *Histogramming [34]*. This technique attempts to reduce communication cost by collecting histogram information rather than collecting raw data for centralized processing. The sink uses histogram information to extract data distribution from the network and filters out the non-outliers.
- *Kernel functions [35]*. This technique requires no a priori known data distribution and uses kernel density estimator to approximate the underlying distribution of sensor data. Thus, each node can locally identify outliers if the values deviate significantly from the model of approximated data distribution. A value is considered as an outlier if the number of values being in its neighbourhood is less than a user-specified threshold. This technique can also be extended to high-level nodes for identification of outlier in a more global perspective.

3. *Evaluation of Statistical-Based Techniques:* Statistical based approaches are mathematically justified and can effectively identify outliers if a correct probability distribution model is acquired. Moreover, after constructing the model, the actual data on which the model is based on is not required.

5.2. Nearest Neighbour-Based Approaches

Nearest neighbour-based approaches are the most commonly used approaches to analyze a data instance with respect to its nearest neighbours in the data mining and machine learning community. They use several well-defined distance notions to compute the distance (similarity measure) between two data instances ([37], [38]). A data instance is declared as an outlier if it is located far from its neighbours. Euclidean distance is a popular choice for univariate and multivariate continuous attributes.

Branch et al. [39] propose a technique based on distance similarity to identify global outliers in sensor networks. This technique attempts to reduce the communication overhead by a set of representative data exchanges among neighbouring nodes. Each node uses distance similarity to locally identify outliers and then broadcasts the outliers to neighbouring nodes for verification. The neighbouring nodes repeat the procedure until all the sensor nodes in the network eventually agree on the global outliers. This technique can be flexible in respect to multiple existing distance-based outlier detection techniques.

Zhang et al. [40] propose a distance-based technique to identify *n* global outliers in snapshot and continuous query processing applications of sensor networks. This technique reduces communication overhead as it adopts the structure of aggregation tree and prevents broadcasting of each node in the network [39]. This technique considers only one-dimensional data and the aggregation tree used may not be stable due to the dynamic changes of network topology.

Zhuang et al. [27] present two in-network outlier cleaning techniques for data collection applications of sensor networks. One technique uses wavelet analysis specifically for outliers such as noises or occasionally appeared errors. The other technique uses dynamic time warping (DTW) distance-based similarity comparison specifically for outliers that are erroneous and last for a certain time. In this technique, each node transforms raw data into the wavelet time-frequency domain and identifies the high-frequency data measurements as outliers and corrects them using proper wavelet coefficients. The long segmental outliers can be detected and removed by comparing the similarity of two sensing series of the neighbouring nodes within two forwarding hops.

1. *Evaluation of Nearest Neighbour-based Techniques:* Nearest neighbour-based approaches do not make any assumption about data distribution and can generalize many notions from statistical-based approaches. However, these techniques suffer from the choice of the appropriate input parameters. Additionally, in multivariate data sets it is computationally expensive to compute the distance between data instances and as a result these techniques lack scalability.

5.3. Clustering-Based Approaches

Clustering-based approaches are popular approaches within the data mining community to group similar data instances into clusters with similar behaviour. Data instances are identified as outliers if they do not belong to clusters or if their clusters are significantly smaller than other clusters. Euclidean distance is often used as the dissimilarity measure between two data instances.

1. *Evaluation of Clustering-Based Techniques:* Clustering based approaches do not require a priori knowledge of the data distribution and are capable of being used in an incremental model, i.e., new data instance can be fed into the system and being tested to find outliers. However, these techniques suffer from the choice of an appropriate parameter of cluster width. Additionally, computing the distance between data instances in multivariate data is computationally expensive.

5.4. Classification-Based Approaches

Classification approaches are important systematic approaches in the data mining and machine learning community. They learn a classification model using the set of data instances (training) and classify an unseen instance into one of the learned (normal/outlier) class (testing). The unsupervised classification-based techniques require no knowledge of available labelled training data and learn the classification model which fits most of the data instance during training. The one-class unsupervised techniques learn the boundary around the normal instances while some anomalous instance may exist and declare any new instance falling outside this boundary as an outlier. Classification-based approaches are categorized into support vector machines (SVM)-based and Bayesian network-based approaches based on type of classification model they use.

