Volume 5, Issue 4, April 2015 ISSN: 2277 128X
International Journal of Advanced Research in
Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Review: on Data Mining and Data Gathering Algorithm in WSN

Vickey Sharma
Department of ECE, KUK,
Haryana, India

***Abstract**— Data mining in sensor networks is the method of selecting application-oriented standards and patterns with acceptable accuracy from a continuous, fast, and probably non ended flow of data streams from sensor networks. In this case, all data cannot be stored and must be processed quickly. Data mining techniques has to be sufficiently quick to process high-speed arriving data. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. The main objectives of this review to study and analyse various data collection and data mining algorithms.*

***Keywords**— Cluster Head, Base station, Sensor node, K-means data relay algorithm, Optimal terminal assignment based path algorithm.*

I. INTRODUCTION

A. Wireless Sensor Networks

A wireless sensor network is a collection of sensor nodes organized into a network. A sensor is a small device which observes the environment of physical parameters such as pressure, temperature, relative humidity, sound, motion, vibration or pollutants, at different locations. WSN is highly distributed networks of wireless sensor nodes, used in large numbers to monitor the environment or system. the transmission range of sensor nodes is limited and they organize themselves in an ad hoc manner, which means that two sensor nodes that cannot reach each other directly transmit on other sensor nodes to relay data between them .in general ,data packets from the source node have to traverse multiple hops before they reach the destination [2].

Fig. 1 shows a WSN which is composed of thousands of sensing nodes highly deployed in a large geographical area and with one or few BS or sink nodes that connect a sensor network to the users via the internet or other networks. Each sensor node consists of a CPU which can be a micro-processor or microcontroller for data processing, memory for storage, a RF transmitter and a receiver with an Omni-directional antenna and a source for power like solar cells or batteries. Each of these sensor nodes receive data, process the data and route that data to the sink node by multi-hopping [11].

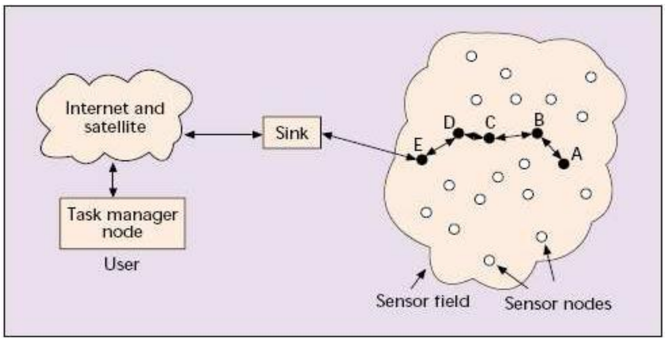

Fig. 1 Structure of WSN

B. Data collection and routing mechanisms for WSNs

The conventional data collection and routing mechanisms for WSNs can be broadly divided into two categories:

1) Event-based data collection: In event-based data collection, sensors are responsible for detecting and reporting a specific event to one or more sinks [5].

2) Periodic data collection: In periodic data collection, all sensor nodes periodically change their observations to the sink based on the latest information of the interested data. In addition, in multi hop-relay data delivery, typical traditional WSNs relay sensor observations to a sink via a tree-based structure. However, multi hop-relay approaches inevitably involve huge amounts of data exchange between nodes, in addition many overheads to maintain the network architecture [5].

C. Data Mining in WSN

Data mining in sensor networks is the method of selecting application-oriented standards and patterns with acceptable accuracy from a continuous, fast, and probably non ended flow of data streams from sensor networks. In this case, all data cannot be stored and must be processed quickly. Data mining methods has to be sufficiently fast to process high-speed arriving data. The current data mining methods are meant to handle the stationary data and use the multistep methods and multi scan mining algorithms for reviewing constant data-sets. Therefore, new data mining methods are not suitable for handling the large quantity, high dimensionality, and distributed nature of the data generated by the WSNs. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use [4].

