International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-7 Issue-4s2, December 2019

# Computational Approach to Overcome Overlapping of Clusters by Fuzzy k-Means

**Katikireddy Srinivas, K.V.D. Kiran**

***Abstract:*** ***Of all the clustering algorithms, the frequently employed methods of partitioning algorithms include k-means, medoids and certain modifications. For K-means, a centroid represents the mean or median point of a group and for K-medoids, wherein a medoid represents the most central point of a data group. We present a hybrid method with both algorithms; k-medoids and k-means to cluster a dataset of thyroid disease drugs and the program is run to generate clusters centred on k-means and k-medoids, followed by enhancing the outcome by implementingfuzzy k-means. Clusterability was carried out by Hopkins statistic and cluster validity by Nbclust resulted in k=3. Both the methods resulted in clusters with negative silhouettes, however, hybrid clustering algorithm resulted in partial overlapping of data points, hence fuzzy k-means algorithm was applied on sub-set of dataset. Finally, of all the six fuzzy algorithms studied, fkm algorithm displayed superior separation of clusters with well-defined data points.***

***Keywords:*** ***k-means, k-medoids, fuzzy k-means, clustering, thyroid disease***

## I. INTRODUCTION

Clustering algorithms has been categorized as Exclusive, Overlapping, Hierarchical and Probabilistic Clustering. In exclusive method, data assembled in an elitemanner, and if a datum has a place with a clear bunch then it couldn't be incorporated into another group[1]. The clustering by overlapping methodutilizes fuzzy sets to group information, so each point may have a place with at least two groups with various degrees of participation. The most mainlyreported and normally utilized dividing techniques arek-means, k-medoids, and other varieties. Partitional clustering techniques strategies make a one-level apportioning of the information. For K-means, a centroid represents the mean or median point of a group of points. For K-medoids, a medoid represents the most representative point of a group of points. K-Means clustering [2] is also an iterative clustering procedure, but it predefines the number of clusters that will be in the dataset. PAM stands for “partition around medoids” [3]. The method intends to discover an arrangement of items called medoids that are halfway situated in groups. The objective of the algorithmic method is to reduce the objectdissimilaritieswith respect to their nearby selected datum. The structure of k-medoids is nearly similar to that of k-means[4]. The cluster representative is the one data point which is located central in the cluster. Any two objects distance is calculated and the onehaving minimum dissimilarity when compared to all other objects is chosen as the center.PAM is susceptible but tough to noise as well as outliers than k means because medoids contemplates marginal distance which isolates it from alternate objects[5]. Large amounts of data are collected and presented in literature. Consequently, unsupervised machine learning tools (i.e, clustering) for discovering knowledge becomes more and more important for big data analyses. Clustering corresponds to a set of tools used in order to classify data samples into groups (i.e clusters). Each groups contains objects with similar profiles. The observation being classified into groups necessitates few methods for measuring the distance between observations, which means no unsupervised machine learning algorithms can take place without some notion of distances. The selection of distance measure iscrucial step in clustering [6]. It characterizes how the likeness of two components (x, y) is computed and it will impact the state of the groups. The most generally utilized and acknowledged technique is Euclidean separation measure. The estimation of separation measures is personally identified with the scale on which estimations are made. Therefore, factors are frequently scaled (i.e. standardized) before estimating the dissimilarities [7]. Generally variables are scaled to have standard deviation one and mean zero.The goal is to make the variables comparable and they will have equal importance in the clustering algorithm. This is especially prescribed when factors are estimated in various scales. The standardized data is a methodology broadly utilized with regards to gene examination before grouping [8].In this work, we present a hybridized program encompassing both k means and medoids algorithm to cluster a dataset of thyroid disease drugs and the program is run to generatedata groups based on the algorithm, thereby refining the outcome based on fuzzy kmeans.

## II. MATERIALS AND METHODS

### 2.1 Dataset

Nearly 189 drugs as dataset was utilized where they are reportedas thyroid inhibitors, downloaded from Malady cards database [9].It was observed that few drugs come under other disease conditions; however,involved in the dataset because they are known to representin severalother diseases including thyroid disease. Experimentally drugs and which are in pre-clinical stages are excluded from the dataset.

