IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-10, Oct 2018 E-ISSN: 2347-2693

Study of Use of Classification Techniques in WSN Data Mining for Resource Optimization

B. A. Parbat [1*], R. K. Dhuware[2]

[1*]Dept. Of Computer Science, R. T. M. Nagpur University, Nagpur, India
[2] Dept. of Computer Science, Dhote Bandhu Science College, Gondia (M.S.), India

[*]*Corresponding Author: bhawana.parbat@gmail.com, Tel.: +07974391994*

Available online at: www.ijcseonline.org

Accepted: 20/Oct/2018, Published: 31/Oct/2018

Abstract— With the wide application of Wireless Sensor Network Technology, a large volume of data is generated. For extracting knowledgeful, understandable and valid patterns from this data, data mining techniques are used. This Wireless Sensor Network Data Mining may use Centralized Mining Approach or Distributed Mining approach. Distributed mining, mining is applied on sensor nodes. After that mined data are sent to sink node. But, in centralized approach whole data from sensor nodes are collected at sink node then mining is applied on dataset. This paper focuses on Centralized Data Mining Approach to mine dataset. Here, Classification Techniques, SVM (support Vector Machine) and KNN (K-Nearest Neighbour), are applied on this collected dataset with taking concentration on optimization of CPU cycle as compressible resource. For this execution time to classify data is used here. For this real dataset, it is resulting that KNN is giving better performace than SVM. The dataset is gathered from a real time data acquisition system based on wireless sensor network that is implemented using XBee Digi modules and open source hardware platform Arduino. It is trying to make a hybrid framework, combination of Distributed Approach and Centralized Approach, for this real time deployment of WSN as a future work.

Keywords— Wireless Sensor Network Data Mining, Centralized Mining Approach, SVM, KNN, Resource Optimization

I. INTRODUCTION

In recent years, Wireless Sensor Networks (WSNs) have attracted rich research efforts. It consists of sensors that are integrated with a physical environment. These sensors are small in size, and capable of sensing physical phenomenon and processing them. Wireless radio transceiver, actuators, small CPU and memory are equipped in a sensor node. A sensor node may have one or more sensors. These sensors can measure pressure, sound, temperature, light, humidity, vibration, etc. Wireless Sensor Nodes are deployed in a mode of area where physical communication is not possible like in open, unsupervised, hostile environment. It provides the facility to capture information in both outdoor and indoor applications. Currently, WSN has been considered as one of the most important technology that is absolutely necessary in our daily uses. WSNs have been successfully applied for object tracking [1, 2], military [3, 4], habitat monitoring [5,6], environment monitoring [7–9], disaster management [10], as well as smart environments etc. WSN produces a large dataset. The capabilities for collecting and storing data have far outpaced someone's abilities to analyze, summarize, and extract knowledge from these data.

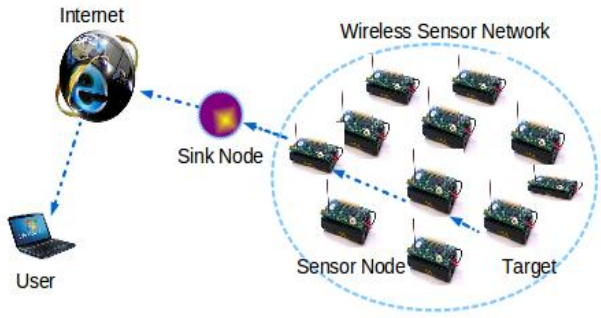

Figure 1 Wireless Sensor Network

The well organized discovery of potentially useful, intelligible, previously unknown, and valid patterns in large datasets is known as Data Mining. With the recent technological advancement, the wireless sensor nodes are getting smaller, but Wireless Sensor Networks are getting larger. In recent scenarios, WSNs contain thousands of nodes and may be millions of nodes in the future. For dealing with the huge amount of data generated by these special kinds of wireless sensor networks, one approach is to use Data Mining techniques. The raw data collected from WSNs, can be examined in detail and transformed to useful information

through data mining. By using data mining techniques, automated or human induced tactical/strategic decision can be provided. As most of the areas will be ruled by wireless sensors in coming years, so it will be challenging to select a correct data mining technique. Here, time monitoring is the essential requirement. Usually the data obtained from wireless sensor network are stored in the central base station in the form of numeric data. Classification is one of the major tasks of data mining. Classification of sensor data is an important research problem in WSNs.

