Evaluation of Clustering Algorithms for Cluster Heads Selection

Dong-Chul Park

***Abstract*—**Unsupervised competitive learning algorithms for clustering of sensor nodes in wireless sensor networks are evaluated with a large scale data set in this paper. The Centroid Neural Network (CNN) is compared with Fuzzy c-Means (FCM) algorithm in determining cluster heads among given sensor nodes. The cluster heads are combined with Low Energy Adaptive Clustering Hierarchy (LEACH) for minimizing battery (energy) consumption in sensor nodes. Initial results show that the CNN can be effectively used for the selection of cluster heads and reduces total energy consumption of the network.

***Keywords*—**centroid neural network, sensor nodes, cluster head, Fuzzy c-Means.

I. Introduction

THE objective of clustering algorithms is to group of similar objects in applications. Various methods have been proposed for data clustering [1]-[4]. Among these methods, k-means algorithm and Self Organizing Map (SOM) [4] have been most widely used in practice because of their simplicity. Fuzzy C-Means (FCM) clustering algorithm has been shown to be advantageous over traditional clustering approaches based on crisp clustering [2]. Lloyds batch k-means algorithm and MacQueens adaptive k-means algorithm are considered as the basis for the SOM [4]. Basically, SOM finds a neuron which is the closest to a given input data and updates the location of the neuron and its neighbors. Parameters for the initial learning coefficient and the total number of iterations for a given set of data should be determined in advance. However, it is not easy to determine a priori the best set of parameters for a given set of data.

The centroid neural network (CNN) is based on the observation that synaptic vectors converge to the centroids of clusters as learning proceeds in conventional unsupervised competitive learning algorithms such as SOM [5]-[7]. The centroid can minimize the mean-squared error of the vector quantization. One of the advantageous features in CNN algorithm is that the CNN does not require any parameters such as the initial learning coefficient and the total number of iterations. The CNN calculates its learning coefficients in each representation of datum. The CNN can also reward and punish by learning coefficients for winners and losers, respectively. CNN does not require the total number of iteration in advance, either

Sensor nodes in Wireless Sensor Network (WSN) collects various information including remaining battery life and transmits collected information to the base station. WSN requires an effective routing protocol. One of the main objectives for various routing protocols is minimizing total energy consumption at sensor nodes[8]-[12]. LEACH can minimize the energy consumption by selecting the cluster heads stochastically [10]. However, this stochastic selection of cluster heads can be improved by using computational intelligence methods [11]-[12]. This paper addresses a comparative study on the application of FCM and CNN to selecting cluster heads for WSN

The rest of this paper is organized as follows: Section II summarizes FCM, CNN, and LEACH. Section III presents experiments and results on several data sets for performance evaluation between FCM and CNN. Section IV concludes this paper.

II. Related Works

A. Fuzzy C-Means Algorithm

The FCM algorithm has been successfully applied to various clustering problems. The FCM algorithm attempts to partition a finite collection of elements $\vec{X} = \{\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_N\}$ into a collection of C fuzzy clusters. Bezdek first generalized the fuzzy ISODATA by defining a family of objective functionsJ_m, $1 < m < \infty$, and established a convergence theorem for that family of objective functions [2]. The objective function for FCM is defined as :

$$J_m(U, \vec{v}) = \sum_{i=1}^{C} \sum_{k=1}^{N} \mu_{ik}^m \, |\vec{x}_k - \vec{v}_i|^2 \tag{1}$$

where $|\cdot|^2$ denotes Euclidean distance measure, $\vec{x}_k$ and $\vec{v}_i$ are the input data, k, and cluster prototype, i, respectively. μ_{ik} is the membership grade of the input data $\vec{x}_k$ to the cluster $\vec{v}_i$, and m is the weighting exponent, $m \in \{1, \cdots, \infty\}$, while N and C are the numbers of input data and clusters, respectively.

