

# A Review of Clustering Technique Based on Different Soft Computing Approach

Mr. Gajendra Dangi, Ms.Malti Nagle, Mr.Tarique Zeya Khan

**Abstract:** Grouping of data is major issue in current scenario of diverse data space. For the grouping of data data mining provides clustering and classification technique. the clustering technique is better than classification technique in case of unknown features of data. the process of clustering technique always faced a problem of number of iteration and selection of centre point. for the reduction of the number of iteration and centre point selection used soft computing technique, such as fuzzy logic, genetic algorithm and neural network. The process of fuzzy logic, the clustering technique process is enhanced but certain limitation. The fuzzy logic reduces the number of iteration and maximized the value of fuzzy membership. In this paper present the review of soft computing based clustering technique.

**Keywords:** Data Mining, Clustering, Soft Computing, Fuzzy Logic

## I. INTRODUCTION

Grouping is an unsupervised learning approach. Clustering includes the gathering of comparable data designs that seem to same characteristic groups together. Grouping has prescient power permitting one to anticipate those data designs that have a similar bunch will have comparative properties. An outline of groups can be conveyed utilizing bunch focuses, which are delegates of the bunches [1,2]. Furthermore, the fascinating data designs that merit encourages consideration can be highlighted when there is an inability to fabricate a decent group demonstrates. Besides, the focused learning in grouping implies that bunches contend to claim data patterns [3]. They are especially inspired by the Fuzzy C-implies (FCM) grouping methods. Due to its idea of fluffy participation, FCM can bargain more adequately with exceptions and to perform enrolment reviewing, which is imperative by and by[4]. they report their commitment to bunch examination, which is another grouping methods that decides consequently the ideal number of bunches. This method is enlivened from the work introduced and enhances the regular model choice process by decreasing the arbitrariness in the instatement of group focuses toward the start of each clustering stage. The definition of a group and choice of the best data portrayal are controlled by appropriate selection of elements, likeness measure, number of bunches and bunch legitimacy, all of which can't be effectively decided. There is no all-around great portrayal; the decision of portrayal must be guided by the area learning"[5]. Along these lines, the comprehension of the data itself and shouldn't something be said about the data to be recovered is essential. Outer clustering assessment is a viable route for the legitimacy of grouping results, and successful model is vital to decently looking at the exhibitions of various clustering methods or tuning parameters a clustering methods. Customary grouping assessment estimations essentially worry about separation between classes or inside a class[6]. In the writing, a few methodologies were created in the endeavor to accomplish bunches of data from a prepared SOM. The general approach utilizes a two-level process. In the first level, a SOM is prepared. The second level can include another

SOM, making a various levelled SOM or some fresh clustering strategy which permits connecting codebook cells of the first-level SOM in greater groups. In different works, the SOM fuzzification issue was specifically examined utilizing FCM methods or other fluffy strategies as second level clustering [11]. Fuzzy frameworks and their applications to data grouping are exceedingly dissected research themes. Most present applications depend on fuzzy deduction framework (FIS) approaches utilizing IF-THEN leads in both Mandami or Takagi-Sugeno-Kang frameworks. Participation capacities and standards are produced utilizing specialists' learning or acquiring data from data [12]. the rest of paper describe as section II related work. In section III. Clustering technique and soft computing. in section IV problem formulation. In section V conclusion & future work.

## II. RELATED WORK

Minyar Sassi Hidri and Mohamed Amine Baatout Et al. [1] creators examined, the fundamental objective of this paper is to build up an improved fluffy grouping calculation which decides the ideal number of bunches in light of the discovery of the two greatest qualities to part the most noticeably bad bunches. Test results and examinations are given to delineate the execution of the new part technique contrasted with the iterative one. Examination concerning the utilization of the new part calculation to manage the measurement lessening issue is a promising road. For example, choice of proper measurements (highlights) for an administered order issue can be performed by progressively applying the bunching calculation to various blends of the element measurements. The bunching results can be assessed in view of various criteria to choose the best mix.

