Chapter 127
Wireless Sensor Network Distributed Data Collection Strategy Based on the Regional Correlated Variability of Perceptive Area

Yongjun Zhang and Enxiu Chen

Abstract On account of the energy saving problem in wireless sensor network (WSN), this paper proposes the method that classifies sensing area according to the similarity of variability of sensing in terms of fuzzy clustering. Through the head of a cluster in a similar area to sample data to represent approximately collection of the regional sensing data. We through a test manifests data collection of partitioned similar sensing area will implement well in data monitoring and it will reduce the total energy consumption, reduce the computing task for data fusion and lengthen the WSN life circle.

Keywords Fuzzy clustering · Similar sensing area · Energy saving

127.1 Introduction

It is the hot research direction in wireless sensor network (WSN) that is how to utilize the energy of every sensor point effectively to prolong the life cycle of WSN as well as equilibrium of all energy in WSN. LEACH [1] is a typical hierarchical routing protocol. Later the scholarship of MIT Hein Zelman proposes the algorithm of LEACH-C [2] that is based on LEACH algorithm. It solves the problem that hasn't considered the reminder energy in a cluster head election to improve the quality of generated clusters. HEED [3] algorithm through a calculation to broadcast AMRP and to become temporary cluster head with the

Y. Zhang (✉)
Shandong Institute of Commerce and Technology, China National Engineering Research Center for Agricultural Products Logistics, Jinan, China
e-mail: zyj420177@163.com

E. Chen
National Engineering Research Center for Agricultural Products Logistics, Jinan, China

S. Li et al. (eds.), *Frontier and Future Development of Information Technology in Medicine and Education*, Lecture Notes in Electrical Engineering 269,
DOI: 10.1007/978-94-007-7618-0_127,

probability of initialization $CH_p = \max(C_p \times \frac{Er}{Em}, p_{\min})$, which select the head of a cluster in terms of reminder energy as first factor and communication in the network as a second factor. The algorithm of HEED is different to the algorithm of LEACH in criteria of cluster head selection and the mechanism of competition, it had improved in speed of clustering and pick more reasonable cluster head to transmit data by considering of inner cluster communication and reminder energy in order to make whole WSN more equilibrium. On the base of above mentioned viewpoints, there are someone propose a method to guarantee the optimized strategy of minimum energy cluster head selection [4, 5], which is based on twice the selection of cluster head and this method makes full consideration of energy parameter and hot zone of WSN. What's more, there are someone use K-MEANS [6, 7] to select a cluster head. These methods, however, are only considered the problem of energy and cost of routing without considering the similarity of sensing in some areas. This paper uses fuzzy cluster method to partition similar area of sensing and use HEED algorithm to select a cluster head to transmit sampling data to SINK point. Therefore, under the condition of request of accuracy and real-time is not high, we can partition similar area to reduce transmission quantity effectively, to reduce the computation of data fusion [8, 9], to cut down the total energy consumption and finally to prolong the life cycle of the whole network.

127.2 Concerned Question Description and Definition

For the area of sensing, the sensing data may have similar change regulation. We can use the similarity of change regulation of sampling data to utilize fuzzy clustering to partition the whole area into several similar areas. In every similar area, we can use HEED algorithm to select a cluster head and utilize it as representation to monitor neighbourhood variation. When inner similar area's real sensing data deviate from the data sampled by cluster head, which exceed threshold *MaxError*. The system should reparation and select the head of clusters to collect the district sensing data. In there, sensing vector is $X = \{p^{(1)}, \ldots, p^{(i)}, \ldots, p^{(n)}\}$ which denote the n sensing data with m sensing results from time t_x to t_y. For example, vector with order number is $p_i = \{p_1^{(i)}, \ldots, p_m^{(i)}\}$, the raw sensing data can be expressed by matrix $\underset{T\in[t_x,t_y]}{SensoringMatrix}[n][m]$.

