Two-Level Clustering Hierarchies using Fuzzy clustering in Wireless Sensor Networks

OSAMAH MASHAQBEH [1], KHALED BATIHA [1], WAFA ALSHARAFAT [2]
[1] Department of Computer Science, Al al-Bayt University, Mafraq, 25113, JORDAN
[2] Department of Information Systems, Al al-Bayt University, Mafraq, 25113, JORDAN

Abstract: Fuzzy algorithms are highly regarded for their simplicity, efficiency, and rapid implementation. They play a crucial role in clustering and classification tasks, as seen with C-means and K-means algorithms. In this paper, we introduced a two-level clustering hierarchy that uses fuzzy clustering techniques. The proposed work proved more effective than well-established methods such as K-means and C-means in handling nonlinear network clusters. We evaluated our proposed work using iris datasets and randomly generated datasets on a multicore system. The outcomes demonstrated that our research techniques yielded adequate performance results for both datasets compared to the K-means and C-means methods.

Keywords: Fuzzy clustering, Head cluster, iris, Wireless Sensor Network.

Received: July 15, 2023. Revised: October 29, 2023. Accepted: December 29, 2023. Published: March 5, 2024.

1. Introduction

Clustering is a useful technique for collecting and organizing similar data into groups to solve communication limitations. Besides network environment, various fields have applied clustering algorithms such as biology, image processing, speech recognition, and psychology. These clustering algorithms differ in clustering methods that depend on density, priority, and grid. According to that, clustering algorithms can be divided into two categories: those that start from an overall perspective, such as the CURE algorithm[1], and those that begin with individual elements, such as the CHAMELEON algorithm[2].

Wireless Sensor Networks (WSNs) are autonomous sensor systems distributed by space with freely moved nodes and can be selected as a host or router. Every node in WSN is managed by a central node called a head cluster (CH). Numerous models have been proposed for WSNs to select CH according to certain characteristics as in [2]. The appropriate selection of CH will enhance the WSN's performance and effectiveness. Thus, CH can be selected according to its geographical coverage or cost-effectiveness. In WSN, one CH or more can be selected. If one CH is selected, it is responsible for collecting and processing data from all nodes in the network. In the case of multiple CHs, each CH collects and processes data for a nearby group of nodes.

The hierarchical cluster structure for a WSN consists of two layers. The first layer consists of several CHs which are responsible for collecting data from neighboring nodes. In the second layer, the process of selecting CH is based on several factors such as power consumption, network topology, and optimization objectives. Data aggregation at the CH level, can be used to reduce data traffic on the network.

Accordingly, this paper focuses on using Fuzzy clustering to create two-level cluster hierarchies that manage efficiently CHs.

2. Related Work

Different classification techniques have been used to cluster nodes where each node belongs to a specific group of nodes under CH's responsibility. Fuzzy clustering was developed by Dunn in 1973. Fuzzy clustering concerns about assigning each node to multiple clusters. Also, this technique was improved by Bezdek in 1981[4]. In this paper, fuzzy clustering by the Local Approximation of Memberships algorithm (FLAME) is applied for clustering nodes.

Clustering concerns about creating groups of nodes, referred to as clusters. Each cluster has more similar nodes than other cluster nodes. Clustering is also known as data segmentation and separating large datasets and nodes into similar clusters, depending on their similarity. Various research methods have been used to investigate different cluster models, such as connectivity clusters and clusters based on centroids. Connectivity-based clustering is also known as hierarchical clustering which

groups nodes based on their distance. In centroid-based clustering, each cluster is represented by a single central vector. The best example of this model is k-means clustering, which provides a formal definition of problem optimization depending on finding the k-centers of the assigned data. Hard clustering model: each node will either belong to a cluster or will not belong to any cluster. Fuzzy clustering is the shape of clustering in which every node in the network can be assigned to multiple clusters, whereas in regular clustering, each node is assigned to only one cluster. Cluster is identified as an application requirement or data that is used in WSNs, and these measures can be distance measurement, density, and connectivity. The fuzzy term was used with the 1965 suggestion proposal by Lotfi Zadeh [5].
Many researchers have used the fuzzy method as an artificial intelligence technique, accompanied by the k-nearest algorithm, to make clustering efficient in WSN environments.