The objective function for FCM is minimized when higher membership grades are assigned to objects which are closer to their centroid [2]. By applying the Lagrange multiplier to minimize the objective function, the center prototypes and membership grades can be iteratively updated as follows:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{C} \frac{\|\vec{x}_k - \vec{v}_i\|^2}{\|\vec{x}_k - \vec{v}_j\|^2}} \tag{2}$$

$$\vec{v}_i = \frac{\sum_{k=1}^{N} \mu_{ik}^m \vec{x}_k}{\sum_{k=1}^{N} \mu_{ik}^m} \tag{3}$$

The FCM finds the optimal values of cluster centers iteratively by applying Eq. (2) and Eq. (3) in an alternating fashion.

Dong-Chul Park is with the Department of Electronics, Myongji University, YongIn, Rep. of KOREA (phone: +82-31-330-6756, fax: +82-31-3306977, parkd@mju.ac.kr)

B. Centroid Neural Network

The Centroid Neural Network (CNN) is an unsupervised competitive learning algorithm [5]. It finds the centroids of clusters at each presentation of the data vector. The CNN introduces the winner neuron and the loser neuron. When a datum x is given to the network at the epoch (k), the winner neuron at the epoch (k) is defined as the neuron with the minimum distance to x. The loser neuron at the epoch (k) to the datum x is defined as the neuron that was the winner of x at the epoch $(k-1)$ but is not the winner of x at the epoch (k). The CNN updates its weights only when the status of the output neuron, (winner, loser, or neutral), has been changed when compared to the status from the previous epoch.

When an input vector x is presented to the network at iteration n, the update equations for winner neuron j and loser neuron i in CNN can be summarized as

$$\mathrm{w}_j(n+1) = w_j(n) + \frac{1}{(N_j+1)}[x(n) - w_j(n)] \quad (4)$$

$$\mathrm{w}_i(n+1) = w_i(n) - \frac{1}{(N_i-1)}[x(n) - w_i(n)] \quad (5)$$

where $\mathrm{w}_j(n)$ and $\mathrm{w}_i(n)$ are the winner and loser neurons with N_i and N_j data, respectively.

The CNN has several advantages over conventional clustering algorithms. Unlike SOM, the CNN requires neither a predetermined learning gain schedule nor the total number of iterations for clustering. It converges to sub-optimal solutions while conventional algorithms such as SOM may give unstable results depending on the learning gain schedule and the total number of iterations. A pseudo code for CNN is given in Fig. 1. More detailed description on the CNN can be found in [6]-[8].

C. Low-energy adaptive clustering hierarchy

Low Energy Adaptive Clustering Hierarchy (LEACH) is a communication protocol which is combined with clustering and routing protocol in wireless sensor network. In order to improve the life time of sensor nodes in a WSN, LEACH intends to minimizes the total energy consumption. LEACH protocol divides nodes into groups and each group has one cluster head (CH). CH collects data from its node members and propagates the acquired information to the base station. This protocol adopts a stochastic process in determining CHs among sensor nodes. Note that all the nodes in the cluster are selected with the equal probability. After selecting CHs, CHs transmits

```
Algorithm CNN(C,D)        [C : number of clusters, D : number of data vectors]
  [initialize weight w1 and w2]
  Find the centroid, c, of all data vector
  Initialize w1 and w2 around c with small ε:
  w1 := c+ε, w2 := w2 −ε
  k := 2, epoch := 0
  for (k ≤ C)
    do
      loser := 0
      for (n = 1) to D
        Apply a data vector x(n) to the network
        Find the winner neuron, j, using Divergence distance for 1 ≤ j ≤ k.
        if (epoch ≠ 0) then Set i is winner neuron, i, for x(n) in previous epoch.
        if (i ≠ j), then neuron i, is loser neuron.
        if (epoch = 0 or i ≠ j)
          Run UpdateCNN-WeightMean(wj, wi, epoch)
          loser := loser +1
        endif
      endfor                          [check for all data]
      epoch := epoch + 1
    while loser ≠ 0
    if k ≠ M
      split the most erroneous group, j, by adding a small vector, ε, nearby group j
      wj+1 = wj + ε
    endif
    k := k+1
  endfor
end

Procedure UpdateCNN-WeightMean(wj, wi, epoch)
  [ Update the winner neuron, wj, and neuron, wi]
  Update winner neuron : wj(n+1) = wj(n)+(x(n)−wj(n))/(Nj+1)
  if epoch ≠ 0                    [loser neuron is occurred only when epoch ≠ 0]
    Update losing neuron : wi(n+1) = wi(n)−(x(n)−wi(n))/(Ni−1)
  endif
end
```

