A Comprehensive Survey on Data Mining Techniques in Wireless Sensor Networks

C. Sudha[1*], A. Nagesh[2]

[1]Dept. of CSE, Mahatma Gandhi Institute of Technology, Hyderabad, India
[2]Dept.of CSE, Mahatma Gandhi Institute of Technology, Hyderabad, India

Abstract- Wireless sensor networks (WSN) have emerged as one of the most exciting fields in Computer Science research nowadays. A WSN is a collection of sensors that are incorporated with a physical domain. These sensors are little in size, and equipped for detecting physical wonder and handling them. The most vital reason for conveying the WSNs-built up applications is to make the ongoing determination which has been turned out to be extremely testing due to the absolute asset restricted processing, imparting limits, and the giant amount of speedy changed information created by WSNs. This motivates to investigate a novel and fitting data mining procedure equipped for extricating learning from enormous volume and an assortment of persistently arriving data from WSNs. In this paper diverse existing data mining strategies received for WSNs are inspected with various grouping, assessment approaches. Based on the barriers of the existing process, an adaptive data mining structure of WSNs for future research are proposed.

Keywords – Wireless sensor network,Data mining.Sensor

I.   INTRODUCTION

A Wireless Sensor Network (WSN) is a wireless network which comprises of a large number of small sensor devices that are densely deployed and useful in different areas like sensing humidity [1], pollution, monitoring bridge, object tracking, data analysis etc. In most of the application, efficient and real time monitoring are essential requirements. In a few applications, this huge amount of information processing needs to be carried out at very high speed. This, subsequently, has developed the need for new efficient systems, stage, and necessitated instruments to manage this enormous some sensory information which is typically organized or unstructured. The growth of wireless sensor networks was initially motivated by military applications such combat zone reconnaissance, national security, medical services, home machines ecological checking, and numerous fields, where human mediation is greatly unsafe. Essentially WSNs have a few attributes which incorporate sensor hub organization, restricted battery power and memory, single or various base station (BS), hub dynamicity, no worldwide one of a kind ID, application mindfulness, ardent sensor nodes, reliable topology change, single-jump transmission or multisource transmission, information repetition, data redundancy etc.. With the recent technological advancement, the sensor nodes are getting smaller, but wireless networks are getting larger, currently containing thousands, but in

future it will be of millions of nodes. Dealing with the huge volume of information produced by these sensor networks creates a very serious challenge, one logically approached through the use of Data Mining techniques. Furthermore, as the communication consumes a significant amount of energy in a WSN, minimizing it is another major challenge. Other challenges in WSNs are design and development issues, efficient data collection at cluster head and provide backup if a failure occurs in the cluster head [2].
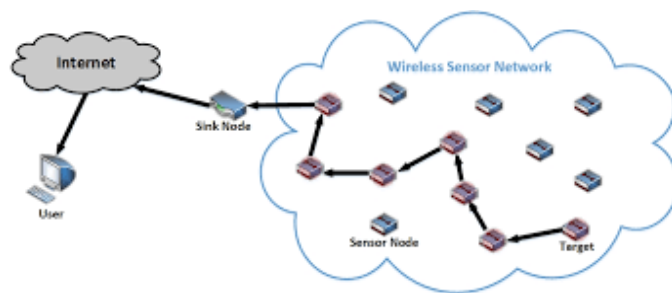

Fig 1. WSNs Architecture

This survey paper is prepared as follows. After the introduction in section I, Process of Data in WSNs, the distinction between traditional data and WSNs data mining process and Data mining challenges in WSNs is discussed in

section II. In section III, literature survey in this section analyzed a set of published research. In section IV, Limitations Of Existing Data Mining Techniques For Wsns finally, the paper ends with the conclusion in section

