Int. J. Advance Soft Compu. Appl, Vol. 12, No. 2, July 2021
ISSN 2074-8523; Copyright © ICSRS Publication, 2021
www.i-csrs.org

Novel Fuzzy Based Density Based Clustering Algorithm for Effective Cluster Prioritization in WSN

N.Gokilavani, B.Bharathi

Research Scholar, Department of Computer Science and Engineering, Sathyabama University, Chennai, India
E-mail: gokilavaniphd15@gmail.com

Professor, Department of Computer science and Engineering, Sathyabama University, Chennai, India.
E-mail:bharathi005@gmail.com

Abstract

***Background:** In WSN (Wireless Sensor Network), various nodes are interrelated with one another for transmitting the data from source node to destination node through various intermediary nodes. Clustering in WSN permits various categorised structures to be constructed on the sensors and this allows the effective utilization of many resources (for instance-power, frequency spectrum and so on). The senor nodes have to arrange themselves into clusters so as to solve the effect of scaling in WSN. Prioritization is a more important term for efficient cluster formation. Clustering on the basis of prioritization is significant and vital as it expands the data transmission rate.*

***Objective:** To perform effective clustering and return the prioritized clusters through the use of novel Fuzzy based Density based Clustering algorithm (FDC) and efficiently remove the data points that do not belong to the cluster. The performance analysis is performed to check the efficiency and effectiveness of the proposed FDC. Methodology: A simulation has been carried out where a test case generation is implemented after taking the dataset as input. The predefined neighbour data points are counted. This counting is executed to obtain the potential cluster seeds. Then the cluster function is expanded. After the expansion, cluster members are grouped efficiently based on the fuzzy rule for returning the prioritized clusters through the use of the proposed novel FDC.*

Results: The performance of the proposed system is analysed in terms of Jaccard Coefficient and Rand to assess its efficiency. Thus the results attained from analysis revealed that through the use of the proposed FDC, the clusters are effectively prioritized.

Keywords: *WSN, Fuzzy based Density based Clustering, Cluster prioritization.*

1 Introduction

WSN (Wireless Sensor Networks) is an infrastructure less and self-configured wireless networks for monitoring the environmental or physical conditions that include sound, pressure, temperature, vibration, pollutants or motion and to collaborately transfer the data via the network to sink or main location. Clustering is the main technique to enhance the lifetime of network in WSNs. It includes cluster formation by grouping the sensor nodes and selecting CHs (Cluster Heads) for all clusters. Moreover a heavy traffic load lead to unanticipated node death due to depletion in energy resource in few network regions that is hotspots that result in disruption of network service. This issue is critical particularly for scenarios of data gathering in which CHS are accountable for forwarding and collecting sensed data to BS (Base Station). The network workload has been distributed uniformly among nodes to evade hotspot problem. In this study [1], a clustering algorithm has been introduced that choose nodes with maximum remaining energy as candidate CHs in each region. Among this, the best nodes are chosen as the final CHs. Additionally to solve hotspot issue, this study applied fuzzy logic for CH node's cluster radius adjustment that depend on few local information that involve distance to local density and BS. Simulation has been carried to assess the performance the proposed system. Simulation results revealed effective results. Thus this study presented a clustering technique called Density based Clustering (DC). This algorithm is an unsupervised learning technique which detect various clusters or groups in the data that rely on concept that cluster residing in a data space of WSN is a contiguous region corresponding to high density point that is segregated from other similar clusters by contiguous regions corresponding to low point density. The DC works by identifying regions where points are separated by regions that are sparse or empty and where the points are concentrated. Data Points that do not belong to the cluster are considered as noise. Initially test case generation is performed after taking the dataset as input. Test case generation is the procedure of test generation that is suitable for a specific system. The predefined neighbour data points are counted to choose the potential cluster seeds. The cluster members are grouped based on fuzzy rule and the prioritized cluster is returned. Finally the proposed system is analysed to assess the efficiency.

The major contributions of this study are listed below:

- To focus on the efficient cluster member grouping and to count the predefined neighbour data points based on fuzzy rule.

- To perform efficient clustering and prioritization of clusters by the use of novel FDC (Fuzzy based Density based Clustering) algorithm.
- To conduct performance analysis to exhibit the efficiency of the proposed system.