D. Clustering in wireless sensor network

In clustering, the sensor nodes are partitioned into unlike clusters. Each cluster is supervised by a node referred as cluster head (CH) and other nodes are referred as cluster nodes. Cluster nodes do not commune directly with the sink node. They have to pass the collected data to the cluster head. Cluster head will collect the data, received from cluster nodes and transmits it to the base station. Thus minimizes the consumption of energy and number of messages communicated to BS. Also numbers of active nodes in communication are decreased. The result of clustering the sensor nodes is to extend the network lifetime [4].

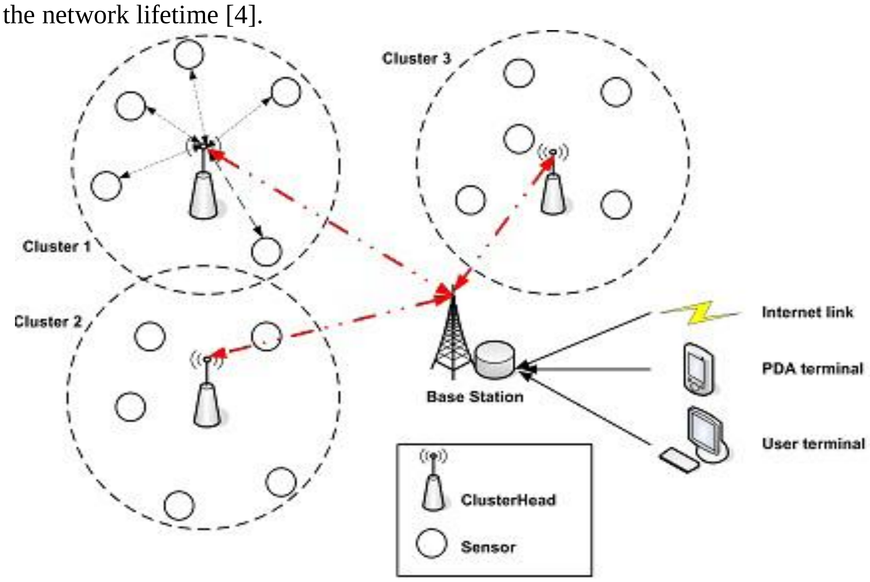

Fig. 2 Clustered Sensor Network

1) Sensor Node: It is the primary component of wireless sensor network that has the ability of processing, sensing, routing, etc.

2) Cluster Head: The Cluster head (CH) is considered as a leader for that specific cluster. And it is responsible for distinct activities carried out in the cluster, such as data gathering, transmission of data to BS, planning in the cluster, etc.

3) Base Station: BS is considered as a primary data collection node for the entire sensor network. It is the bridge (via communication link) between the sensor network and the end user. Normally this node is designed as a node with no power constraints.

4) Cluster: It is the structural unit of the network, created to make the Communication easy in the sensor network.

E. Sensor Networks Applications

Sensor nodes can be used for continued sensing, event discovery, event ID, position sensing, and local influence of actuators. The approach of micro-sensing and wireless connection of these nodes promises many new applications. We classify the applications into environment, military applications, home applications, health and other commercial applications. It is possible to spread this classification with more categories such as chemical processing, space exploration and hazard relief [13].

1) Military applications: Wireless sensor networks can be an essential part of military communications, military command, computation, information, investigation, and targeting (C4ISR) systems. The rapid distribution, self-management and fault tolerance characteristics of sensor networks make them a very promising sensing technique for military C4ISR. Since sensor networks are based on the complex distribution of disposable and low-cost sensor nodes, damage of some nodes by hostile actions does not affect a military operation as much as the destruction of a conventional sensor, which makes sensor, networks conception a better approach for war. Some of the military functions of sensor networks are monitoring equipment accessories, friendly forces and ammo; battlefield observation; reconnaissance of opposing forces and terrain; targeting; battle damage assessment; and nuclear, biological and chemical (NBC) attack detection and reconnaissance.

2) Environmental applications: The environmental applications include trailing the movements of birds, small animals, and insects; monitoring environ- mental conditions that affect crops and livestock; irrigation; macro instruments for large-scale Earth monitoring and planetary exploration; chemical/ biological detection; precision agriculture; biological, Earth, and environmental monitoring in marine, soil, and atmospheric contexts; forest fire detection; meteorological or geophysical research; flood detection; bio-complexity mapping of the environment; and pollution.