II. DATA MINING PROCESS FOR WSNs

The major four processes of data mining techniques used for WSNs are frequent mining, sequential mining, clustering and classification [3]. All these techniques can be used in centralized as well as distributed environment in wireless sensor networks. A general survey of data mining techniques is found in [4]. In this research paper. Lots of surveys are available on each of the above techniques discussed in [4].For example, in [5] frequent pattern mining over data stream is proposed. The clustering techniques examined in these papers totally focus on architecture and management of the network and not in order discovery from arriving data. In [7], the author examined A New Spectral Classification for Robust Clustering in Wireless Sensor Networks. Thus, selected data mining algorithms found in Wireless Sensor Networks are discussed here. Now in the literature survey section we will highlight some of the existing solutions with special highlighting on their techniques used in order to develop the concert of wireless sensor networks.

Table 1.Distinction between traditional data and WSNs data mining process

	WSNs Data	Traditional Data
Data Type	Dynamic	Static
Processing Architecture	Distributed	Centralized
Memory Usage	Restricted	Unlimited
computational power	Weak	High
Data Flow	Continuous	Stationary
Data Length	Unbounded	Bounded
Energy	Limited	No constraints

It might be found from Table 1 that Traditional data mining is unified, computationally expensive, and eager about transactional data. It promptly gathers information on the imperative site which isn't limited by the method for computational resources. In evaluation with data sets, the WSNs information streams dependably in programs with different refresh rates. Because of tremendous amount and high stockpiling rate, it's impossible to retailer the entire WSNs data or to look over it in excess of one example. These attributes of sensor data and the correct outline issues of sensor systems make basic information mining approaches testing. As an outcome, it's important to improve data mining processes that can break down and strategy WSNs data in multidimensional, multilevel, single-pass, and on-line way

A.Challenges
i)Resource Restriction:
The sensor nodes are resources in phrases of power, memory, communique bandwidth, and computational energy. The primary challenge faced by data mining strategies for WSNs is to assure the mining accuracy, condition whilst retaining the useful resource consumption of WSNs to a minimal.

ii)Rapid and Enormous Arrival of Data:
The characteristic idea of WSNs data is its over the high speed. In numerous spaces, learning arrives faster than we are skillful to mine. Moreover, spatiotemporal implanting of sensor information plays a fundamental position in WSNs application. This may at times reason numerous established information preparing strategies to perform inadequately on spatiotemporal sensor knowledge. The task for data mining procedures is the means by which to adapt to the enduring, fast, and adjusting learning streams and moreover discover how to contain client connection throughout rapid data arrival.

iii) Web based Mining:
In WSNs, condition learning is topographically dispensed, inputs arrive reliably, and later information devices could substitute the result focused on more established data observably. The majority of data mining Techniques that break down learning in a disconnected strategy don't meet the necessity of dealing with dispersed circle data. Subsequently, a wander for information mining strategies is the correct method to framework apportioned spilling information on the web.

iv) Displaying changes of mining Results after some time:
At the point when the information creating wonder is changing after some time, the separated model whenever ought to be cutting-edge. On account of the congruity of data streams, a few specialists have expressed that taking pictures the other of mining comes about is more essential regarding this matter than the mining result. The examination inconvenience is the best move toward to model this adjustment in the result

v)Data Revolution:
Considering the way that sensor nodes are restricted as far as bandwidth, rebuilding formed information over the system simply isn't practical. Favorable position structure change is an important inconvenience. Subsequent to extricating mannequin and examples locally from WSNs data, the yield is exchanged to the base station. The assignment for information mining framework is the most ideal approach to effectively speak to data and scholarly examples over system for transmission.

vi)Energetic Network Topology:

Sensor organizes conveyed in, most likely cruel, uncertain, heterogeneous, and dynamic situations. Besides, sensor hubs may simply exchange among unmistakable areas at any factor after some time. Such dynamicity and heterogeneity build up the unpredictability of outlining a fitting data mining process for WSNs.