Fig. 1 A pseudocode of Centroid Neural Network [6]

an information about the CH nodes to other non-CH nodes [10].

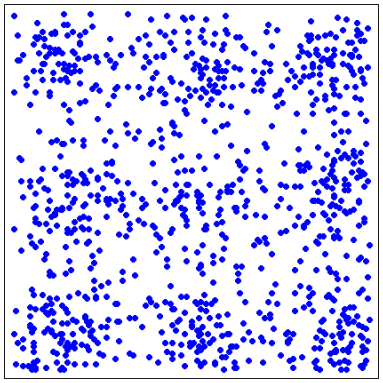
LEACH shows improvements on minimizing the energy consumption of wireless sensor nodes by simply selecting the cluster heads stochastically, there is a room for improvement because the CH selection process is rather random and it ignores physical locations of sensor nodes. We may conclude that the total energy consumption required for CHs and cluster members may be larger than an optimal case when the CHs are the centers of sensor nodes in a cluster. In order to improve this situation, FCM and CNN are applied to cluster selection process. By selecting CHs near the center of clusters, the energy is optimized locally in each group and thus the energy consuming can be decreased in the whole networks rather than randomly selecting the clusters as in LEACH.

III. Experiments and Results

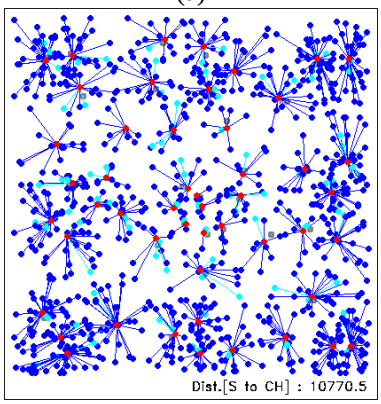
For experiments, the same setup as in LEACH is used [10]. We randomly generate 1,000 sensor nodes that locate at (x, y) in a certain node range $0 \le x \le 200$ and $0 \le y \le 200$ and BS locate at (100,350). Each data message is 500 bytes over 1 Mb/s of bandwidth. The header for each packet is 25 bytes. The cost of communication energy when transmitting a message with size of l bit through d distance is evaluated as follows:

$$E_{Tx}(l,d) = \begin{cases} lE_{elec} + l\varepsilon_{fs}(d^2), & d < d_0 \\ lE_{elec} + l\varepsilon_{mf}(d^4), & d \ge d_0 \end{cases} \quad (6)$$

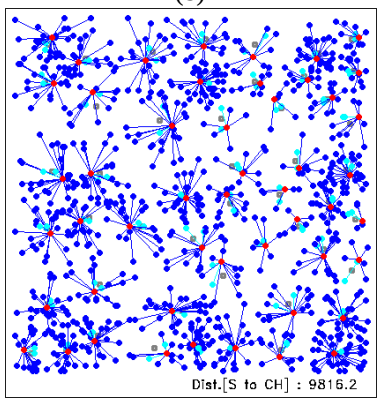
and for a received message:



(a)



(b)



(c)

Fig. 2 Examples of experiment data and results: (a) location of sensor nodes, (b) results of cluster head locations for FCM, and (c) CNN (red dots: current CHs, light blue dots: past CHs, blue dots: never been CHs)

$$E_{Rx}(l) = E_{Rx-elec(l)} = lE_{elec} \quad (7)$$

where d_0 is a threshold distance. ε_{fs} and ε_{mf} are the amplifier energy respect to the free space model or the multipath model. E_{elec} is the electronics energy. In our experiments, these parameters are set as: E_{elec} = 50 nJ$/bit$, ε_{mf} = 0.0013 $pJ/bit/m^4$, ε_{fs} = 10 $pJ/bit/m^2$ and d_0= 86.202. The same parameter set is also used in [10],[12].