1. *Support Vector Machine-Based Approaches:* SVM techniques separate the data belonging to different classes by fitting a hyperplane between them which maximizes the separation. The data is mapped into a higher dimensional feature space where it can be easily separated by a hyperplane. Furthermore, a kernel function is used to approximate the dot products between the mapped vectors in the feature space to find the hyperplane. Rajasegarar et al. [42] propose a SVM-based technique for outlier detection in sensor data which uses one-class quarter-sphere SVM to reduce the effort of computational complexity and locally identify outliers at each node.
2. *Bayesian Network-Based Approaches:* Bayesian network-based approaches use a probabilistic graphical model to represent a set of variables and their probabilistic independencies. They aggregate information from different variables and provide an estimate on the expectancy of an event to belong to the learned class. They are categorized as naive Bayesian network, Bayesian belief network, and dynamic Bayesian network approaches based on degree of probabilistic independencies among variables.
   - *Naive Bayesian Network models*. This technique [24] maps the problem of learning spatio-temporal correlations to the problem of learning the parameters of the Bayesian classifier and then uses the classifier for probabilistic inference. Each node locally computes the probabilities of each of its incoming readings being in all subintervals (classes) divided from the whole values interval. If the probability of a sensed reading in its class is smaller than that of being in other classes, it is considered as an outlier.
   - *Bayesian Belief Network models*. This technique [23] uses BBN to capture not only the spatiotemporal correlations that exist among the observations of sensor nodes but also conditional dependence among the observations of sensor attributes. Each node trains a BBN to detect outliers based on behaviours of its neighbours' readings as well as its own reading. An observation is

considered as outlier if it falls beyond the range of the expected class. Compared to naive Bayesian networks, this technique improves the accuracy in detecting outliers as it considers conditional dependencies among the attributes.

- *Dynamic Bayesian Network models*. This technique [43] uses DBNs to fast track changes in dynamic network topology of sensor networks. It identifies outliers by computing the posterior probability of the most recent data values in a sliding window. The data measurements that fall outside the expected value interval are considered as outliers.

3. Evaluation of Classification-based Techniques: Classification-based approaches provide an exact set of outliers by building a classification model to classify. However, a main drawback of SVM-based techniques is their computational complexity and the choice of proper kernel function. Learning the accurate classification model of a Bayesian network is challenging if the number of variables is large in deployed WSNs.

5.5. Spectral Decomposition-Based Approaches

Spectral decomposition-based approaches aim at finding normal modes of behaviour in the data by using principle components. Principal component analysis (PCA) is a technique that is used to reduce dimensionality before outlier detection and finds a new subset of dimension which capture the behaviour of the data. Chatzigiannakis et al. [26] propose a PCA-based technique to solve data integrity and accuracy problem caused by compromised or malfunctioning sensor nodes. This technique uses PCA to efficiently model the spatio-temporal data correlations in a distributed manner and identifies local outliers spanning through neighbouring nodes. Each primary node offline builds a model of the normal condition by selecting appropriate principal components (PCs) and then obtains sensor readings from other nodes in its group and performs local real-time analysis. The readings that significantly vary from the modelled variation value under normal condition are declared as outliers.

1. Evaluation of Spectral Decomposition-Based Techniques: Principal component analysis-based approaches try to capture the normal pattern of the data using the subset of dimensions and can be applied to high-dimensional data. However, selecting suitable principle components, which is needed to accurately estimate the correlation matrix of normal patterns, is computationally very expensive.

VI. CONCLUSION

In this paper, we address the problem of outlier detection in WSNs and provide a technique-based classification framework to categorize current outlier detection techniques designed for WSNs. We also introduce the key characteristics and brief description of current outlier detection techniques using the proposed classification framework and provide an evaluation for each technique. The shortcomings of existing techniques for WSNs clearly calls for developing outlier detection technique which takes into account multivariate data and the dependencies of attributes of the sensor node, provides reliable neighbourhood, proper and flexible decision threshold, and also meets special characteristics of WSNs such as node mobility, network topology change and making distinction between errors and events.

