

HASIL CEK_Bapak Sugiyarto_3

by Cek_bapak Sugiyarto Bapak Sugiyarto

Submission date: 19-Apr-2021 09:23AM (UTC+0700)

Submission ID: 1563006068

File name: 17012020_MEDIA_STATISTIKA_SGY_ANNISA.docx (94.57K)

Word count: 2649

Character count: 15955

COMPARATIVE STUDY OF DISTANCE MEASURES ON FUZZY SUBSTRUCTIVE CLUSTERING

Annisa Eka Haryati¹, Sugiyarto Surono²

¹ Magister Pendidikan Matematika, Universitas Ahmad Dahlan

² Matematika, Universitas Ahmad Dahlan

²
e-mail: sugiyarto@math.uad.ac.id

DOI: 10.14710/medstat.XX.X.XX-XX

Article Info:

Received:

Accepted:

Available Online:

Keywords:

Fuzzy Subtractive Clustering,
Hamming, Minkowski Chebyshev

Abstract: Clustering is a data analysis process which applied to classify the unlabeled data. Fuzzy clustering is a clustering method based on membership value which enclosing set of fuzzy as a measurement base for classification process. Fuzzy Subtractive Clustering (FSC) is included in one of fuzzy clustering method. This research applies Hamming distance and Minkowski Chebyshev combination distance as a distance parameter in Fuzzy Subtractive Clustering. The objective of this research is to compare the output quality of the cluster from Fuzzy Subtractive Clustering by using Hamming distance and Minkowski Chebyshev combination distance. This research shows that the Partition Coefficient value resulted on Fuzzy Subtractive Clustering by applying Minkowski Chebyshev distance generally higher than the application of Hamming distance. This concludes that the output quality of the cluster by using Minkowski Chebyshev distance is better.

1. INTRODUCTION

Clustering is a data analysis process which applied to classify the unlabeled data (Gan et al., 2007). In the optimum classification, each set of data will have a high percentage of similarity on certain cluster. Principally, cluster analysis will classify the given set of data which have high similarity into the same cluster and the dissimilar set of data into different cluster (Rencher, 2016).

Fuzzy clustering is a clustering method which applied to determine the set of data into certain cluster based on their membership value (Sharma & Verma, 2019). There are no certainty on each data to be included in one certain cluster. Hence, each data will have the possibility to be included as a member in different cluster (Jang et al., 2005). There are few known method for fuzzy clustering process, which are Fuzzy C-Means method and Fuzzy Subtractive Clustering.

Fuzzy Subtractive Clustering is one of the clustering method where the number of cluster classification is still undetermined yet. The basic concept of this method is to establish the coordinate of each data which possess the highest number of density value. Set of coordinate which have the highest number of neighboring coordinates will be chosen as the cluster centroid and its density values will get reduced (Sangadji et al., 2018). In this method, the obtained number of cluster are affected by certain parameter, which is the radius (Dyvak et al., 2018).

Few researches has been done by applying Fuzzy Subtractive Clustering, such as the research which done by (Ghane'i Ostad et al., 2018) to determine the overlapped community on LBSN. In this research, every cluster centroid will be defined by the application of fuzzy Subtracting Clustering based on each set of data potentials. This research concludes that the given method have higher accuracy than the previous used method.

The research which analyzed by (Salah et al., 2019) is applying subtractive classification which combined with swam particle optimization based on fuzzy classifier. Furthermore, (Abdolkarimi & Mosavi, 2020) also do a research related with fuzzy inference system for subtractive clustering which combined with wavelet for improving the navigation system. Other than that, (Banteng et al., 2019) and (Zeng et al., 2019) also applying subtractive clustering to analyze ad hoc cellular network based on information criteria for forecast process based on linear combination with independent variables.

Another research also worked by (Benmouiza & Cheknane, 2019) with the objective to forecast the solar radiation value by applying subtractive clustering, fuzzy c-means, and network partition. Three clustering methods used in this research is done for classifying every given set of data into certain classification so that each data included on certain classification have the high percentage of similarity to improve the understanding of the correlation between each data and to simplify the forecasting process.

