

Collaborative Data Processing in WSN Using Voronoi Fuzzy Clustering

S. Nithya Kalyani, E. Sasikala B. Gopinath

S. Nithya Kalyani*

Department of Information Technology
K.S.R College of Engineering
Tiruchengode, India
*Corresponding author: nithyakalyani.me@gmail.com

E. Sasikala

Department of Information Technology
K.S.R College of Engineering
Tiruchengode, India

B. Gopinath

Department of Electrical and Electronic Engineering
Vivekanandha Institute of Engineering and Technology for Women
Tiruchengode, India

Abstract: In this paper, developed a novel Voronoi Fuzzy Clustering (VF) algorithm for energy efficient collaborative data aggregation in wireless sensor network. VF algorithm is fusion of Voronoi diagram and modified Fuzzy C- Means with respect to distance and Quality of Service. Here throughput, delay time and delivery ratio are considered as QOS parameters. Once clustering of sensor nodes is completed then data management technique such as data aggregation or compression is done for further decision making in sink node. Data mining clustering algorithm reduces overall transmission of data from each sensor to the sink node thus energy spent by individual sensor node is minimized. The cluster heads collects all sensed information from their respective cluster members and performs data aggregation or compression before transmitting the data to the sink node. Finally, the simulations are performed and the results are analyzed within the simulation set up to determine performance of the proposed algorithm in the sensor network. Our proposed approach has achieved 60% efficiency when compare with the K means algorithm.

Keywords: Clustering, Data aggregation, Compression, Voronoi fuzzy clustering algorithm, Energy, QOS, Throughput, Delivery ratio.

1 Introduction

Cluster aggregation is essential techniques that naturally reduce energy costs in wireless sensor networks (WSN) without compromising the quality of data delivery. The process of separating the sensor nodes into groups are called as clustering. A number of challenges involved in clustering. At first, the clusters themselves have to be identified. Secondly, cluster heads have to be chosen. Thirdly, routes have to be discovered from every node to their cluster head, and finally, the cluster heads have to proficiently relay the data to the sink node. This paper focuses all the later three problems. The foremost problem is defined by the application domain.

Data aggregation is another vital function in WSN to reduce the consumption of energy due to the number of limitation of WSN. The key idea of this process is to eliminating redundancy in data, minimizing the number of transmissions via integrate all the incoming data in cluster head from diverse sources and enroute it to the sink. This focuses in data-centric approach. Aggregation algorithms are limited to application requirement that is either in time or energy performance

A WSN consists of tiny sensing devices, which normally run on battery power. Sensor nodes are densely deployed in the region of interest. Each device has sensing and wireless communication capabilities, which enable it to sense and gather information from the environment and then send the data and messages to other nodes in the sensor network or to the remote base station [4]. Wireless sensor networks have been envisioned to have a wide range of application in both military and civilian domains [1]. Due to the limitation of sensor node researchers have designed lots of energy-efficient routing protocols to prolong the lifetime of sensor networks [2]. In this paper, the Fuzzy C-Means (FCM) clustering algorithm is modified by incorporating with Voronoi diagram for clustering the sensor node based on the distance and the Quality of Services.

2 VF Algorithm for Cluster Data Aggregation

The ultimate aim of the proposed approach is to reduce the consumption of the energy and increase the life time of WSN using collaborative data processing. In order to reduce the energy of the every node, clustering algorithm is used where the sensor nodes are clustered and subsequently, the cluster heads are selected. The cluster head is used to transmit the sensed data of sensor nodes to the base station (sink node), subsequently from the cluster head, the data is processed by aggregation or compression. In this approach, the process is separated into two phases.

2.1 VF algorithm to select the cluster head

The wireless sensor nodes are located at different places and the sensors are clustered using the Voronoi fuzzy clustering algorithm. Initially, apply the Voronoi tessellation to the sensor nodes based on the position and energy of the sensor. Subsequently use fuzzy clustering algorithm to cluster the nodes as it's easy to cluster and select the cluster head after applying the Voronoi. For finding the membership function, select any number of Voronoi cells as a cluster head and calculate the membership for every Voronoi cells with the assumed cluster head. The Voronoi cell goes to cluster which has highest value membership function. With the help of the membership function, the cluster head among the clusters are calculated.