REFERENCES

1. X. Ma, D. Yang, S. Tang, Q. Luo, D. Zhang, and S. Li, Online Mining in Sensor Networks, *IFIP international conference on network and parallel computing*, Vol. 3222, pp. 544-550, 2004.
2. S. Subramaniam, T. Palpanas, D. Papadopoulos, V. Kalogerakiand, and D. Gunopulos, Online Outlier Detection in Sensor Data using Nonparametric Models, *J. Very Large Data Bases*, VLDB 2006.
3. A. Perrig, J. Stankovic, and D. Wagner, Security in Wireless Sensor Networks, *CACM*, Vol. 47, No. 6, pp. 53-57, 2004.
4. F. Martincic and L. Schwiebert, Distributed Event Detection in Sensor Networks, *Proc. International Conference on Systems and Networks Communication*, pp. 43-48, 2006.
5. V. Hodge and J. Austin, A Survey of Outlier Detection Methodologies, *Artificial Intelligence Review*, Vol. 22, pp. 85-126, 2003.
6. D.M. Hawkins, *Identification of Outliers*, London: Chapman and Hall, 1980.
7. V. Barnett and T. Lewis, *Outliers in Statistical Data*, New York: John Wiley Sons, 1994.
8. Y. Zhang, N. Meratnia, and P.J.M. Havinga, A Taxonomy Framework for Unsupervised Outlier Detection Techniques for Multi-Type Data Sets, *Technical Report*, University of Twente, 2007.

