International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307, Volume-1, Issue-2, May 2011

Outlier Detection Using Support Vector Machine in Wireless Sensor Network Real Time Data

M. Syed Mohamed, T. Kavitha

Abstract---**Outlier detection has many important applications in sensor networks, e.g., abnormal event detection, animal behavior change, intruder detection etc. Outliers in wireless sensor networks (WSNs) are sensor nodes that issue attacks by abnormal behaviours and fake message broadcasting. The probable sources of outliers include noise and errors, events, and malicious attacks on the network. Wireless sensor networks (WSNs) are more likely to generate outliers due to their special characteristics, e.g. constrained available resources, frequent physical failure, and often harsh deployment area. In this project we motivate our technique in the context of the problem of outlier detection. This paper is going to present the real time network outlier detection method in the wireless sensor networks. We proposed the technique to classify the sensor node data as local outlier or cluster outlier or network outlier using Standard Support Vector Machine classification method which is one of the best classification methods among the various outlier detection methods. If the data is classified as network outlier then it may be due an event otherwise if it is classified as a cluster outlier then it is an error in the cluster due to some environmental factor or network otherwise is an error in the sensor node due to some defect in that sensor. Experiments with real data show that our method is efficient and accurate to detect the outliers in real time. The real time data are collected from the Sensor Scope system and implemented using MATLAB.**

Index Terms---**Outliers, Support Vector Machine (SVM), Wireless Sensor Network (WSN)**

I. INTRODUCTION

The objective in outlier detection is to find data elements those are most deviated from the remaining data based on some measure. Outliers arise due to mechanical faults, changes in system behavior, fraudulent behavior, human error, instrument error or simply through natural deviations in populations.

Manuscript received April, 2011

M.Syed Mohamed, M.E. Scholar (CSE), Anna University of Technology, Tirunelveli, Tamilnadu, India. (e-mail: seyadumohamed@gmail.com)

T.Kavitha, Assistant Professor, Department of Computer Science and Engineering, Anna University of Technology, Tirunelveli, Tamilnadu, India.. (e-mail: kavi_gce@rediffmail.com)

A. Wireless Sensor Network

Sensors integrated into structures, machinery, and the environment, coupled with the efficient delivery of sensed information, could provide tremendous benefits to society. Wireless sensing networks can eliminate costs, easing installation and eliminating connectors. The ideal wireless sensor is networked and scaleable, consumes very little power, is smart and software programmable, capable of fast data acquisition, reliable and accurate over the long term, costs little to purchase and install, and requires no real maintenance.

B. Architecture of Wireless Sensor Network

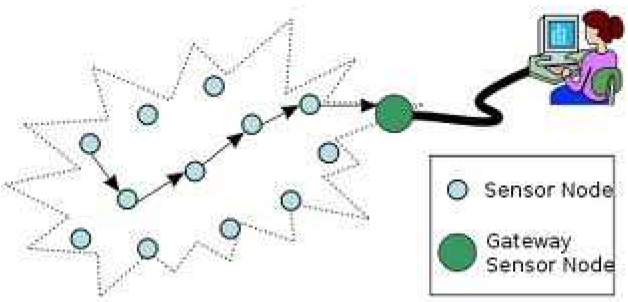

Fig 1: **Architecture of Wireless Sensor Network [11]**

1. In WSN, the tiny sensors are deployed all over the Field.
2. The Networks are usually composed of few sinks and large quantity of sensor nodes.
3. Sensor nodes are organized into clusters.
4. Each node has a corresponding cluster header.
5. Each sensor node can measure temperature, smoke and relative humidity.
6. Nodes location information can be obtained by equipments such as Global Positioning System (GPS).

C. Outliers

1. Type of Outliers

Local Outliers---Each node identifies the anomalous values only depending on its historical values [1]-[3]. Each sensor node collects readings of its neighboring nodes to collaboratively identify the anomalous values.

Global Outliers---The identification of global outliers can be performed at different levels in the network. In a centralized architecture, all data is transmitted to the sink node for identifying outliers.

2. Sources of Outliers

Errors---An error refers to a noise-related measurement or data coming from a faulty sensor [12],[13]. Outliers caused by errors may occur frequently, while outliers caused by events tend to have extremely smaller probability of occurrence.

