Journal of Scientific & Industrial Research
Vol. 74, January 2015, pp. 38-42

Analysis of Node Clustering Algorithms on Data Aggregation in Wireless Sensor Network

S Nithyakalyani[1]* and B Gopinath[2]

*[1]Department of Information Technology, K.S.R College of Engineering, Tiruchengode, India
[2]Department of Electrical and Electronic Engineering, Vivekanandha College of Engineering for Women, Tiruchengode, India

Received 20 July 2012; revised 24 August 2013; accepted 10 November 2014

One of the most important constraints to be studied in Wireless Sensor Networks (WSNs) is its life time. There are two typical data mining processes that support to reduce the energy consumption of WSN is clustering and data summarization. One of the primary goals of node clustering in WSN is in-network preprocessing that aims to obtain qualified information and to limit the energy consumed. A clustering algorithm is composed of three parts first electing cluster head (CH), selection of cluster membership and transferal data from members to CH.CH relays only one of the aggregated or compressed data packet to sink/ base station. In this paper a brief comparative study is made from different research proposals, which suggests different cluster head selection approaches for data aggregation. The algorithms under this study are Voronoi based K-means clustering algorithm, Voronoi Fuzzy C-means clustering algorithms and Voronoi based Genetic clustering algorithm. Significant factors for evaluating and comparing these algorithms are defined, analyzed and summarized. It has been assumed that the sensor nodes are randomly distributed and are not mobile, the coordinates of the base station (BS) and the dimensions of the sensor field are known.

Keywords: Wireless sensor network, Clustering algorithms, Voronoi diagram,k-means, Fuzzy, Genetic, Data aggregation.

Introduction

Recent advances in wireless communications leads to Wireless Sensor Network WSN. It is a collection of autonomous sensor nodes that are operated for monitoring the environments. Each sensor node consists of tiny hardware components with tiny OS, memory, sensing device, ADC (Analog to Digital Converter), a central processing unit (CPU) and a small battery power. Usually it communicates in short range distances over a radio frequency channel and these devices are small in size. The working procedure of this tiny node is sensing, processing and communicating data to BS[1].The development of wireless sensor networks was originally motivated for military applications such as battlefield surveillance, national security, health care, home appliances environmental monitoring, and many other fields, where human intervention is extremely dangerous, and difficult. Basically WSNs possess some characteristics which include node deployment, limited battery power and memory, single or multiple base Station (BS), node dynamicity, no global unique ID, application awareness, immovable sensor nodes, consistent topology change, single-hop transmission or multi-hop transmission, unpredictable sensor nodes (due to node disaster), data redundancy, election of CH, and data aggregation in CH. The lifetime of WSN is measured by the over-all time before the first sensor node runs out of power. Majority of sensor networks applications fall into the questioning applications for future prediction. For which data mining a vital tool to continuously collect and integrate the data. The WSNs amazing features direct us to innovative research challenges in several data mining process. The traditional data mining methods are centralized and computationally expensive. Data aggregation in wireless sensor network is very much attracted by research communities nowadays to prolong the network lifetime. Data aggregation algorithms are frequently measured by executing the algorithm several rounds. In every round, data from all the sensor nodes are gathered and then forwarded to the base station. Data aggregation algorithms are classified based on the type of communication architecture that it employs such as clusters, grid, chain, connected dominating sets and trees [3-11].

*Author for correspondence
E-mail: nithyakalyani.me@gmail.com

Need for Clustering in WSN

An arrangement of sensor nodes into different virtual groups is known as clustering. Each cluster comprises of CH and its members. A CH generally serves like a leader for its cluster, performing intra cluster transmission arrangement, data sending, and so on. The cluster heads can summarize the data and send it to the data center or BS as a single packet, thus reducing the overhead from packet headers. The CHs rotate randomly in each epoch within the network for load balancing.

In each round of the cluster formation phase, the network need to select cluster heads and transfers the aggregated data to BS. For electing a cluster head, the following questions are to be considered.

