A REVIEW ON USE OF DATA MINING METHODS IN WIRELESS SENSOR NETWORK

Aditya Kumar Naik[1], Rajendra Kumar Dwivedi[2]
[1,2]Dept. of CSE, MMMUT

Abstract

These Wireless sensor network is one of the most emerging technology which is being used in computer science. WSN consists of small sensor nodes which can sense and communicate in a wireless fashion in a defined network and they get the data to sink. The sink is connected to gateway and collected data are sent to the base station. In current situation there are thousands of sensor nodes which in future it may be millions of sensor nodes which will work together depending on the situation. Sensors have many constraints like limited resource, memory, computation power etc. Data mining is used to evaluate data from several angles or dimensions and is the process of finding pattern in a large relational database. The purpose of using data mining is to increases the efficiency of the sensor nodes. We can have refined data at the sink by suppressing the unwanted data using data mining techniques. This can improve the energy efficiency or the battery power of the sensor nodes. This paper is focused on distinct techniques of data mining which have been adopted by wireless sensor networks showing their advantage and disadvantage with respect to the conditions

Index Terms: Data Mining Technique, Wireless Sensor Network, Sensors

I. INTRODUCTION

In Today's scenario wireless sensor network has become very essential in our daily uses, it is considered as one of the most important technology for the twenty-first century without WSN our work would have been very difficult. WSN can reach into those areas where human cannot even think to reach like disaster areas, space, sea etc. WSN has earned terrific attention from both industry and academic all over the world. Wireless sensor network consist of thousands of low cost nodes, low-power, and multifunctional senor that are deployed in a region of interested area which run on battery. The sensor nodes generally are small in size each node has one or more embedded microcontroller i.e. CPU or DSP chip set. These nodes have sensing capabilities along with communication capabilities. These nodes interact over a short area through a wireless medium which are organized into ad-hock network. Wireless sensor network has following advantage and disadvantage [16]:

- *Data Redundancy*: In the implicated region sensor nodes are arranged densely in order to fulfill the desired task. Interrelation is achieved by these sense nodes

- *No Universal Identification*: It is very difficult to maintain high overhead for the identification purpose and very difficult to build a global clock.

- *Frequent Topology Changes*: Generally the node topology or network changes very rapidly due to the failure of nodes or damage caused, channel fading and death of battery.

- *Unreliable Sensor Node*: In most cases the sensor which are arranged are unreliable as they are deployed in a very disturbing

environment and they are prone to failure or physical damage.

- *Denser Node Deployment*: Generally sensor node are arranged in a very off-key environment where humans cannot survive and mostly the sensor gets damaged or fail to function properly. Therefore this reason only the nodes are deployed in a very dense form that is thousands of nodes together.

WSN is a very powerful tool which helps to monitor, understand and manage processes [12]. WSN has many application areas which are as follows:

A. *Environment Monitoring*
In this sensor are deployed in the environment and their functions are to monitor the environment parameters and conditions in real time.

- *Disaster Monitoring*: Here the sensors are densely deployed in the affected area or in area which is prone to disaster so that we can monitor the area in term of loss of infrastructure and human lives.

- *Air Quality Management*: Here the sensors are arranged on ground, such that the quality of air can be monitored which will help us in controlling the air pollution.

B. *Military Application*
WSN can be rapidly be deployed in battle field or the hostile region which will help in monitoring the activities of our enemy or intruder. Sensor are deployed in battle field in order to monitor or to get real time activities of our army like number of tanks, vehicles, weapons, etc. Therefore the sensor provide the correct data and people can make decision in real time.

C. *Other Utilities*
Sensors also helps in monitoring the municipal water supply as sensor are deployed in each joint so that we can get a clear data whether the water is leaking or not. It also helps in monitoring the electricity grid and streetlights such that if any light is non-functional can be repaired.

D. *Health Analysis*:
Here the sensors are deployed in dense so that these wireless sensors can be used in monitoring of patient activities. These wireless sensors can be set in patient home and the activities of patient can be monitored, such that when a patient falls or is ill it can inform the doctor. It can also be helpful in medicine purpose by reminding the patient to take medicine at a fixed allotted time.