III. DATA MINING TECHNIQUES FOR WSNs

A. Frequent Pattern Mining

Frequent pattern mining is utilized to discover the establishment of factors that co-emerge every now and again inside the dataset. The aim is to locate the most extreme energizing relations among factors. Traditional common pattern mining algorithms are the CPU and the I/O in depth, making it extremely rich to mine dynamic nature of WSN measurements. Not at all like the mining static database, dynamic nature of WSNs measurements caused the inspect of on-line mining of frequent item set. As a final product, traditional frequent mining algorithms are changed in venture with the nature of WSNs data [8]. The basic frequent pattern mining approach is an affiliation rule mining strategy. The main perceived affiliation rule mining set of rules is Apriori. It is constructed absolutely in light of the degree-sensible applicant age and check philosophy by method for making various outputs over databases.

B. Sequential Pattern Mining

Frequent pattern mining has been drawn out to find more mind boggling structure comprehensive of sequential pattern mining. It decides frequent subsequences as patterns in a series database. A series database stores a measure of records, where in all data are successions of organized occasions, with or without actual notions of time. A substantial number of actual global space names comprising of consumer profiling, medication, adjacent climate estimate, and Bioinformatics show a characteristic inclination to be demonstrated by the method for arrangements of occasions/objects related with each extraordinary. This enormous sort of uses of sequential pattern mining raise this hell one of the vital subjects in WSNs facts mining as demonstrated by the research efforts created in the most

recent years. The sequential simple mining systems in sensor group fundamentally in view of both traditional sequential pattern mining alongside Apriori-like algorithm [9], Apriori-primarily based methods: GSP PSP [10], and test blast strategies: FreeSpan and PrefixSpan [11] or a couple of new arrangement of guidelines are contrived specifically to artistic creations with the sensor group environment.

C. Clustering

Clustering is unsupervised learning, were given information is classified into subsets with the goal that every subset speaks to a group which has particular properties [12]. It has been estimated a valuable method particularly for applications that expect adaptability to the substantial number of sensor nodes. Clustering additionally wires aggregation of information with a specific end goal to survey the general transmitted data. In the present written works, issues identified with bunching are tended to by node grouping or data clustering. As of late, gigantic realities of node grouping calculations have been intended for WSNs. These bunching procedures extensively change in their destinations relying upon the hub misuse and bootstrapping plans, the sought after network architecture, the cluster head of the network operation. In spite of the fact that node might be identified with data clustering, for instance, considering information, likeness of the neighboring node, numerous well known node clustering algorithm that parcel the sensor nodes into various little groups and choose a group set out toward each gathering don't utilize the data mining procedures straightforwardly. In this examination, we just spotlight on data clustering procedures for predictive data mining and discover data relationships between the nodes.

D.Classification

Classification-based approaches [13] have adapted the traditional classification techniques such as decision tree-based, rule-based, nearest neighbor-based, and support vector machines-based techniques based on the type of the classification model that they used.

Table 2. Study On Data Mining Techniques In WSNs

S.NO	Name of Author & Year of Publication	Techniques used	DM Method	Node Task	Application Area	Limitation
1.	Halatchev and Gruenwald [14]. [2005]	Data Stream Association Rule Mining	Apriori like	Sense and Send	Traffic Monitoring	Ignore the sensor that reports different values

2.	Jiang, N.[15] [2007]	CRAM	FP growth based	Sense and send	Data Analysis	Inefficient for handling high speed data.
3.	Esposito, F., Basile, T. M. A., Di Mauro[16] [2010]	Relational Framework	Apriori Like	Sense and send	Environmental monitoring	Memory and time consuming
4	Khawaja, F.MavHome[17] [2003]	Episode Discovery	Generalized sequential pattern(GSP)	Sense and send	Inhabitants behavior prediction	Inefficient for complex activities
5.	Yeo, M. H., Lee, M. S., Lee, S. J., Yoo, J. S[18] [2008]	DCC	Data clustering	Data suppression	Generic WSNs Application	High Clustering rate
6.	Taherkordi, A., Mohammadi, R., Eliassen[12] [2008]	Clustering Sensory data	K-Means	Data Summarization	Data Analysis	In efficient For Large WSNs
7	Chikhaoui, B., Wang, S., Pigot, H.A[19] [2010]	Person identification algorithm	Decision Tree	Sense and Send	Healthcare	Does not guarantee the correctness
8.	Pigot, H.A[20][2008]	Fuzzy predictor model	Fuzzy rules	Sense and send	Health care	In efficient for complex scenariors