- Who will initialize the cluster head selection?
- What are the parameters necessary for deciding the role of a sensor node?
- Which sensor nodes will be nominated as cluster heads?
- Does it require re-initiation of cluster formation process?
- Whether the selected cluster heads are evenly distributed?
- Whether it assures load balancing of clusters?
- Which data transmission is appropriate for large network, Single-hop or Multi-hop?
- Clustering is very much essential for WSNs for the following reasons.
- Reduces energy consumption by improving bandwidth utilization.
- Load balancing for even distribution of sensors.
- Re-clustering when fault tolerance occur CH transmits only aggregated data to the data sink node
- Reducing number of nodes taking part in transmission to BS by electing CH
- Scalability for large number of nodes.
- Reduces data delay by regulating the number of hops.
- Efficient utilization of resources.
- Depending upon the application here the sensors is employed, the above clustering objectives are considered.

Classification of Clustering Features

The clustering algorithms of WSNs are generally categorized based on the classification of clustering features. In this paper such features are used to categorize and compare the clustering algorithms.

- Based on clustering properties are:
  - Number of Cluster
  - Data transmission
  - Number of cluster member
- Based on Cluster head Abilities:
  - Dynamicity
  - Types of the node
  - Role
- Based on CH Selection Process:
  - Selection Approach
  - Purpose of node Clustering
  - CH Selection
  - Complexity
  - Purpose of data aggregation

Need for Voronoi structure in WSN

Voronoi diagram is a well-known computational geometry structure, often used in WSN for coverage problem and to find optimal cluster. It partitions the given area into a number of Voronoi cells in such way that there exists only one sensor in each cell. Voronoi diagram can act as fundamental sampling method for WSN coverage, with sensor acting as sites. If the entire Voronoi polygons vertices are covered then the region is said to be completely covered. Let S= $\{p_1, p2,...p_n\}$ be a set of points. Voronoi region $V(p_i)$ is calculated using

$V(p_i)=\{x:|p_i-x| \leq |p_j-x|, \text{ for all } j\neq i\}$.

Need for sensor fusion in WSN

There are several terminologies like data fusion, sensor fusion, information fusion, data aggregation, data integration and data summarization to describe the term fusion. Sensor fusion is the process of integrating sensory data from multiple-sensor that aims to obtain the information with high quality, consistency, and accuracy. The information combined from the different sensors improves the understanding of our environment and provides some basic knowledge for planning and decision-making about the environment. Finding an optimal solution to sensor fusion is an NP –hard problem. However this process minimizes the number of transmission with little energy consumption. Improper and aggressive fusion may lead to more energy consumption and degradation of communication quality, such as latency, robustness and accuracy. Data fusion is a method which is executed in cluster head by applying the maximum, minimum and average functions. As a result, cluster head only transmits the useful information via aggregated data to the sink node periodically thus reducing congestion and its linked problems.

Classification of WSN Topology

There are four main classifications of WSN topologies. The first one is single-hop flat model, the second one is single-hop cluster model, the third one is multi-hop flat model and the fourth is multi-hop cluster model. If the networks are single-hop models then all sensor nodes transmit their data directly to the sink node. These architectures are not suitable for large-scale areas since transmission cost becomes so expensive for energy consumption. In the multi-hop flat model, all nodes share the same information like routing tables, thus overhead and energy consumption are increased and utilization of resources is not much. Then finally in the multi-hop clustering model, sensor nodes maintain low overhead and energy consumption due to data aggregation process is done in CH and CH relays only aggregated data to BS. In the multi-hop clustering model, resources can be allocated orthogonally to each cluster to reduce collisions between clusters and can be reused cluster by cluster. As a result, the multi-hop clustering model is appropriate for the sensor network deployed in remote large-scale areas.This paper reviews and compares Data relay K-Means clustering algorithm, Fuzzy C-Means clustering algorithm and Voronoi based Genetic clustering algorithm for cluster head selection approach for data aggregation in WSN. The performance measures of these algorithms are analyzed, implemented, investigated and results are plotted and discussed.