All the applications which are mentioned above are reliable and real-time monitoring is the essential requirements. All the application generate huge volume of data which are fast changing and heterogeneous in nature. All the data is collected and filtered into useful information or pattern by using data mining techniques. As in upcoming years the whole world will be ruled by wireless sensors, so it will be very crucial to choose a correct data mining technique which is a big challenge in WSNs.

The pattern of paper is as follows: section 2 explains about data mining technique. Section 3 explains about WSN in brief. Section 3 deals with the literature survey and section 4 describe the conclusion of the whole paper.

II. DATA MINING TECHNIQUES

Data mining is elucidated as heart of enlightenment discovery process. It can be briefed as the process of gathering facts from different area and refining it into appropriate knowledge. It can also be defined as the evocation of hidden predictive knowledge from a huge database. Technically, from a large relational databases a search is performed among several field areas to obtain a useful pattern which can be used in future. The motive of using data mining is to help companies to focus on the most important data in their data warehouses. The main idea of data mining technique is to extract data from large dataset and convert into some useful pattern for future use. Data mining is divided into two models, descriptive and predictive. All Data mining techniques fall under these two categories [13].

A. *Predictive model*
The primary goal of using this exemplary is that we can predict the future result than the current situation. It falls under the supervised learning and the predicted output can be numeric as well as in categorized form, as it always predicts the target value.

B. *Descriptive model*
This method is generally used to generate correlation, frequencies, cross tabulation etc. It is used to discover regularities in the data and uncover patterns. From bulk of data, a search is performed for finding interesting subgroup patterns.
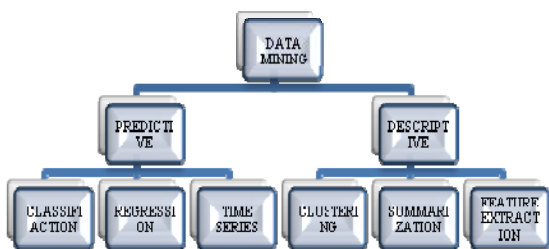
Figure-1 Data Mining Models

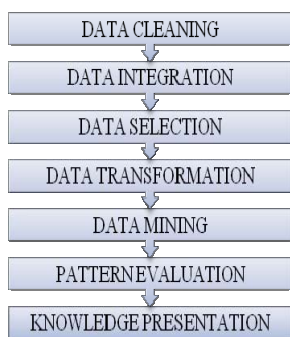The main steps are for knowledge discovery stage is as follows:

Figure-1 Stages for Knowledge Discovery

1. *Data Cleaning*: It is the first step in which inconsistent data and noises are removed.

2. *Data Integration*: It is the second step where the combination of multiple data source is done.

3. *Data Selection*: It is the third step where necessary data are searched from the database to reach the goal.

4. *Data Transformation*: It is also called as data consolidation. In this stage the selected e data are transformed into useful pattern for mining

5. *Data mining*: It is the most crucial stage in which the data mining techniques are applied on transformed data so that the tendency of the extracted pattern can be determined..

6. *Pattern Evaluation*: It is the sixth step in which only the important pattern which holding accurate information are chosen.

7. *Knowledge Presentation*: It is the final stage in which the mined data are provided to the user using virtualization and knowledge representation techniques.