In Fuzzy Subtractive Clustering method, the similarity measurement to determine the number of points which have the highest number of neighboring points is needed. The most frequently used distance parameter to determine the similarity measurement is Euclidean distance. In (Rezaei & Rezaei, 2019), the most known distance measurement to measure the distance between two set of fuzzy clusters is Hamming distance. This statement became the basis on the application of Hamming distance for this research. Other than the Hamming distance, the Minkowski Chebisev combination distance application also had been analyzed by (Rodrigues, 2018) for classifying process by using K-Nearest Neighbors (KNN) and (Surono & Putri, 2020) for clustering with Fuzzy C-Means and Principal Component Analysis application. Based on the statement below, this research will compare the achieved result through the application of Hamming distance and Minkowski Chebisev combination distance for clustering process using Fuzzy Subtractive Clustering method.

2. LITERATURE REVIEW

2.1. Fuzzy Subtractive Clustering

Fuzzy clustering is a clustering method based on the membership values which enclose fuzzy cluster as a clustering measurement basis. Every set of data are provided with probability value for their classification into existing group, this means that this each of data is not absolutely included in only one cluster and they will have a probability value to be classified in different cluster which have the highest percentage of membership level.

Fuzzy Subtractive Clustering (FSC) is one of the fuzzy clustering method where the number of existing cluster is unestablished yet. The basic concept of this method is to determine every coordinates of each data which have the highest density with their neighboring points. The coordinate with the highest number of neighboring points will be used as the cluster centroid. Then, the density level of the coordinate which used as a cluster centroid will be reduced and algorithm will determine the different coordinates which have the highest number of neighboring points to become another cluster centroid. This process will be executed until every set of coordinates have been tested.

3. RESEARCH METHODS

3.1. Fuzzy Subtractive Clustering Steps

The method used in this research is Fuzzy Subtractive Clustering (FSC) with distance parameters, namely Hamming distance and Minkowski Chebisev combination distance. The obtained clusters will be evaluated using the Partition Coefficient to see the quality of the clusters. The data used in this study is hypertension patient data. The variables used were age (X_1), gender (X_2), systolic pressure (X_3), diastolic pressure (X_4), and body weight (X_5). Information about the methods used in this study will be described as follows:

- Determine the parameter values, which are, r (radius), q (squash factor), accept ratio, reject ratio.
- Transform the given set of data into fuzzy number by applying membership function (Mahajan & Gupta, 2019) (Debnath & Gupta, 2019):

$$\mu(x) = \begin{cases} 1 & x \leq a \\ e^{-\left(\frac{x-a}{b-a}\right) - e^{-s}} & a \leq x \leq b \\ 0 & x \geq b \end{cases} \quad (1)$$

- Determine the potential of every given set of data D_i ; $i = 1, 2, 3, \dots, n$ with steps as follows:

First, measure the distance for every given set of data by applying:
Hamming distance (Chen & Deng, 2020)

$$Dist_{ij} = \left(\frac{|\mu_A(u_i) - \mu_B(u_i)|}{r} \right) \quad (2)$$

Minkowski Chebisev combination distance (Rodrigues, 2018) :

$$Dist_{ij} = \left(\frac{w_1 \sqrt[p]{\sum_{m=1}^k |x_m - y_m|^p} + w_2 \max_{m=1}^k |x_m - y_m|}{r} \right) \quad (3)$$

Second, determine the initial potential of each data by applying:

$$D_i = \sum_{k=1}^n e^{-4(\sum_{j=1}^m Dist_{ij}^2(x_i))} \quad (4)$$

- Find the data coordinates with the highest level of potential:
 $M = \max[D_i | i = 1, 2, \dots, n]$; for the first iteration.
 $Z = \max[D_i | i = 1, 2, \dots, n]$; for the second, third, and the following iteration.
- Measure the ratio (R) of the cluster centroid candidates by applying the following equation:

$$R = \frac{Z}{M} \quad (5)$$

In initial iteration, value $Z=M$.

- f. Check the suitability of the cluster centroid candidates to become the cluster centroid by using 3 condition as follows:

Condition 1: If $\text{ratio} > \text{accept ratio}$, the cluster centroid candidates are suitable to become the cluster centroid.