1.1 Organization of the paper

The paper is organised in the following mentioned ways. Initially, the basic ideas about clustering in WSN, importance of clustering and prioritization is discussed in section I. The existing methods for cluster prioritization along with its limitations is specified in section II. The proposed system is afforded with novelty to efficiently prioritize the clusters which is described briefly in section III. The results obtained from the proposed system is discussed in section IV. Finally, an overall summary about the proposed system in cluster prioritization is explained in section V.

2 Review of Existing Work

Clustering analysis has wide range of data analysis applications that involve machine learning, data mining and information retrieval. Practically, many cluster algorithms lack certain metrics such as varied shapes and densities, noise effects and cluster overlaps and so on. To resolve this issue this study [2] introduced a simple but efficient DCF (Density based Clustering Framework) and executed a clustering algorithm on the basis of DCF. A raw dataset has been segregated in DCF by neighbourhood density estimation model into core and non-core points. Hence the core points have been initially clustered as they indicate the dense or centre region of the structure of the cluster. DCF categorizes the non-core points to preliminary clusters in an organized order. Experimentally the proposed algorithm has been compared with the existing DBSCAN and Dp algorithms on real-world and synthetic data sets. The results revealed the efficiency of the proposed algorithm. The future work include the use of the current clustering approaches to suit the framework to enhance the performance as well as compare DCF with various existing methods [3]. Additionally the selection of metrics (k) in neighbourhood valuation model has been studied. Similarly this study [4] introduced a scheme called DBCP (Density Based Controller Placement) that utilizes a switch clustering algorithm based on density to divide the network into various sub networks. The ideal count of controllers is attained on the basis of density based clustering. The performance of DBCP has been evaluated on a dataset of publicly existing network topologies. The results revealed that DBCP affords improved performance than the traditional methods in terms of propagation latency, time consumption as well as fault tolerance [5]. In addition OD (Outlier Detection) aids in detecting intrusion, errors, noise, fraud, effects etc. In this study [6], a novel OD process labelled DBSCAN-OD (Density Based Spatial Clustering of Applications) with noise has been introduced on the basis of the DBSCAN algorithm. Two procedures have been integrated in accordance with DBSCAN methodology that include parameter computation and class identification in STD (Spatial Temporal Databases). Simulation has been undertaken to assess the performance of the proposed system. The results revealed the efficiency of the proposed methodology in detecting outliers. Additionally this study [7] developed an approach that is parameter free. Here points that lie close to one another have been detected to form cluster and the whole count of clusters has been exposed. Clusters that have been well separated are detected in the outlier's presence. On the other hand, the final count of clusters have been assessed from the identified clusters for not well segregated dataset. These are combined to yield final clusters. The performance of the proposed methodology have been assessed via experiments. The experimental results revealed that the proposed system is effective. Minimizing the complexity of the algorithm has yet to be implemented. Furthermore this study [8] explored an innovative computational technique based on DCF model and SimRank for miRNA DAP (Disease Association Prediction). The AUC and case studies recommended the outstanding performance of the proposed model in the miRNA disease prediction [9].

Additionally test case prioritization has been a general scheme to enhance the fault detection rate. In this study [10], requirement information and log output have been utilized to generate a prioritization method that is based on cluster. This study employed the technique for evaluation to device regression in progress. A simplified dataset have been generated and employed the equivalent prioritization method to show the significance of fault knowledge. The outcomes of the retrieved dataset represent an enhancement in the fault detection rate. In spite of the constraints, the proposed technique has been a solid foundation for exploration in future. Similarly this study [11] explored an ARS (Adaptive Random Sequences) scheme that is depended on clustering methods via the use of black-box information. The proposed approach can try to create varied neighbouring test cases as much as possible. Studies have been conducted experimentally to validate the proposed method. The results attained from the experiments exhibit high efficiency and improved early fault detection probability than the existing methods. Further research has to be conducted into OOS test case characteristics in near future. The sampling approach has to be enhanced for TCP test cases optimization. On the other hand this study [12] introduced a fuzzy logic model to select the cluster head. The introduced model utilized five descriptors to conclude the chance for individual node to get converted to a CH (Cluster Head). Residual energy, density, distance from BS, compacting and location suitability are the several descriptors that have been introduced. The proposed method has been compared with the various existing methodologies. Simulation has been carried out to assess the proposed system's efficiency. The results attained from simulation exhibit improvement in energy efficiency in accordance with energy consumption and network lifetime balancing amongst sensor nodes for varied topologies and network sizes. The results revealed an average enhancement in terms of initial node dead as well as half nodes dead. Likewise this study [13] explored a Cytoscape plugin (CytoCluster) by incorporating six clustering algorithms. Here the users have the ability to choose various clustering algorithms based on its requirements. The main intention of the six algorithms have been to identify the functional modules or protein complexes. The proposed system can be easily expanded as well as various functions and clustering algorithms can be included in the plugin. Thus this revealed the efficiency of the proposed method. Similarly in this paper [14], hybrid algorithms based on WGW (Whale and Grey Wolf) optimization algorithms have been proposed for prioritizing the requirements of software. The proposed method has been analysed and the results revealed the accuracy of the proposed system [15]. In the same way, this paper [16] surveyed the parallel clustering optimization in terms of throughput that exploit commonalities as well as reuse for multithreading scalability among variant computations. Similarly this study [17] recommended an innovative OD algorithm to conquer the error. The results determined the effectiveness of the proposed system.