IV. LIMITATIONS

Tables 2 display the traits of data mining strategies designed for WSNs. It is found from the relative evaluation that the existing techniques have the subsequent shortcomings

i) Most of the techniques do now not keep in mind the heterogeneous data and anticipate that the sensor information is homogenous [14]. In a few times, homogenous information cannot make a contribution efficiently in the direction of real-time preference.

ii) The greater part of the procedures just consider the spatial, or transient or spatiotemporal connections among sensor data of adjacent nodes and does never again bear in mind the property reliance among sensor nodes. This in the flip will build the computational unpredictability and declines the precision of mining approach.

iii) The dominant part of system utilizes centralized approach in which all data are transmitted to the sink node for discovering patterns. These procedures cause a mess verbal trade overhead and defer the response time. While the methodologies that utilized dispensed structure, streamline response time and power intake, they have the indistinguishable issue as that of the brought together

approach if the aggregator/bunch head has a huge assortment of hubs underneath its enrollment.

iv) The systems assessed using systematic displaying utilized certain rearrangements and presumption to evaluate the general execution of the proposed strategy. Such suppositions and disentanglements may moreover cause dubious outcomes with controlled confidence. None of the proposed approach is assessed with the guide of the utilization of genome organization. Albeit genuine arrangement is mind boggling, expensive, and tedious, rectify results can handiest be gained by means of the use of genuine organization.

v) Some strategies used the artificial data. Although artificial data are effortless to be had, there always been probabilities that consequences generated on artificial data are not correct.

V. CONCLUSION

The rising want to use for the data mining Technique, within the field of WSNs resulted within the improvement of several algorithms. Every this kind of algorithms rectifies positive issues identified with the right WSNs kind and readiness. In this survey paper, we examined, talked about, and thought about the associated existing study processes. We establish that the strategy designed for mining sensor data in the network area is gainful for taking constant decision and

capacity essential for development of capable component for data storage, recovery, retrieval, and Transaction process in imperative perspective. In addition, we've provided a normal investigation and audit of the past examinations and their limits that could offer bits of knowledge for end users in making the utilization of or building up a proper records mining technique and fitting innovation for WSNs. We trust that WSNs applications will turn out to be additional develop and famous with the headway of sensor innovation, and sensor data will turn out to be more measurements rich. Mining procedures will then be extremely huge, so that you can behavior advanced analysis, which includes figuring out trends and locating interesting patterns for this reason improving WSNs execution and activity. The reason to give this paper is to invigorate distractions in using and forming the past research into rising applications.

VI REFERENCES

[1] Azhar Mahmood, Ke Shi, Shaheen Khatoon, Mi Xiao "Data Mining Techniques for Wireless Sensor Networks: A Survey" International Journal of Distributed Sensor Networks July 2013.

[2] Szewczyk, R., Osterweil, E., Polastre, J., Hamilton, M., Mainwaring, A., Estrin, D.Habitat monitoring with sensor networksCommunications of the ACM200447634402-s2.0-424311408710.1145/990680.990704 Google Scholar, Crossref, ISI

[3]S.Stankovic, O.Rakocevic, N. Kojic, D.Milicev, "A Classification and Comparison of Data Mining Algorithms for Wireless Sensor Networks", lCIT2012,978-1-4673-0342-2112 IEEE.

[4]V. Maojo and J. Sanandré, "A survey of data mining techniques," Medical Data Analysis, Lecture Notes in Computer Science, vol. 1933, pp. 17–22, 2000.

