ISSN 2047-3338

An Improved Data Mining Technique and Online Mining for Energy Efficiency in Wireless Sensor Networks: A Comparative Review

Mohd Muntjir[1], Ahmad Tasnim Siddiqui[2] and Mohd Rahul[3]
[1,2,3]College of Computers and Information Technology, Taif University Taif, Kingdom of Saudi Arabia
[1]muntjir.m@gmail.com, [2]tasnim5@yahoo.com, [3]mohd.rahul@gmail.com

***Abstract*—Data mining in large sensor networks just starts to attract interest. Finding patterns in such an environment is both compelling and challenging. We also outline our preliminary proposal of solutions to these problems. There has recently been a considerable amount of research work done on using data compression techniques to minimize the volume of the transmitted traffic, and the consequently assist in reducing power consumption levels in Wireless Sensor Networks. A Wireless Sensor Network is the network of small sensor nodes, which are energy constraint devices and have limited data computational and transmission control. The nodes are organized in the form of cluster around a Base Station. In this paper we have define how to improve the energy efficiency of the network by using data mining technique by discovery, usage, and understanding of patterns and knowledge, mining moving object data, mining text, Web, and other unstructured data.**

***Index Terms*– Wireless Sensor Network, Data Mining, Clustering and Energy Efficiency**

I. INTRODUCTION

RECENT technology advances have qualified the development of small, battery-powered, wireless sensor node. These little sensor nodes, operational with sensing, calculation, and communication competences, can be organized in large numbers in wide geographical areas to display, detect and report time-critical occurrences. Subsequently, wireless networks containing of such sensors create exciting prospects for extensive, data-intensive measurement and investigation presentations. In many of these purposes, it is necessary to mine the sensor readings for patterns in real time in order to make intelligent decisions straightway [1].

In this paper, we study the challenges, problems, and possible solutions of online mining and data mining for sensor networks. Research on data mining has been productive; conversely, online mining for sensor networks faces numerous new experiments. Main, sensors have thoughtful resource constraints including battery lifetime, communication bandwidth, and CPU capability and storage [2]. Another, sensor node flexibility increases the intricacy of sensor data because a sensor may be in a dissimilar neighborhood at any point of period [7], [19]. Third, the sensor data come in time-ordered streams over systems. These experiments make conventional mining techniques unsuitable, because conventionally mining is consolidated, computationally exclusive, and intensive on disk-resident transactional statistics. In response to these experiments, we suggest to develop mining techniques that are unambiguously geared for sensor network surroundings [5]. This sensor node takes the information from the atmosphere, practices the data and transmits the data throughout radio signal. Typically, these nodes have inadequate battery life due to some limitations such as limited battery life, very low memory, low data rates, low bandwidth dispensation, variable link quality and little computing competence [6]. Despite these restrictions, the sensors were employed in the large numbers, they postulate with a very real picture of the field being perceived [20].

WSN is a collection of minor, frivolous sensor nodes organized in large numbers to control the ambient circumstances. Collection of sensor nodes is screamed cluster.

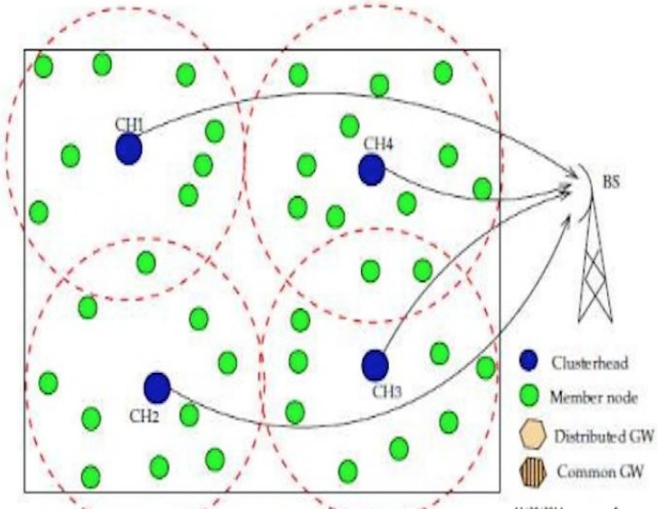

Fig. 1. Data communication in clustered network

Each cluster contains of one or more cluster heads. Since their cluster members the cluster head folds all the statistics. The composed information is routed to the Base Station. The Base Station is a immovable node, which is qualified to receive and transmit the data within the entire system. The number of cluster head selection differs on the number of sensor nodes. Choosing more than one cluster head enthusiastically controls energy depletion.