Consequently this study [18] proposed VAC-BNR (Vector Angle Cluster and Bridge Nodes based Routing) protocol that depend on clustering of bridge nodes and direction vector angle. The experimental analysis has been carried out to assess the efficiency of the proposed method. The results attained from experiments revealed that the proposed system has the ability to enhance the data package transmissibility and minimize transmission delay. Likewise this study [19] developed an EEDCF (Energy Efficient Distributed Clustering on the basis of Fuzzy) with non-uniform distribution. In the proposed approach, individual sensor node computes the probability of existing as CH based on fuzzy. The results attained from experiments revealed that the proposed algorithm is effective than the existing methodologies in terms of energy consumption, data transmission as well as network lifetime. Similarly this study [20] proposed a group pf index structure that is based on graph to quicken the DBSCAN clustering as well as its neighbour search operations. Group method is vigorous to noise by early trimming of outliers with few or zero distance computations. The proposed method has to be extended in near future with large datasets using distributed and parallel types of the proposed system by including HPC (High Performance Computing) methods. Besides this study [21] intended to prioritize test cases in a way the testing effort significantly gets minimized. On the other hand the code coverage retains the same. This is achieved via clustering approach by selecting test cases from individual cluster by assuring uniform code coverage distribution. The proposed system has been analysed and results revealed its efficacy. In addition this study [22] proposed an effective algorithm that has the capacity to accelerate the DBSCAN and enhance its performance in identifying neighbouring clusters with varied border points and densities. The proposed system also executed the insensitivity and parameter reduction to the starting points. The proposed system has been experimentally tested. The results revealed that the proposed methodology is scalable and effective that can be employed to datasets with high complex structures. Automatic detection of single parameter (k) has yet to be performed. Likewise a DBC (Density Based Clustering) algorithm that is NG-DBSCAN have been proposed in this study [23]. The proposed methodology functions on symmetric distance computation and arbitrary data. The distributed design of proposed methodology cause it to be scalable for huge datasets and thus affords clustering results of high quality. For instance this study [24] studied the accident patterns in Metropolis. Various non-spatial and spatial datasets have been collected, analysed and processed. Based on these metrics, WSI (Weighted Severity Index) have been produced. DBCTAR (Density Based Clustering for Traffic Accident Risk) has been undertaken to support the BSS (Black Spots Severity) distribution. The obtained results include service area analysis, shortest path analysis, accident spot vulnerability and severity level traversal along the metropolis on the basis of WSI [25].

3 Proposed Work

The proposed system takes dataset as input for prioritizing the cluster. Various steps are involved in this process for cluster prioritization. Initially a test case is generated. Then the predefined neighbour data points are counted to obtain the potential cluster seeds. Additionally the cluster function is expanded and the members in the cluster are grouped based on fuzzy rule. Finally the prioritized cluster is returned through the proposed Fuzzy based Density Based clustering algorithm (FDC). Performance of the proposed system is also analyzed. The overall view of the proposed system that is explained above is given in the below figure.1.