9. V. Chandola, A. Banerjee, and V. Kumar, Anomaly Detection: A Survey, *Technical Report*, University of Minnesota, 2007.
10. P.N. Tan, M. Steinback, and V. Kumar, *Introduction to Data Mining*, Addison Wesley, 2006.
11. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, San Francisco, 2006.
12. J. Chen, S. Kher, and A. Somani, Distributed Fault Detection of Wireless Sensor Networks, *Proc. 2006 workshop on dependability issues in wireless ad hoc networks and sensor networks*, pp. 65-72, 2006.
13. X. Luo, M. Dong, and Y.Huang, On Distributed Fault-Tolerant Detection in Wireless Sensor Networks, *IEEE Trans. Comput.*, Vol. 55, No. 1, pp. 58-70, 2006.
14. M. Ding, D. Chen, K. Xing, and X. Cheng, Localized Fault-Tolerant Event Boundary Detection in Sensor Networks, *Proc. IEEE Conference of Computer and Communications Societies*, pp. 902- 913, 2005.
15. B. Krishnamachari and S. Iyengar, Distributed Bayesian Algorithms for Fault-Tolerant Event Region Detection in Wireless Sensor Networks, *IEEE Trans. Comput.*, Vol. 53, No. 3, pp. 241- 250, 2004.
16. A.P.R. Silva, M.H.T. Martins, B.P.S. Rocha, A.A.F. Loureiro, L.B. Ruiz, and H.C. Wong, Decentralized Intrusion Detection in Wireless Sensor Networks, *Proc. 1st ACM international workshop on Quality of service & security in wireless and mobile network*s, pp. 16-23, 2005.
17. V. Bhuse, and A. Gupta, Anomaly Intrusion Detection in Wireless Sensor Networks, *J. High Speed Networks*, Vol. 15, No. 1, pp. 33-51, 2006.
18. M. Zoumboulakis and G. Roussos, Escalation: Complex Event Detection in Wireless Sensor Networks, *Lecture Notes in Computer Science*, Springer Berlin/Heidelberg, pp. 270-285, 2007.
19. P.N. Tan, Knowledge Discovery from Sensor Data, *Sensors*, 2006.
20. I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, Wireless Sensor Networks: A Survey, *J. Computer Networks*, Vol. 38, No. 4, pp. 393-422, March, 2002.
21. M. M. Gaber, Data Stream Processing in Sensor Networks. In J. Gama and M. M. Gaber, *Learning from Data Streams Processing Techniques in Sensor Network*, pp. 41-48. Springer Berlin Heidelberg, 2007.
22. P. Sun, Outlier Detection in High Dimensional, Spatial and Sequential Data Sets, *Doctoral dissertation*, University of Sydney, Sydney, 2006.
23. D. Janakiram, A. Mallikarjuna, V. Reddy, and P. Kumar, Outlier Detection in Wireless Sensor Networks using Bayesian Belief Networks, *Proc. IEEE Comsware*, 2006.
24. E. Elnahrawy and B. Nath, Context-Aware Sensors, *Proc. EWSN*, 2004.
25. S.R. Jeffery, G. Alonso, M.J. Franklin, W. Hong, and J. Widom, Declarative Support for Sensor Data Cleaning, *International Conference on Pervasive Computing*, pp. 83-100, 2006.
26. V. Chatzigiannakis, S. Papavassiliou, M. Grammatikou, and B. Maglariset, Hierarchical Anomaly Detection in Distributed Large-ScaleSensor Networks, *Proc. ISCC*, 2006.
27. Y. Zhuang and L. Chen, In-Network Outlier Cleaning for Data Collection in Sensor Networks, *Proc. VLDB*, 2006.
28. M. Markos, S. Singh, Novelty Detection: A Review-Part 1: Statistical Approaches. *J. Signal Processing*, Vol. 83, pp. 2481-2497, 2003.
29. M. Markos, S. Singh, Novelty Detection: A Review-Part 2: Neural Network based Approaches. *J. Signal Processing*, Vol. 83, pp. 2499-2521, 2003.
30. W. Wu, X. Cheng, M. Ding, K. Xing, F. Liu, and P. Deng, Localized Outlying and Boundary Data Detection in Sensor Networks, *IEEE Trans. Knowl. Data Eng.*, Vol. 19, No. 8, pp. 1145-1157, 2007.
31. L.A. Bettencourt, A. Hagberg, and L. Larkey, Separating the Wheat from the Chaff: Practical Anomaly Detection Schemes in Ecological Applications of Distributed Sensor Networks, *Proc. IEEE International Conference on Distributed Computing in Sensor Systems*, 2007.
32. Y. Hida, P. Huang, and R. Nishtala, Aggregation Query under Uncertainty in Sensor Networks, 2003.
33. M.C. Jun, H. Jeong, and C.C.J. Kuo, Distributed Spatio-Temporal Outlier Detection in Sensor Networks, *Proc. SPIE*, 2006.
34. B. Sheng, Q. Li, W. Mao, and W. Jin, Outlier Detection in Sensor Networks, *Proc. MobiHoc*, 2007.
35. T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos, Distributed Deviation Detection in Sensor Networks, ACM *Special Interest Group on Management of Data*, pp. 77-82, 2003.
36. S. Papadimitriou, H. Kitagawa, P.B. Gibbons, and C. Faloutsos, LOCI: Fast Outlier Detection using the Local Correlation Integral, *International Conference on Data Engineering*, pp. 315-326, 2003.
37. E. Knorr and R. Ng, Algorithms for Mining Distance-Based Outliers in Large Data Sets, *International Journal of Very Large Data Bases*, pp. 392-403, 1998.
38. S. Ramaswamy, R. Rastogi, and K. Shim, Efficient Algorithms for Mining Outliers from Large Data Sets, *ACM Special Interest Group on Management of Data*, pp. 427-438, 2000.
39. J. Branch, B. Szymanski, C. Giannella, and R. Wolff, In-Network Outlier Detection in Wireless Sensor Networks, *Proc. IEEE ICDCS*, 2006.
40. K. Zhang, S. Shi, H. Gao, and J. Li, Unsupervised Outlier Detection in Sensor Networks using Aggregation Tree, *Proc. ADMA*, 2007.
41. S. Rajasegarar, C. Leckie, M. Palaniswami, and J.C. Bezdek, Distributed Anomaly Detection in Wireless Sensor Networks, *Proc. IEEE ICCS*, 2006.
42. S. Rajasegarar, C. Leckie, M. Palaniswami, and J. C. Bezdek, Quarter Sphere Based Distributed Anomaly Detection in Wireless Sensor Networks, *Proc. IEEE International Conference on Communications*, pp.3864-3869, 2007.
43. D.J. Hill, B.S. Minsker, and E. Amir, Real-Time Bayesian Anomaly Detection for Environmental Sensor Data, *Proc. 32nd Congress of the International Association of Hydraulic Engineering and Research*, 2007.