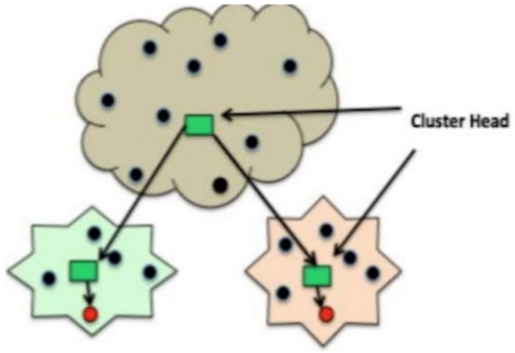

Fig. 2. Cluster Head

A. Clustering in WSN

The categorized network structure authorize grouping of sensor nodes into cluster and to allocate a particular task to the sensor in the cluster, formerly moving communication to the higher levels. WSN clustering technique helps to accomplish high-energy effectiveness and convince about the long network period. In the categorized manner each cluster has a cluster leader, that supervised the entire cluster node known as cluster head and numerous common sensor nodes are the members of clusters. The CH nodes conduct the data to the sink either straight or via intermediate interaction with other cluster head node. Hence, In order to balance energy ingesting amongst all network nodes, nearby is sequence of cluster head in WSN [10].

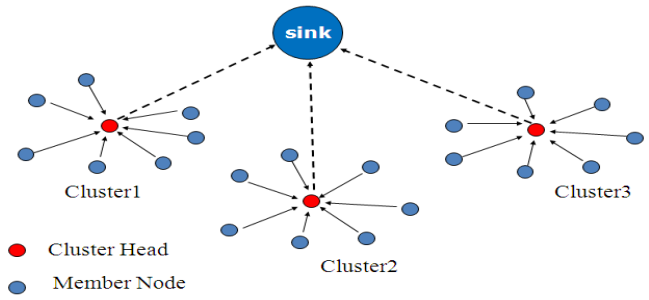

Fig. 4. Sink

B. Cluster Support Sensor Networks

Clustering sensors into the group is more adaptive and well-organized methodology is adopted for routing and for node communication by many conclusions on sensors energy contemplation [25].

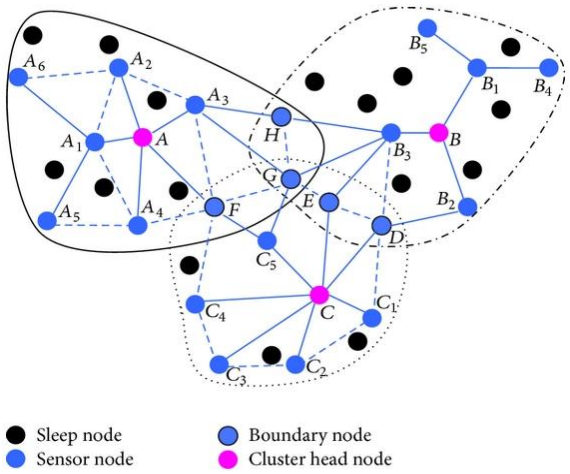

Fig. 3. Cluster -based Sensor network

C. *Data Mining*

A number of scientists determined the requirement of such need and thus was stood the science of data mining. WSN with large number of sensor nodes are used to observer the sensing arena. Sensor network data-transfer consistency and Sensor node energy competence are the primary design constraints. Particular of the applications may need heterogeneous sensor nodes with dissimilar sensing phenomena or the different hardware features.

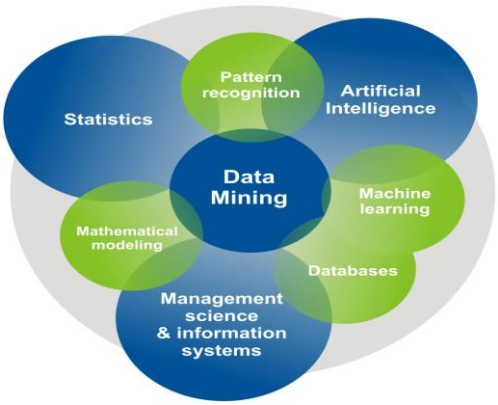

Fig. 5. Data Mining

Use of Data Mining for Energy Efficiency

Usually, a sensor node is the small hardware device containing of a sensing unit, and a power unit, along with processing unit, communication unit that are used for detecting, communication resolutions and data administering. This acts as border between the sensor nodes and clients. The individuality of the sensor node lies in its small size and the frivolous. Though, there are lots of restraints such as limits on resources in terms of dynamism, bandwidth computational swiftness, and recollection. Since of these constraints the connections between the sensors are limited to the short distances and low data charges. These sensors are used in extensive range of applications and real time applications