[5]J. Cheng, Y. Ke, and W. Ng, "A survey on algorithms for mining frequent itemsets over data streams," Knowledge and Information Systems, vol. 16, no. 1, pp. 1–27, 2008.

[6]Emad M. Abdelmoghith, "A Data Mining Approach to Energy Efficiency in Wireless Sensor Networks", 2013 IEEE 24th International Symposium on Personal, Indoor and Mobile Radio Communications: Mobile and Wireless Networks, 978-1-4577-1348-4/13, 2013 IEEE.

[7]Brahim Elbhiri , Sanaa El Fkihi, "A New Spectral Classification for Robust Clustering in Wireless Sensor Networks", WMNC'2013, 978-1-4673-5616-9/13, 2013 IEEE.

[8] Agrawal, R., Srikant, R.Fast algorithms for mining association rulesProceedings of the 20th International Conference Very Large Data Bases (VLDB '94)1994Citeseer487499 Google Scholar

[9] Agrawal, R., Srikant, R.Mining sequential patternsProceedings of the IEEE 11th International Conference on Data EngineeringMarch 19953142-s2.0-0029212693 Google Scholar

[10]Srikant, R., Agrawal, R.Mining sequential patterns: generalizations and performance improvementsProceedings of the Advances in Database Technology (EDBT '96)1996117 Google Scholar

[11]Masseglia, F., Cathala, F., Poncelet, P.The PSP approach for mining sequential patternsPrinciples of Data Mining and Knowledge Discovery1998176184 Google Scholar, Crossref

[12] Taherkordi, A., Mohammadi, R., Eliassen, F.A communication-efficient distributed clustering algorithm for sensor networksProceedings of the 22nd International Conference on Advanced Information Networking and Applications Workshops/Symposia (AINA '08)March 20086346382-s2.0-4912455110.1109/WAINA.2008.130 Google Scholar, Crossref

[13] Sharma, L. K., Vyas, O. P., Schieder, S.Nearest neighbour classification for trajectory dataInformation and Communication Technologies2010101180185 Google Scholar, Crossref

[14] Halatchev, M., Gruenwald, L.Estimating missing values in related sensor data streamsProceedings of the 11th International Conference on Management of Data (COMAD '05)2005 Google Scholar

[15] Jiang, N.Discovering association rules in data streams based on closed pattern miningProceedings of the SIGMOD Workshop on Innovative Database Research2007 Google Scholar

[16] Esposito, F., Basile, T. M. A., Di Mauro, N., Ferilli, S.A relational approach to sensor network data miningInformation Retrieval and Mining in Distributed Environments2010163181 Google Scholar, Crossref

[17] Khawaja, F.MavHome: an agent-based smart homeProceedings of the 1st IEEE International Conference on Pervasive Computing and Communications (PerCom '03)March 20035215242-s2.0-33746769349 Google Scholar

[18] Yeo, M. H., Lee, M. S., Lee, S. J., Yoo, J. S.Data correlation-based clustering in sensor networksProceedings of the International Symposium on Computer Science and its Applications (CSA '08)October 20083323372-s2.0-5664909764310.1109/CSA.2008.21 Google Scholar, Crossref

[19] Chikhaoui, B., Wang, S., Pigot, H.A new algorithm based on sequential pattern mining for person identification in ubiquitous environmentsProceedings of the 4th International Workshop on Knowledge Discovery form Sensor Data (ACM SensorKDD '10)2010Washington, DC, USA2028 Google Scholar

[20] Chikhaoui, B., Wang, S., Pigot, H.A A fuzzy predictor model for the occupancy prediction of an intelligent inhabited environmentProceedings of the IEEE International Conference on Fuzzy Systems (FUZZ '08)June 20089399462-s2.0-5524908923410.1109/FUZZY.2008.4630482 Google Scholar, Crossref