**Manuscript published on 30 December 2019.**
***Correspondence Author(s)
**Katikireddy Srinivas,** Research Scholar (13303051), Department of CSE, KL Deemed to be University, Vaddeswaram, Andhra Pradesh.
**Dr.K.V.D. Kiran,** Professor, Department of CSE, KL Deemed to be University, Vaddeswaram, Andhra Pradesh, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

### 2.2 R

Outcome of data in biology realized the significance of data mining in the framework of convolution of bio systems. Integration of data made few aspects conceivable with the initiation of computer-aided tools, statistical modules and related softwares [10].

R program is freely available software available in an environment employing object oriented programming mainly focussed for statistical computing as well as graphics. NumerousR packages have been developed since many years to report the precise needs of data integration, process and analysis, manipulating data etc by means of statistical methods, such as network modelling, clustering and graph visualization etc. [11].

### 2.3 Hopkins statistic

*Hopkins statistic* [12] assesses the tendency to cluster a dataset which measures the possibility that a given dataset is created by uniformly divided data that means it would test the data spatial randomness [13].

*Hopkins statistic* (H) is considered as mean distance between nearer neighboring data points in a dataset which is divided by the summation mean distances of nearer neighbors in real as well as through the virtual dataset. If *Hopkins statistic*value is near to zero, then the null hypothesis can be rejected and determine that the dataset is considerably clusterable.

### 2.4 Hybrid clustering

K-means cluster method is easiest and the most broadly utilized dividing technique for part a dataset into an arrangement of k groups.The technique utilizes Euclidean separation measures between information focuses to decide the within and the between-group similitudes. The PAM calculation depends on the look for k agent objects or medoids among the perceptions of the dataset. These perceptions ought to speak to the structure of the information. In the wake of finding an arrangement of k medoids, k groups are developed by appointing every perception to the closest medoid. The objective is to discover k agent objects which limit the whole of the dissimilarities of the perceptions to their nearest delegate protest. For a given bunch, the aggregate of the dissimilarities is ascertained utilizing Manhattan distance.

### 2.5 Fuzzy k-means algorithm

The clusterings generated by the k-means techniquecan be referred as either "hard" or "crisp" clusters, since any element vector x either is or isn't an individual from a specific cluster. This is in disparity to "soft" or "fuzzy" methods, wherein a component vector *x* can have a level of enrollment in every cluster. The fuzzy-k-means technique of Dunn and Bezdek (14) permits every component vector *x* to represent a gradation of affiliation in each cluster.

## III. RESULTS AND DISCUSSSION

Clusters and clustering is a technique of data exploration utilized for determining assemblies or arrangement in a dataset. The maximum frequently employed partitioning algorithmic aproaches are K-means and K-medoids clustering [15] or Partitioning Around Medoids [16].In general, clustering process is characterized as gathering objects in sets, wherein the objects in a particular group are as alike as possible,however, objects from diversegroups are as dissimilar as likely. A good clustering will produce groups with a high intra-class likeness and a low between class similitude. The dataset considered here, columns represent variables and rows are treated as observations (i.e., samples). Data was examined before applying separation measures and consequently some spellbinding insights, for example, the mean and the standard deviation of the factors were processed. From the examination it was seen that the factors have extensive contrast in means and differences. This is because of the way that the factors are estimated in various units. Consequently, they ought to be standardized to make them practically identical. Standardization comprises of changing the factors to such an extent that they have mean zero and standard deviation one. To assess the cluster ability of the dataset, *Hopkins statistic*was employed where the value is significantly < 0.5, indicating that the data is highly clusterable.

### 3.1Cluster Validity

In a book written by Theodoridis and Koutroubas (2008) [17], three ways to deal with research on validity of clusters are defined. The first depends on outer criteria, which comprise in contrasting the consequences of group examination with remotely referred class labels. The second methodology depends on interior criteria, which utilize the data got from inside the grouping procedure to assess how well the consequences of cluster analysis fit the information without reference to outside data. The third methodology of depends on relative criteria, which comprises in the assessment of a bunching structure by contrasting it and other clustering schemes. However, in this work, NbClust was engaged which wasintegrated with 30 validity indices to examine the numbers of groups in a given dataset. Therefore, from this analysis, the outcome signified that about 13 different index programssuggestedthree clusters as optimumwhereas eleven index programs suggested twogroups and 4 indices reportedfour clusters. As per the majority ruling method, the best output referred to three groups. Hence, it can be established that the optimum numbers of clustering groups, *k* for the given dataset comprised ofvarious drugs involved in thyroid disease was three cluster results. So, an initial k=3 value was utilized to achieve k-means, PAM as well as hybrid algorithm on the thyroid dataset.