such as security perseverance, military presentations, habitat observing, and nuclear power plants. Sensors collect useful information in a timely manner and send to a consolidated node named base station. Although, the data being interpreted into useful knowledge due to their complexity and a lack of operational data analysis procedures.

Mining Data from Sensor Networks

This is particularly true in much technical, business, engineering requests. Sensor networks are finding number of applications in many areas, including smart structures, battlegrounds, and the human physique. Furthermore, sensors must process a constant stream of data in WSN. Data mining in wireless sensor networks is a stimulating part, as the algorithms need to work in embarrassed and extremely demanding atmosphere of sensor networks [37].

The mobility of the nodes, can be controlled or uncontrolled. Further, the relay node with controlled mobility passages with a chosen mobility pattern in the area of reflection, whereas the relay nodes with uncontrolled mobility may track any mobility pattern for their movement in the area of statement. Controlled mobility can increase the enactment of wireless sensor systems in terms of energy depletion and packet delay [33].

Energy Saving Models for WSN

On these days, wireless sensor networks are being used for a fast-growing number of different application fields. Monitoring the sensor network involves the frequent acquisition of measurements from all sensors in WSN. Since sensor data achievement and communication are the main sources of power consumption and sensors are battery-powered, and another important issue in this context is energy saving during data assortment. Thus, the challenge is to prolong sensor lifetime by reducing communication cost and computation energy. Relatively than directly querying all network nodes, only the illustrative sensors are queried by reducing the communication and computation with power costs. Enlarging the lifetime of the network through minimizing the energy is an important challenge in WSN.

Sensors cannot be easily swapped or recharged due to their ad-hoc deployment in menacing environment.

The SeReNe structure is stands for selecting representatives in a sensor Network (SeReNe) are an atmosphere for the empathy of energy saving models to competently query sensor systems. Figure demonstrations the SeReNe framework combined into sensor network construction. From the sink, Sensor nodes regularly obtain queries.

The query is applying by each node over its sensor data and the result refers to the base station by means of a multi-hop data gathering protocol. Previously transmitting the query consequences, the sensor data can be sometimes condensed by compression procedures. SeReNe produces sensor network models by means of two steps: 1) Selection of sensor representatives, and 2) Correlation study.

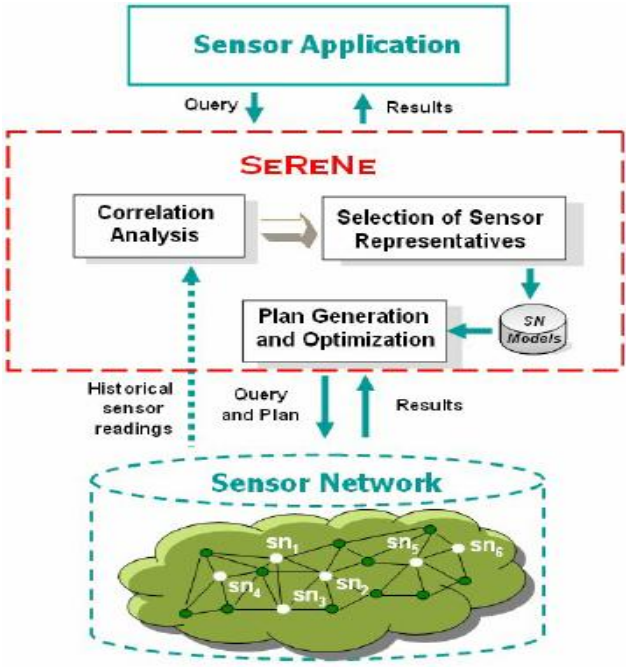

Fig. 6. Architecture of the SeReNe framework

II. DESIGN OF ONLINE MINING

In this section, we identify the subsequent three problems of online mining.