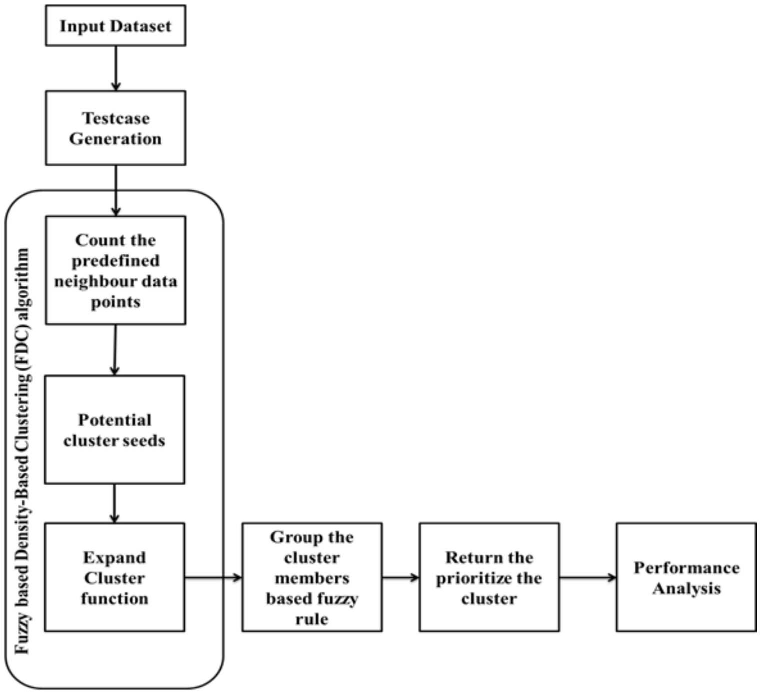

Fig.1. Overall view of the proposed FDC

Thus the proposed system is based on fuzzy and density based algorithm. It is briefly explained below:

3.1 FDC

From the past few years, various techniques are proposed for fuzzy clustering. Moreover investigators have been working to enhance them or introduce innovative techniques to overcome the existing limitations. The aim of this section is to discuss the core clustering algorithms that integrate amongst fuzzy and density based clustering algorithms (FDC). This algorithm is similar to DBSCAN. The major difference is instead of utilizing a distance based function, it utilizes the fuzzy rule to efficiently find the neighbouring data points. The FDC algorithm is briefly given in below table.1.

Table 1: FDC algorithm

FDC algorithm
Input: data set - $\boldsymbol{\pi}\ \mathbf{1}, \boldsymbol{\pi}\ \mathbf{2}$ Output: A partition of $\boldsymbol{K} = \boldsymbol{K}\mathbf{1}, \ldots, \boldsymbol{Kn}$ of n clusters. 1. Select parameters $\boldsymbol{\pi}\ \mathbf{1}$ and $\boldsymbol{\pi}\ \mathbf{2}$. 2. Note all points unclassified and set s=1. 3. Identify an unclassified fuzzy point $\boldsymbol{q}$ having parameters $\boldsymbol{\pi}\ \mathbf{1}$ and $\boldsymbol{\pi}\ \mathbf{2}$. 4. Select $\boldsymbol{q}$ to be classified. Begin with new cluster $\boldsymbol{Cs}$. Assign $\boldsymbol{q}$ to cluster $\boldsymbol{Cs}$. 5. Make an empty set of seeds $\boldsymbol{T}$. Determine complete unclassified clusters in the set$\boldsymbol{N}(\boldsymbol{q}, \boldsymbol{\pi}\ \mathbf{1})$, insert every points into the set $\boldsymbol{T}$. 6. Acquire a point $\boldsymbol{r}$ in the set $\boldsymbol{T}$, note $\boldsymbol{r}$ to be classified, assign $\boldsymbol{r}$ to the cluster $\boldsymbol{Cs}$, and eliminate $\boldsymbol{r}$ from the set $\boldsymbol{T}$. 7. Analyse if $\boldsymbol{r}$ is a fuzzy core point with parameters $\boldsymbol{\pi}\ \mathbf{1}$ and $\boldsymbol{\pi}\ \mathbf{2}$. In such a case include every unclassified points in the set $\boldsymbol{N}(\boldsymbol{q}, \boldsymbol{\pi}\ \mathbf{1})$ to the set $\boldsymbol{T}$. 8. Repeat 6th and 7th step until the collection of seeds is empty. 9. Discover an innovative fuzzy core point with parameters $\boldsymbol{\pi}\ \mathbf{1}$ and $\boldsymbol{\pi}\ \mathbf{2}$. Repeat steps 4 to 7. 10. Note all points that do not fit to any cluster as noise.