Pseudo code:

Input:

No. of sensor $S = \{s_1, s_2, \dots, s_i\}$, Position and energy of every sensors $s_i = (x_i, y_j, z_k)$

Algorithm:

Step 1: Find the distance for each s_i with r : $d(s_i, r) = \sqrt{(x_s - x_r)^2 + (y_s - y_r)^2 + (z_s - z_r)^2}$

Step 2: $\{r | d(s_i, r) \leq d(s_j, r), i \neq j\}$, where $r =$ set of neighboring sensor

Step 3: $r = (r_1, r_2, \dots, r_n)$

Step 4: $V(s_i) = \frac{d(s_i, r_j)}{2}$

Step 5: Get the value of "c"

Step 6: Select cluster heads ch_y

Step 7: Find the member ship function $\mu_{xy} = \frac{1}{\alpha + \beta} [\alpha (D_{xy}) + \beta (Q_{xy})]$

Step 8: Find distance membership function using $D_{xy} = \left(\frac{\|v_x - ch_y\|}{\sum_{k=1}^c \|v_x - ch_k\|} \right)$

Step 9: Find QOS membership function using $Q_{xy} = \frac{\sum_{z=1}^Q v_x^y(z)}{(\sum_{k=1}^c \prod_{z=1}^Q v_x^k(z))}$

Step 10: Find max of μ_{xy} for each s_i

Step 11: $s_i \max(\mu_{xy})$

Step 12: Find $ch_y = \frac{\sum_{x=1}^N \mu_{xy}^m \cdot s_i}{\sum_{x=1}^N \mu_{xy}^m}$

- Step 13: Update ch_y
- Step 14: Go to step 5 until ch_y is placed in same $V(s_i)$
- Step 15: Each s_i send data $s_i(d)$ to corresponding cluster head ch_i
- Step 16: The cluster head ch_i receives $D = (d_1, d_2, \dots, d_i)$
- Step 17: Select Option 1. Data aggregation 2. Data compression
- Step 18: If (1)
- Select option a. Avg b. Max c. Min
 - a. $avg = \frac{\sum_{n=1}^n(d_i)}{N}$
 - b. $max = \max(d_i)$
 - c. $min = \min(d_i)$
- Step 19: If (2)
- a. Find probability of received data $P(d_i) = \frac{F(d_i)}{N}$
 - b. Sort the data into ascending order $P(D) = (P(d_1), P(d_2), \dots, P(d_i))$
 - c. Add the first two data $(P(d_1) + P(d_2))$
 - d. Merge the id of the data
 - e. Repeat the process 2 and 3 until getting only 2 probability of data
- Step 20: The result of (18) and (19) goes to sink node.

Let S sensor nodes is defined by $S = \{s_1, s_2, \dots, s_i\}$ and $(1 \leq i \leq n)$ is the total number of sensors in the wireless network. Each sensor node is defined by the position values (x coordinate and y coordinate) and the energy value (z). The position and the energy of the sensor nodes are given by $s_n = (x_i, y_j, z_k)$. A Voronoi diagram is a set of triangles, called Voronoi triangles. The Voronoi triangles generation is based on the distance and the energy of the sensor node. Find the neighbors of the each sensor node based on the minimum distance as in Step 2,3,4, where $d(s_i, r)$ is the distance from point s_i to r , predicted using Step 1. That is r is the set of sensors closer to s_i than to any other s_j .

2.2 Calculation of membership function

From the entire sensor nodes, clustering is performed in the sensor nodes based on the membership function. To find the membership function of each sensor node, let assume certain the number of sensor node as cluster head. In order to find the membership function, first choose 'c' cells as a Cluster Head (CH) that is $c = 2$ and assume any two cells as cluster head $y_1 = ch_1 = (x_{i1}, y_{j1}, z_{k1})$ and $y_2 = ch_2 = (x_{i2}, y_{j2}, z_{k2})$. With the help of assumed cluster head calculate the membership function of each sensor point. Based on the membership functions the nodes are clustered. In this paper, the membership function is attained based on two functions first one is based on the distance and second one is based on the QOS values. The equation in Step 7 is use for finding the membership function, where μ_{xy} is the membership function of x^{th} sensor node with respect to the y^{th} cluster head node, D_{xy} is the distance between x^{th} sensor node to the y^{th} cluster head node, Q_{xy} is the value of the QOS of x^{th} sensor node with respect to the y^{th} cluster head node, α is the weightage of distance (user defined), β is the weightage of QOS (user defined).

2.3 Distance based membership function

Based on the Euclidian distance between the cluster head and sensor node distance between the two points are calculated. This calculation describes, how much faraway the each sensor nodes from the each cluster head node. The Step 8 is used for calculating the distance based membership function.