Limin Fu et al. [6] performed a Fuzzy clustering by Local Approximation of Memberships (FLAME) on DNA microarray data. They divided the approach into two main categories; first, defining the (gene or cluster) neighbor and identifying objects with "archetypal" called CSO (Cluster Supporting Objects) and constructing clusters; second, assigning each object by its membership of its neighboring objects, the membership spreads from the CSO to all neighbors. The results showed that FLAME performs well on large datasets and recognizes nonlinear relationships and non-global clusters. In addition, FLAME automatically defines cluster numbers and identifies cluster outliers.
In [7] Zhu and his colleagues have proposed the Modified Fuzzy C-means clustering algorithm based on motion similarity, which improves the stability of the dynamic topology network of large-scale mobile wireless sensor clusters. Simulation results show that nodes with similar motion states are more likely to cluster together, and the modified fuzzy C-means algorithm improves the communication link holding time of nodes in the cluster.
Yadav et al. [8] improved the efficiency of the existing k-means algorithm. They used the previous iteration data in the current and next iterations to cluster the network nodes, which reduced the computational complexity of the k-means. The results showed that the improved clustering of k-means has less computation time and better accuracy than the enhanced k-means (EKM) method.

Raval et al.[9] improved the k-means technique to determine the initial centroid and assign the center data to its closest cluster. In addition, this step produces higher accuracy with less time complexity than the traditional k-means method.
Chao et al. [10] have used the concept of clique in the k-nearest. The clustering in this algorithm is based on gathering nodes into a cluster. If the cluster meets the requirement, then it will be considered; otherwise, the clustering will be repeated until it meets the algorithm requirements. The result showed that the algorithm gets a lower Error Rate than the k-means algorithms ,which shows the advantage of the k-nearest neighbor clique clustering (KNNCC) algorithm.

In [11], the researcher discussed the problem of energy efficiency in wireless sensor networks and highlighted the importance of clustering in reducing power consumption and increasing the network's life span. The simulation results show that without clustering, the network's lifetime becomes three times shorter, with nodes farthest from the base station dying at an early stage due to higher energy requirements for direct communication.

Despite the importance of the aforementioned studies, this study presents a different perspective. As mentioned above, these previous studies combined the same type of nodes and chose one head cluster for each cluster. However, this study will perform the clustering by using FLAME. The clustering will be conducted on each cluster, and finally, multiple head clusters of well-known data will be selected.

3. The Proposed Clustering Method for WSN

The classification of nodes in the network requires the use of classification techniques to detect to which group the node belongs on the network. Fuzzy clustering is a form of clustering in which every node in the network can be assigned to multiple clusters. Fuzzy clustering was developed by Dunn in 1973 and improved by Bezdek [4]. In this study, we use Fuzzy clustering by Local Approximation of Memberships algorithm to construct two layers of network topology. The experiments with our study were applied to the iris dataset, which is the most used Clustering method.

3.1 Clustering:

Clustering is the task of making groups of data from data sets. Each cluster, or group of data, has nodes that are more similar than the other nodes in other clusters. A cluster is also called data or node segmentation because the clustering does the partitions for large data or nodes into similar clusters according to their similarity.

3.2 Fuzzy clustering:

Fuzzy clustering is a clustering shape in which every node in the network can be assigned to multiple clusters, as compared to regular clustering in which each node is assigned to only one cluster. Clusters are identified by application requirements or data that is used, and these measures can be distance measurement, density, and connectivity. The fuzzy term was first used in 1965 as proposed by Lotfi Zadeh[5].

Fuzzy Clustering using local Approximation membership is a clustering algorithm that defines the cluster by the dense point of the dataset. Fu and Medico used this algorithm in biology [6], due to its simplicity, better performance, and system strength.