Data mining consists of five major elements [15]:-

1. Extract, transform, and load transaction data onto the data warehouse system.

2. Store and manage the data in a multidimensional database system.

3. Provide data access to business analysts and information technology professionals.

4. Analyze the data by application software.

5. Present the data in a useful format, such as table.

There are four types of data mining techniques which are commonly used in WSN which uses both centralized and distribute approach is shown below:-
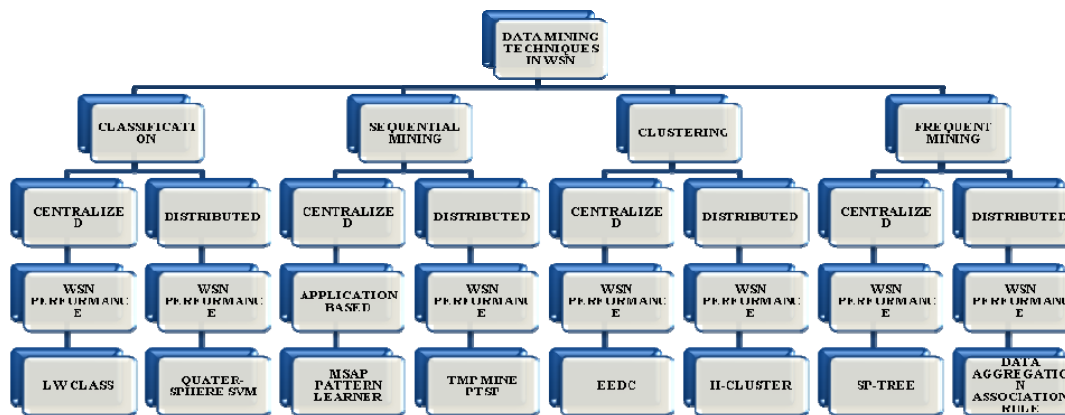

Figure-3 List of Data mining techniques in WSN

There are three essentials elements for data mining without these nothing can be achieved, they are as follows;

1. *Data*: are any facts, numbers, or text that can be processed by a computer and these growing amounts of data in different formats and different databases are stored. The data which defines other data is called met data such as design of logical database.

2. *Information*: Any data which provide information in form of pattern or relationship.

3. *Knowledge*: The patterns which are formed or gained are converted into knowledge for future use.

III. WIRELESS SENSOR NETWORKS

A wireless sensor network (WSN), it is a wireless network which dwell of partially disperse autonomous equipments that uses sensors to monitor environmental and real condition. A WSN is a collection of large sensor nodes which can be in hundreds or even thousands of small, cheap nodes which are deployed into a network at certain location. These sensors have the capability to sense, process and communicate to its peer in order to work together in a cooperative manner. A sensor network consist of thousands of battery which sense huge bulk of data and send it back to the sink through gateway for processing purpose by using different data mining techniques. Each node consists of processing capability in form of microcontrollers that is CPU or DSP chip sets. WSNs have different endowment like denser level of node deployment, sever energy, computation power, data generated by sensor nodes are bulk in size, fast changing and have very less resource constraints [14]. In figure shown below consist of thousands of sensor node which are arranged in a desired location, the jobs of senor nodes are to collect bulk of data and transfer it to number of gateway. They get the data to a sink and the sink is wired to a gateway over a network, all the data are transfer to the base station where useful data are sorted for future work. The nodes communicate with each other in a wireless manner which follows ad-hoc network. At present WSNs consist of thousands of nodes which in future may contain more than millions of nodes. In order to solve huge volume of data we need to solve all the major issues of the sensors like processing capability, limitation in memory, etc.
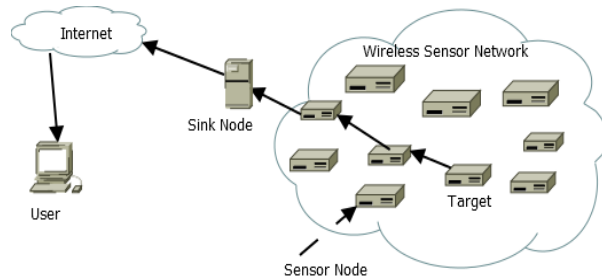

Figure-4 Wireless Sensor Network

Other issues these nodes are, in future we have to deal with heterogeneous data like sound, smell, taste, image, audio, video, etc. In this kind of data, significant amount of energy is consumed. There are basically two types of approaches in WSNs which are mentioned below:

Table 1 Difference between Distributed and Centralized Sensor

Application	Distributed	Centralized
Energy	Limited	No constraints
Data Length	Unbounded	Bounded
Response Time	Real time	Non- real time
Update Speed	High	Low
Number of Passes	Single	Multi-pass
Data Flow	Continuous	Stationary
Data Type	Dynamic	Static

A. *Centralized Approach*

Generally the data mining technique uses centralized approach. In this approach data are collected directly at a central location which is not protected or restricted by any computational resources. This approach involves the selection of cluster head by the base station itself. It is done on the basis of amount of residual energy and distance of sensor nodes from base station. In today's scenario the data which are generated are very fast developing and are huge in volume as the memory storage is limited. The main problem in this approach is, all the data collected are transferred to the central node and if the central node fails then the whole network will fail.

