



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

(Open Access, Double Blind Peer-reviewed, Refereed and Indexed Journal)

www.iasir.net

A Statistical Approach to Fuzzy Number Clustering: A Review

Aditya Kumar Anand

Research Scholar, Department of Mathematics

L. N. M. U., Darbhanga, Bihar, INDIA.

Abstract: This paper begins with the literature review of the fuzzy clustering, cluster process, and cluster methods. Cluster is said to be formed when the members in a group shares common characteristics among themselves. In this paper, clustering of different sets of fuzzy number has been performed using two way method. In the first step, degree of similarity is calculated between the fuzzy number sets. In the second step, statistical clustering is performed depending upon the degree of similarity value obtained in first step. Same cluster is formed if the value obtained is very near to each other. It is found that, this method can be a practical approach to make a fuzzy cluster.

Keywords: Fuzzy Cluster, Cluster Method, Process, Necessity, Algorithm, Degree of Similarity

I. Introduction

A Cluster is said to be formed, if the members present in the group shares the common characteristics with each other. A member in a cluster do not dissimilar to another member present in the group. There may be many cluster present in a cluster, which can be very similar to each other, i.e., they resemble to other members in the group than the member present in other group. Cluster Analysis is a convenient method for identifying the clusters. It is also used to group the variables present in a cluster into homogeneous and the distinct groups.

Cluster Analysis can be understood easily with the use of the following real time example. In a company, there is mandatory for its employee to be in the office for 9 hours; that is every employee swipe in and swipe out time should be equal to or more than 9 hours, if the time detail for each employee is not equal to or less than 9 hours than the escalations can happen which on further escalations can leads to the termination of the employee.

Now, let assume that there are 100 employee present in this office and the table below present the time detail of group of 20 employee in each Clusters is shown in Table 1.

Table 1 Timing Details of Employees

S. No.	Cluster	Swipe-In (Before)	Swipe-Out (After)	Office Hours
1.	Cluster 1	8:00 AM	5:00 PM	9
2.	Cluster 2	9:00 AM	6:00 PM	9
3.	Cluster 3	10:00 AM	7:00 PM	9
4.	Cluster 4	11:00 AM	8:00 PM	9
5.	Cluster 5	12:00 PM	9:00 PM	9

In Table 1, it is seen that each employee is in the same company and they completes 9 hours in the office but still, the company has categorized them into different clusters depending on their entry office hours. Here, grouping employee into one cluster is easy because of accurate data captured through the RFID card as well the time of entry but in real scenario, when it comes to variables than it if one of the difficult task to handle. To minimize the complexity, different algorithms has been designed and many works are going on around the globe to optimize the cluster analysis process of variables and the objects.

In this paper first of all introduction to cluster has been done. In section II, necessity of cluster has been discussed. Section III details about the cluster process. In section IV cluster procedure has been briefly described. Cluster Algorithm, degree of similarity and the cluster formation has been presented in section V, section Vi and section VII respectively. Further section presents the conclusion and references.

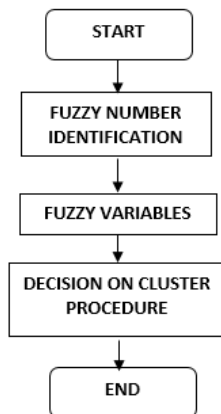
II. Necessity

The objective of cluster analysis is the classification of the variables into a group to form a cluster depending upon the similarity among them. Cluster analysis also helps in organizing the data present in the group. It is one of the unsupervised method which do not use any prior class identifiers. The main potential of clustering is to detect the underlying structure in data, not only for classification, but also for model reduction and optimization.

III. Process

Clustering begins with the identification of the group of fuzzy numbers that are very similar to each other. After which, decision on cluster variable is decided. When the cluster variables are decided, then the cluster procedure is chosen, which will be implemented to form the cluster. These steps are very crucial for the analysis, as different procedures require different decisions prior to analysis. Figure 1 shows the flow chart representation of the cluster process.

Figure 1 Flowchart of Cluster Process



IV. Cluster Procedure

In this section, Hierarchical and partitioning method of cluster procedure has been described.

A. Hierarchical Cluster Procedure

Hierarchical clustering is the process of cluster method in which tree-like cluster structure is formed. Most of the hierarchical method forms the agglomerative clustering meaning the cluster is consecutively formed from objects. It starts with individual objects representing a clusters and gradually these clusters are merged on the basis of characteristics and similarity. It follows a bottom-up approach. Another, way to make a cluster can be based on top-down approach. This method is called as divisive clustering. In this approach, first all objects are into one cluster then they are separated or splitted on the basis of similar characteristics. In hierarchical clustering, once a members is assigned to a cluster it can't be re-assigned to another cluster unlike the partitioning method of clustering.

B. Partitioning Cluster Procedure

This is the another procedure of cluster procedure In this cluster procedure, first the object is assigned to number of cluster then successively reassigned to other clusters to minimize the within-cluster variation, which is basically the (squared) distance from each observation to the center of the associated cluster. If the reallocation of an object to another cluster decreases the within-cluster variation, this object is reassigned to that cluster.

V. Cluster Algorithm

This step is applied, once the cluster procedure has been decided. In this step, decision on which algorithm has to be applied is taken. There are numerous of agglomerative procedures. They are differentiated on the basis of they define the distance from a newly formed cluster to a certain object, or to other clusters in the group. Some of the most important agglomerative clustering procedures are as follows:

- i. **Single linkage (nearest neighbor):** The distance between two clusters is the shortest distance between any two members in the two clusters.
- ii. **Complete linkage (furthest neighbor):** The distance between two clusters is based on the longest distance between any two members in the two clusters.
- iii. **Average linkage:** The distance between two clusters is defined as the average distance between all pairs of the two clusters' members.
- iv. **Centroid:** In this approach, first the centroid of each cluster is calculated. Then the distance between the two clusters equals the distance between the two centroids.
- v.