Telmo M. Silva Filho, Bruno A. Pimente, Renata M.C.R. Souza and Adriano L.I. Oliveira Et al. [2] they present two halves and half strategies for fluffy bunching that intend to manage these inadequacies. The techniques, alluded to as FCM-IDPSO and FCM2-IDPSO, join FCM with a current rendition of PSO, the IDPSO. The trials considered

the talked about strategies and some current PSO-based fluffy bunching techniques. The outcomes have demonstrated that both IDPSO-based strategies examined in this paper accomplished better outcomes in regards to the ARI, foundation J and number of cycles than FCM-PSO, with FCM2-IDPSO being the best strategy for all datasets. FCM2-IDPSO achieved a normal relative lessening of 44, 9% for the quantity of emphases, when contrasted with FCM-PSO. Matched Wilcoxon marked rank tests were made to confirm these outcomes

Grigorios Tzortzis and Aristidis Likas Et al. [3] as indicated by scientists, preparing includes a min-max issue that is iteratively illuminated, where the weights are refreshed in the augmentation venture to precisely reflect the changes of the bunches at every cycle. In addition, they have displayed a procedure for changing the p type to the fundamental dataset properties, so that the inherent gathering structures can be identified, which enormously encourages the utilization of their calculation. To reach dependable determinations, MinMax k-Means was broadly tried on different datasets. Comes about show its power over awful introductions and its efficacy, with respect to most cases it beats each of the three thought about strategies, to be specific k-Means, k-Meanspp and pifs k-Means. In general, MinMax k-Means gives off an impression of being an extremely aggressive and simple to utilize strategy for managing the affectability to introduction of k-Means.

S. Revathi and Dr. T. Nalini Et al. [4] they talked about Clustering is the way toward gathering of information, where the gathering is set up by discovering similitudes between information in view of their attributes. Such gatherings are named as Clusters. The execution of the different grouping calculations is contrasted in view of the time brought with frame the assessed bunches. The trial aftereffects of different bunching calculations to frame groups are delineated as a diagram. As the quantity of bunch increments progressively, an opportunity to frame the groups likewise increments. The most remote first bunching calculation takes not very many seconds to group the information things though the straightforward K-Means sets aside the longest opportunity to perform grouping. Therefore, it is exceptionally hard to utilize basic K-Means grouping calculation for huge datasets.

Daphne Teck Ching Lai and Jonathan M. Garibaldi Et al. [5] they tried different things with changing measures of marked information and they assess classification exactness utilizing cross-approval. It was found that classification precision expanded utilizing 15 or 17 bosom disease biomarkers. Utilizing SVM-RFE and CFS, enhanced classification exactness was found on three UCI datasets, Arrhythmia, Cardiotocography and Yeast. They utilized a scope of highlight choice methods and connected it with ssFCM to order the NTBC and 3 UCI datasets. ssFCM with NB-RFE delivered the best classification exactness for NTBC, superior to without highlight determination. they have utilized ssFCM with NB-RFE to recognize critical components, diminishing the quantity of elements from 25 to 15. they likewise

connected ssFCM with highlight choice to different datasets, making the system broader.

Daphne Teck Ching Lai and Jonathan M. Garibaldi Et al. [6] In this work, two varieties of semi-directed Fuzzy c-implies (ssFCM) calculations are investigated to arrange the Nottingham Tenovus Breast Cancer dataset into a similar six subgroups. Three trials were directed utilizing the two ssFCM calculations and the outcomes are assessed by utilizing rater assertion measures. The ssFCM calculations identified the six classes of bosom malignancy be that as it may, is in low concurrence with Soria's classification. The outcomes are looked at in view of the level of concurrence with classifications by Soria et al. These outcomes are additionally contrasted and those utilizing un-managed bunching methods, which they physically dole out names utilizing visual assessment. Pedrycz97 could recognize the six classes however the level of assertion is much lower than MCBIC and KM.

Janmenjoy Nayak, Bighnaraj Naik and H.S. Behera Et al. [7] they presented a thorough review on FCM and its applications in over one decade has been done in this paper to demonstrate the efficiency and relevance in a blend of areas. Additionally, another aim of this review is to urge new analysts to make utilization of this basic calculation in critical thinking. This paper makes a brief review worried with FCM, from 2000 to 2014 so as to close how FCM and its applications have created, amid this period. In all actuality, the classification of the considerable number of procedures and their applications depend on the FCM catchphrase file hunt and digests of the articles, assembled for this exploration.

Jun Chin Ang, Andri Mirzal, Habibollah Haron and Haza Nuzly Abdull Hamed Et al. [8] This paper gives a survey on the flow and significant element choice explores in quality expression microarray investigation. It additionally talks about the difficulties and issues confronted with a specific end goal to accomplish better maladies expectation or new sicknesses disclosure. To adequately manage these issues, the choices made in each phase of highlight determination are critical. A wealth of quality choice methodologies has been planned by specialists, yet this paper infers that there are as yet many open doors for further change.