Fig. 2 shows the location of sensor nodes and CHs and grouping information for FCM and CNN. In Fig. 2, the selected CHs are in red dots while light blue dots represent the node that

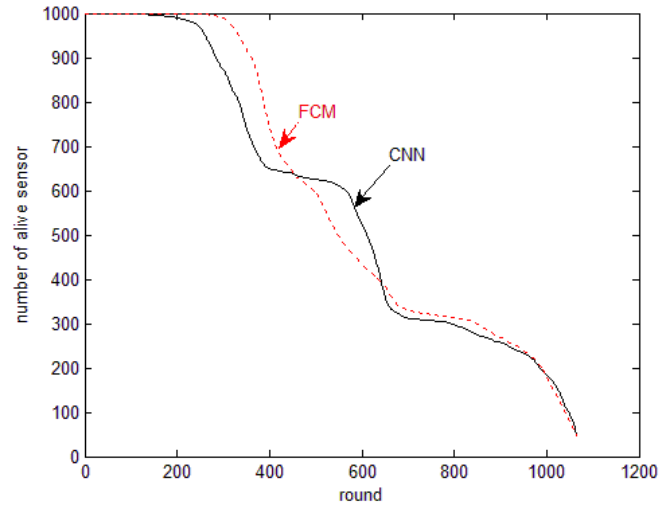


Fig. 3 Comparison of the numbers of alive sensors after rounds for the protocols based on FCM and CNN.

has been selected as CHs in previous rounds. Blue dots are the nodes that never been selected as CHs in previous rounds and are the candidates for future CHs. The total distances from the chosen CHs to the nodes in their groups in these instances are 10,770.5 and 9816.2 for FCM and CNN, respectively. This clearly shows the difference of energy consumptions required for different protocols. Similar results are obtained in different sensor node distributions.

Another important criterion in evaluating different routing protocols is the numbers of alive sensors after rounds. In our experiments, the number of alive sensors are calculated each round and shown in Fig.3. The results shows that FCM-based protocol gives a favorable performance over CNN-based one in terms of the first death round while CNN-based protocol gives a favorable performance over FCM-based one in terms of the half death round.

Another important criterion in evaluating different routing protocols is the computational speed. In order to evaluate speed for selection process of cluster heads in FCM and CNN, the CPU times are calculated under the following computing environment:

- CPU: Intel(R) Core(TM) i5-2400, 3.1㎓ CPU,
- RAM size: 2 GB, OS: Window 7 Enterprise, 64bit.

Table I summarizes the clustering time required for different protocols. Note that LEACH does not require clustering time because it selects CHs randomly. On average, FCM requires 1,200.4 mS while CNN finds clusters in 579.2 mS. This clearly shows that CNN reduces the clustering time about 52% when compared to FCM.

TABLE I
AVERAGE CPU TIME REQUIRED FOR CLUSTERING (SEC)

Data	Gaussian
FCM	1.200
CNN	0.579

IV. CONCLUSIONS

A comparative study on the application of clustering algorithms to clustering sensor nodes in wireless sensor networks is reported in this paper. The clustering algorithms

used in this paper are Fuzzy C-Means algorithm and Centroid Neural Network. The protocols based on clustering algorithms find clusters of sensor nodes optimally by adopting Fuzzy C-Means algorithm and Centroid Neural Network and follow the selection method for cluster heads from LEACH algorithm. Experiments are performed on example platform used in LEACH algorithm in order to evaluate performances of the different algorithms. A sensor network map with 1,000 nodes is designed for experiments and evaluated the number of nodes alive after a period. The results show that the CNN shows very compatible performance with FCM. The computing speeds for the selection of cluster heads for each round are also evaluated for FCM and CNN. The results show that the CNN can save the clustering time required for the selection of cluster heads about 52%. From experiments and results, we can conclude that the CNN can be more effective than conventional FCM in terms of total energy consumption of the network and system life time while the CNN can find its cluster heads 2 times faster than FCM.