A. *Detection of Sensor Data Irregularities*

This problem is particularly important in the sensor network setting because it can be used to recognize abnormal or interesting events or faulty sensors in WSN. We chance this problem into two smaller difficulties. One is to detect irregular patterns of numerous of sensory attributes and the other to detect irregular sensory data ofa single attribute with respect to time. The irregular multi-attribute pattern detection problem has the hypothesis that there are some normal patterns among multiple sensory characteristics, which is true in some normal marvels. Once these normal patterns are broken anywhere, the irregularity is detected and informed.

B. *Detection of Irregular Patterns*

It's proposed a new method named pattern variation discovery to solve this difficulty. This method works in the following four steps: i) Selection of a reference frame.
This frame comprises of the directions with which we want to look for irregularities among multiple sensory features. It is also possible to determine the reference frame that results in a lot of irregularities. Moreover, ii) Definition of normal patterns. Hence,this definition can be models of numerous sensory attributes among multiple attributes. Furthermore,iii) Incremental maintenance of the normal patterns. When a sensor gets a new round of recitations, the normal patterns are accustomed incrementally. The last is, iv) Discovery of irregularity. When a normal pattern is broken at some point along the reference frame an irregularity emerges. Its mean, the pattern variation occurs. The problem then develops to discover the irregular matrix among a set of matrix. An

irregular matrix signifies that at the resultant time point, the distribution pattern of all the sensory attributes on all the nodes are irregular in sensors. Since our approach involves a lot of comparisons between matrices we propose to use the technique of (SVD) Singular Value Decomposition [4]. SVD is a powerful data decrease and approximation procedure, which removes the useful features of a matrix in matrices [36].

C. Detection of Irregular Sensor Data

We offer to detect irregular single-attribute sensor data with respect to time or space by construction model. For sequential irregularities in sensor data, which we build a model of the sensory data as the readings of a node comes in. Once some reading substantially affects the coefficients of the model then it is identified as an irregularity in WSN [30]. If some readings of a node differ from what the model expects based on the readings of the neighboring nodes, an anomaly is detected. In order to reduce resource ingesting, it may define the neighboring nodes to be those only a single hop away on the network system. As a node moves geologically, the parameters of its model is incrementally accustomed [31].

D. Clustering of Sensor Data

This method works as follows. First onte the cluster the sensor data along each sensor attribute distinctly. All caused clusters form a set of clusters which we call the Cluster Set. Another second, build a bipartite graph G with the Sensor Set and the Cluster Set being the two vertex sets [32].

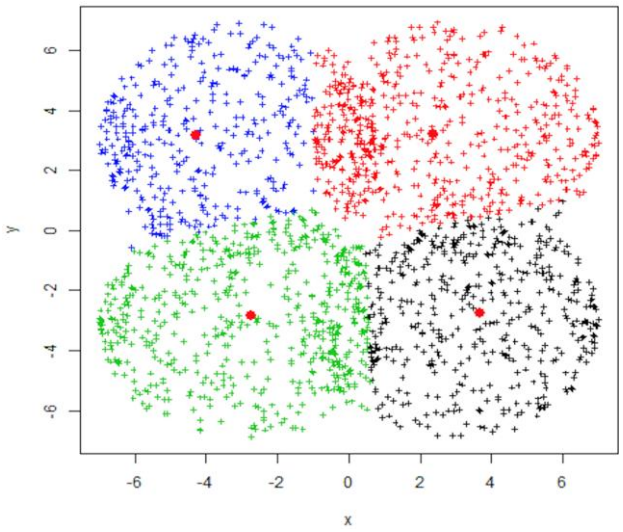

Fig. 7. Multi-dimension clustering

Additionaly, the role of a illustrative node can be rotated among the nodes in one group for load assessment [35].

E. Discovery of Sensory Attributes Correlations

Here its being treat readings of each sensory attribute as a data stream i.e. a sequence of data items xi at the order number i. Furthermore, as we are involved in correlation between data changes of various sensory attributes, we replace each sensory data value xi with its difference from its former data item, and $\Delta i = xi - xi\text{-}1$. Hence, a time series of sensor data is characterized as a sequence of Δi. Let S1, …, Sm-1, and Sm be a collection of m sensor data streams, repersents to each for one feature. Thus,one way of representing these data streams is to use a matrix A with time points and attributes being row and column index. Furthermore, we can then collect data by correlated attributes or correlated time points [24], [26] in this matrix systems. Recall that we propose to use the technique Singular Value Decomposition for matrix reduction in pattern variation discovery in WSN [35]. Hence, this technique can also be used to discover the best subspace that identifies the strongest linear correlations in the fundamental data set [10]. Moreover, (Singular Value Decomposition) SVD efforts to recognize similarity patterns of related values in the A matrix and the similarity of each row with the patterns in pattern variation [4]. Instead, we can consider sensory attributes correlations as inter-transaction association rules [13]. However in the context of a large-scale mobile sensor network, this problem is more complex and stimulating than traditional market basket analysis in WSN [29].