This work is based on study of use of Wireless sensor network Data Mining Techniques. There may be two approaches of data mining in WSN. One may be Distributed Approach and another may be Centralized Approach. In this work, centralized approach of data mining is used. Data sensed by sensor nodes are sent to sink node, and data mining is applied on this gathered data. Among data mining techniques, this work is based on study of use of classification techniques for resource optimization. This work emphasises on execution time as a resource to classify the WSN data by relating execution time to compressible resource, CPU Cycle.

A. Wireless Sensor Network Data Mining Techniques

Data mining is a process of finding valid patterns and knowledge from huge amounts of data. A data source may include database, data warehouse, the Web, the information repository or data that are streamed into the system dynamically [11].

Data mining Techniques that are applied on WSN can be referred as Wireless Sensor Network Data Mining. As in the previous paragraph, it may be concluded that data may be collected or streamed for mining. So, the data mining for WSNs may be one of the following approaches:

1) **Centralized Approach:** In this approach whole data of WSN are collected at sink node, and data mining is applied on this collected data. It is a traditional data mining approach.
2) **Distributed Approach:** This is framework is completely different from centralized approach of mining. Here, data mining is applied on data streams that are coming online. It accomplishes data mining operations based on the type and the available distributed resources.

Rest of the paper is organized as follows, Section I contains the introduction of WSN, Data Mining, WSN Data Mining. Section II contains the related work of Wireless Sensor Network Data Mining. Section III is Methodology that contains Centralized Data Mining Approach of WSN Data Mining, Implemented Real Time System and Dataset, Resource Optimization and experimentation, Section IV describes Results and Discussion, Section V concludes Research Work with Future Scope.

II. RELATED WORK

Wireless Sensor Network Data Mining is emerging as a novel area of research and it offers wide application areas.

In [12], Azhar Mahmood et.al, presented a survey on Data Mining Techniques for WSN. In this paper, comparision of data mining techniques for WSNs is given with Processing Architecture and Data Mining Methods. In this paper, it is clear that in centralized approach whole raw data gathered from WSNs are transmitted to central node (sink node) where offline mining can take place for analysis.

In [13], Rekha Sunny T et.al, presented a survey on Distributed Data Mining in P2P Networks. Centralized Approach of Data Mining is also specified in this paper. It specifies that in this approach all data are collected into a central site, then an algorithm is run against the data for mining.

Prachi Singh, [14], presented literature review to find behavioural patterns between the sensor data in Wireless Sensor Network. They also suggested two possible methodologies for data extraction from WSN. One is the direct transmission, in which without any optimization the data is transmitted to the central site. And in another methodology, optimization is tried by each node before sending to central site.

In [15], Azzedine Boukerche et.al, proposed a data mining algorithm to take out the frequent patterns efficiently. And comparison of the performance of mining algorithm with FP-Growth also presented in this paper. This paper also says about existence of Centralized methodology to extract dataset from central site before applying data mining techniques.

In [16], Azhar Mahmood et.al, presented a review article on mining of data generated from sensor networks. It is concluded in past research the selected existing mining research approaches for mining of sensor networks are: Centralized Data Mining Approach and Distributed Data Mining Approach.

Aditya Kumar Naik et.al, [17], presented a review on use of data mining methods in WSN. This paper also explains about two data mining approaches for WSN: Centralized and Distributed.

In[18], C. Sudha et.al, inspected and presented a survey on existing data mining strategies for WSNs with various grouping , assessment approaches.

In [19], Mirjana Maksimović et.al, presented comparative study of data mining techniques for WSN data for fire

detection. Classification techniques had been implemented in a data mining tool, WEKA, on a formed dataset in this paper. The purpose was to apply the learning algorithms and to select the best one for prediction purpose.