A test case dataset is taken as input. Initially two parameters (π 1 and π 2) are selected. Then all the unclassified points are noted and set s=1 is assigned. An unclassified fuzzy point q is identified and selected q for classification. Begin with new clusters Cs and q is assigned to Cs. Subsequently, an empty set of seeds T is made. Complete unclassified clusters in the set$N(q, \pi\ 1)$ is determined and every points are inserted into the set T. A point r in the set T is acquired, r is assigned to cluster Cs and r is eliminated from the set T. An analysis is performed to find if r is a fuzzy core point with parameters π 1 and π 2. If so then every unclassified points in the set $N(q, \pi\ 1)$ is included to the set T. The steps 6 and 7 is repeated until the collection of seeds is empty. Finally new fuzzy points that have parameters π 1 and π 2 are found and steps 4 to 7 is repeated. Hence all the points that do not fit to any of the clusters are noted as noise.

3.2 DC Algorithm

DC (Density based Clustering) algorithm is a data clustering algorithm. It takes the fundamental ideas into account. The fundamental idea is the clusters residing in the data space are dense regions that are segregated by minimum object density regions. It groups the organized points that are packed closely, marking the points of outliers which lie single in regions of low-density. It takes two parameters as input which is given below:

ϕ-Neighbourhood – Objects residing within a radius of ϕ from an object.

"High density" - ϕ-Object's neighbourhood consists of at least MinimumPts of objects.

The DC algorithm is briefly given in the below table.2.

Table 2: DC algorithm

DC Algorithm
Step1: DC (E, eps, MinimumPts) $\boldsymbol{M}$ = 0 for each unvisited point $\boldsymbol{Q}$ in dataset $\boldsymbol{E}$ note $\boldsymbol{Q}$ as visited NearPts = regionQuery($\boldsymbol{Q}, \boldsymbol{eps}$) if sizeof (NearPts) < MinimumPts note $\boldsymbol{Q}$ as NOISE else $\boldsymbol{M}$ = next cluster expandCluster($\boldsymbol{Q}$, NearPts, $\boldsymbol{M}$, $\boldsymbol{eps}$, MinimumPts) Step 2: expand Cluster($\boldsymbol{Q}$, NearPts, $\boldsymbol{M}$, eps, MinimumPts) add $\boldsymbol{Q}$ to cluster $\boldsymbol{M}$ for each point $\boldsymbol{Q}$' in NearrPts if $\boldsymbol{Q'}$ is not visited note $\boldsymbol{Q'}$ as visited NearPts' = regionQuery($\boldsymbol{Q'}, \boldsymbol{eps}$) if sizeof (NearPts') >= MinimumPts NearPts = NearPts coimbed with NearPts' if $\boldsymbol{Q'}$ is not yet member of any cluster add $\boldsymbol{Q'}$ to cluster $\boldsymbol{M}$ regionQuery ($\boldsymbol{Q}, \boldsymbol{eps}$) Step 3: return all points within $\boldsymbol{Q'}$s eps-neighborhood (including $\boldsymbol{Q}$)

Initially a data is obtained from the input (test file) by the module DC (E, eps, Minimum Pts) where E is the filename. The minimum count of points and (ϕ) neighbourhood are mentioned when running the file. The results are investigated for varied ϕ and Minimum Pts. Once the data is fetched, compute_ dbscan is utilized for assigning label to individual data points. It also takes the neighbours into account which are density reachable. Based on this clustering is performed. A module named expand Cluster is called when the neighbouring points have the ability to be density reachable. Subsequently, the neighbours of the particular data points are added to the similar cluster. A function known as region Query is called to undertake this process. This function affords the neighbours of a particular data point which reside at ϕ distance from a point thereby fulfilling the conditions of Minimum Pts. Finally, labelling is performed for the overall data set with a cluster number that matches it and if it is noise then it is denoted by 1. This list of labels is later utilized to calculate the external index to view the clustering outcome.