The Fuzzy clustering can be divided into three major stages [12, 13]:

- Step one: Extract the data structure from the dataset and classify network nodes to belong to one of three sets; inner, outer, or rest.
- Step Two: Initialization of fuzzy membership for each of the three sets; outer, inner, and rest.
- Step Three: Build the network cluster by fuzzy membership.

4. Experimental Results and Evaluation

4.1 Simulation Software (MATLAB)

All simulations are performed using MATLAB software installed on an Asus desktop computer with Intel core i5, 6 GB of RAM, and 2TB of Hard Disk, running on Microsoft Windows 10 and Microsoft Visual Studio 2017 to develop the C++ code.

4.2 Clustering Measurements

To analyze the performance of our research method, we applied it to two kinds of data sets; dummy dataset, and Iris data sets. The measurements used in our experiment are error rate, accuracy, precision, Euclidean distance, and Manhattan Distance between nodes [14, 15].

1. Error Rate:

The calculation of the Error rate depends on the number of misclassified nodes and the total number of tested nodes in the dataset [15], or false positive divided by True Positive and False Negative as defined in Table (5-1), as equation qualified as in (1).

$$\text{Error Rate}=\frac{\text{False positive}}{\text{True Positive +False Negative}} \qquad (1)$$

2. Accuracy Rate:

The Accuracy was defined as the proportion of correct predictions of the size of the actual dataset. Accuracy calculated as in (2):

$$\text{Accuracy} = \frac{umber\ of\ true\ node\ position}{total\ number\ of\ node} * 100 \qquad (2)$$

3. Precision:

Precision is defined as the proportion of positive nodes that are correctly classified [12], as defined in Table I. Precision is calculated as in (3).

$$\text{Precision} = \frac{true\ positve}{true\ postive + false\ postive} \qquad (3)$$

TABLE I. MEASUREMENT PARAMETERS

Parameter	Definition
True Positive (TP)	A node in same class in the same cluster
False Positive (FP)	Node in different classes in the same cluster
False Negative (FN)	A node in different classes in a different cluster

4. Distance Measurements:

It includes two Distance Measurements:

i. Manhattan Distance:

Manhattan distance calculates the absolute differences between the coordinates of two nodes [16], as the equation qualified in (4).

Manhattan distance = (x1-x)+(y1-y) (4)

ii. Euclidean Distance:

Euclidean distance calculates the root of square difference between the coordinates of two nodes [16], Euclidean is calculated as in (5):

$$\sqrt{((x_1 - x) + (y_1 - y))^2} \qquad (5)$$

4.3 Dataset:

The research method was tested on Iris's flower data set or Fisher's Iris data set from the UCI datasets (UCI, 2017). It has been introduced by the biologist Ronald Fisher in his paper, which applied multiple measurements in taxonomic problems [17]. Iris's data set became a test case for classification techniques, which included three sets of 50 instances. The three sets are called setosa, versicolor, and virginica [17]. Each set has four numeric attributes, Sepal length, Sepal width, Petal length, and Petal width.

4. 4 Simulation Results and Evaluation of the Proposed Clustering Algorithm

Several experiments were conducted, and the initial results indicate that the best value of the inner cluster is 95 instances in terms of accuracy, error rate, and precision.

The early results indicate that the proposed fuzzy clustering achieved better performance when the experiments were carried out using the iris dataset and the randomly generated dataset.

5. Conclusion

In this paper, fuzzy clustering has been presented. By using fuzzy clustering, nodes are divided into three types: Inner, Outer, and the Rest, then Initialization of fuzzy membership by the location of the node. Each Inner node in the network is set to be a main head cluster upon which to build the network cluster by fuzzy membership. The network is built and each cluster is built a priority graph to connect each node to its priority value. Then the distance between each node and the leading head cluster in the cluster is calculated and then the nodes are divided into Inner, Outer, and the Rest by the priority value.