Events---An event is defined as particular phenomena that change the real-world state, e.g., forest fire, chemical spill, air pollution, etc.

D. Advantages of SVM-Based Classification Technique

It is the popular classification-based approaches in the data mining and machine learning communities [6]-[10]. It is widely used to detect outliers due to the following main advantages:

1. Do not require an explicit statistical model
2. Provide an optimum solution for classification by maximizing the margin of the decision boundary
3. Avoid the dimensionality problem

II. PROBLEM DESCRIPTION

Sensors in the WSN are organized into clusters which are collection of nearest sensors. Our aim is to classify each sensor data in all the clusters are classified as a node outlier or cluster outlier or network outlier. We are initially collecting 'n' sensor data for training in real time and finding each sensor radius and its mean radius. Also we are finding median radius of all the sensors in that clusters. This process is performed real time as a training value. The mean radius of each sensors and median value of the cluster is set to initial threshold values.

In the testing phase, when a new data is received from a sensor, that value is compared with the initial threshold values. If the new data is less than the initial threshold value, it is classified as a normal data and then training set is also updated with the new data by inserting new data and eliminating earliest data and compute mean radius and median radius as the new threshold values. Otherwise, if the new sensor data is more than the mean radius of that sensor and less than the median value of that cluster then it is classified as node outlier. If the new data is more than both thresholds it is classified as cluster outlier and that data is ignored for updating the training set. Finally, collecting all the cluster outlier data and it is compared with the mean cluster radius. If the cluster outlier data is less than mean radius of cluster radius, it is classified as cluster outlier otherwise it is classified as network outlier which may be an event. This process is depicted in the following diagrams.