III. LITERATURE REVIEW

QingquanZhanga et.al. (2009) suggest a novel gradient-based method which uses the statistical methods to evaluate the position of a stationary aim. Mobile nodes can be directed towards the aim using the shortest path.

Xueli Shen et.al. (2011) present the radio propagation path loss and the previous weighted centroid location algorithm is improved. The weighted factor in the algorithm is retrieved, which makes positioning correctness and positioning error of the unknown nodes healthier.

Cesare Alippi et.al.(2009) uses effective energy administration strategies should be include policies for an economic use of energyhungry sensors which become one of the components affecting the network lifetime.

Balasubramaniam Natarajan et.al (2011) develops a novel optimal control theory based formulation of this sensor deployment problem. Manipulating affinity between the problem at the hand and linear quadratic regulator, and an analytical solution is tested and consequent.

Nithyakalyani and S. Suresh Kumar et.al. (2012) presents K-Means Data Relay (KMDR) clustering algorithm for grouping the sensor nodes there by decreasing number of nodes transmitting data to sink node,it intensifies the network performance by reducing the communication directly above. The contribution of K-MDR is to reduce power ingesting; the simulation experimental results show that the time efficiency of the algorithm is accomplished.

Hongbo Jiang, Chonggang Wang et.al (2011) develops an energy-efficient framework for the clustering-based data collection in WSNs by integrating adaptively enabling & disabling prediction structure. The framework is clustering constructed. All sensor nodes are signifying by the cluster head in the cluster and collects data values from all of them. Furthermore, to the realize prediction techniques efficiently in WSN, the author present adaptive scheme to control prediction used in our framework, analyse the presentation trade-off between limiting prediction cost and reducing statement cost, and design algorithms to exploit the profit of adaptive scheme to enable & disable prediction processes.

V. Karthik et.al.(2013) proposed the data collection method includes deployment of the multiple mobile robots whose goal is to gather data from the nodes whose energy level is below the threshold value. In planned hybrid scheduling and navigation of mobile robots justified by both the amalgamation of location and time based approaches with multiple region development.

LuigiCoppolino, SalvatoreD'Antonio et.al. (2013) propose a lightweight and hybrid, distributed Intrusion Detection System for WSN. This Intrusion Detection System uses both anomaly-based and misuse-based detection methods. It improve Central Agent which achieves highly accurate intrusion detection by using the data mining methods, and a number of the Local Agents running lighter anomaly-based detection systems.

M. Vijayalakshmi , V. Vanitha et.al. (2013) proposed Prediction and Clustering methods which use the temporal correlation among sensor data, provide a chance for reducing energy consumption of the continuous sensor data assortment. Hence, it can achieve prolongs and stability network generation. An adaptive scheme is presented which is used to control the prediction and analyse the performance trade-offs between reducing prediction cost and communication cost. Furthermore,design the algorithms to take the benefit of adaptive scheme to enable or disable prediction processes.

Shruti Kukreja, Prof Poonam Dabas et.al. (2015) propose a refine clustering method for improving the network era. Unlike LEACH where the clustering process is based on the random selection of cluster heads the author use k-nearest neighbour (k-NN) to create cluster in the network then the author use local heuristic search technique(search to find CHs). So that minimum energy is consumed in any broadcast [27].