### 3.2 k-means algorithm

The k-means approach is a partitioning problem, wherein the data segregatedas groups with everyrepetition of the algorithm. Since the assignments were started at random, n start = 25 is specified, which means that the program shallattempt 25 various random starting points and then chose the result with lowermost within cluster disparity(Figure 1).A better cluster shall result in values with minimum within ss and bigger between ss which further relies on the sum of k clusters selectedoriginally. Henceforth, low within ss and high between ss for k=3 was obtained.



Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

Within cluster sum of squares by cluster:
[1] 674.8948 1555.1192 2486.3922
(between_SS / total_SS= 24.0 %)
>km.res$betweenss
[1] 1487.594

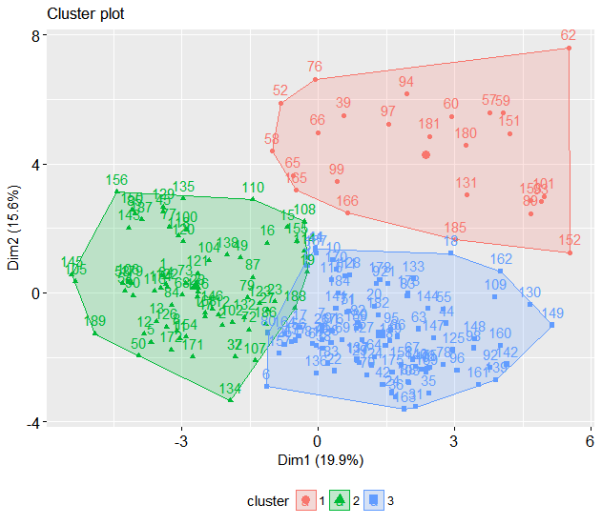


**Figure 1: Output of 3 clusters and cluster centers obtained from kmeans program.**

### 3.3 Partitioning Around Medoids

It was reported that outliers influence the outcome of k-means cluster result which would otherwiseinfluence the task of cluster annotations. Hence, anew,strong algorithm is presented by PAM algorithm, also referred as k-medoids.

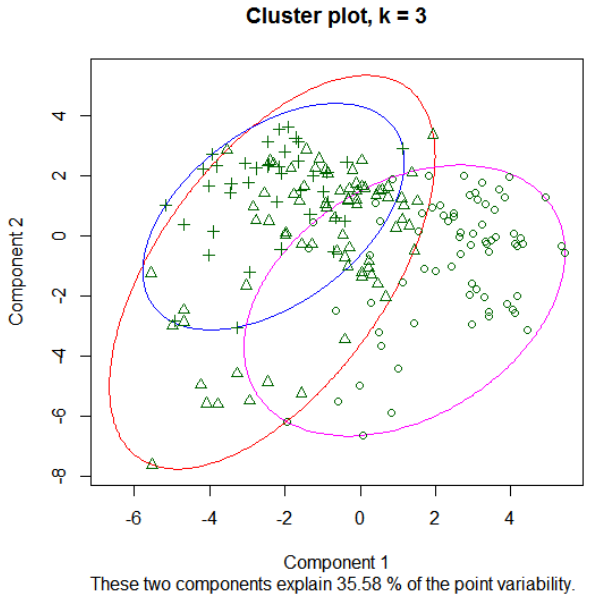


**Figure 2: Cluster plot obtained for PAM algorithm.**

From both methods, it was observed that some samples have a negative silhouette. This means that they are not in the right cluster. On contrast between k-means versus PAM, k-meansensuedsilhouettes around 13 which are negative whereas PAM resulted in 27, respectively.

About 189 data grouped as 3 clusters by kmeans and PAM clustering is compared below and the disparity in subjecting groups is demonstrated.