A. Architecture of the Proposed Method

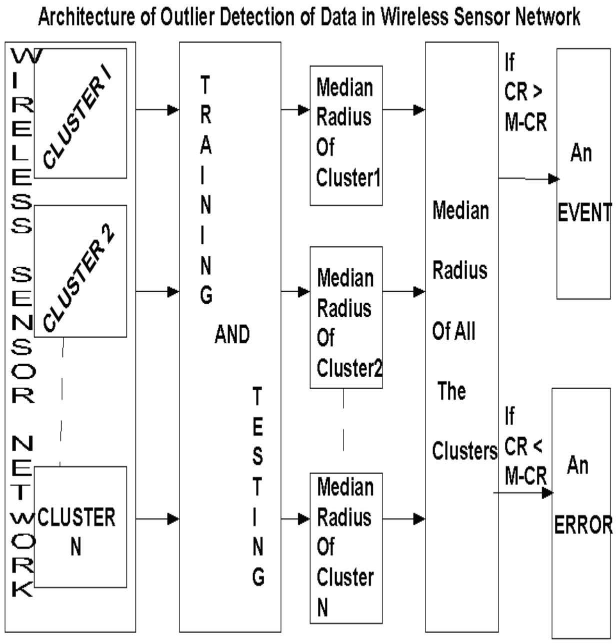

Fig.2: **Complete Architecture of Outlier Detection of Data in WSN**

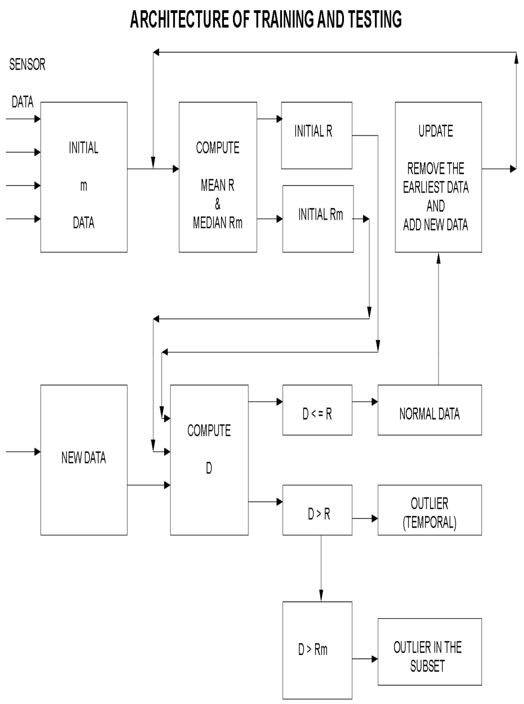

Fig. 3: **Architecture of Training and Testing of Data in WSN (Phase I)**

The Fig.3 explains the training and testing procedure. Initial 'n' sensor data is taken for testing and finding mean value of 'n' sensor data and it is considered as initial R and also finding similar process for all the sensor data and then finding median value of those sensors which is considered as median of sensors in a cluster as initial Rm. When a new data coming from a sensor its value is compared with its initial value and it is classified as normal or outlier as shown in the Fig.3. If the new data is normal then the initial R and Rm values are updated otherwise it keeps earlier values. In phase-I, the new sensor data is classified as node outlier or cluster outlier. If the data is classified as cluster outlier then it is given to the phase-II for further classification of data as error or an event.

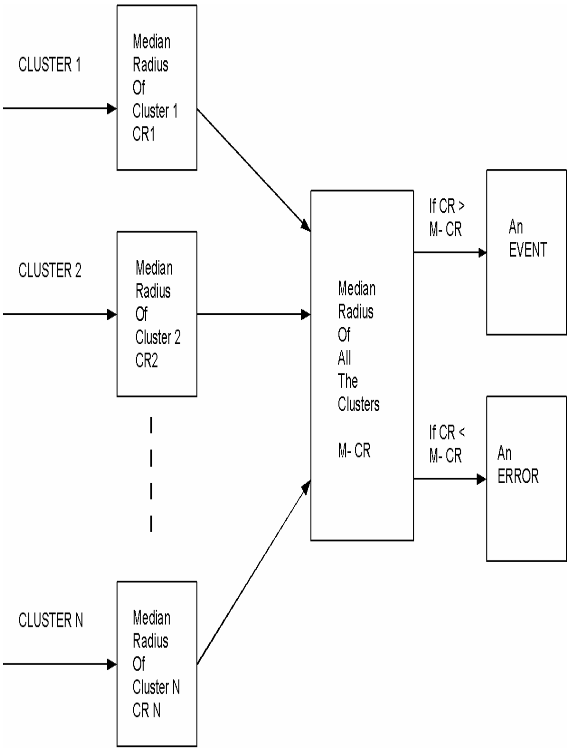

Fig.4: **Architecture of Global / Cluster Outlier Detection of Data in WSN (Phase II)**

The Fig.4 explains the process of classify the cluster outlier as an event or an error in the network by comparing with all the cluster data as shown in the Fig.4. If the sensed data classified as an event then the rescue process will start otherwise rectify the error in the network.

III. RESULTS AND DISCUSSION

A. Experimental Dataset

The real data are collected from a closed neighborhood from a WSN deployed in *Grand-St-Bernard* [4] as shown in Fig.5. The results are obtained using MATLAB.