3) Health applications: The health applications provides interfaces for the disabled; integrated patient monitoring; diagnostics; drug administration in hospitals; monitoring the movements and internal processes of insects or other small animals; telemonitoring of human physiological data; and tracking and monitoring doctors and patients inside a hospital.

4) Home automation: Due to advance in technology, sensor nodes and actuators can be buried in equipment, such as refrigerators, vacuum cleaners, VCRs and microwave ovens. These sensors inside the domestic devices can interact with each other and with the external network via the Internet or Satellite. They allow end users to manage home devices locally and remotely more easily.

5) Smart environment: The design of smart environment can have two different viewpoints, i.e., human-centered and technology-centered. For human-centered, a elegant environment has to adapt to the needs of the end users in terms of input/ output possibilities. For technology-centered, newly hardware technologies, networking explanations, and middleware services have to be developed. A scenario of how sensor nodes can be used to create a smart environment is described in. The sensor nodes can be embedded into furniture and equipments, and they can interact with each other and the room server. The room server can also commune with other room servers to learn about the services they offered, e.g., scanning, faxing, and printing. These room servers and sensor nodes can be integrated with existing embedded devices to become self-management, self-control, and adaptive systems based on control theory models. Another example of smart environment is the ‘‘Residential Laboratory’’ at Georgia Institute of Technology. The computing and sensing in this environment has to be consistent, insistent, and transparent.

II. LITERATURE SURVEY

The various schemes used for data collecting and data mining are described below:

A. Robust Data Collection in WSN with Mobile Sinks

A method for the robust data collection in wireless sensor network is provided by the .Shrikant D. Dhamdhere in 2014 [1] called Optimal Terminal Assignment based Path (OTABP) that increases data collection and minimize the energy consumption as well as optimizing the assignment of sensor nodes to the sub sinks. OTABP method based on efficiently assignment of the member node to sub sinks. Propose work also implements approximate algorithm to solve the sub sinks selection and member node assignment problem. In this paper the author aim is to collect maximum data efficiently in wireless sensor networks by utilizing minimum energy using Optimal Terminal Assignment based Path (OTABP)of controllable movable sink.

B. Energy Saving In WSN Using Burst Algorithm

This method is provided by S.Punithashankari in 2013 [2]. A wireless sensor network is composed of a large number of low-cost devices distributed over a geographical area. Sensor nodes have finite processing capabilities therefore simplified protocol architecture should be designed so as to make communications simple and useful. Moreover, the power supply unit is based on a battery that has limited energy the networks should be aimed at minimizing the consumption of energy. A forwarding scheme for wireless sensor network designed at combining low computational complexity and high performance in terms of energy efficiency and accuracy. The technique relies on a packet-separation algorithm based on the BURST process. The burst performance is obtained by Shortest Path with Load Balancing (SP).The energy efficiency is low in this splitting process compare to those lowest paths. To approximation of the mean energy degradation factor achievable with the forwarding scheme should be derived and the reassembling is done in end side, the values are coded in block code and simulated using MATLAB simulator. The proposed algorithm out performs in terms of energy saving and increases the network lifetime.

C. Data Collection in WSN using Mobile Robots

This method for data collection using mobile robots is given by V. Karthik in 2013 [3]. Since Data collection process is one of the important aspects in the design consideration for future analysis in wireless sensor networks. Energy utilization by parent node increases due to continuous forwarding of sensed data from their respective child nodes especially based on the tree topology. Once the energy in the parent nodes was totally spent off, some of the child nodes get unique/ separated from the sink node. The proposed data collection method involves deployment of multiple mobile robots whose responsibility is to collect the data from the nodes whose energy is below the maximum value. The position of mobile robots to collect the data from partitioned nodes usually achieved by time and location based schemes. In projected hybrid scheduling, the navigation of mobile robots scheduled by both the combination of time and location based approaches with multiple region scheduling. In large network scenario, the mobile robot gets more burdens due to its more responsibilities to visit all partitioned nodes. So the entire scenario is divided into different regions and the deployment of multiple mobile robots is based on the needs. Therefore, the potency of sensed data collected by the base station or sink node from partitioned/islanded WSN is improved doubly using multiple MR. Through simulation is under the context of NS-2 simulator, the outcomes from various conditions show that proposed multiple mobile robots can improve the performance of collecting the sensed data in large-scale sensing fields and also it improves the lifetime of the sensor nodes.