IV. RELATED WORK

A lot of work has been invented in the areas of sensor networks, data streams and data mining, but little work has been done at the intersection of these parts. [6]. Although, existing sensor databases deficiency support for complex and online mining processes. There is prevalent literature regarding irregularity detection [12], [21].Sensor databases and query processing methods have been proposed for obtaining and managing sensor data in WSNs [14], [15]. There is also initial work on modeling sensor data that includie a distributed model based on kernel density estimators [16] and a distributed regression infrastructure [9]. While the clustering problem has been widely considered [5. Existing sensor network clustering methods [3], [8], [22] largely concern about the distance among nodes and the network topology, and not sensory data. Just, the clustering problem has also been studied in data stream [1], [18]. Furthermore,there is some work on correlated data items [11] with respect to their retrieves in order to improve data convenience in sensor networks. Further, in comparison, we focus on finding out correlations among sensory values in sensors [28]. Moreover,a related problem is identifying correlations among streams [10]. In other hand, there has also been initial work on online analytical processing and mining for data streams [17], [18], [23].

V. CONCLUSION

Sensor nodes are capable of transmitting and sensing. Sensors collect large amount of data in a highly distributed method. The composed data contain all the information about the area. But sometimes it happens that the users are need only the specific data and rest of information is treated as inappropriate. In future, same can be used to excerpt the desired information from the set of the large data. We have recognized the challenges for online mining and data mining in large-scale mobile sensor network atomosphere. The main apprehension is to satisfy the mining accuracy necessities while maintaining the resource ingesting to a minimum. We provide preliminary deliberations towards solving these problems. Hence,we believe that the patterns discovered can not only enable the applications to gain insight into the sensor data but also be used to tune the system presentation. As future a work, we will careful more about energy-awareness, adaptivity and fault-tolerance of online mining for sensor networks in addition to a further study of our planned methods.

ACKNOWLEDGMENT

We would like to thank our friends and colleagues, and the university who provided us an efficient support to work on this atmosphere and good infrastructure. We would also like to thank to all the previous researchers who worked very hard and helped others to comprehend the subject of Wireless Sensor Networks System.

REFERENCES

[1] Cesare Alippi, Giuseppe Anastasi, "Energy Management in Wireless Sensor Networks with Energy-hungry Sensors", IEEE Instrumentation and Measurement Magazine,April 2009.

[2] QingquanZhanga, Gerald E. Sobelmana, "Gradient-based target localization in robotic sensor networks", Pervasive and Mobile Computing" , February 2009.

[3] Xueli Shen and Huan Zhang*, "Improvement of Centroid Location Algorithm for Wireless Sensor Networks",International Conference on Computer Science and Information Technology,2011.

[4] Balasubramaniam Natarajan, Ahmad Ababnah, "Optimal Control-Based Strategy for Sensor Deployment",IEEE transactions on systems, man, and cybernetics ,January2011.

[5] Hongbo Jiang, Chonggang Wang, "prediction or not? an energy-efficient framework for clustering-based data collection in wireless sensor networks",IEEE transactions on parallel and distributed systems,June2011.

[6] Nithyakalyani and 2S. Suresh Kumar, "Data Relay Clustering Algorithm for Wireless Sensor Networks: A Data Mining Approach",Journal of Computer Science 8 (8), 2012.

[7] Luigi Coppolino, Salvatore D'Antonio, "Applying Data Mining Techniques to Intrusion Detection in Wireless Sensor Networks",Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, 2013.

[8] V. Karthik, "Region Based Scheduling With Multiple Mobile Robots for Data Collection Strategies in Wireless Sensor Networks", International Journal of Innovative Research in Science, Engineering and Technology,July 2013.

[9] M. Vijayalakshmi , V. Vanitha, "Cluster based adaptive prediction scheme for energy efficiency in wireless sensor

networks", International Research Journal of Mobile and Wireless Communications, June 2013.

[10] Shruti Kukreja, Prof Poonam Dabas, " Clustering in WSN Using Data Mining and Classification Technique",International Journal of Advanced Research in Computer Science

[11] Anisha kamboj , TanuSharma, An Enhanced Data Mining Technique for Energy Efficiency in WSN, International Journal of Engineering Trends and Technology (IJETT) – Volume 36 Number 3- June 2016,pp-137-143

[12] Xiuli Ma1 , Dongqing Yang1 , Shiwei Tang1 , Qiong Luo2 , Dehui Zhang1 , Shuangfeng Li1," Online Mining in Sensor Networks", http://opendl.ifip-tc6.org/db/conf/npc/npc2004/MaYTLZL04.pdf

[13] C.C.Aggarwal, J.Han, J.Wang, P.S.Yu. A Framework for Clustering Evolving Data Streams.VLDB2003.