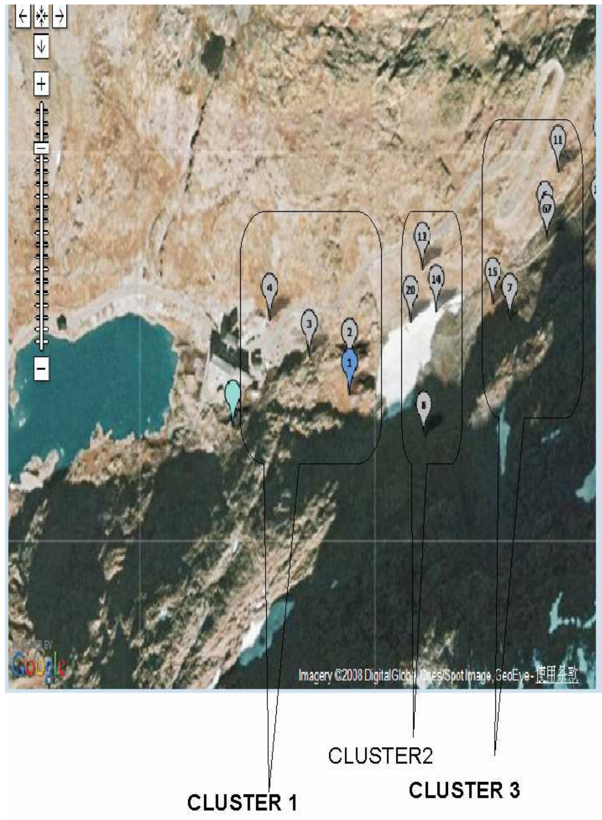

Fig. 5: **Grand-St-Bernard Deployment**

B. Result Analysis

The Fig.6 represents the training set of node data. It finds some outliers and normal data. The Fig.7 represents the classified data of sensor node which identifies local outliers of sensor node. The Fig.8 represents the comparison of node outliers in cluster-1. The Fig.9 represents the comparison of node outliers in cluster-2. The Fig.10 represents the comparison of outliers in all the clusters. In cluster-1 the node-4 and in cluster-2 the node-2 has more outlier data may be due to error in those sensor nodes. The cluster-1 has more outliers compared with cluster-2.

As per our experimental results, new sensor data are effectively classified as normal or event because of real time training sets which are updated for every normal sensor data which leads to some computational complexity but yields better results.