VI. Degree of Similarity

A similarity measure is a function $S: TF[0,1] \times TF[0,1] \rightarrow [0,1]$; indicating the degree of similarity between two fuzzy numbers. This value must match the intuitive perception that exist between the fuzzy numbers. The closer this value is to 1, the more similar the fuzzy numbers will be to each other.

Let, there be two trapezoidal fuzzy numbers represented as $A = (a_1, a_2, a_3, a_4, h_A)$ and $B = (b_1, b_2, b_3, b_4, h_B)$.

Let $0 \leq x \leq 1$ and $0 \leq y \leq 1$. The similarity between x and y is defined as:

$$S(x, y) = 1 - |\sqrt{x} - \sqrt{y}|$$

With this method of calculating the similarity measure, small absolute of difference corresponds to a high degree of similarity, whereas two components with a large absolute of difference are considered to be very dissimilar.

Now, the degree of similarity between the fuzzy numbers can be calculated using the formula:

$$S(A, B) = \left(S(h_A, h_B) \times \prod_{i=1}^4 S(a_i, b_i) \right)^{\frac{1}{3}}$$

VII. Cluster Formation

In this section, clustering of 14 different fuzzy sets presented in Table 2 has been performed. As discussed in the earlier section, two way approach has been implemented in order to form the cluster of fuzzy sets. In the first step degree of similarity is obtained using the definitions and methods described in the section VI. Depending upon the obtained value of the degree of similarity, clustering of fuzzy numbers has been performed. Table 3 presents the cluster formed using the method described.

Table 2 Fuzzy Sets of Numbers

S. No.	A	B
1.	(0.2, 0.3, 0.5, 0.6; 1.0)	(0.2, 0.3, 0.5, 0.6; 1.0)
2.	(0.2, 0.3, 0.5, 0.6; 1.0)	(0.2, 0.4, 0.4, 0.6; 1.0)
3.	(0.2, 0.3, 0.5, 0.6; 1.0)	(0.6, 0.8, 0.8, 1; 1.0)
4.	(0.2, 0.3, 0.5, 0.6; 1.0)	(0.6, 0.7, 0.9, 1; 1.0)
5.	(0.2, 0.3, 0.5, 0.6; 1.0)	(0.2, 0.3, 0.5, 0.6; 0.8)
6.	(0.5, 0.5, 0.5, 0.5; 1.0)	(0.5, 0.5, 0.5, 0.5; 1.0)
7.	(0.5, 0.5, 0.5, 0.5; 1.0)	(0.5, 0.5, 0.5, 0.5; 0.8)
8.	(0.3, 0.3, 0.3, 0.3; 1.0)	(0.5, 0.5, 0.5, 0.5; 1.0)
9.	(0.3, 0.3, 0.4, 0.4; 1.0)	(0.5, 0.5, 0.7, 0.7; 1.0)
10.	(0.2, 0.2, 0.3, 0.3; 1.0)	(0.5, 0.5, 0.6, 0.6; 1)
11.	(0.1, 0.2, 0.3, 0.4; 1.0)	(0.4, 0.5, 0.6, 0.7; 1)
12.	(0.2, 0.3, 0.4, 0.5; 1.0)	(0.2, 0.2, 0.2, 0.5; 1)
13.	(0.2, 0.3, 0.4, 0.5; 1.0)	(0.1, 0.4, 0.4, 0.4; 1)
14.	(0, 0, 0.3, 0.3; 1.0)	(0.2, 0.4, 0.4, 0.5; 1)
15.	(0.2, 0.3, 0.5, 0.6, 1)	(0.3, 0.4, 0.4, 0.5, 1)

Table 3 Cluster Formation on Basis of Degree of Similarity

Cluster	Fuzzy Set	S(A,B)
A.	1.	1
	6.	1
B.	5.	0.964
	2.	0.947
	7.	0.946
	13.	0.903
	12.	0.902
	15.	0.893
C.	8.	0.794
	9.	0.765
D.	10.	0.69
	11.	0.678
	3.	0.652
	4.	0.655
	14.	0.539

VIII. Conclusion

In this paper, review about the fuzzy clustering has been presented. It begins with the knowledge about the cluster analysis and the necessity of clustering. Process of fuzzy cluster has also been explained. Two step method has been used to cluster variables. Two different cluster methods i.e., hierarchical and partitioning cluster procedure has been described. Different algorithm has also been explained. A new cluster approach on the basis of degree of similarity has been proposed.

IX. References

- [1] M. S. Aldererfer and R. K. Blasfield, "Cluster Ananlysis", Sage Publications, Los Angeles, 1985.
- [2] M. R. Anderberg, "Cluster ANalysis for Applications", Academic Pree, New York, December 1773.
- [3] B. S. Everitt, S. Landau, and M. Leese, "Cluster Ananlysis", Arnold Publishers, London fourth edition, May 2001.
- [4] D. Fisher, "Iterative Optimization and Simplification of Hierarchical Clusterings", Journal of Artificial Intelligence Research, 4:147-179, 1996.
- [5] B. Zhang, M. Hsu and U. Dayal, "K. Harmonic Means- A data clustering Algorithm", Technical Report HPL 1999-124, Hewlett Packard Laboratories, Oct. 29, 1999.
- [6] M. Sarstedt and E. Mooi, A Concise Guide to Market Research, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-53965-7_9, Springer-Verlag Berlin Heidelberg 2014.