Clustering vector - kmeans:
[1] 2 2 3 3 2 3 3 2 3 3 2 2 2 3 2 2 3 3 2 3 3 3 2 3 3 3 2 2 3 3 3 2 3 2 3 3 2
[38] 3 1 3 3 3 3 3 2 3 2 3 2 2 3 1 2 2 3 3 1 1 1 1 3 1 3 3 1 1 3 2 3 3 3 2 2 3
[75] 3 1 2 3 2 3 3 3 3 2 2 3 2 3 1 2 3 3 1 1 3 3 1 3 1 2 1 2 3 2 2 2 2 2 3 2 3
[112] 2 2 2 3 2 3 2 3 2 2 2 2 3 3 2 3 3 2 3 1 2 3 2 2 3 3 2 3 3 3 3 2 3 2 2 3 3
[149] 3 2 1 1 1 2 2 2 3 3 3 3 3 3 3 3 1 1 3 2 3 3 2 2 3 2 3 3 3 3 2 1 1 3 3 3 1
[186] 2 2 2 2
Clustering vector - PAM:
[1] 1 1 2 2 1 3 2 1 2 2 1 1 1 2 2 1 1 3 1 1 2 3 2 3 2 2 2 1 2 2 3 1 2 1 3 3 1
[38] 2 1 2 2 2 2 2 1 3 1 2 1 1 2 1 1 1 2 2 2 1 2 2 2 2 3 3 1 1 3 1 2 2 3 2 1 2
[75] 3 1 1 3 1 2 2 2 2 1 1 2 2 3 2 1 2 3 3 1 2 3 2 3 2 1 2 1 3 1 1 1 2 1 3 1 2
[112] 1 1 2 2 1 3 1 3 1 1 1 2 2 3 1 3 3 1 3 3 1 3 2 1 2 2 1 3 3 3 3 1 2 1 1 2 3
[149] 3 1 2 2 2 1 2 1 1 2 3 3 2 3 3 2 1 1 2 1 2 2 1 1 2 1 3 2 2 1 1 2 2 3 2 2 2
[186] 2 1 1 1

Finally, the cluster sizes resulted for kmeans algorithm and PAM methods revealed that both the methods resulted in varied clusters.

**Clusters defined by k-means clustering:**

| Group | | Data | average silhouette width |
|---|---|---|---|
| 1 | 1 | 25 | 0.17 |
| 2 | 2 | 70 | 0.18 |
| 3 | 3 | 94 | 0.10 |

**Clusters defined by PAM clustering:**

| Group | Data | average silhouette width |
|---|---|---|
| 1 | 1 | 72 | 0.14 |
| 2 | 2 | 76 | 0.04 |
| 3 | 3 | 41 | 0.11 |

### 3.4 Hybrid kmeans-PAM Clustering Algorithm

Algorithms implemented in various clustering procedures split the dataset as abundant assemblies that generally outcomes in dispensing the data points in few groups representing a notch of similarities as conceivable and the points in other clusters demonstrate dissimilar nature. By and by, these two standard grouping techniques have their very own preferences and confinements. Thus, a novel crossover approach is executed to blend the best ofk-means and PAM clustering. continues in three phases. First it figures k starting medoids as k-groups on the underlying dataset. At that point the PAM group focuses are calculated followed by processing k-means by utilizing cluster centers as the initial *k.*

The 3 clustering resulted groups acquired utilizing k=cluster centers, which are the three groups of PAM calculation, brought about totally three diverse bunch sizes 77, 25 and 87 separately. Curiously, the negative outlines got from mixture strategy are 11 against 13 from k-implies alone and 27 by PAM technique, which recommends the way that half breed technique is useful in removing data from clusters (Table 1).


*Published By:*
*Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)*
*© Copyright: All rights reserved.*

**Table 1: Three cluster groups appeared from kmeans and PAM methods.**

| | Group1 | Group2 | Group3 | Size of Hybrid cluster |
|---|---|---|---|---|
| Group1 | 59 | 18 | 0 | **77** |
| Group2 | 9 | 14 | 2 | **25** |
| Group3 | 4 | 44 | 39 | **87** |
| Size of PAM cluster | **72** | **76** | **41** | |

It is worth to take note of that the individual k-means calculation could group dataset as 25, 70 and 94 gatherings while the hybrid kmeans- kmedoids brought about comparable cluster of size 25 and remaining being 87 and 77. This data can be seen graphically (Figure 3).