D. Sensory Data Collection in WSN with Mobile Sinks

An algorithm for the sensory data collection in WSN is provided by the Sherin Mathew in 2013 [4]. Recently, there has been a rapid growth in the wireless communication technique. Mobile sinks can be mounted upon urban vehicles with fixed trajectories provide the ideal infrastructure to effectively recover sensory data from such unique wireless sensor network fields. Current methods uses either single-hop transfer of data from SNs that lie within the MS's range or heavy involvement of network periphery nodes in recovery of data, processing, buffering, and delivering task. The author proposed a technique aims at minimizing the overall network overhead and energy dissipation associated with the multi hop data recovery process while also ensuring balanced energy utilization among SNs and extended network lifetime. This is achieved through making cluster structures that consist of member nodes that direct their measured data to their assigned cluster head (CH). CHs perform data filtering upon raw data exploiting potential spatial-temporal data redundancy and forward the filtered information to appropriate end nodes with sufficient remaining energy, located in closeness to the MS's route. Simulation results show the tremendous performance of our proposed algorithm to strike the appropriate performance in the energy consumption and network lifetime for the wireless sensor networks.

E. A Novel Information Acquisition Technique for Mobile-Assisted WSNs

This technique is proposed by Fan Bai in 2012 [5] called an adaptive data-harvesting approach for mobile-agent-assisted data collection in wireless sensor networks (WSNs) inspired by social Ecology. In this paper the author uses marginal value theorem, the author divide the entire sensor field into small patches and gather the correlated data from each patch. Each observation X collected by a given sensor node to be considered as a marginal information source with a relative standard deviation $\sigma(x|Y,I)$, where Y is a set of previously collected observations by the mobile agent, and I is the background knowledge obtained from the sensor field. The mobile agent estimates the comparison based on the available knowledge gathered from the current patch and the previous patches and then chooses the next visiting sensor node. The next node contain the maximum information gain obtained until $\sigma(x|Y,I)$ is smaller than a predefined threshold (TH). Since in a dynamically changing environment, the interaction varies among different patches, an effective way to recognize the correlation model is the key to efficient data harvesting. The projected estimation technique of the marginal value theorem, which is called EMVT, is used to maintain the fidelity of the interested data with relatively fewer collected sensor observations.

F. Bounded Relay Hop Mobile Data Gathering in WSNs

This method for data gathering is given by Miao Zhao in 2012 [6] by employing mobile collectors that collects data via short-range communications. To follow saving maximum energy at sensor nodes, automatically, a mobile collector should pass through the transmission range of each sensor in the field such that each data packet can be directly transmitted to the mobile collector without any relay. Though, this approach may lead to certainly increased data gathering latency due to the low moving velocity of the mobile collector. Providentially, it is observed that data collecting latency can be effectively shortened by performing proper local aggregation via multi hop transmissions and then uploading the aggregated data to the mobile collector. In such method, the number of bounded transmission hops should not be arbitrarily large as it may increase the energy consumption on packet relays, which would badly influence the overall efficiency of mobile data gathering. In this paper, the author study the trade-off between energy saving and data gathering latency in mobile data gathering by exploring a balance between the relay hop count of local data aggregation and the moving tour length of the mobile collector. The author proposes a polling-based mobile gathering technique and formulates it into an addition problem, named bounded relay hop mobile data gathering (BRH-MDG). Specifically, a subset of sensors will be selected as polling points that buffer locally aggregated data and upload the data to the mobile collector when it reaches. Until, when sensors are associated with these polling points, it is certain that any packet relay is bounded within a given number of hops. The author then gives two efficient algorithms for selecting polling points among sensors. The effectiveness of this approach is validated through extensive simulations.