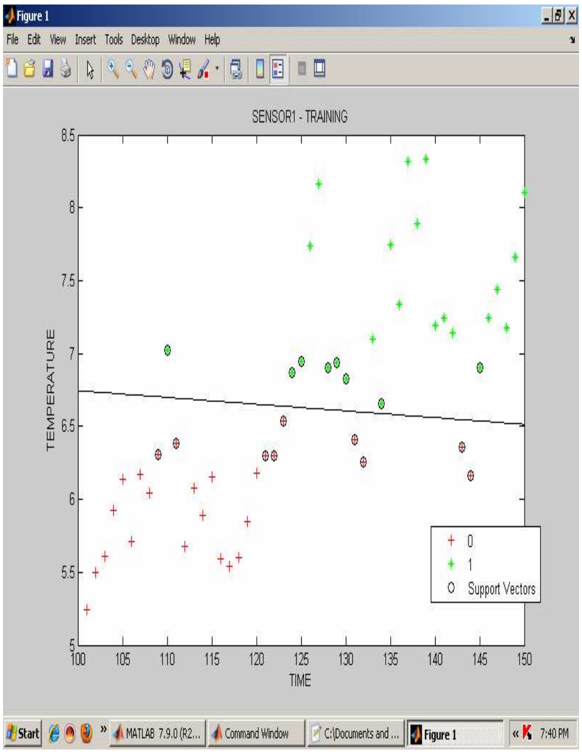

Fig.6: **Training Set of Sensor Node**

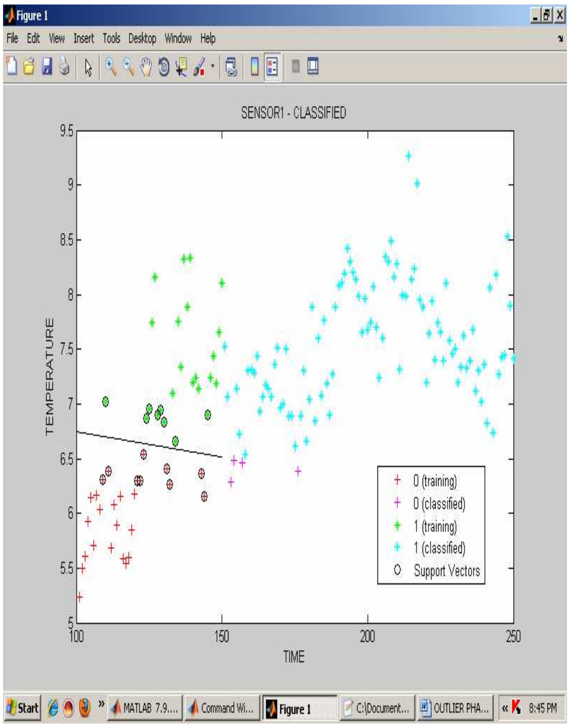

Fig. 7: **Testing of Sensor Node Data**

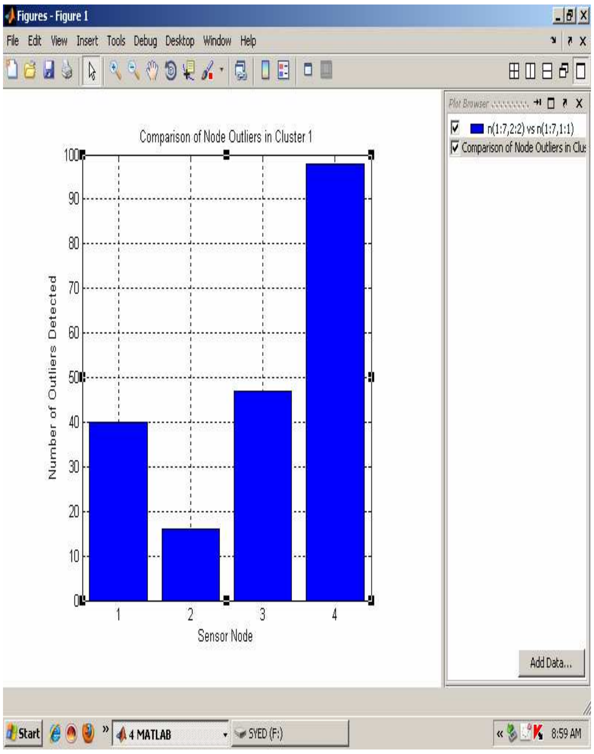

Fig. 8: **Comparison of Node Outliers in Cluster 1**

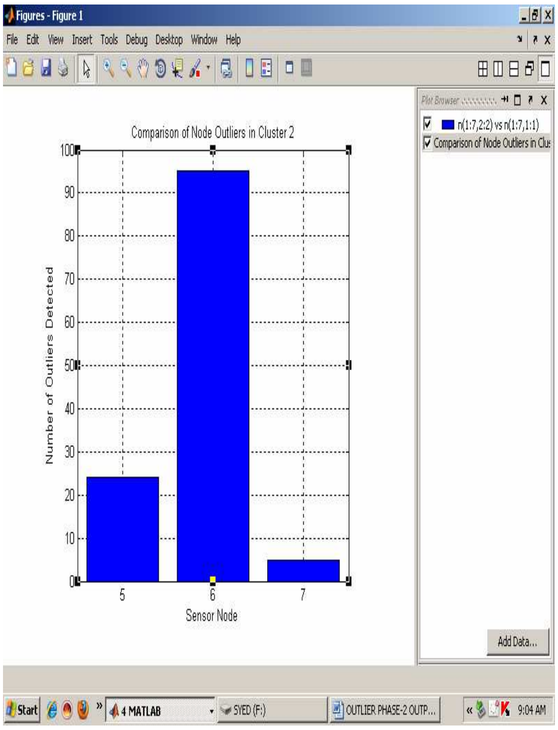

Fig. 9: **Comparison of Node Outliers in Cluster 2**

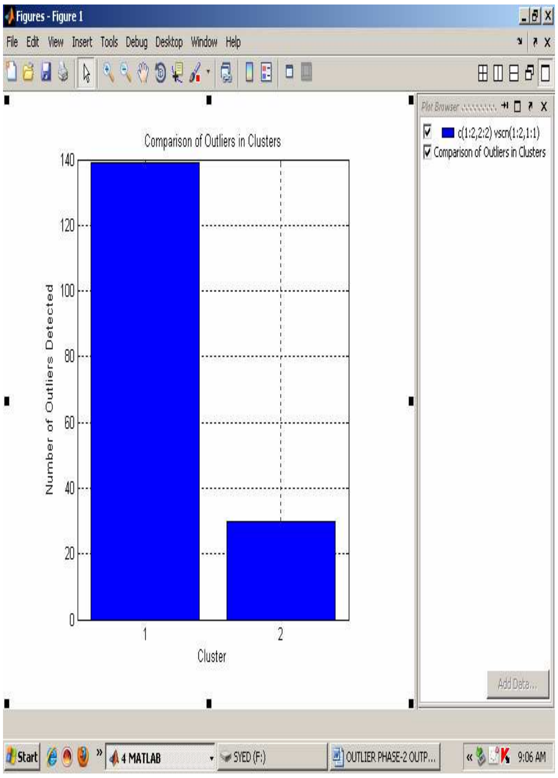

Fig. 10: **Comparison of Outliers in Clusters**

IV. CONCLUSION

In this paper, we have examined the real time wireless sensor data as outlier or normal based on SVM based outlier detection technique. We also classified the outlier data as node or cluster or global outliers. We compared the performance of the technique with synthetic data and real data of the SensorScope System. We found that our method finds the outlier with better accuracy and minimum false alarm rate. In our experiments, we find some computational complexity exists because of updating the training set in real time. In future, we find the solution for minimizing the computational complexity.

V. REFERENCES

[1] Chandola V, A. Banerjee, and V. Kumar. Outlier detection: A survey. *Technical Report, University of Minnesota*, 2007

[2] M. Syed Mohamed and T. Kavitha. Real Time Outlier Detection in Wireless Sensor Networks. *International Journal of Latest Trends in Computing (E-ISSN: 2045-5364) 114 . Volume 2, Issue 1, March 2011*

[3] Ding M, D. Chen, K. Xing, and X. Cheng, "Localized fault-tolerant event boundary detection in sensor networks," in *Proceedings of IEEE Conference of Computer and Communications Societies*, 2005, pp. 902-913.

[4] *http: //sensorscope.epfl.ch/index.php/MainP age.*

[5] Luo X, M. Dong, and Y. Huang, "On distributed fault-tolerant detection in wireless sensor networks", *IEEE Transactions on Computers,* vol. 55, pp. 58-70, 2006.

[6] M.Mohamed Sathik, M.Syed Mohamed and A. Balasubramanian. Fire Detection Using Support Vector Machine in Wireless Sensor Network and Rescue Using Pervasive Devices. International Journal of Advanced Networking and Applications 636 Volume:02, Issue:02, Pages:636-639 (2010).