Voronoi Based Clustering For Data Aggregation in WSNs

In this section, three Voronoi based clustering algorithms for data aggregation of the research paper are deliberated with respect to cluster feature classifications and network model. Initial assumptions made for all three network models are:

- Network is homogeneous and deployed randomly.
- All sensor nodes have same facilities for sensing, computation and communication.
- For long range transmissions, the cluster-heads are nominated for performing transmission.
- Each sensor node relays the sensory data directly to the CH for further computation.
- The sensor can recognize or sense any incident within its sensing range.
- Coverage is predictable based on sensing range.
- Every sensor node location is known to it by some method such as the global positioning system or during initial deployment.
- All sensor nodes have the same initial energy level.

Voronoi based k-means clustering Algorithm

Voronoi based k-means clustering Algorithm (VkMC) is developed to group the sensor nodes for energy efficient data communication. In k arbitrary points are picked as the cluster head by the sink node[12]. Cluster members are obtained for each CH based on distance metric. Then in each CH data aggregation process as mention in section 1 are done for limiting the energy spent in transmission. In case of smaller network, clustering leads to high cost if k is predefined, and in such case, clusters need to be limited.

Voronoi Fuzzy Clustering Algorithm

Voronoi Fuzzy Clustering Algorithm (VFCA) is developed to address the problem of optimal cluster head selection approach for data aggregation [13]. However it improves the reliability of the detected information by the sensor. Data aggregation is performed in each CH to reduce the communication overhead in the network and to increase the network life time by decreasing the energy consumption for data transmission. VFCA is the fusion of Voronoi structure and Fuzzy C-Means Clustering algorithm. This algorithm first creates a balanced set of clusters by randomly selecting C cluster heads. Then applies Voronoi structure to partition the given C region into Voronoi cells. The member nodes inform their data to the respective CHs. Then CHs aggregate the data and send it to central base or sink node. The CHs is changed at frequent intervals in order to balance the network load. The aggregation process is carried to remove the redundancy of the sensory data and the processed high quality data is then sent to the sink node. The objective of this approach is to reduce the consumption of the energy and to increase the lifespan of the sensor nodes in the network.

Voronoi based Genetic clustering Algorithm

The wireless sensor nodes are arbitrarily deployed in the sensing environment and the sensors are grouped using the VGC. Initially, Voronoi diagram for the given field based on the position and energy of the sensor is constructed. Later the genetic clustering algorithm is applied to form set of cluster. Once clusters are created using this algorithm [14], sensed data are transmitted from cluster members to cluster head (CH) where data fusion is done to reduce the energy consumed by each node. Finally the cluster head transmits the fused information to the base station which is useful for further decision making in sink.

Experimental Setup and Comparison of Results

The sensor nodes are randomly deployed in the field and the base station is situated outside the sensing region. The number of times the nodes transmit their messages to the CH and CH transmits the message to the base station from the beginning of the network life is defined as the number of transmissions and it is used to measure the network lifetime for different clustering methods. Wireless channel is assumed to be ideal and so there is no retransmission of control packets because of collision. Network model is assumed as Table 1, and for energy calculation, number of bits n = 4000bits, E_{elec} = 50nJ / bit and Transmitter amplifier $€_{amp}$ = 10 pJ / bit / m2 are assumed. The energy consumed by a radio to run the transmitter and receiver electric circuit is given in equation (1) and (2). Energy consumed for data aggregation is given in equation (3).

The energy for data transmission is calculated by the equation,

$$ET(n,d) = n * E_{elec} + n* €_{amp} *d^2 \quad \ldots (1)$$

The energy required to receive n - bits is,

$$ER(n) = E_{elec} * n \quad \ldots (2)$$

Data aggregation energy is,

$$EAG = 5nJ/bit \quad \ldots (3)$$

All the three algorithms are implemented using MATLAB. The performance of the algorithms are analyzed using the evaluation metrics including running time and size of received data in the sink node based on the number of cluster head in the wireless sensor network. Objective of node clustering for WSN is achieved using data mining techniques i.e clustering and data aggregation. Voronoi based cluster head selection approach for data aggregation in WSNs are used for coverage aware energy efficacy