Armin Daneshpazhouh and Ashkan Sami Et al. [9] They talked about technique fills in as takes after. To start with, some solid negative occasions are separated by a kNN-based calculation. This paper gives an exhibit of the SSODPU strategy, which not the same as previous works, handles the issue of recognizing exceptions with just few named positive cases. The outcomes confirm that this approach outperform the past works in the writing. As indicated by the outcomes from various informational indexes with various disseminations, it is understood that the talked about strategy works better on unequal informational indexes as opposed to informational collections with a similar number of anomalies and inliers. Feng Zhao, Jiulun Fan and Hanqiang Liu Et al. [10] In this paper, they parameter r is controlled by the brilliant area technique. Furthermore, a novel dark level histogram is

developed by utilizing the self-tuning non-nearby spatial data for every pixel, and afterward fluffy c-implies grouping calculation with the ideal determination construct stifled system is executed in light of this histogram. The self-tuning non-nearby spatial data of a pixel is gotten from the pixels with a comparative neighborhood configuration to the given pixel and can safeguard more data of the picture than the spatial data got from the pixel's neighborhood window. This technique is connected to Berkeley and other genuine pictures intensely tainted by clamor. The picture division tests exhibit the predominance of the talked about strategy over other fluffy calculations.

Gustavo J. Meschino, Diego S. Trance like states, Virginia L. Ballarin, Adriana G. Scandurra and Lucía I. Passoni Et al. [11] In this paper, they consolidate both methodologies. The vast majority of the past chips away at fluffy grouping depend on fluffy induction frameworks, yet they talked about the plan of another bunching framework in which they utilize predicate fluffy rationale to play out the grouping errand, being consequently planned in light of information. Given a datum, degrees of truth of fluffy predicates related with each group are processed utilizing consistent participation capacities defined over information highlights. The etymological interpretability of the predicates permits clarifying the qualities of the groups found, and it constitutes a significant commitment of this paper. The approach utilizing just a single compound predicate clarifying each bunch is the best decision for phonetic translation. That is the reason they are continuing along these lines, scanning for other combination plans.

Tong Zhang, Long, Chen and C. L. Philip Chen Et al. [12] they characterized new administrators are specifically intended to scan for the ideal bunch focuses of shadowed fluffy C-implies. Considering the neighborhood spatial data in picture division methodology, another division calculation called I-Ching spatial shadowed fluffy C-Means (IC-SSFCM) is talked about. Customary division approaches in view of Fuzzy C-Means (FCM), Shadowed Fuzzy C-Means (SFCM), and Spatial Shadowed Fuzzy C-Means (SSFCM) are contrasted and the talked about technique. The exploratory outcomes demonstrate that the talked about ICSSFCM is extremely efficient approach in handling the covering fragments, as well as in smothering the clamor in pictures.

Yingdi Guo, Kunhong Liu, Qingqiang Wu, Qingqi Hong and Haiying Zhang Et al. [13] in this paper, they outline a Spatial Distance Weighted Fuzzy C-Means calculation, named as SDWFCM, to manage this issue. This calculation can completely utilize spatial elements to allocate tests to various groups, and it just needs to ascertain the enrollments one time, which diminishes the running time incredibly contrasted and other spatial fluffy C-implies calculations. The tests are done in light of genuine oil geography information and counterfeit information, and the outcomes demonstrate that SDWFCM can accomplish better execution contrasted and conventional bunching technique, and their spatial group lists can give the evaluation of bunches by mulling over

spatial structure successfully. Creators talked about another fluffy K-mean calculation for grouping spatial information, named as DWSFCM. It can be connected to successfully manage spatial components and normal elements in the meantime.

Olimpia I. Boycott, Adrian I. Boycott and Delia A. Tuse Et al. [14] The fluffy grouping is an imperative device to recognize the structure in information, along these lines they apply the Fuzzy C-Means Algorithm to get a fluffy segment of an arrangement of characteristics. The principle benefit is connected with the inferring of the administrative choices which turn out to be more refined because of the fluffy approach. In view of the thought created in the present paper, it would be extremely intriguing to apply the same Fuzzy C-implies Algorithm in these cases. Then again, they can conclude that it is uncertain whether a decent IPA implies a dividing of the execution significance plane in two, three, four, nine or another number of sets. they trust that the number depends rather from the information. Such an outcome would be profitable for deciding the ideal number of properties from the perspective of their significance identified with the issue under review, an imperative subject of verbal confrontation in the current writing.