4 Results and Discussion

The proposed system is executed to prioritize the clusters and this methodology is evaluated in terms of performance metrics such as Jaccard Coefficient and Rand. The results are specified below and comparative analysis have also been made which is discussed below:

4.1 Performance analysis

As mentioned earlier, the proposed system is evaluated using performance metrics (Jaccard Coefficient, Rand) to confirm its efficiency. It is precisely explained below:

4.1.1 Jaccard Coefficient

The Jaccard Coefficient refers to a static that is utilized for estimating the diversity and similarity of sample sets. It is given by:

$$Jaccard\ Coefficient = \frac{|M_{11}|}{|M_{11}|+|M_{10}|+|M_{01}|} \quad (1)$$

Here $|M_{11}|$ - Represents the set consisting of (agree, similar cluster)

$|M_{10}|$ - Represents the set consisting of (disagree, dissimilar clusters)

$|M_{01}|$ - Represents the set consisting of (disagree, dissimilar clusters)

4.1.2 Rand Index

It calculates the similarity measure amongst two clusterings by taking all counting pairs and sample pairs which are allocated in similar or dissimilar clusters in true and predicted clusterings.

$$Rand\ Index = \frac{|Agree|}{|Agree|+|Disagree|} = \frac{|M_{11}|+|M_{00}|}{|M_{11}|+|M_{00}|+|M_{10}|+|M_{01}|} \quad (2)$$

Here,

$|M_{11}|$ - Denotes the set consisting of (agree, similar cluster)

$|M_{10}|$ - Denotes the set consisting of (disagree, dissimilar clusters)

$|M_{01}|$ - Denotes the set consisting of (disagree, dissimilar clusters)

$|M_{00}|$ - Denotes the set consisting of (agree, dissimilar clusters)

4.2 Dataset description

The test cases are generated for parameter requirement. It is the method of writing SQL (Structured Query Language) test cases for confirming and analysing the functionalities of the database. The motive of this process is to assess the outcome against probable results. Three test cases are generated where a random generation occurs through the proposed FDC.

4.3 Comparative analysis

The performance of the existing and proposed method is compared to determine the Rand index and Jaccard Coefficient to return the prioritized clusters. It is given in below table.3, table.4, table.5 and table.6.

Table 3: Determination of Jaccard Coefficient and Rand index for the proposed FDC

FDC (PROPOSED)		
eps	The Jaccard Coefficient	Rand Index
1	0.6319	0.8151
1.2	0.7825	0.7333
1.3	0.7749	0.7155
1.4	0.7545	0.6731
1.5	0.7321	0.6134

Table 4: Determination of Jaccard Coefficient and Rand index for the existing K-MEANS

K- MEANS		
eps	The Jaccard Coefficient	Rand Index
1	0.5885	0.3515
1.2	0.5806	0.6991
1.3	0.57999	0.5935
1.4	0.5794	0.5904
1.5	0.57888	0.529

Table 5: Determination of Jaccard Coefficient and Rand index for the existing Fuzzy Clustering

Fuzzy Clustering		
eps	The Jaccard Coefficient	Rand Index
1	0.5728	0.7462
1.2	0.5712	0.6363
1.3	0.57066	0.6329
1.4	0.5689	0.6193
1.5	0.5685	0.616

Table 6: Determination of Jaccard Coefficient and Rand index for the existing Hierarchical agglomerative clustering

HIERARCHICAL AGGLOMERATIVE CLUSTERING		
eps	The Jaccard Coefficient	Rand Index
1	0.5678	0.6095
1.2	0.5668	0.5942
1.3	0.5644	0.5924
1.4	0.5626	0.5875
1.5	0.5598	0.5823

From the above table.3, 4, 5 and 6, it has been found that the value of Jaccard Coefficient and rand index is high than the state-of-the-art methods thereby showing that the performance of the proposed FDC is efficient than the traditional K-means, fuzzy clustering and hierarchical agglomerative clustering method.

The efficiency of the proposed system than the exiting methods in terms of Jaccard Coefficient and rand index is shown in the below figure.2 and figure.3.