Table 1—Network model

Parameter	Values
Cluster Head	5,10,15
Population Size	400,800,1200
Network Grid	(0,0)-(1000,1000)
Sink node / Base Station	(1050, 1050)
Sensing Range	2m
Data Packet Size	100 bytes
Broadcast Packet	25 bytes
Packet Header Size	25 bytes
Round (Tcluster)	300 TDMA frames
Initial Energy	1 J/battery
Channel Bandwidth	1 Mbps
Crossover Type	One point
Mutation Rate	0.1
Selection Type	k-means
Crossover Rate	0.8

Comparative analysis of Voronoi based clustering algorithms for data aggregation

Analysis of Network Lifetime

Figure 1 exemplifies that the simulation time taken for the first node to die for network containing of 400 sensor nodes with cluster head is equal to 5. The proposed Voronoi based clustering algorithms show high performance during the first node failure when compared with existing algorithms. The round at which the first node dies will be decreased when the transmission range is large.

Analysis of Running Time

The running time is evaluated by keeping the number of sensor nodes (400) as constant and by varying the number of cluster heads each time. Figure 2 portrays the running time of all clustering algorithms. And it confirms that the V-FCM algorithm takes less time to execute when compared to the other algorithms though the execution time of

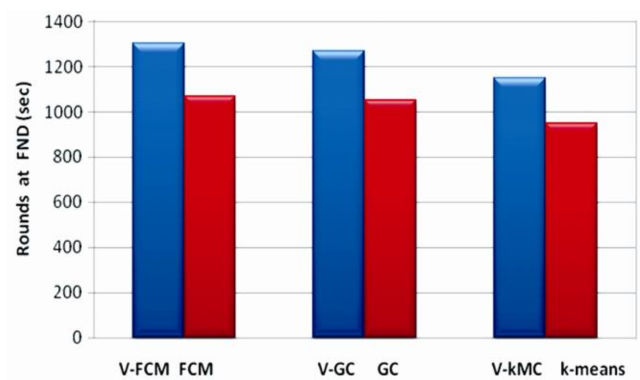

Figure 1—Round at which FND VS Algorithms

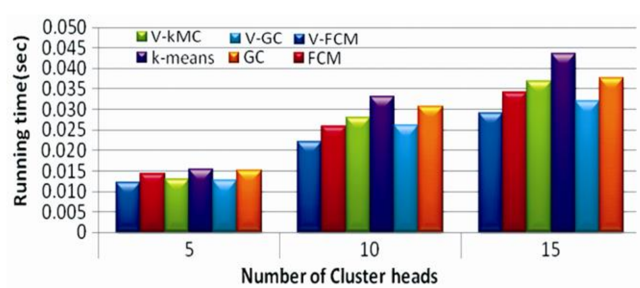

Figure 2—Running Time of Clustering Algorithms VS Cluster Heads

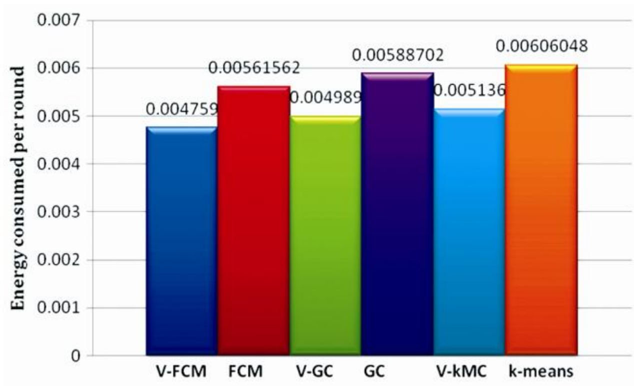

Figure 3—Energy Consumption per Round VS Clustering Algorithms

V-GC is minimized when compared to GC in the proposed method and V-GC produces the best global optimal clustering. Thus the running time of the all the algorithms are directly proportional to the number of cluster heads.