G. Data Relay Clustering Algorithm for Wireless Sensor Networks

This approach for data mining in WSNs is provided by S. Nithyakalyani in 2012 [7]. At present sensors are very necessary for today life to observe environment where human cannot get involved very often. Wireless Sensor Networks (WSN) are used in many real world applications like traffic control, environmental monitoring, trajectory monitoring. It is more difficult for sensor network to sense and collect a large amount of data which are incessant over time, which need to be delivered to sink for further decision making procedure. Clustering of sensors data act as a nucleus job of data tapping. A clustering in wireless sensor network involves choosing cluster heads and assigning cluster members (sensors) to it for efficient data relay. The constraints in power supply, limited communication bandwidth, storage resources are the major challenges in WSN facing today. The author proposed K-Means Data Relay (KMDR) clustering algorithm for grouping sensor nodes there by reducing number of nodes transmitting data to sink node, it decreases the communication overhead and in this manner increase the network performance. Furthermore Conserve and Observe Modes (COM) algorithm reduces the number of nodes within the cluster there by without compromising the coverage face major challenges such as constriction in power supply, limited bandwidth and storage resources region of it. The impact of K-MDR is to decrease power consumption finally the simulation experimental results show that the time efficiency of the algorithm is achieved.

H. Nearest Neighbour Classification for Wireless Sensor Network Data

This technique for sensors data mining is provided by Khushboo Sharma in 2011 [8]. Advances in wireless technologies have led to the development of sensor nodes that are capable of sensing, processing, and transmitting. They collect large amounts of sensor data in a highly decentralized manner. Classification is an important task in data mining.

In this paper a Nearest Neighbours Classification technique is used to classify the Wireless Sensor Network data. The experimental investigation by the author yields a significant output in terms of the correctly classified success rate being 92.3%.

I. Data Collection Using Mobile Robot In Wireless Sensor Networks

A innovative data-collecting technique using a mobile robot to acquire sensed data from a wireless sensor network (WSN) that possesses partitioned/islanded WSNs is proposed by the Tzung-Cheng Chen in 2011 [9]. This algorithm permits the improvement of data collecting performance by the base station by identifying the locations of partitioned/islanded WSNs and navigating a mobile robot to the preferred location. To detect the locations of the separated/islanded WSNs, two control methods, a global- and local-based method, are designed. Accordingly, the navigation techniques of the robot can be scheduled based on time and location using three scheduling techniques: location based, time based, and dynamic moving based. With these techniques, the mobile robot can collect the sensed data from the separated/islanded WSNs. Therefore, the potency of sensors data collected by the base station in separated/islanded WSNs is improved. Through simulation under the context of an ns-2 simulator, the results, from various conditions, show that the collecting techniques proposed can dramatically improve sensed data-collecting performance in partitioned or islanded WSNs.

J. Cluster Based Approximate Data Collection in Wireless Sensor Network

This method is provided by Ashu Gupta [10]. Data collection is a fundamental task in wireless sensor networks. Sensor networks are composed of sensor nodes which cooperatively send sensed data to the base station. As sensor nodes use battery for driven, an effective usage of power is essential in order to use network for long time, hence it is needed to minimize data traffic inside sensor networks. In this paper, the author focuses on approximate data collection that will enhance the lifetime of the wireless sensor network. The key idea of Approximate Data collection is to divide a network of sensor nodes into clusters, in which each cluster will have numbers of cluster members and a single cluster head. Each cluster member will have a set of sensed raw data to find out the local approximate data on each cluster member within the cluster, and will send these approximate data to the cluster head. After collecting the local approximate data from all cluster members within the cluster, the cluster head accomplish the global data estimation, and send this comparative data to the base station.