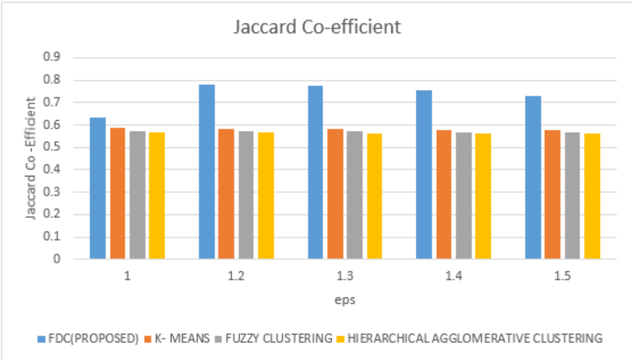

Fig.2. Comparative analysis of the Proposed and Existing methods in terms of Jaccard Coefficient

In the above figure.2, when epsilon value (eps=1) then the Jaccard Coefficient is 0.7 for the proposed system, 0.59 for K-Means, 0.56 for fuzzy clustering and 0.55 for hierarchical clustering. In this way, for different eps value, the Jaccard Coefficient varies. The results revealed that high Jaccard Coefficient is achieved for the proposed system thereby confirming the efficiency of the proposed system.

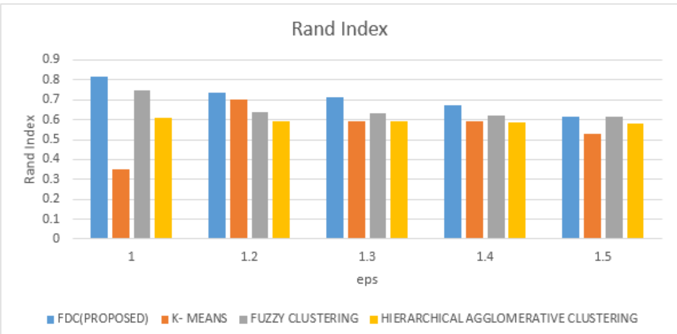

Fig.3. Comparative analysis of the Proposed and Existing methods in terms of Rand index

In the above figure.3, when epsilon value (eps=1) then the Rand index is 0.9 for the proposed system, 0.3 for K-Means, 0.7 for fuzzy clustering and 0.61for hierarchical clustering. In this way, for different eps value, the Rand Index varies. The results revealed that high Rand Index is achieved for the proposed system thereby confirming the efficiency of the proposed system.

5 Conclusion

The clustering is more essential to improve the network's lifetime in WSN. Hence an efficient clustering is implemented through the fuzzy rule. Prioritization of clusters is indispensable for effective clustering. Thus this study presented an effective novel FDC (Fuzzy based Density based Clustering) algorithm. A test case is automatically generated by DC (Density based Clustering) algorithm. Then the predefined neighbouring data points are calculated to obtain the potential cluster seeds. The cluster members are grouped successfully and the clusters are prioritized through the proposed FDC. Hence a successful and effective cluster prioritization is attained. The performance metrics (Jaccard Coefficient and Rand) is evaluated to check the proposed system's efficiency. The performance analysis explored better outcome by effectively returning prioritized clusters.

References

[1] P. Neamatollahi and M. Naghibzadeh, "Distributed unequal clustering algorithm in large-scale wireless sensor networks using fuzzy logic," *The Journal of Supercomputing,* vol. 74, pp. 2329-2352, 2018.

[2] J. Lu and Q. Zhu, "An effective algorithm based on density clustering framework," *Ieee Access,* vol. 5, pp. 4991-5000, 2017.

[3] A. Nishad and S. Abraham, "SemTraClus: an algorithm for clustering and prioritizing semantic regions of spatio-temporal trajectories," *International Journal of Computers and Applications,* pp. 1-10, 2019.

[4] J. Liao, H. Sun, J. Wang, Q. Qi, K. Li, and T. Li, "Density cluster based approach for controller placement problem in large-scale software defined networkings," *Computer Networks,* vol. 112, pp. 24-35, 2017.

[5] M. Du, S. Ding, and Y. Xue, "A robust density peaks clustering algorithm using fuzzy neighborhood," *International Journal of Machine Learning and Cybernetics,* vol. 9, pp. 1131-1140, 2018.

[6] A. Abid, A. Kachouri, and A. Mahfoudhi, "Outlier detection for wireless sensor networks using density-based clustering approach," *IET Wireless Sensor Systems,* vol. 7, pp. 83-90, 2017.

[7] B. Mukhoty, R. Gupta, K. Lakshmanan, and M. Kumar, "A parameter-free affinity based clustering," *Applied Intelligence,* vol. 50, pp. 4543-4556, 2020.

[8] X. Li, Y. Lin, C. Gu, and Z. Li, "SRMDAP: SimRank and density-based clustering recommender model for miRNA-disease association prediction," *BioMed research international,* vol. 2018, 2018.