Condition 2: If $\text{reject ratio} < \text{ratio} \leq \text{accept ratio}$, further suitability test will be executed to determine whether the cluster centroid candidates are suitable to become the new cluster centroid. If the result of suitability test still below the minimum requirement for cluster centroid suitability value, the iteration process will be terminated because there are no further data available for further consideration. The steps for executing the second condition process are:

$$Md = -1$$

for $k = 1, 2, \dots, p$, where p = the number of cluster.

$$Sd_k = \sum_{j=1}^m \left(\frac{V_j - C_{kj}}{r} \right)^2 \quad (6)$$

If $(Md < 0)$ or $(Sd_k < Md)$, then $Md = Sd_k$,

$$Mds = \sqrt{Md};$$

Where Mds is the nearest distance between the coordinates of the cluster centroid candidates to the cluster centroid. If $(\text{ratio} + Mds) \geq 1$; the cluster centroid candidates are suitable to become the new cluster centroid. Meanwhile, if $(\text{ratio} + Mds) < 1$; then, the cluster centroid candidates are not suitable and won't be reconsidered as the new cluster centroid (potential of the data is set to 0)

Condition 3: if $\text{ratio} \leq \text{reject ratio}$, there will be no further data consideration to become the cluster centroid candidates and the iteration process will be terminated.

- g. If the new cluster centroid has already acquired, the data potential around the existing cluster centroid will be reduced by using equation:

$$D_i^t = D_i^{t-1} - D_{cki} \quad (7)$$

where

$$D_{cki} = Z * e^{-4 \left[\sum_{j=1}^m \left(\frac{C_{kj} - x_{ij}}{r * q} \right)^2 \right]} \quad (8)$$

- h. Retransform the cluster centroid into the its data original form:

$$x = (a - b) \ln(\mu - \mu e^{-s} + e^{-s}) + a \quad (9)$$

- i. Measure the membership degree by using equation:

$$\mu_{ki} = e^{-\sum_{j=1}^m \left(\frac{x_{ij} - C_{kj}}{\sqrt{2} \sigma_j} \right)} \quad (10)$$

where

$$\sigma_j = \frac{r * (X_{max_j} - X_{min_j})}{\sqrt{8}} \quad (11)$$

- j. Determine the validity index by applying Partition Coefficient (Utomo & Marutho, 2018)

$$PC = \frac{1}{N} \left(\sum_{i=1}^N \sum_{j=1}^K \mu_{ij}^2 \right) \quad (12)$$

with N is the number of research object, K is the number of cluster, and μ_{ij} is the membership value of the object “ i ” on cluster centroid “ j ”.

4. RESULT AND DISCUSSION

FSC measurement process will resulting in number of acquired cluster and Partition Coefficient value. The parameter value which applied in this research are $q = 1.25$, accept ratio = 0.8, reject ratio = 0.2, $r = 0.72, 0.79, 0.97, 1.12, 1.31$, and 1.6. Data which used for FSC measurement will be transformed into fuzzy number by applying equation (1) with the result as follows

Table 1 Fuzzy Numbers

X_1	X_2	X_3	X_4	X_5
0	0	0.4609	0.1289	0.4018
0.3913	1	0.5516	0.4785	0.1468
\vdots	\vdots	\vdots	\vdots	\vdots
0.5734	0	0.3775	0.2862	0.6813

Furthermore, clustering by using Hamming distance for each existing radius will resulted in 3 classification; by using $r = 0.72$, 4 clusters will be achieved; by using $r = 0.79$, 3 clusters will be achieved; and by using $r = 0.97$, 2 clusters will be achieved. The achieved cluster centroid are described as follows:

$$C_{0.72} = \begin{bmatrix} 0.4340 & 0 & 0.4609 & 0.2862 & 0.4018 \\ 0.4634 & 1 & 0.4609 & 0.2862 & 0.6501 \\ 0.4196 & 1 & 0.4609 & 0.2862 & 0.1468 \\ 0.3913 & 0 & 0.5144 & 0.1579 & 1 \end{bmatrix}$$

$$C_{0.79} = \begin{bmatrix} 0.4340 & 0 & 0.4609 & 0.2862 & 0.4018 \\ 0.4634 & 1 & 0.4609 & 0.2862 & 0.6501 \\ 0.3114 & 1 & 0.3008 & 0.1289 & 0.0775 \end{bmatrix}$$

$$C_{0.97} = \begin{bmatrix} 0.3373 & 0 & 0.4101 & 0.3775 & 0.4785 \\ 0.4634 & 1 & 0.4609 & 0.2862 & 0.6501 \end{bmatrix}$$