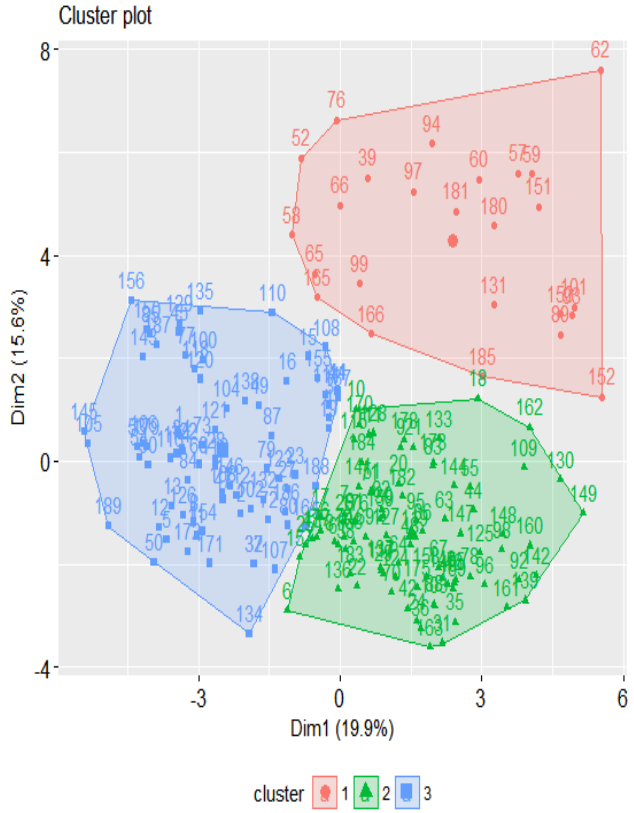


**Figure 3: Visual representation of 3 clusters and centers resulted from hybrid kmeans-kmedoids method.**

It was observed from the plots of hybrid clustering algorithm that the data points at the edge of clusters 2 and 3 were found to be overlapped and efficient clustering was not possible when plots are visualized.

Hence, work was initiated to introduce k-means procedure coupled with fuzzy algorithm. Fuzzy being soft clustering procedure mixed with hard clustering k-means, reported as fuzzy k-means (FKM) algorithm in order to produce meaningful clusters.

**Table 1: A subset of 189 thyroid drug-parameters dataset selected for fuzzy k-means analysis.**

| **Drug** | **Polar Surface Area** | **Refractivity** |
|---|---|---|
| Propylthiouracil | 41.13 | 48.9 |
| Hydrocortisone | 94.83 | 97.4 |
| Prednisone | 91.67 | 97.57 |
| Sevoflurane | 9.23 | 23.3 |
| Etoricoxib | 59.92 | 95.04 |
| Propofol | 20.23 | 56.42 |
| Remifentanil | 76.15 | 100.56 |
| Methylprednisolone | 94.83 | 103.04 |
| Menthol | 20.23 | 47.45 |
| Acetylcholine | 26.3 | 51.35 |
| Benzocaine | 52.32 | 47.53 |
| Dexamethasone | 94.83 | 102.49 |
| Triamcinolone | 115.06 | 99.38 |
| DinoprostTromethamine | 97.99 | 100.47 |
| Diclofenac | 49.33 | 75.46 |
| Travoprost | 96.22 | 127.86 |
| Timolol | 79.74 | 83.92 |
| Bimatoprost | 89.79 | 122.83 |
| Latanoprost | 86.99 | 124.34 |
| Dexmedetomidine | 28.68 | 62.98 |
| Metoprolol | 50.72 | 76.7 |
| Tizanidine | 62.2 | 64.77 |
| Clonidine | 36.42 | 59.09 |
| Methyltestosterone | 37.3 | 89.07 |
| Estradiol | 40.46 | 79.9 |

### 3.5 Fuzzy k-means algorithm:

A subset comprising 25 data points (Table 1) from the 189 thyroid drug-parameter dataset was subjected to fkm algorithm and the output graph is reported in Figure 4. It is evidenced that the program is able to cluster 3 sets with clear demarcation.