[7] Rajasegarar S., C. Leckie, M. Palaniswami, and J. C. Bezdek. Quarter sphere based distributed anomaly detection in wireless sensor networks. *IEEE International Conference on Communications*, June 2007.

[8] Tax D.M.Jand R. P. W. Duin. Support vector data description. *Machine Learning*, 54(1):45-66, 2004.

[9] J.Platt, "Fast training of support vector machine using sequential minimal optimization," Advances in Kernel Methods: support vector machine, MIT Press, Cambridge, MA, 1998

[10] Jair Cervantes, Xiaoou Li , Wen Yu " Multi-Class SVM for Large Data Sets Considering Models of Classes Distribution"

[11] http://en.wikipedia.org/wiki/Wireless_sensor_network

[12] Zhang Y, N. Meratnia, and P. Havinga. An online outlier detection technique for wireless sensor networks using unsupervised quarter-sphere support vector machine. *ISSNIP 2008*

[13] Zhang Y., N. Meratnia, and P. Havinga. Outlier detection techniques for wireless sensor networks: A survey. *Technical Report, University of Twente*, 2007.

SHORT BIODATA OF THE AUTHORS

Mr.M.Syed Mohamed has received his MCA degree in 2002 and M.Phil in 2006 from Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India. He is now doing M.E. Computer Science and Engineering in Anna University of Technology, Tirunelveli, Tamilnadu, India. His areas of interest are Wireless Sensor Networks and Data Mining. He has published many papers in National and International conferences and also in the International Journals in areas such as Wireless Sensor Networks and Data Mining. He is also having twenty years of teaching experience.

Mrs. T. Kavitha obtained her B.E degree in Computer Science and Engineering from MK University in 1999 and M.E degree in Govt. college of Engineering, Tirunelveli, Tamilnadu. Right from 2005 she is in the Department of Computer Science & Engineering under various designations. She is now working as Assistant Professor in the Department of CSE at Anna University Tirunelveli, Tirunelveli, Tamilnadu, India. Her research interests are Wireless Sensor Networks, Mobile Ad hoc Networks, and Image Processing. She has presented many papers in National and International Conferences and also in International Journals in the same area. She is a member of ISTE.