This approach is mostly inappropriate due to scalability issue like limited communication bandwidth, power supply and less storage space.

B. *Distributed Approach*
This approach is better than centralized approach as the data is not collected at a central node, if one node fails it will not have any effect on any other nodes. In this method the resources are need to be self organized. The reliability of senor nodes can be improved by the inducement of distributed architecture. This approach helps in saving energy up-to a great extent. This technique is used in order to control the issues like:-[15]

• Sensor nodes are prone to failure

• Better collection of data at sensor nodes

• To provide backup with nodes if failure occurs in the central node.

Table 2 Literature survey on Data Mining Techniques in WSNs

Sr.No	Name of Author	Year	Reference	Technique Used	Pros	Cons
1.	Boyer and Masdari	[2015]	[1]	Fuzzy C-Mean(FCM)	Life of the network is improved.	Dead nodes are reduced.
2.	Dhamdhere and .Shrikant	[2014]	[2]	Optimal Terminal Assignment Based Path.(OTABP)	Life of the network is improved and energy consumption is less.	Does not support dynamic sub-sink.
3.	Saana and Brahim ELbhiri	[2013]	[3]	Spectral Classification for Robust Clustering in Wireless sensor network	Lifetime of the network is improved and consumes less energy.	Other technique may be more efficient than the proposed method.
4.	S.Suresh and Nithyakalyni	[2012]	[4]	K-Mean Data Relay(KMDR) Clustering algorithm	Reduce power consumption and communication overhead.	Limited Bandwidth.
5.	Kushboo Sharma	[2011]	[5]	Nearest Neighbor Trajectory Classification (NNTC) algorithm.	High classification accuracy	Distance similarity problem with different framework.
6.	Sumathy and Ahamed	[2010]	[6]	Novel hybrid routing protocol	Reduced packet for transmission and increased the lifetime of the network.	Not suitable for all architecture.
7.	Ahmed and Lee	[2009]	[7]	Sensor Pattern Tree(SP-Tree)	Efficient FP-growth based and require less memory consumption	Communication efficiency of the system decreases.
8.	Chunyu Ali and Longjiang	[2009]	[8]	Distributed clustering algorithm H-cluster.	Lowest data loss rate and has best clustering quality.	Data loss increases for large scale.
9.	Zhen Yand and Tianjing Wang	[2008]	[9]	Location-Aware-Based Data Clustering (LABCD) algorithm	Reduce number of transmission cost	Works well with highly efficient data
10.	Tahekordi and Eliassen	[2008]	[10]	K-Mean Clustering algorithm	Reduce number of transmission message & power consumption	Extra money required for cluster head formation.
11.	S.Samarah and A.Boukerche	[2007]	[11]	Distributed Data Extraction Methodology.	Reduce the data size to half.	Cost for the buffer node increases.

In [1] author Boyer and masdari in proposed the Fuzzy C Mean (FCM) algorithm. The author has used the combo of FCM and LEACH protocol to reduce the energy use in the wireless sensor network by introducing hybrid protocol in which the sensor node uses optimized energy in contrast to the LEACH energy method and it helps in increasing the lifetime of the network, although the dead nodes are reduced in contrast to LEACH.

In [2] author .Shrikant & Dharmdhere in proposed the Optimal Terminal Assignment Based Path (OTABP). This method is focused on conveniently assignment of the section nodes to the sub-sinks. It helps in increasing the data lot and it reduces the energy utilization as well as amend the stint of sensor nodes to the sub-sinks. It also increases the network lifetime. The main issue is that it does not work well with dynamic sub-sink.

In [3] author Sanna and Brahim Elbhiri proposed a new technique Spectral Classification for Robust Clustering in Wireless Sensor Networks (SCRC-WSN). This protocol works on centralized approach, it uses graph theory technique with a goal to separate the network in a fixed matchless number of clusters. The simulation result spectacle that the proposed algorithm has shown more energy efficiency distribution compared to the LEACH and lifetime of the network is improved. The main issue is that other spectral technique may be more efficient than this proposed technique.