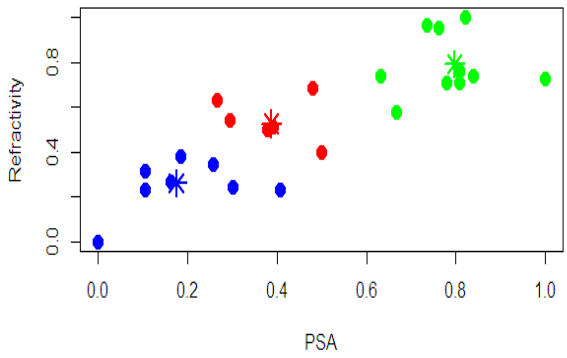


**Figure 4: Dataset clusters via FKM algorithm showing 3 distinct clusters.**


Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

### 3.6 Fuzzy k-means via entropy regularization

The method is able to perform the fuzzy k-means clustering with entropy regularization. The entropy regularization avoids using artificial fuzziness parameter *m*. Instead of *m*, a degree of fuzzy entropy *ent*, similar to the notion of temperature in statistical physics, is provided [18]. An excitingstuffregarding the fuzzy k-means via entropy regularization is that the models are gotten as weighted means with weights equivalent to the enrollment degrees (instead of to the participation degrees at the intensity of m as is for the fuzzy k-means). It is observed from Figure 5 that few objects from one cluster appeared in other cluster groups.

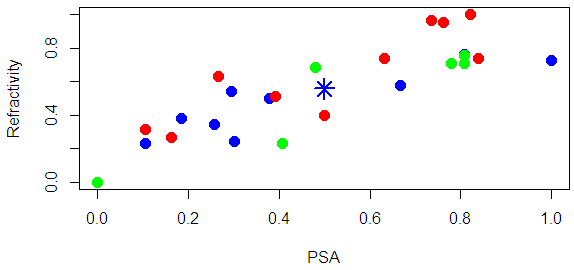


**Figure 5: Fuzzy k-means with entropy regularization showing 3 clusters but few objects found to be mixed with other cluster groups.**

### 3.7 Fuzzy k-means via entropy regularization plus *noise cluster*

The entropy regularization avoided using the artificial fuzziness parameter *m*. The noise cluster is an additional cluster (with respect to the k standard clusters) such that objects recognized to be outliers are assigned to it with high membership degrees [19].

From figure-6 it is evident that Fuzzy k-means with entropy regularization and noise cluster is able to group objects effectively, however, the method failed to reproduce the result with respect to few objects being distributed in other clusters.

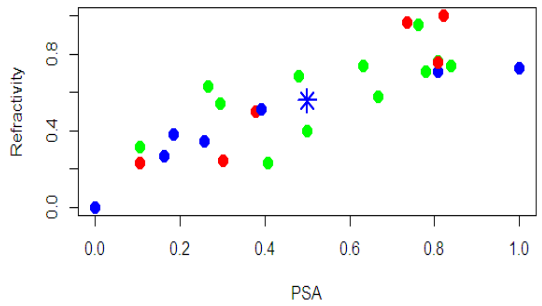


**Figure 6: Fuzzy k-means with entropy regularization and noise cluster showing 3 clusters with one object appearing in another cluster.**

### 3.8 Gustafson and Kessel-like fuzzy k-means

The program performs the Gustafson and Kessel-like fuzzy k-means clustering algorithm and is convenient to determineclusters of non-spherical size [20]. The program was able to determine better cluster groups except at two data points of cluster3 (Figure 7).

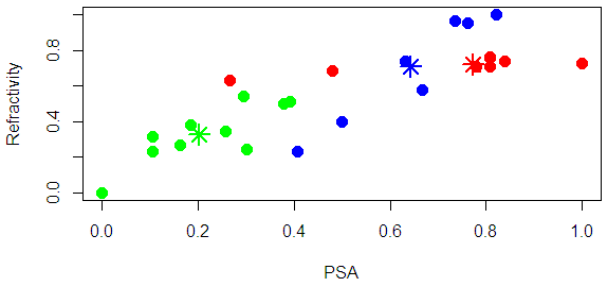


**Figure 7: Gustafson and Kessel - like fuzzy k-means clustering algorithm showing distributed data**

### 3.9 Gustafson and Kessel-like fuzzy k-means via entropy regularization

The program performs the Gustafson and Kessel - like fuzzy k-means clustering algorithm with entropy regularization [21]. The methodpermits to evadeutilizing the artificial fuzziness parameter *m*. If standardization is set to *stand=1*, the algorithm runs based on standard data. Figure 8 suggested that the data was discrete and the program unable to identify and cluster better possibilities.