Consider Table 1, which describes the positions of the each sensor node, form that select any number of the nodes as a cluster head first. Based on the selected cluster head, distance based membership function is found. Here, $C1 = (9.5, 1.3)$, $C2 = (8.8, 2.5)$ are select as a cluster head, based on that the distance based membership functions are calculated for other sensors. The distance based membership function using Step 8 for each sensor with respect to the $C1$ and $C2$ is calculated and the results are provided in Table 2.

Table 1: The positions of the sensor nodes

	x	y
V1	2.5	2
V2	4.1	1.3
V3	5.4	0.8
V4	9.5	1.3
V5	8.8	2.5

Table 2: Distance based membership of the sensor nodes

	V1(2.5,2)	V2(4.1,1.3)	V3(5.4,0.8)	V4(6.8,3.0)	V5(8.8,2.5)
V1(2.5,2)	0	0.2	0.3	0.4,	0.6
V2(4.1,1.3)	0.2	0	0.1	0.2	0.4
V3(5.4,0.8)	0.3	0.1	0	0.1	0.3
V4(6.8,3.0)	0.4	0.2	0.1	0	0.2
V5(8.8,2.5)	0.6	0.4	0.3	0.2	0

2.4 QOS based membership

Consider the following Table 3 that contains the values of the QOS of the each node with respect to each cluster head. Each of the sensor nodes has different quality parameter with respect to application where the sensors are applied. For this work QOS functions are throughput, delivery ratio, delay time are taken. Based on those QOS values of the nodes the cluster head can be obtained. Each of the node have the time slots at the beginning time, after generation of data packet the nodes are send to the cluster head within the next time slot, if the packets are not deliver in the particular time period then it marked as the delayed packet. The delay time is calculated by the ratio of the received time to the sending time of the node; the result of this subtracted by 1 that is a delay time.

Each of the nodes has the number of packets to send to the cluster head while sending the data packet to the cluster head some of the packets are not receive in the cluster head because of the interruption, noise. The ratio of number of received data packet in the cluster head to the number of sent data packet from the node.

Each of the nodes generates the data packets and sends it to the cluster head, flow of the data packet is varies from node to node with respect to the cluster head. The throughput of a node is calculated using number of packets sends to the cluster head in a particular time interval and it is different for each node with respect to the cluster.

The membership value based on the QOS values of the node with respect to the cluster heads are discovered by Step 9 and it is given in Table 4.

Table 3: The QOS values of each node with all possible cluster heads

	V1(2.5,2)	V2(4.1,1.3)	V3(5.4,0.8)	V4(6.8,3.0)	V5(8.8,2.5)
V1(2.5,2)	0,0,1	0.6,0.4,0.7	0.2,0.6,0.8	0.5,0.8,0.3	0.4,0.4,0.5
V2(4.1,1.3)	0.2,0.5,0.8	0,0,1	0.4,0.8,0.2	0.3,0.5,0.6	0.7,0.4,0.5
V3(5.4,0.8)	0.4,0.8,0.2	0.4,0.5,0.6	0,0,1	0.2,0.5,0.8	0.4,0.4,0.5
V4(6.8,3.0)	0.4,0.2,0.8	0.5,0.5,0.4	0.4,0.5,0.4	0,0,1	0.4,0.2,0.8
V5(8.8,2.5)	0.4,0.8,0.2	0.3,0.4,0.5	0.6,0.4,0.5	0.4,0.2,0.8	0,0,1

Table 4: The QOS membership of the nodes

	C1	C2
V1	0.55	0.44
V2	0.46	0.53
V3	0.48	0.51
V4	0.51	0.48
V5	0.45	0.54

With the help of D_{xy} and Q_{xy} values calculate the membership function using Step 7. Here equal weightage is given for both the distance and the QOS and calculated values are given in Table 5. The values of the weightage depends on the user. If the user wants to group the sensor nodes mainly based on distance then the user give the value of α is greater than β , if the user wants to group the sensor node based on the QOS then the value of β is greater than α .