Definition 1 sensing similarity is $\underset{p_i,p_j\in X}{SimSensoring}(p_i, p_j)$: suppose there are two sensor points p_i and p_j in some area *SimArea* and for the given sensor point set $X = \{p^{(1)}, \ldots, p^{(n)}\}$ has no objects with m dimensions, we define similarity of sensing as follows:

$$SimSensoring(p_i, p_j) = \sigma d_s(p_i, p_j) + (1 - \sigma)\left|\cos(p_i, p_j)\right| \qquad (127.1)$$

Definition 2 max error of sensing *MaxError*: as a classified similar area *SimArea*, we can find two sensor points p_i and p_j they have max dissimilarity, so the *MaxError* in this area satisfy that it $\underset{p_i, p_j \in SimArea}{Min}\{SimSensoring(p_i, p_j)\}$ is true it implicates value $\frac{\sqrt{\frac{1}{m}\sum_{k=1}^{m}(p_i - p_j)^2}}{Max\{p\} - Min\{p\}}$ is close to α, α is max tolerance of real measurement error and range of *MaxError* is (0,1).

Definition 3 similar sensing area (SSA): SSA is a set of sensing points that satisfy the condition $A_j(P_i) = \{P_i | Min\{SimSensoring(P_i, C_j)\} > \delta\}$, C_j is a regional centre of $A_j(P_i)$, so the whole sensing area can be expressed by $sensorArea = \bigcup_{j=1}^{N} A_j(P)$ and for $\forall A_i(P), \forall A_j(P) \in sensorArea$ there is the expression $A_i(P) \cap A_j(P) = \phi$ is correct.

127.3 Distributed Data Collection Strategy

Data collection in the similar sensing area in this paper we can concisely define as SSADC. Firstly, WSN becomes ad hoc network use hierarchical method to collect data from sensing area. Next, using fuzzy clustering to partition data area by using the changing of value of points. Finally, we can use HEED algorithm to select the head of cluster in certain partitioned area. It can represent changing of data by sampling the monitoring areas' data. The criteria for electing of cluster head is to take reminder energy as main factor and to take the density of sensing points as the second factor. With the passage of time, similar area may be changed more or less. Therefore when the sensing error exceeds a certain value, which should do a new fuzzy clustering and reselecting the head of cluster to finish the job of data collection. The algorithm of data collection by using fuzzy clustering as follows:

Method:

Step 1 During the period $[t_x, t_y]$, through statistic we record the sensing data by X, $p_i \in SSA$ (similar sensing area);

Step 2 As for vectors $X = \{p_i^{(j)}\}\big|_{1 \le i \le m, 1 \le j \le n}$ that with n elements and each has m dimensions, for standardized $p_i^{(k)}$ that can be standardized by $p_i^{(k)} = \frac{p_i^{(k)} - \underset{1 \le i \le m}{Min}\{p_i^{(k)}\}}{\underset{1 \le i \le m}{Max}\{p_i^{(k)}\} - \underset{1 \le i \le m}{Min}\{p_i^{(k)}\}}, 1 \le k \le n,$ to compact data into (0,1) and store the result in $M_{n \times m}$;

Step 3 According to the sensing similarity $SimSensoring(p_i, p_j)$, compute the fuzzy similar Matrix $SimMatrix[n][n]$;

Step 4 For the given level of similarity λ, through the maximum tree algorithm of fuzzy clustering to get clustering result $C = \{c^{(1)}, \ldots, c^{(k)}\}$;

Step 5 If for every clustering result there exist two max dissimilar points that satisfy $\frac{\sqrt{\frac{1}{m}\sum_{k=1}^{m}(p_i - p_j)^2}}{Max\{p\} - Min\{p\}} < \alpha$, then set $\lambda = \lambda + h$ (h is step length) and go to step (4) to redo cluster analysis;

Step 6 For every $c_i \in C$ in clusters result C, we use HEED method to elect the head of cluster in every c_i to sample data and transmit it;

Step 7 At an interval of time β, the system will check periodically if there exist two max dissimilar points $\underset{p_i, p_j \in SimArea}{Min}\{SimSensoring(p_i, p_j)\}$ it satisfies $\frac{\sqrt{\frac{1}{m}\sum_{k=1}^{m}(p_i - p_j)^2}}{Max\{p\} - Min\{p\}} > \alpha$, then go to Step 1 to recompute similar area;

Step 8 The end of this algorithm.

For the convenience of data querying, we build a hash table $I(C)$ to index the final clustering result. In the similar sensing area to take HEED algorithm to select the head of clustering for balance the energy and represent a similar areas' sensing data.