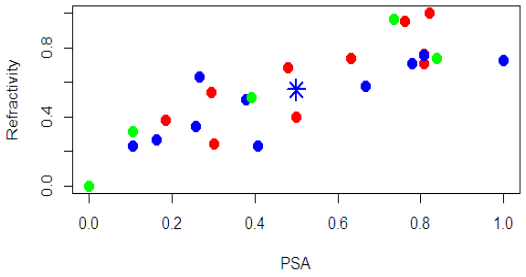


**Figure 8: Gustafson and Kessel-like fuzzy k-means clustering algorithm with entropy regularization showing distributed clusters.**

### 3.10 Gustafson and Kessel-like fuzzy k-means using entropy regularization plus noise cluster

The program runs the Gustafson and Kessel-like fuzzy k-means clusters using entropy regularization and noise cluster which is different from fuzzy k-means, and the method identifies non-spherical clusters.

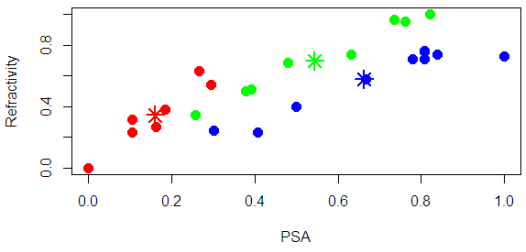


**Figure 9: Gustafson and Kessel - like fuzzy k-means clustering algorithm with entropy regularization and noise cluster resulted in better clusters.**


*Published By:*
*Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)*
*© Copyright: All rights reserved.*

Of all variations in FKM algorithms presented here, only natural FKM algorithm is able to produce estimated three better cluster solutions. Hence, it should be noted that testing all possibilities should be made before proceeding with allied variations of algorithms.

### 3.11 Validation Studies

Several validation studies has been proposed in clustering datasets, however, not all validation procedures are required in one aspect. Hence, following few validation statistics are presented here, given in Table 2.

**Table 2: Validation indexes tested for fuzzy k-means**

| Algorithm | PC | PE | MPC | SIL | SIL.F | XB |
|---|---|---|---|---|---|---|
| Value (Max/Min) | Max | Min | Max | Max | Max | Min |
| Fkm | 0.7716 | 0.4205 | 0.6574 | 0.6863 | 0.7465 | 0.1519 |
| fkm.ent | 0.3333 | 1.0986 | 0.0000 | ND | ND | 3.9646 |
| fkm.ent.noise | 0.3333 | 1.0986 | 0.0004 | ND | ND | 3.4139 |
| fkm.gk | 0.7848 | 0.3985 | 0.6772 | 0.2449 | 0.3364 | 2.5180 |
| fkm.gk.ent | 0.3333 | 1.0986 | 0.0000 | ND | ND | 1.7843 |
| fkm.gk.ent.noise | 0.8971 | 0.0631 | 0.8457 | 0.1077 | 0.0947 | 2.3210 |

Partition Coefficient index (PC), Partition Entropy index (PE), Modified partition coefficient index (MPC), Silhouette index (SIL), Fuzzy silhouette index (SIL.F), Xie and Beni index (XB)

From table 2, it is evidenced that the fuzzy k-means algorithm resulted in parameter values within the limits. On a comparative note, examining Table 2, suggests that the validity index values of fkm-gk with entropy-noise algorithm appear to be within the limits for first 3 indices such as PC, PE and MPC against fkm algorithm validity index values. Similarly, SIL, SIL.F and XB validity index tests are passed by fkm algorithm than fkm-gk with entropy-noise algorithm.

However, it is worth to note and compare Figures 4 and 9, given here. Careful observation of both the plots revealed that the 3 clusters appear in fkm algorithm and fkm-gk with entropy-noise are convincing. The only difference is in the range of data points being considered as a cluster. However, both the methods reported well defined clusters. Comparatively, fkm algorithm displayed superior separation of clusters with well-defined data points.