The C matrixes that mentioned below are the result of cluster centroid for each applied radius. The $C_{0.72}$ matrix points the matrix of the acquired cluster centroid with radius 0.72. The first row on $C_{0.72}$ matrix shows the first cluster centroid, the second row shows the second cluster centroid, the third row shows the third cluster centroid, and the fourth row shows the fourth cluster centroid.

Furthermore, the application of Minkowski Chebisev combination distance will resulted in 3 classification for each radius; by using $r = 1.12$, 4 clusters will be achieved; by using $r = 1.31$, 3 clusters will be achieved; and by using $r = 1.6$, 2 clusters will be achieved. The achieved cluster centroid are described as follows:

$$C_{1.12} = \begin{bmatrix} 0 & 0 & 0.4609 & 0.1289 & 0.4018 \\ 0.5900 & 1 & 0.4609 & 0.1289 & 0.5900 \\ 0.5092 & 0 & 1 & 1 & 0.6196 \\ 0.7894 & 0 & 0.7571 & 0.2862 & 0.2439 \end{bmatrix}$$

$$C_{1.31} = \begin{bmatrix} 0 & 0 & 0.4609 & 0.1289 & 0.4018 \\ 0.5900 & 1 & 0.4609 & 0.1289 & 0.5900 \\ 0.5092 & 0 & 1 & 1 & 0.6196 \end{bmatrix}$$

$$C_{1.6} = \begin{bmatrix} 0 & 0 & 0.4609 & 0.1289 & 0.4018 \\ 0.7508 & 1 & 0.6501 & 0.4785 & 0.7132 \end{bmatrix}$$

The C matrixes that mentioned below are the output of cluster centroid for each applied radius. The $C_{1.12}$ matrix points the matrix of the achieved cluster centroid with radius 0.72. The first row, the second row, the third row, and the fourth row on $C_{0.72}$ matrix shows the first, second, third, and fourth cluster centroid respectively.

Afterwards, the achieved number of clusters are evaluated by applying Partition Coefficient value. Then, the result of the Partition value for FSC method with the application of Hamming distance and Minkowski Chebisev combination distance is compared. The result comparison is shown in figure 1.

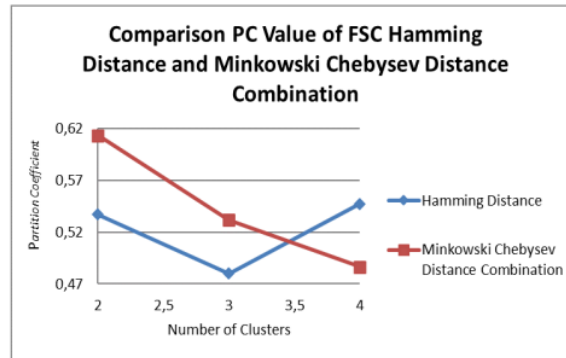


Figure 1 Partition Coefficient Value Comparison of Fuzzy Subtractive Clustering

Based on figure 1, the achieved Partition Coefficient value for Hamming distance application are lower than the Partition Coefficient value for Minkowski Chebisev combination distance application. The Partition Coefficient value from the Hamming distance for 2 achieved clusters is 0.5369, for 3 achieved clusters is 0.4801, and for 4 achieved clusters is 0.5473. At the same time, the Partition Coefficient value from Minkowski Chebisev combination distance for 2 achieved clusters, 3 achieved clusters, and 4 achieved clusters are 0.6135, 0.5319, and 0.4867 respectively.