We seek to guarantee scalability, and adaptively, which means when the number of nodes increases the network will not be overloaded or have a poor performance. The results of the experiment show that our research method can give consistency and accuracy in the distribution of a node with a low error rate.

Our result demonstrates that two-level fuzzy is adequate for clustering. Furthermore, changing the network states, and head cluster failure will be investigated as future enhancements.

References

[1] Y.Qian, Q.Shi and Q. Wang, “ CURE-NS: a hierarchical clustering algorithm with a new shrinking scheme,” International Conference on Machine Learning and Cybernetics (ICMLC), Beijing, China, 2003.

[2] Y. Zhang, Sh. Ding, L.Wang, Y.Wang, and L Ding “ Chameleon algorithm based on mutual k-nearest neighbors,” Applied Intelligence (2021) 51:2031–2044 https://doi.org/10.1007/s10489-020-01926-7, 2021.

[3] Dajin Wang, An Energy-efficient Clusterhead Assignment Scheme for Hierarchical Wireless Sensor Networks, Int J Wireless Inf Networks, Published online: 15:61–71, 2008.

[4] J.C. Bezdek, “ Pattern Recognition with Fuzzy Objective Function Algorithm,” Advanced Applications in Pattern Recognition (AAPR), 1981.

[5] Z., Lotfi. "Fuzzy Logic and Computational Intelligence." IEEE Intelligent Systems 26.4 (2011): 11.

[6] [6]Limin Fu and Enzo Medico, FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data, BMC Bioinformatics 2007 8:3, 2007.

[7] P. Zhu, Y. Shen, X. Shi and Y. Wang, "Modified Fuzzy C-means Clustering Algorithm for Mobile Wireless Sensor Networks Based on Motion Similarity," *2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, Chongqing, China, 2022, pp. 997-1003, doi: 10.1109/ITAIC54216.2022.9836881.

[8] Asmita Yadav and Sandeep Kumar Singh, An Improved K-Means Clustering Algorithm, International Journal of Computing Academic Research (IJCAR) ISSN 2305-9184, Volume 5, Number 2 (April 2016), pp.88-103. 2016

[9] Unnati R. Raval, Chaita Jani, Implementing & Improvising K-means Clustering Algorithm, International Journal of Computer Science and Mobile Computing, Vol.5 Issue.5, May- 2016, pg. 191-203. 2016

[10] Chao Qu, Ruifen Yuan, Xiaorui Wei, KNNCC: An algorithm for k-nearest neighbor clique clustering, Machine Learning, and Cybernetics (ICMLC), 2013 International Conference, 14-17 July 2013.

[11] T. M. Tatarnikova, “ Wireless Sensor Network Clustering Model,” pp 240-243, doi: 10.1109/SCM58628.2023.10159119, 24 May 2023.

[12] K.Alzboon, J. Al-Nihoud, W.Alsharafat,, “ Novel Network Intrusion Detection Based on Feature Filtering Using FLAME and New Cuckoo Selection in a Genetic Algorithm,” Appl. Sci. 2023, 13, 12755. https://doi.org/10.3390/ app132312755.

[13] L. Fu, E. Medico, “ FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data,” BMC Bioinform. 2007.

[14] W. Alsharafat, A. Alshdifat, Kh. Batiha, A. Alsukkar, “ A New Crossover Methods and Fitness Scaling for Reducing Energy Consumption of Wireless Sensor Networks,” IEEE Access, Vol.10, 2022. Digital Object Identifier 10.1109/ACCESS.2022.3203696.

[15] Mohammad F. Eltibi, Wesam M. Ashour, Initializing K-Means Clustering Algorithm using Statistical Information, International Journal of Computer Applications (0975 – 8887)Volume 29– No.7, September 2011.

[16] Archana Singh, Avantika Yadav, Ajay Rana, K-means with Three different Distance Metrics, International Journal of Computer Applications (0975 – 8887) Volume 67– No.10, April 2013.
[17] https://archive.ics.uci.edu/: last visited on 12/1/2024.