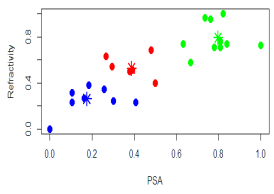


**Figure 4: Dataset clusters via FKM algorithm showing 3 distinct clusters.**

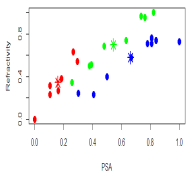


**Figure 9: Gustafson and Kessel - like fuzzy k-means clustering algorithm with entropy regularization and noise cluster resulted in better clusters.**

## IV. CONCLUSION

From both individual k-means and k-medoids methods, it was observed that some samples reported negative silhouettes. On comparison between k-means and PAM, the former resulted in 13 negative silhouettes whereas PAM method resulted in 27 negative silhouettes and similar is the observation with cluster size. Moreover, overlapping of clusters was observed in each case as well as in hybrid method. Hence, a set of six fuzzy algorithm variants studied on a subset of thyroid dataset resulted in 3 distinct clusters by fuzzy k-means followed by Gustafson and Kessel - like fuzzy k-means with entropy regularization and noise cluster algorithm.

## REFERENCES


1. Hastie T, Tibshirani R, Friedman J. Unsupervised learning. In The elements of statistical learning 2009 (pp. 485-585).Springer New York
2. J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability"*, Berkeley, University of California Press, 1:281-297
3. Kaufman, L. and Rousseeuw, P.J. (1987), Clustering by means of Medoids, in Statistical Data Analysis Based on the L1 - Norm and Related Methods, edited by Y. Dodge, North- Holland, 405–416
4. Hartigan, J. A.; Wong, M. A. (1979), "Algorithm AS 136: A K-Means Clustering Algorithm". Journal of the Royal Statistical Society, Series C 28 (1): 100–108
5. Milligan GW, Cooper MC. An examination of procedures for determining the number of clusters in a data set. Psychometrika. 1985 Jun 27;50(2):159-79
6. Fraley C, Raftery AE. How many clusters? Which clustering method? Answers via model-based cluster analysis.The computer journal. 1998 Jan 1;41(8):578-88
7. MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, eds L. M. Le Cam & J. Neyman, 1, pp. 281–297. Berkeley, CA: University of California Press.
8. Halkidi M, Batistakis Y, Vazirgiannis M. On clustering validation techniques.Journal of intelligent information systems. 2001 Dec 1;17(2):107-45
9. http://www.malacards.org/
10. Nolwenn Le Meur and Robert Gentleman. Analyzing Biological Data Using R: Methods for Graphs and Networks.Chapter 19
11. Huber W, Carey VJ, Long L, Falcon S, Gentleman R. (2007) Graphs in molecular biology. BMC Bioinformatics, 8(6):S8
12. Hopkins, Brian; Skellam, John Gordon (1954)."A new method for determining the type of distribution of plant individuals".Annals of Botany. Annals Botany Co. 18 (2): 213–227
13. Banerjee, A. (2004). "Validating clusters using the Hopkins statistic". IEEE International Conference on Fuzzy Systems: 149–153
14. J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms* (Plenum Press, New York, 1981
15. MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, eds L. M. Le Cam & J. Neyman, 1, pp. 281–297. Berkeley, CA: University of California Press.
16. Kaufman, L. and Rousseeuw, P.J. (1990). Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York
17. Theodoridis S, Koutroubas K (2008).Pattern Recognition.4th edition. Academic Press.
18. Li R., Mukaidono M., 1995. A maximum entropy approach to fuzzy clustering. Proceedings of the Fourth IEEE Conference on Fuzzy Systems (FUZZ-IEEE/IFES '95), pp. 2227-2232
19. Dave' R.N., 1991. Characterization and detection of noise in clustering. Pattern Recognition Letters, 12, 657-664
20. Gustafson E.E., Kessel W.C., 1978. Fuzzy clustering with a fuzzy covariance matrix. Proceedings of the IEEE Conference on Decision and Control, pp. 761-766
21. Ferraro M.B., Giordani P., 2013. A new fuzzy clustering algorithm with entropy regularization. Proceedings of the meeting on Classification and Data Analysis (CLADAG).




*Published By:*
*Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)*
*© Copyright: All rights reserved.*