127.4 Experimental Results and its Analysis

We use MATLAB7.0 to analyse 1100×100 m agriculture plant area sensing data. In the monitoring area we randomly and uniformly laid out 300 soil humidity sensors, the collecting point SINK locates in coordination (0,100). We deploy the WSN to compare the performance of our method for collecting data to LEACH, LEACH-C and HEED. Suppose the sensor's initial energy is 0.5 J, $E_{elec} = 50$ nJ/bit, $\varepsilon_s = 0.\ 0013$ pJ/(bit.m4), $E_{com} = 5$ nJ/bit, the length of the message is 500 Bytes, the delay of sending and receiving is 25 μs. The test length is 1200 s and the intervals to repartition the similar area is 120 s.

Through the experiment in Fig. 127.1 we can see HEED method forming speed of the cluster has certainly improved compared to LEACH and LEACH-C, which thought the cost of the inner cluster communication after forming a cluster and select a cluster head that is more suitable to do the task of data transmit. Because this paper also uses HEED to choose the head of cluster in similar area it will be optimized in each partitioned field. What's more, we utilize cluster head to transmit sensing data to represent regional data it will save a lot of cost of energy in data transmission.

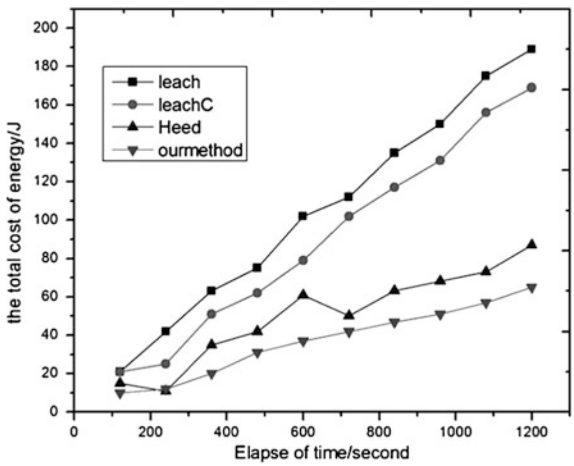

Fig. 127.1 The comparison of 4 cluster head selection methods in total energy cost

127.5 Conclusion

Through the fuzzy clustering method to split the whole monitoring area into many similar sensing areas. As for each similar sensing region we elect the head of cluster by using HEED algorithm to sampling data to represent the areas' sensing data. By using a distributed data collection method to collect monitoring data with less quantity of communication, to decrease the cost of energy in the whole network, to decrease data fusion calculation to improve the reliability of data, finally we can prolong the life cycle of WSN.

References

1. Heinzelman WR, Chandrakasan A, Balakrishnan H (2000) Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of the 33rd Hawaii international conference on system sciences
2. Heinzelman W, Chandrakasan A, Balarkrishnan H (2002) An application-specific protocol architecture for wireless microsensor networks. IEEE Trans Wireless Commun 1(4):660–670
3. Younis O, Fahmy S (2004) Distributed clustering in ad-hoc sensor networks:a hybrid. Energy-efficient approach, Ieee transactions on mobile computing 3:366–379
4. Peng W, Edwards DJ (2010) K-means like minimum mean distance algorithm for wireless sensor networks [C] 2010. In: 2nd international conference on computer engineering and technology IEEE pp120–124.
5. Zheng J, Wang P, Li C (2010) Distributed data aggregation using Slepian-wolf coding in cluster-based wireless sensor network. IEEE Trans Veh Technol 59(5):2564–2574
6. Chen HF, Mineno H, Mizuno T (2008) Adaptive data aggregation scheme in clustered wireless sensor networks. Comput Commun 31(15):3579–3585
7. Nurdin HI, Mazumdar RR, Bagchi A (2009) Reduced-dimension linear transform coding of distributed correlated signals with implete observations. IEEE Trans Inf Theory 55(6):2848–2858

8. Jiang HB, Jin SD, Wang CG (2011) Prediction or not? An energy-efficient framework for clustering-based data collection in wireless sensor network. IEEE Trans Parallel Distrib Syst 22(6):1064–1071
9. Biswas P, Qi H, Xu Y (2008) Mobile agent-based collaborative sensor fusion. Inf Fusion 9(3):399–411