### III. CLUSTERING TECHNIQUE & SOFT COMPUTING

In this section discuss the clustering technique and soft computing. The clustering technique divided into two groups one is hard clustering and other is soft clustering. The process of hard clustering take more time and more numbers of iteration, instead of soft clustering technique [13,14].

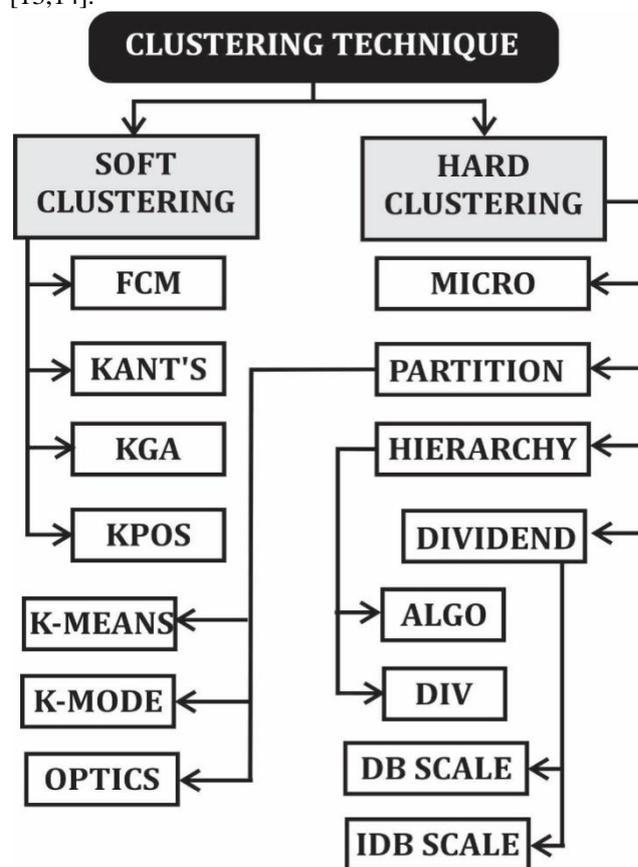


Figure 1 Shows That Different Clustering Technique

The soft computing offering various algorithm such as fuzzy logic, neural network and optimization algorithm. the fuzzy logic and optimization algorithm play an important role for enhancing the performance of clustering technique. the fuzzy logic gives the concept of approximation of centre data and cluster process. The optimization technique also used for the enhancement of clustering technique. the optimization technique offering various algorithm such as genetic algorithm, ant colony optimization, particle swarm optimization and many more optimization technique[11]. Artificial ants used in algorithm are named as software ants or agent and number of agents expressed With R. Ants start with empty solution strings and in the first iteration the elements of the pheromone matrix are initialized to the same values. With the progress of iterations, the pheromone matrix is updated depending upon the quality of solutions produced. At the end of any iteration level each agent generates the solution using the in-order resultant from updated pheromone matrix. In Ant Colony Optimization, a number of artificial ants (here data packets) build solution to the considered optimization problem at hand and exchange information on the quality of these solutions via a communication scheme that is pheromone deposit on the path of the journey performed by it[12]. PSO (Particle swarm optimization) is an evolutionary computation technique, and firstly introduced by Kennedy and Eberhard in 1995. Like genetic algorithm, the study of bird's prey behavior is an iteration-based optimization tool. In the PSO algorithm has been widely used in function optimization, neural network training, fuzzy system control and other applications of genetic algorithms [15]

#### IV. PROBLEM FORMULATION

Clustering or combination of data collections into abstractly meaningful clusters is a well-study problem. Clustering mean to find the intrinsic structure of data by organizing data objects into likeness groups or clusters. The fuzzy clustering [13] is the basic clustering algorithm to remove the clusters based on objects density. In this algorithm, first the number of objects present within the neighbour region (member function) is computed. If the neighbour objects count is below the given threshold value, the object will be marked as Noise, or else the new cluster will be formed from the core object by finding the group of densities connected objects that are maximal with admiration to density reachable. Some problem given below [12,14]

1. Maximized the value of member function
2. Maximized the number of iteration
3. Loss of data
4. Content validation of cluster
5. Inter cluster merging problem

#### V. CONCLUSION & FUTURE WORK

The grouping of data is big challenge in unsupervised learning. The unsupervised learning group the data on the bias of similarity and iteration. The increased number of iteration raised the problem of data loss and content validation of cluster. In the process of review found that various authors used the soft computing technique for the minimization of iteration and loss of data. In process of clustering the selection of centre point is also major issue. The selection of centre in clustering technique used random methods. The area of future research includes the validation of content and minimization of iteration during the clustering process.