In [4] author S.Suresh and Nithyakalyani proposed a new clustering algorithm that is K-Mean Data Relay (KMDR). This method helped in reducing the power consumption, it also reduces the communication overhead and increased the lifetime of the network. All these things were achieved by grouping the sensor nodes together, their by cutting the count of nodes which are broadcasting data to the sink node. The main issues are the constraint in power supply, storage resources, bandwidth and limited communication.

In [5] author kushboo Sharma in 2011 proposed the NNTC (Nearest Neighbor Trajectory Classification) algorithm. This algorithm is a very lazy learner as it does not process the pattern during training period, as the training phase is trivial. This method has a very high classification accuracy and the classification is done in a very efficient manner. The main issue is that it has a distance similarity and integration problem with different framework.

In [6] author V.Sumathy et.al proposed novel compound routing technique for data gathering in agriculture application. It is based on Cluster Head routing and Rapid Spanning Tree. This method has reduced the power consumption with heterogeneous data. It also reduces the packets required by the network convergence time and the lifetime of the network is increased. The main issue is that this method may not be suitable for different network architecture.

In [7] author Ahmed and Lee proposed a new tree-based data structure called Sensor Pattern Tree (SP-Tree). This technique is used for mining association rule for WSN data. The SP-Tress can be put together by a single scan over the sensor epochs and also enables an profitable FP-gain-based mining method. The main issue is that for high numbers of sensors the communication efficiency of the system decreases.

In [8] author Chunyu Ali and Longjiang Guo proposed a new method distributed clustering H-Cluster algorithm. It follows distributed approach for clustering. It insures that the elucidation of data clustering mostly taking place in a near-by geographical area. This algorithm catches a set of sensor data as input. The sensor data are generally collected from the sensor nodes from start time to the current time. This technique has achieved the best clustering quality, the highest energy efficiency and lowest data loss rate. The main issue is that the amount of data loss increases for large scale wireless sensor network.

In [9] author Zhen yang and Tianjing Wang proposed a (LABDC) algorithm to analyze the structural correlation of sensor data. It is a lossy mechanism which reduces the number of transmission and the transmissions cost and give accurate query result to the users. This algorithm exploits spatial and non-spatial characteristics of sensor data in wireless sensor network. The main issue is it works well only with highly efficient sensor data.

In [10] author Tahekordi and Eliassen proposed A distributed version of K-Mean clustering algorithm and collected data towards the sink. It

reduced the number of communication transmission and as a result the time and power consumption of sensor nodes is minimized. It also maintains the clustering quality by iteration. The main issue is that it has a serious chain reaction on communication efficiency when the numbers of sensors are very large and requires extra money for cluster head formation.

In [11] author S.Samarah and A.Boukerche proposed Distributed data extraction methodology. It is a eminent data mining method which is agile to achieve a time relation amid sensor device in a define network. This method is used to solve the association rules mining problem that allow traditional mining algorithm. It is able to decrease the number of exchanged message during the transmission and data size by half. The main issue is that the cost for node for buffer increases and also delay increases for some critical messages which result in loss of time.

IV. CONCLUSION

In this paper, we have explained different classification of data mining methods which have been espoused for Wireless sensor networks which uses different classification and evaluation approaches. We inspected that each data mining technique mentioned here works on either centralized or distributed approach are discussed however the main consideration of the approach was on information extraction and analysis on data with respect to WSNs. We have analyzed, detailed and briefly elaborated both advantage and disadvantage of each algorithm. All of the algorithms which have been discussed above are capable of solving different types of issues related to WSNs types and applications. We have observed that mining technique developed is helpful for taking real-times decision. However it has few issues focusing on the limitation of WSNs like limited communication bandwidth, limitation in memory capacity, processing power, CPU capacity and storage. Therefore it is very necessary to upgrade the performance of upcoming WSNs sensors using an appropriate data mining techniques in both centralized and distributed approaches.