[14] I.F.Akyildiz, W.Su, Y.Sankarasubramaniam, and E.Cayirci. Wireless Sensor Networks: A Survey. Computer Networks, Vol.38, No.4, pp.393-422, 2002.

[15] S.Bandyopadhyay and E.J.Coyle. An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks. IEEE INFOCOM 2003.

[16] D.Barbara, W.Dumouchel, C.Faoutsos and P.Haas. The New Jersey Data Reduction Report. IEEE Data Engineering Bulletin, Vol.20, No.4, pp.3-45, 1997.

[17] V.Estivill-Castro. Why So Many Clustering Algorithms----A Position Paper. SIGKDD Explorations. Vol.4, Iss.1, 2002.

[18] D.Estrin, R.Govindan, J.Heidemann, and S.Kumar. Next Century Challenges: Scalable Coordination in Sensor Networks. MobiCOM 1999

[19] M.J.Franklin. Challenges in Ubiquitous Data Management. Informatics 2001.

[20] S.Ghiasi, A.Srivastava, X.Yang, and M.Sarrafzadeh. Optimal Energy Aware Clustering in Sensor Networks. Sensors 2002, 2, pp.258-269.

[21] C.Guestrin, P.Bodik, R.Thibaux, M.Paskin and S.Madden. Distributed Regression: an Efficient Framework for Modeling Sensor Network Data. IPSN 2004.

[22] S.Guha, D.Gunopulos, and N.Koudas. Correlating Synchronous and Asynchronous Data Streams. KDD 2003, pp. 529-534.

[23] T.Hara, N.Murakami, and S.Nishio. Replica Allocation for Correlated Data Items in Ad Hoc Sensor Networks. SIGMOD Record, Vol.33, No.1, pp.38-43, March 2004.

[24] E.M.Knorr and R.T.Ng. Algorithms for Mining Distance-based Outliers in Large Datasets. VLDB 1998.

[25] H.Lu, L.Feng, J.Han. Beyond Intra-transaction Association Analysis: Mining Multi-dimensional Inter-transaction Association Rules. ACM Transaction on Information system, Vol.18, No.4, pp.423-454, 2000.

[26] S.Madden, M.J.Franklin, J.M.Hellerstein, and W.Hong. TAG: A Tiny Aggregation Service for Ad-hoc Sensor Networks. In Symposium on OSDI, 2002.

[27] S. Madden, M.J.Franklin, J.M.Hellerstein, and W.Hong. The Design of an Acquisitional Query Processor for Sensor Networks. SIGMOD 2003.

[28] T.Palpanas, D.Papadopoulos, V.Kalogeraki, and D.Gunopulos. Distributed Deviation Detection in Sensor Networks. SIGMOD Record, Vol.32, No.4, Dec.2003

[29] T.Palpanas, M.Vlachos, E.Keogh, D. Gunopulos, and W.Truppel. Online Amnesic Approximation of Streaming Time Series. ICDE 2004.

[30] N.H.Park and W.S.Lee. Statistical Grid-based Clustering over Data Streams. SIGMOD Record, Vol.33, No.1, March 2004.

[31] F.Perich, A.Joshi, T.Finin, and Y.Yesha. On Data Management in Pervasive Computing Environments. IEEE Transactions on Knowledge and Data Engineering. Vol. 16, No. 5, May 2004.

[32] G.J.Pottie and W.J.Kaiser. Wireless Integrated Network Sensors. Communications of the ACM. Vol.43, No.5, pp.51-58, May 2000.

[33] Muntjir, Mohd. "Novice Evaluation and Comparative Survey on Database Management System, Data Warehousing and Data Mining." International Journal of Computer Applications 136.10 (2016): 39-45.

[34] S.Sarawagi, R.Agrawal, and N.Megiddo. Discovery-driven Exploration of OLAP Data Cubes. EDBT 1998.

[35] O.Younis and S.Fahmy. Distributed Clustering in Ad-hoc Sensor Networks: A Hybrid, Energy-efficient Approach. IEEE INFOCOM 2004.

[36] Muntjir, Mohd. "Implementation of Data Mining in Wireless Sensor Networks: An Integrated Review." International Journal of Computer Science and Network 5.4 (2016): 692-698.

[37] Muntjir, Mohd, and Ahmad Tasnim Siddiqui. "E-Commerce Framework Based on Evaluation of Data Mining and Cloud Computing." International Journal of Computer Science and Information Security 14.4 (2016): 286.