ACKNOWLEDGMENT

This work was supported by the IT R&D program of The MKE/KEIT (10040191, The development of Automotive Synchronous Ethernet combined IVN/OVN and Safety control system for 1Gbps class). The author would like to thank E.-A. Park and Sang-Chul Lee and members of ICRL for their help in preparing this manuscript.

REFERENCES

[1] J. Hartigan, *Clustering Algorithms*. New York, Wiley, 1975.
[2] J.C. Bezdek, "A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.PAMI-2(1), 1980, pp. 1-8.
[3] Q. Liang and J.M. Mendel, "MPEG VBR Video Traffic Modeling and Classification Using Fuzzy Technique," *IEEE Trans. Fuzzy Systems*, vol. 9(1), 2001, pp. 183-193.
[4] T. Kohonen, "The Self-Organizing Map," *Proc. IEEE*, vol. 78(9), 2002, pp. 1464-1480.
[5] Dong-Chul Park, "Centroid Neural Network for Unsupervised Competitive Learning," *IEEE Trans. Neural Networks*, vol. 11(2), May, 2000, pp.520-528.
[6] Dong-Chul Park and Y. Woo, "Weighted centroid neural network for edge reserving image compression," *IEEE Trans. Neural Networks*, vol. 12(11), Mar. 2001, pp.1134-1146.
[7] Dong.-Chul Park, O.-H. Kwon, and Jio Chung, "Centroid neural network with a divergence measure for GPDF data clustering," *IEEE Trans. Neural Networks*, vol. 19(6), June 2008, pp. 948-957.
[8] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Communications*, vol.1(4), Oct. 2002, pp. 660-670.
[9] M. Sharma and K. Sharma, "An Energy Efficient Extended LEACH," *in Proc. IEEE CSNT Conference*, May 2012, pp. 377-382.
[10] Asaduzzaman and Hyung-Yun Kong, "Energy Efficient cooperative LEACH Protocol for Wireless Sensor Networks," *IEEE Trans. Communications and Networks*, vol. 12(4), Aug, 2010, pp. 358-365.
[11] C.P. Subha, S. Malarkan, and K. Vaithinathan, "A Survey On Energy Efficient Neural Network Based Clustering Models In wireless Sensor Network," *in Proc. IEEE ICEVENT Conference*, Jan, 2013, pp. 1-6.
[12] E. Park , K. Lee, D. Park , and S. Min, "Centroid Neural Networks for Wireless Sensor Networks," *in Proc. of WCSN*, Dec. 2013.

Dong-Chul Park (M'90-SM'99) received the B.S. degree in electronics engineering from Sogang University, Seoul, Korea, in 1980, the M.S. degree in electrical and electronics engineering from the Korea Advanced Institute of Science and Technology, Seoul, Korea, in 1982, and the Ph.D. degree in electrical engineering, with a dissertation on system identifications using artificial neural networks, from the University of Washington (UW), Seattle, in 1990. From 1990 to 1994, he was with the Department of Electrical and Computer Engineering, Florida International University, The State University of Florida, Miami. Since 1994, he has been with the Department of Electronics Engineering, MyongJi University, Korea, where he is a Professor. From 2000 to 2001, he was a Visiting Professor at UW. He is a pioneer in the area of electrical load forecasting using artificial neural networks. He has published more than 130 papers, including 40 archival journals in the area of neural network algorithms and their applications to various engineering problems including financial engineering, image compression, speech recognition, time-series prediction, and pattern recognition. Dr. Park was a member of the Editorial Board for the IEEE TRANSACTIONS ON NEURAL NETWORKS from 2000 to 2002.