III. DATA MINING AND DATA GATHERING ALGORITHM

A. Data gathering algorithm

The various data gathering algorithm are given in table below:

SPT-DGA	Shortest Path Tree Based Data Gathering Algorithm
PB-PSA	Priority Based PP Selection Algorithm
SHDG	Single Hop Data Gathering Algorithm
CME	Controlled Mobile Element Scheme
MR	Mobile Robot

B. Data Mining Algorithm

The various data mining algorithm are given below in table:

K-MDR	K-Means Data Relay Algorithm
COM	Conserve And Observe Mode Algorithm
DT	Decision Tree Algorithm
NNTC	Nearest Neighbour Trajectory Classification Algorithm
OTABP	Optimal Terminal Assignment Based Path Algorithm

IV. CONCLUSION

The conventional data collection and routing mechanisms for WSNs can be broadly divided into two categories: In event-based data collection, sensors are answerable for detecting and reporting a specific event to one or more sinks. In periodic data collection, all sensor nodes periodically update their observations to the sink based on the latest information of the interested data. In multi hop-relay data delivery, distinctive traditional WSNs relay sensor clarification to a sink via a tree-based structure. However, multi hop-relay approaches inevitably involve huge amounts of data exchange between nodes, in addition many overheads to maintain the network architecture.

REFERENCES

[1] Shrikant D. Dhamdhere, "Robust Data Collection in Wireless Sensor Networks with Mobile Sinks", International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014.
[2] S.Punithashankari, "Energy Saving In Wireless Sensor Network Using Burst Algorithm", IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), Volume 5, Issue 1 (Jan. - Feb. 2013).
[3] 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, Vol. 2, Issue 7, July 2013.
[4] Sherin Mathew, "A Novel Approach for Sensory Data Collection in Wireless Sensor Networks with Mobile Sinks", International Journal of Engineering Science and Innovative Technology (IJESIT), Volume 2, Issue 2, March 2013.
[5] Fan Bai, "A Novel Information Acquisition Technique for Mobile-Assisted Wireless Sensor Networks", IEEE Transactions on Vehicular Technology, Vol. 61, No. 4, May 2012.
[6] Miao Zhao, "Bounded Relay Hop Mobile Data Gathering in Wireless Sensor Networks", IEEE Transactions On Computers, Vol. 61, No. 2, February 2012.
[7] S. Nithyakalyani, "Data Relay Clustering Algorithm for Wireless Sensor Networks: A Data Mining Approach", Journal of Computer Science 8 (8): 1281-1284, 2012.
[8] Khushboo Sharma, "Nearest Neighbour Classification for Wireless Sensor Network Data", International Journal of Computer Trends and Technology, volume-2, Issue-2, 2011.
[9] Tzung-Cheng Chen, "On Data Collection Using Mobile Robot in Wireless Sensor Networks", IEEE Transactions On Systems, Man, And Cybernetics—Part A: Systems And Humans, Vol. 41, No. 6, November 2011.
[10] Ashu Gupta and R.Annie Uthra," Cluster Based Approximate Data Collection in Wireless Sensor Network", IJECSE, Volume2, Number 2
[11] Jyoti Rajput," A Survey on Secure Data Aggregation in Wireless Sensor Network", IJARCSSE, Volume 4, Issue 5, May 2014.
[12] Laxmi Choudhary, "Challenges for Data Mining", IJREAS, Volume 2, Issue 2, February 2012
[13] I.F. Akyildiz, "Wireless sensor networks: a survey", Computer Networks (Elsevier) Journal, pp. 393-422, March 2002.
[14] Wendi B. Heinemann, "An Application-Specific Protocol Architecture for Wireless Micro sensor Networks", IEEE Transactions on Wireless Communications, vol. 1, No. 4, October 2002.
[15] Mukesh Kumar Jha, T.P Sharma "Secure Data aggregation in Wireless Sensor Network: A Survey", International Journal of Engineering Science and Technology, ISSN: 0975-5462, Vol. 3 No.3, March-2011.
[16] P. D. Patel, P.B. Lapsiwala, R.V. Kshirsagar "Data Aggregation in Wireless Sensor Network", International Journal of Management, IT and Engineering, vol. 2, Issue 7 July-2012.
[17] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," IEEE Trans. Wireless Communication., vol. 1, no. 4, pp. 660–670, Oct. 2002.
[18] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, "In-network aggregation techniques for wireless sensor networks: A survey," IEEE Wireless Communication , vol. 14, no. 2, pp. 70–87, Apr. 2007.