Table 5: Sensor nodes membership function

	μ_{11}	μ_{12}
V1	0.535	0.455
V2	0.49	0.50
V3	0.50	0.48
V4	0.25	0.74
V5	0.72	0.27

Consider the two membership value μ_{xy_1}, μ_{xy_2} , where x be the sensor node, y_1 and y_2 be the cluster heads. If value of μ_{xy_1} minimum than μ_{xy_2} then the sensor node x moves to the cluster head y_2 else x move to y_1 . In Figure 1 dark shaded cells illustrates the assumed cluster head in a wireless sensor network. After finding the membership value for all Voronoi cells, each and every cell is clustered based on the two assumed cluster heads. After the selection of the cluster head and its members, the cluster head transmits the received sensed data from the members to the sink and this will lead to more of time and energy consumption in CH. To solve this problem, data management process is achieved.

2.5 Data Management process

Data management process involves data aggregation or compression. Data aggregation is a method which aims to reduce the localized congestion problem. It tries to gather knowledge from the sensed data by applying the maximum, minimum, average functions using Step 18.

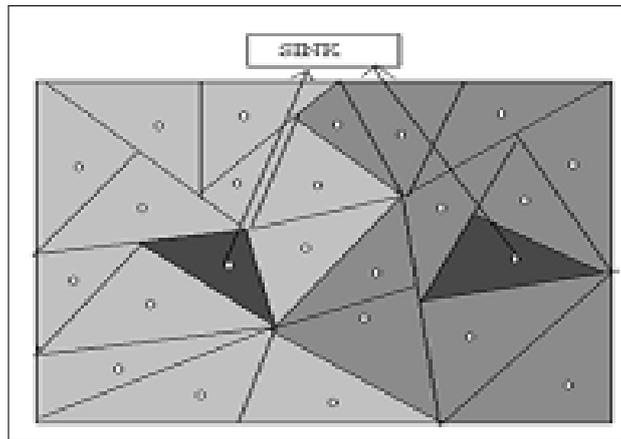


Figure 1: Transmission of aggregated or compressed data to the sink node

It then transmits only the useful information (aggregated data) to the sink node thus reducing congestion and its linked problems. Thus data aggregation helps the life of the wireless network to get increased. The cluster head sends the aggregated data to the sink node for the particular time interval. On the otherhand the collected sensed data are compressed in the cluster head by using the huffman code as in Step 19.

3 Experimental Results and Discussion

The experimental results of the proposed approach to cluster head selection is described in this section. The comparative analysis of the proposed cluster head selection approach and K-Means algorithm is presented. The proposed cluster head selection for data aggregation in wireless sensor network has been implemented using MATLAB and the performance of the proposed system is 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.

3.1 Evaluation of running time by varying the number of the sensor nodes

The VF algorithm is compared with the K-Means algorithm in order to evaluate the performance of the proposed VF algorithm. The following Figure 2 shows that the running time of the VF algorithm and K-Means algorithm. To evaluate the graph we change the number of sensor nodes at each execution and evaluate the time of the both algorithm, the value of the cluster head is constant for both algorithm to find the execution time. Here number of cluster heads is considered as 10 for all execution. The following Figure 2 shows that, VF algorithm needs less time to execute when compared with the K-Means algorithm. the reason behind is the if the membership function value of the sensor node is repeat in a single Voronoi cell then the execution of the membership function for that sensor node get stop since the running time of the VF algorithm get reduced.

3.2 Evaluation of running time by varying the number of cluster heads

Here, the running time is evaluated by keep the number of sensor nodes (1000) at constant and varying the number of cluster heads for each time. The following Figure 3 shows that the running time of VF algorithm and K-Means algorithm and when we see that graph we conclude the VF algorithm takes less to execute when compare to the K-Means algorithm. The running time of

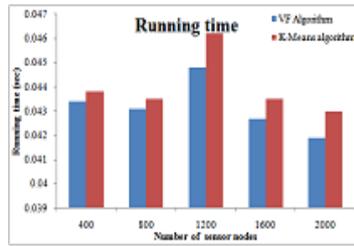


Figure 2: Running time VS No. of nodes

the both algorithm is directly proportional to the number of cluster heads. The calculation of the membership function for each point with respect to the each cluster head takes more time since the running time is increase when the number of cluster head increase.