We postulate that WSNs application will become very famous if the sensor technology improves and the sensor data would become more knowledge rich. The mining methods would become very significant for conducting advanced analysis like discovering patterns and determining tendency thus improving WSNs operation and performance. The major challenge will be related to hardware like compressing storing and filtering of huge volume of data generated by sensor nodes. Also, in future we have to deal with heterogeneous kinds of data like image, sound, smell, location and etc. The main motive to present this paper is to summarize, analyze and extract knowledge from data into useful ideas for better decision making.

REFERENCES

[1] Bouyer, A., Hatamlou, A. and Masdari, M.,. "A new approach for decreasing energy in wireless sensor networks with hybrid LEACH protocol and fuzzy C-means algorithm". International Journal of Communication Networks and Distributed Systems, 14(4), 2015 pp.400-412.

[2] .Shrikant D. Dhamdhere , Shanthi K.Guru "Robust Data Collection in Wireless Sensor Networks with Mobile Sinks" International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, pp.4999-5002

[3] B. Elbhiri, S. E. Fkihi, R. Saadane, N. Lasaad, A. Jorio and D. Aboutajdine, "A new spectral classification for robust clustering in wireless sensor networks", Wireless and Mobile Networking Conference (WMNC), 2013 6th Joint IFIP, Dubai, 2013, pp. 1-10

[4] Nithyakalyani, S; Kumar, S Suresh "Data Relay Clustering Algorithm for Wireless Sensor Networks: A Data Mining Approach" Journal of Computer Science Vol .8, 2012, pp.1281-1284.

[5] Sharma, K., Rajpoot, M. and Sharma, L.K., "Nearest Neighbour Classification For Wireless Sensor Network Data". International Journal of Computer Trends and Technology, Vol 2(2),2011, pp.41-43.

[6] N. P. Karthickraja, V. Sumathy and M. N. Jabeer Ahamed, "A novel hybrid routing protocol for data aggregation in agricultural applications," Communication Control

and Computing Technologies (ICCCCT), 2010 IEEE International Conference on, Ramanathapuram, 2010, pp. 227-231.

[7] S. K. Tanbeer, C. F. Ahmed, B. S. Jeong and Y. K. Lee, "Efficient mining of association rules from Wireless Sensor Networks," Advanced Communication Technology, 2009. ICACT 2009. 11th International Conference on, Phoenix Park, 2009, pp. 719-724.

[8] Guo, L., Ai, C., Wang, X., Cai, Z. and Li, Y, "Real time clustering of sensory data in wireless sensor networks". 2009, December In IPCCC (pp. 33-40).

[9] Tianjing Wang and Zhen Yang, "A Location-Aware-Based Data Clustering algorithm in Wireless Sensor Networks," Communication Systems, 2008. ICCS 2008. 11th IEEE Singapore International Conference on, Guangzhou, 2008, pp. 1-5.

[10] A. Taherkordi, R. Mohammadi and F. Eliassen, "A Communication-Efficient Distributed Clustering Algorithm for Sensor Networks," Advanced Information Networking and Applications - Workshops, 2008. AINAW 2008. 22nd International Conference on, Okinawa, 2008, pp. 634-638.

[11] A.Boukerche and S.Samarah, "An Efficient Data Extraction Mechanism for Mining Association Rules from Wireless Sensor Networks" ICC, 2007,pp.3936-3941

[12] Zheng, J. and Jamalipour, A., "Wireless sensor networks: a networking perspective". John Wiley & Sons,2009

[13] Pradnya P. Sondwale, "Overview of Predictive and Descriptive Data Mining Techniques",IJARCSSE,Vol-5,2015,pp. 262-265

[14] I.F. Akyildiz, Weilian Su, Y. Sankarasubramaniam and E. Cayirci, "A survey on sensor networks," in IEEE Communications Magazine vol. 40, no. 8, pp. 102-114, Aug 2002.

[15] A. R. Raut and S. P. Khandait, "Review on data mining techniques in wireless sensor networks," Electronics and Communication Systems (ICECS), 2015 2nd International Conference on, Coimbatore, 2015, pp. 390-394.

[16] Jun Zheng. "Introduction to Wireless Sensor Networks", Wireless Sensor Networks.