In [20], authors consider a MIMO multicell system where multiple mobile users ask for computation offloading to a common cloud server. For computational resource, CPU Cycle is considered here. Each module to be executed is characterized by the number of CPU cycles necessary to run the module itself.

III. METHODOLOGY

A. Centralized Data Mining Approach of WSN Data Mining:

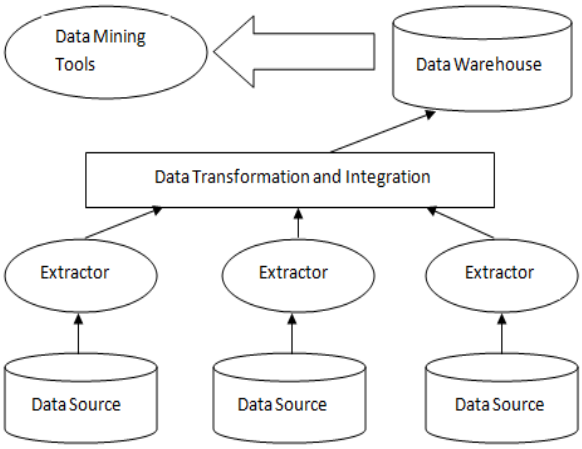

Figure 2 Data Mining –Centralized Approach [19]

B. Implemented Real time System and Dataset

The experiments performed in this paper evaluate the datasets obtained from previously implemented and deployed Real Time Data Acquisition System for WSN using XBee Digi modules and open source hardware platform Arduino. The network consists of three modules: two sensor nodes, Node A and Node B, and a sink node as in the following Figure 2. Each node is equipped with Atmega 328 micro controller, XBee S2 module and DHT11 sensors [21]. Figure 3 and Figure 4 shows Transmitter Module and Receiver Module respectively. This WSN was deployed in a polyhouse. A polyhouse is a type of greenhouse and modern farming that is constructed using Polyethylene as main material which stabilizes the ultraviolet rays and helps in proper photosynthesis in crops [21].

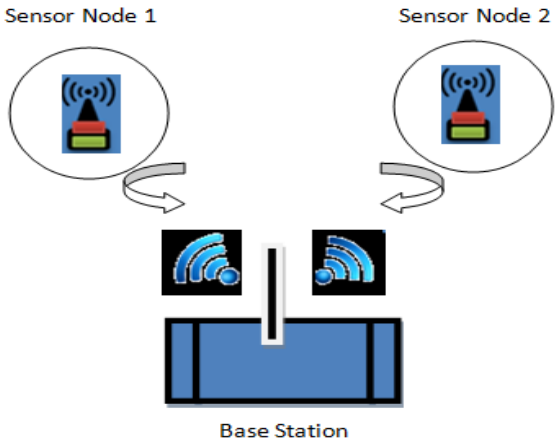

Figure 2 Network Architecture[21]

Figure 3 Sensor Node (Transmitter Module)

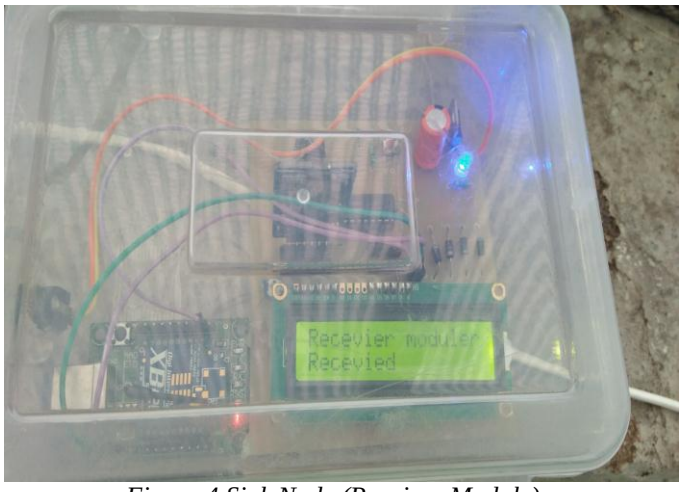

Figure 4 Sink Node (Receiver Module)

Figure 5 Parameter Observations at Polyhouse

For experimental purpose, dataset has been obtained from sink node (Base Station). The data consists of humidity, temperature measurements, Event for Temperature, event for Humidity with specific node ID. For experimentation only data related to temperature are used here.