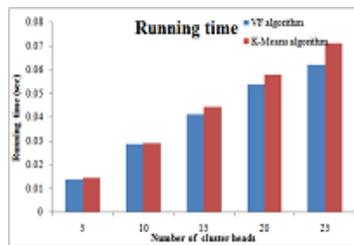


Figure 3: Running time VS No. of clustering head

3.3 Number of received data in sink node by varying the number of the sensor nodes

After the selection of cluster head, the each of the sensor node sends their sensed data to the cluster head. The cluster head received many data's from the sensor nodes, as per the user suggestion (data aggregation or data compression) the cluster head operate the received data and send that data to the sink node. The sink node receives the data from the cluster heads, the Figure 4 shows that the number of received data in a sink node for different number of sensor nodes. Based on the position of the cluster head, the received data of the sink node get varied since the position of the cluster head is different for both VF algorithm and K-Means algorithm. Hence, we take the number of received data in a sink node as a parameter. The following graph describes, the sink node receives more number of data's by using VF algorithm rather than using the K-Means algorithm.

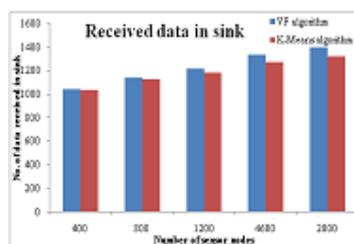


Figure 4: No. of data received in sink node Vs No. of nodes

3.4 Number of received data in sink node by varying the number of the cluster heads

The number of received data in a sink node is varies based on the number of cluster heads in a sensor network. Figure 5 shows that the number of received node in a sink node based on different number cluster heads for both VF algorithm and K-Means algorithm, the sink node receives more number of data's by using VF algorithm rather than using the K-Means algorithm.

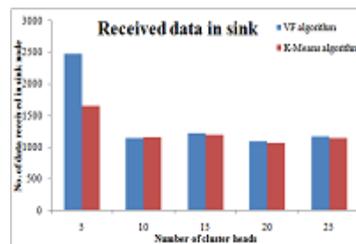


Figure 5: Size of data received in sink node Vs No. of cluster head

4 Conclusions

In this paper, we have presented the approach for energy efficient cluster algorithm for data summarization in wireless sensor network. In order to reduce the energy of the wireless sensor network, we have presented the efficient Voronoi fuzzy clustering algorithm. At first the Voronoi cells are applied for each node in the wireless sensor network based on the energy of the sensor node. Consequently, the cluster head is selected by the fuzzy clustering algorithm. Every node in the wireless sensor network sends their sensed data to the cluster head. The data aggregation and data compression process is done in the cluster head. The cluster head aggregate the received data by taking the value of minimum, maximum and average and uses the Huffman compression algorithm to compress the data and that compressed or aggregated data will send to the sink node. Finally, the experiment has carried out and results ensured that the efficiency (60%) is higher when compare with the K means algorithm.

Bibliography

- [1] Kulik, J.; Heinzelman, W; Balakrishnan, H. (2002); Negotiation-based protocols for disseminating information in wireless sensor networks, *Wireless Networks*, 8:69-185.
- [2] Xianghui Wang; Guoyin Zhang (2007); DECP: A Distributed Election Clustering Protocol for Heterogeneous Wireless Sensor Networks, *Computational Science*, 4489:105-108.
- [3] Kim, J.M; Park,S.H; Han,Y.J; Chung,T.M (2008); CHEF: cluster head election mechanism using fuzzy logic in Wireless Sensor Networks, in *International Conference of Advanced Communication Technology*, 654-659.
- [4] Mohammad Zeynali; Leili Mohammad Khanli; Amir Mollanejad (2009); TBRP: Novel Tree Based Routing Protocol in Wireless Sensor Network, *International Journal of Grid and Distributed Computing*, 2(4): 35-48.

- [5] Ye, M.; Li, C.F.; Chen, G.; Wu, J.(2005); EECS: An Energy Efficient Clustering Scheme in Wireless Sensor Networks, In *Proc. of the IEEE International Performance Computing and Communications Conference*, 535-540.
- [6] Tian, D.; Georganas, N.D.(2002); A Node Scheduling Scheme for Energy Conservation in Large Wireless Sensor Networks, *From Thesis: Multimedia Communications Research Laboratory, School of Information Technology and Engineering, University of Ottawa*.
- [7] Guru, S.M.; Steinbrecher,M; Halgamuge,S; Kruse,R.(2007); *Multiple Cluster Merging and Multihop Transmission*, LNCS 4459: AGPC, Springer, 89-99.
- [8] Qingchao Zheng; Liu,Z.; Liang Xue; Yusong Tan; Dan Chen; Xinping Guan(2010); An Energy Efficient Clustering Scheme with Self organized ID Assignment for Wireless Sensor Networks, *16th IEEE International Conference on Parallel and Distributed Systems*, DOI: 10.1109/ICPADS.2010.83, 635-639.