Analysis of Energy Consumption per Round

Figure 3 denotes that the amount of energy consumed by each algorithm for first round time. Energy consumption is increased when transmission range is increased (i) between sensor nodes or (ii) between sensor nodes to CH or (iii) between CH to BS. So the energy consumption for data aggregation is less than the energy consumption of data communication.

Conclusion

In this paper, comparison between different clustering algorithms for data aggregation using Voronoi diagram are classified and discussed. Algorithms used in this paper for comparative studies are VFCA, VkMC, and VGC. They are indeed robust because they reduce the number of distance calculation in clustering algorithms. Besides, they reduce network traffic and the channel contention. They also facilitate data aggregation for decreasing the network energy consumption. In these clustering algorithms, both the rotation of cluster heads and residual energy are sufficient to balance the energy consumption across the network. Furthermore, this research study analyzes the cluster head energy conservation effectively. The benchmark parameters such as number of nodes, network size, initial node energy, minimum energy, network's threshold and data transfer rate are used for well analyzing the coverage of each algorithm. The performance evaluation of the Voronoi based clustering algorithms shows that these algorithms decreases communication cost when compared to the existing clustering methods. Evidently, the results make it clear that i) Voronoi Fuzzy clustering is the best among the three algorithms implemented. ii) In VkMC and VFCM, the number of cluster heads should be known in prior. The final clusters obtained from Voronoi based *k*-means clustering are not adept at global optimization, they provide only local optimization. iii) Even though the best optimal clusters are obtained from VGC, it does not guarantee execution time.

References

1 A.Taherkordi R, Mohammadi & Eliassen F, A communication-efficient distributed clustering algorithm for sensor networks, *Proc Int Conf Adv Info Newt & Appl,* **1** (2008) 634-638
2 Ma X, Yang D, Tang T, Luo Q & D. Zhang, Online mining in sensor networks *Lect Notes Comput Sci,* **3222** (2005) 544-550
3 Younis O, Krunz K & Ramasubramanian S, Node clustering in wireless sensor networks: recent development & deployment Challenges, *J IEEE Networks*, **20** (2006) 20-25
4 Heinzelman W R, Chandrakasan A & Balakrishnan H, Energy-efficient communication protocol for wireless microsensor networks, *Proc. Int Conf System Sciences*, **8** (2000)10-15
5 Manjeshwar A & Agarwal D P, APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks, *Proc Int S Parallel & Distributed Processing*, **1** (2002) 195-202
6 Ghiasi S, Srivastava A, Yang A & Sarrafzadeh M, Optimal energy aware clustering in sensor networks, *J Sensors*, **2** (2002) 258-269
7 Shilpa Dhar, Energy aware topology control protocols for wireless sensor networks, *Master Thesis*, 2005
8 Khanna R, Liu H & Chen H H, Self-organization of sensor networks using genetic algorithms, *Int J Sensor Networks*, **1** (2006) 241-252
9 Kim J M, Park S H, Han Y J & Chung T M, CHEF: Cluster head election mechanism using fuzzy logic in Wireless Sensor Networks, *Proc Int Conf Adv Commun Techno,***2** (2008) 654–659
10 Patole J R & Abraham J, Design of MAP-REDUCE and K-MEANS based network clustering protocol for sensor networks, *Proc Int Conf Comp Comm and Net Tech*, **1** (2012)
11 Azad P & Sharma V, Cluster head selection in wireless sensor networks under fuzzy environment,
12 *J ISRN Sensor Networks*, **2013** (2013).
13 Nithyakalyani S & Suresh Kumar S, Data Relay Clustering Algorithm for Wireless Sensor Networks: A Data Mining Approach, J Comp Sci, **8** (2012) 1281-1284
14 Nithyakalyani S & Suresh Kumar S, An Approach to Data Aggregation in Wireless Sensor Network using Voronoi Fuzzy Clustering Algorithm , *J Sci Ind Res*, **72** (2013) 287-293
15 Nithyakalyani S & Suresh Kumar S, Energy Efficient Data Aggregation using Voronoi based Genetic Clustering Algorithm in WSN, *Int J Comp App*, **54** (2012) 37-41