C. Resource Optimization

As the sensors have many constraints like limited memory, computation power, energy etc. In Distributed Data mining approach for WSN, the main point of concentration is the optimization of these resources. But in this paper, as the concept of centralized approach of data mining for WSN is used where whole dataset is collected at a single site and then data mining is applied , the point of concentration is resource optimization for mining algorithms.

A resource is a source of materials or services to a system from which a benefit is gained. Any physical or virtual component of restricted amount of availability within a computer system is a system resource, or simply resource. There may be compressible or incompressible resources. Compressible resources are those to which the amount used by a task can be controlled without the need of killing it. Compressible resources can be throttled benignly. CPU cycle is an example of compressible resource [22]. The clock cycle decides the speed of a computer processor, or CPU. The amount of time between two pulses of an oscillator is the clock cycle. The speed of clock is measured in Hz, either megahertz (MHz) or gigahertz (GHz).

Intel (R) Core (TM) i3-4005U CPU @ 1.70GHz Processor is used for implementing both data mining algorithms in this paper. So, it may be said that, for a unit time if execution time is reducing for a task then CPU cycle will also be reduced. Therefore execution time can be measured for considering CPU cycle as resource. And point of concentration can be on an algorithm with minimum execution time.

D. Experimentation

Classification is a data mining technique that analyzes a set of data and generates a set of grouping rules which can be used to classify future data. The purpose is to apply the different classification algorithms for the dataset and then to choose best one for prediction purpose based on execution time i.e. the algorithm that will take less time to classify same data, will be best one between two algorithms for resource (CPU cycle) optimization. In this paper, two algorithms; SVM (Support Vector Machine) and KNN (K-Nearest Neighbour) are implemented in a mining tool. 80% data of this dataset has been used as training dataset and remaining 20% as test dataset.

IV. Results and Discussion

Following section shows the performance of above two algorithms for the dataset in terms of Execution Time, Confusion Matrix and Misclassification Rate.

A. SVM (Support Vector Machine)

a. Execution Time (to classify test dataset)

```
Time difference of 0.3532472 secs
```

b. Confusion Matrix

	Actual		
predicted	2 Fans Start	All Fans Start	All Fans Stop
2 Fans Start	170	190	1
All Fans Start	0	0	0
All Fans Stop	1	0	249

c. Misclassification Rate

```
[1] 0.314239
```

B. KNN (K- Nearest Neighbour)

a. Execution Time (to classify test dataset)

```
Time difference of 0.06201696 secs
>
```

b. Confusion Matrix

poly_test_target	2 Fans Start	All Fans Start	All Fans Stop
2 Fans Start	167	0	4
All Fans Start	61	129	0
All Fans Stop	0	0	250

c. Misclassification Rate


```
[1] 0.106383
>
```

In tabular form the above results for both classification techniques can be summarized as follows:

Table 1.Comparision of Classification Techniques for Specified Dataset

	Execution Time	Misclassified Data	Misclassification Rate
SVM	.3532472	192	.314239
KNN	.06201696	65	.106383

Now the graphical presentations of above results are as follows:

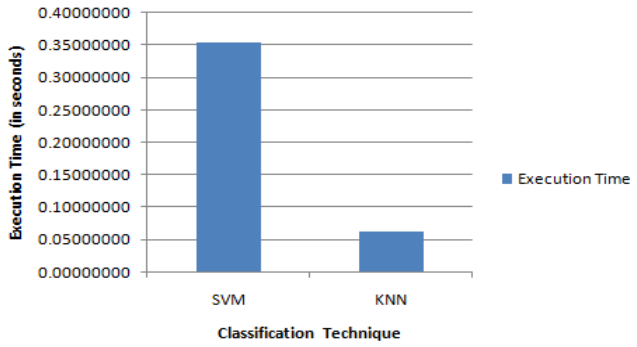

Figure 6 Execution Time for Classification Techniques

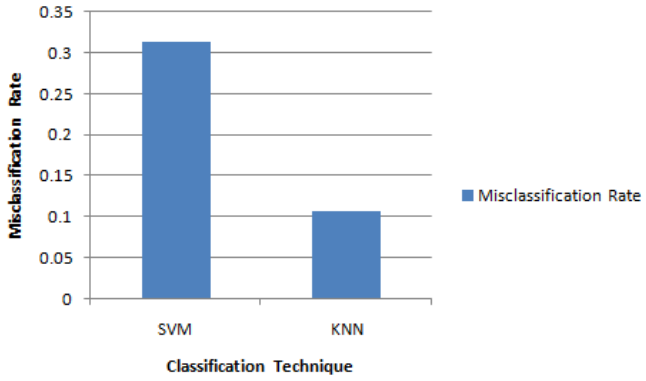

Figure 7Misclassification Rate for Classification Techniques

As CPU Cycle is a computational resource also. Each module to be executed can be characterized by the number of CPU cycles necessary to run the module itself [20]. So, it can be said that for given WSN dataset the algorithm, taking less execution time to classify test dataset, can be considered as better than another for resource optimization. Here, it is found that KNN is performing better than SVM.

V. CONCLUSION AND FUTURE SCOPE

From the above experimental results, table and graphs, it can be concluded that after applying centralized data mining approach of WSN on collected dataset of polyhouse KNN classification algorithm is performing better than SVM for both parameter Execution Time and Misclassification Rate. The results attained by experiments show that the KNN algorithm is more suitable for optimizing CPU Cycle as a Resource, or for Resource Optimization. For future work it is being tried to work on Hybrid Framework of WSN Data Mining by using both concepts of Distributed Data Mining Approach and Centralized Data Mining Approach for resource optimization. For this hybrid framework resource optimization may be applied for WSN and also for general concept like execution time for classification of data by using classification techniques.

ACKNOWLEDGMENT

Our sincere appreciation to the management of Mahavir Iron and Steel Industries (Plant Tissue Culture Division) and my Research Supervisor Prof. R. K. Dhuware who has directed to arrive out with this wonderful piece of research work.

REFERENCES

[1] S. H. Chauhdary, A. K. Bashir, S. C. Shah, M. S. Park, "EOATR: energy efficient object tracking by auto adjusting transmission range in wireless sensor network", Journal of Applied Sciences, vol. 9, Issue. 24, pp. 4247–4252, 2009.

[2] P. K. Biswas, S. Phoha, "Self-organizing sensor networks for integrated target surveillance", IEEE Transactions on Computers,vol. 55, Issue No. 8, pp. 1033–1047, 2006.

[3] J. Yick, B.Mukherjee, D. Ghosal, "Wireless sensor network survey", ComputerNetworks, vol. 52,no. 12,pp. 2292–2330, 2008.

[4] T. Arampatzis, J. Lygeros, S. Manesis, "A survey of applications of wireless sensors and wireless sensor networks", in Proceedings of the 20th IEEE International Symposium on Intelligent Control (ISIC '05), pp. 719–724, June 2005.

[5] A. Rozyyev, H. Hasbullah, F. Subhan,"Indoor child tracking in wireless sensor network using fuzzy logic technique", Research Journal of Information Technology, vol. 3, Issue. 2, pp. 81– 92, 2011.

[6] R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, D. Estrin, "Habitat monitoring with sensor networks", Communications of the ACM, vol. 47, no. 6, pp. 34–40, 2004.

[7] L. T. Lee, C. W. Chen, "Synchronizing sensor networks with pulse coupled and cluster based approaches", Information Technology Journal, vol. 7, Issue 5, pp. 737–745, 2008.