[9] P. Kandil, S. Moussa, and N. Badr, "Cluster-based test cases prioritization and selection technique for agile regression testing," *Journal of Software: Evolution and Process,* vol. 29, p. e1794, 2017.

[10] E. Tumbul, "Test case prioritization in limited environment," Graz University of Technology, 2020.

[11] J. Chen, L. Zhu, T. Y. Chen, D. Towey, F.-C. Kuo, R. Huang, *et al.*, "Test case prioritization for object-oriented software: An adaptive random sequence approach based on clustering," *Journal of Systems and Software,* vol. 135, pp. 107-125, 2018.

[12] A. Hamzah, M. Shurman, O. Al-Jarrah, and E. Taqieddin, "Energy-efficient fuzzy-logic-based clustering technique for hierarchical routing protocols in wireless sensor networks," *Sensors,* vol. 19, p. 561, 2019.

[13] M. Li, D. Li, Y. Tang, F. Wu, and J. Wang, "CytoCluster: A cytoscape plugin for cluster analysis and visualization of biological networks," *International journal of molecular sciences,* vol. 18, p. 1880, 2017.

[14] R. Masadeh, A. Hudaib, and A. Alzaqebah, "WGW: A hybrid approach based on whale and grey wolf optimization algorithms for requirements prioritization," *Advances in Systems Science and Applications,* vol. 18, pp. 63-83, 2018.

[15] S. R. Sree, "In Agile Development, Backlog Prioritization Using UI Feature Usage Measurement via Machine Learning."

[16] M. Gowanlock, D. M. Blair, and V. Pankratius, "Optimizing parallel clustering throughput in shared memory," *IEEE Transactions on Parallel and Distributed Systems,* vol. 28, pp. 2595-2607, 2017.

[17] K. U. Kadam and S. B. Patil, "IDENTIFY OUTLIERS USING LOCAL MINIMA DENSITY IN KOHONEN SELF ORGANIZATION FEATURE MAP," *Journal of Critical Reviews,* vol. 7, pp. 2123-2127, 2020.

[18] Y. Shao, "Clustering algorithm for wireless sensor network to improve the efficiency of acnode," *International Journal of Reasoning-based Intelligent Systems,* vol. 10, pp. 77-82, 2018.

[19] Y. Zhang, J. Wang, D. Han, H. Wu, and R. Zhou, "Fuzzy-logic based distributed energy-efficient clustering algorithm for wireless sensor networks," *Sensors,* vol. 17, p. 1554, 2017.

[20] K. M. Kumar and A. R. M. Reddy, "A fast DBSCAN clustering algorithm by accelerating neighbor searching using Groups method," *Pattern Recognition,* vol. 58, pp. 39-48, 2016.

[21] A. Bala, "A Neighborhood Probability Based Agglomerative Clustering for Test Case Prioritization in Regression Testing," 2018.

[22] Y. Lv, T. Ma, M. Tang, J. Cao, Y. Tian, A. Al-Dhelaan, *et al.*, "An efficient and scalable density-based clustering algorithm for datasets with complex structures," *Neurocomputing,* vol. 171, pp. 9-22, 2016.

[23] A. Lulli, M. Dell'Amico, P. Michiardi, and L. Ricci, "NG-DBSCAN: scalable density-based clustering for arbitrary data," *Proceedings of the VLDB Endowment,* vol. 10, pp. 157-168, 2016.

[24] J. Olusina and W. Ajanaku, "Spatial analysis of accident spots using weighted severity index (WSI) and density-based clustering algorithm,"

Journal of applied sciences and environmental management, vol. 21, pp. 397-403, 2017.

[25] M. Alswaitti, M. Albughdadi, and N. A. M. Isa, "Density-based particle swarm optimization algorithm for data clustering," *Expert Systems with Applications,* vol. 91, pp. 170-186, 2018.

Notes on contributors

N.Gokilavani is a Research Scholar from the Department of Computer Science and Engineering working in Sathyabama University, Chennai, India. I am 8 years of Software Testing experience. My major area of research include software engineering, software testing.

Dr.B.Bharathi is a professor from the Department of Computer science and Engineering working in Sathyabama University, Chennai, India. She has 15 years of teaching experience. Her major areas of research include software engineering, software architectures and Big Data. She has more than 40 publications in refereed journals. She is the reviewer of many refereed journals and also acted as advisory member for various conferences.