#### REFERENCES

- [1] Minyar Sassi Hidri and Mohamed Amine Baatout “2-Maximum spanning fuzzy values-based splitting: an iterative clustering algorithm”, IEEE, 2015, Pp 1-6.
- [2] Telmo M. Silva Filho, Bruno A. Pimente, Renata M.C.R. Souza and Adriano L.I. Oliveira “Hybrid methods for fuzzy clustering based on fuzzy c-means and improved particle swarm optimization”, Expert Systems with Applications, 2015, Pp 6315–6328.
- [3] Grigorios Tzortzis and Aristidis Likas “The MinMax k-Means clustering algorithm”, Pattern Recognition, 2014, Pp 2505–2516.
- [4] S. Revathi and Dr. T. Nalini “Performance Comparison of Various Clustering Algorithm”, International Journal of Advanced Research in Computer Science and Software Engineering, 2013, Pp 67-72.
- [5] Daphne Teck Ching Lai and Jonathan M. Garibaldi “Improving Semi-Supervised Fuzzy C-Means Classification of Breast Cancer Data Using Feature Selection”, IEEE, 2016, Pp 1-6.
- [6] Daphne Teck Ching Lai and Jonathan M. Garibaldi “A Preliminary Study on Automatic Breast Cancer Data Classification using Semi-supervised Fuzzy c-Means”, International Journal of Biomedical Engineering and Technology, 2015, Pp 1-19.
- [7] Janmenjoy Nayak, Bighnaraj Naik and H.S. Behera “Fuzzy C-Means (FCM) Clustering Algorithm: A Decade Review from 2000 to 2014”, Springer, 2015, Pp 133-149.
- [8] Jun Chin Ang, Andri Mirzal, Habibollah Haron and Haza Nuzly Abdull Hamed “Supervised, Unsupervised, and Semi-Supervised Feature Selection: A Review on Gene Selection”, IEEE/ACM, 2016, Pp 971-989.
- [9] Armin Daneshpazhouh and Ashkan Sami “Semi-Supervised Outlier Detection with Only Positive and Unlabeled Data Based on Fuzzy Clustering”, IEEE, 2013, Pp 1-17.
- [10] Feng Zhao, Jiulun Fan and Hanqiang Liu “Optimal-selection-based suppressed fuzzy c-means clustering algorithm with self-tuning non-local spatial information for image segmentation”, Expert Systems with Applications, 2014, Pp 4083–4093.
- [11] Gustavo J. Meschino, Diego S. Comas, Virginia L. Ballarin, Adriana G. Scandurra and Lucía I. Passoni “Automatic design of interpretable fuzzy predicate systems for clustering using self-organizing maps”, Neurocomputing, 2015, Pp 47–59.

- [12] Tong Zhang, Long, Chen and C. L. Philip Chen “Clustering Algorithm Based on Spatial Shadowed Fuzzy C-means and I-Ching Operators”, *INTERNATIONAL JOURNAL OF FUZZY SYSTEMS*, 2016, Pp 1-8.
- [13] Yingdi Guo, Kunhong Liu, Qingqiang Wu, Qingqi Hong and Haiying Zhang “A New Spatial Fuzzy C-Means for Spatial Clustering”, *WSEAS TRANSACTIONS on COMPUTERS*, 2015, Pp 269-281.
- [14] Olimpia I. Ban, Adrian I. Ban and Delia A. Tuse “Importance-Performance Analysis by Fuzzy C-Means Algorithm”, *Expert Systems With Applications*, 2016, Pp 9–16.
- [15] Lai, Daphne Teck Ching and Garibaldi, Jonathan M. and Soria, Daniele and Roadknight, Christopher M. “Classification of breast cancer immune-histochemical data using semi-supervised Fuzzy c-Means”, Springer, 2014, Pp 1-21.