[8] N. Sabri, S. A. Aljunid, B. Ahmad, A. Yahya, R. Kamaruddin, andM. S. Salim, "Wireless sensor actor network based on fuzzy inference system for greenhouse climate control", Journal of Applied Sciences, vol. 11, Issue. 17, pp.3104–3116, 2011.

[9] D. Kumar, "Monitoring forest cover changes using remote sensing and GIS: a global prospective", Research Journal of Environmental Sciences, vol. 5, pp. 105–123, 2011.

[10] Y. C. Tseng, M. S. Pan, and Y.Y. Tsai,"Wireless sensor networks for emergency navigation", Computer, vol. 39, no. 7, pp. 55–62, 2006.

[11] J. Han, M. Kamber, J. Pei,"Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, USA, pp. 8, 2012.
[12] A. Mahmood, K. Shi, S. Khatoon, Mi Xiao, "Data Mining Techniques for Wireless Sensor Networks: A Survey", International Journal of Distributed Sensor Networks, Vol. 2013, 2013
[13] R. Sunny T, S. M. Thampi, "Survey on Distributed Data mining in P2P Networks", dblp computer science bibliography, 2012
[14] P. Singh, "Sensor Association Rules: a Survey", Vol. 9, Issue 9, pp. 67-71, 2014
[15] A. Boukerche, S. Samarah, "A New Representation Structure for Mining Association Rules from Wireless Sensor Networks", PARADISE University of Ottawa, 2007
[16] A. Mahmood, K. Shi, S Khatoon, "Mining Data Generated by Sensor Networks: A Survey", Information Technology Journal, Vol. 11, 2012
[17] A. K. Naik, R. Kumar Dwivedi, "A Review On Use Of Data Mining Methods In Wireless Sensor Network", International Journal Of Current Engineering And Scientific Research (IJCESR), Vol. 3, Issue. 12, 2016
[18] C. Sudha, A. Nagesh, "A Comprehensive Survey on Data Mining Techniques in Wireless Sensor Networks", International Journal of Computer Sciences and Engineering (IJCSE), Vol. 6, Issue. 6, pp. 1523-1527, 2018.
[19] M. Maksimović, V. Vujović, "Comparative Analysis Of Data Mining Techniques Applied To Wireless Sensor Network Data For Fire Detection", Journal of Information Technology and Applications, pp. 65-77, 2013
[19] B.-H. Park, H. Kargupta, "Distributed data mining: Algorithms, systems, and applications" Data Mining Handbook, 2002.
[20] S. Sardellitti G. Scutari S. Barbarossa "Joint optimization of radio and computational resources for multicell mobile-edge computing", IEEE Transactions On Signal And Information Processing Over Networks, Vol. 1, no. 2, pp. 89-103, 2015.
[21] B. A. Parbat, R. K. Dhuware, "Real Time Data Acquisition System for WSN Using Arduino for Polyhouse", International Journal of Computer Sciences and Engineering(IJCSE), Vol. 6, Issue. 8, pp. 608-612, 2018.
[22] L. He, Z. Qiang, W. Zhou, S. o,"A Review of Resource Scheduling in Large-Scale Server Cluster", International Conference on Knowledge Management in Organizations, pp. 494-505, 2016, ISBN 978-3-319-62698-7.

Authors Profile

Bhawana Amit Parbat pursued Bachelor of Science from Pt. Ravishankar Shukla University, Raipur, Chhattisgarh in 2003 and Master of Science also from Pt. Ravishankar Shukla University, Raipur, Chhattisgarh in 2007. She is currently pursuing Ph.D. from Dept. of Computer Science, R. T. M. Nagpur University, Nagpur, India and working as Technical Assistant in Hemchand Yadav University, Durg, Chhattisgarh, India.

R. K. Dhuware is Head of Department of Computer Science, D. B. Science College, Gondia, Maharashtra and Co-ordinator, Special Task Committee, B. Voc. Program, R. T. M Nagpur University, Nagpur, India.