5. CONCLUSION

In this research, the clustering method which being used is Fuzzy Subtractive Clustering with the application of Hamming distance and Minkowski Chebisev combination distance. Then, the Partition Coefficient value which measured from the Fuzzy Subtractive Clustering method with 2 mentioned distance parameter will be compared. Generally, the Minkowski Chebisev combination distance application is resulting in Partition Coefficient value which are better than the Partition Coefficient value acquired from the application of Hamming distance. This conclusion is concluded because of the Partition Coefficient value

resulted from the application of Minkowski Chebisev combination distance are generally higher than the Partition Coefficient value achieved from the Hamming distance application.

REFERENCES

- Abdolkarimi, E. S., & Mosavi, M. R. (2020). Wavelet-adaptive neural subtractive clustering fuzzy inference system to enhance low-cost and high-speed INS/GPS navigation system. *GPS Solutions*, 24(2), 1–17. <https://doi.org/10.1007/s10291-020-0951-y>
- Banteng, L., Yang, H., Chen, Q., & Wang, Z. (2019). Research on the subtractive clustering algorithm for mobile ad hoc network based on the akaike information criterion. *International Journal of Distributed Sensor Networks*, 15(9). <https://doi.org/10.1177/1550147719877612>
- Benmouiza, K., & Cheknane, A. (2019). Clustered ANFIS network using fuzzy c-means, subtractive clustering, and grid partitioning for hourly solar radiation forecasting. *Theoretical and Applied Climatology*, 137(1–2), 31–43. <https://doi.org/10.1007/s00704-018-2576-4>
- Chen, C., & Deng, X. (2020). Several new results based on the study of distance measures of intuitionistic fuzzy sets. *Iranian Journal of Fuzzy Systems*, 17(2), 147–163. <https://doi.org/10.22111/ijfs.2020.5225>
- Debnath, I. P., & Gupta, S. K. (2019). Exponential membership function and duality gaps for i-fuzzy linear programming problems. *Iranian Journal of Fuzzy Systems*, 16(2), 147–163. <https://doi.org/10.22111/ijfs.2019.4549>
- Dyvak, M., Maslyiak, Y., Voytyuk, I., & Maslyiak, B. (2018). Modified method of subtractive clustering for modeling of distribution of harmful vehicles emission concentrations. *CEUR Workshop Proceedings*, 2300, 58–62.
- Gan, G., Ma, C., & Wu, J. (2007). Data clustering: theory, algorithms, and applications. In *Data Clustering: Theory, Algorithms, and Applications* (Issue January 2007). <https://doi.org/10.1137/1.9780898718348>
- Ghane'i Ostad, M., Vahdat Nejad, H., & Abdolrazzaghi Nezhad, M. (2018). Detecting overlapping communities in LBSNs by fuzzy subtractive clustering. *Social Network Analysis and Mining*, 8(1), 1–11. <https://doi.org/10.1007/s13278-018-0502-5>
- Jang, J. S. R., Sun, C. T., & Mizutani, E. (2005). Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence. In *IEEE Transactions on Automatic Control* (Vol. 42, Issue 10). <https://doi.org/10.1109/tac.1997.633847>
- Mahajan, S., & Gupta, S. K. (2019). On fully intuitionistic fuzzy multiobjective transportation problems using different membership functions. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-019-03318-8>
- Rencher, A. C. (2016). *Methods of multivariate analysis* (Second edi, Vol. 4, Issue 1). John Wiley & sons, inc.
- Rezaei, K., & Rezaei, H. (2019). New distance and similarity measures for hesitant fuzzy soft sets. 16(6), 159–176.
- Rodrigues, O. (2018). Combining minkowski and cheyshev: new distance proposal and survey of distance metrics using k-nearest neighbours classifier. *Pattern Recognition*

- Letters*, 110, 66–71. <https://doi.org/10.1016/j.patrec.2018.03.021>
- Salah, H., Nemissi, M., Seridi, H., & Akdag, H. (2019). Subtractive clustering and particle swarm optimization based fuzzy classifier. *International Journal of Fuzzy System Applications*, 8(3), 108–122. <https://doi.org/10.4018/ijfsa.2019070105>
- Sangadji, I., Arvio, Y., & Indrianto. (2018). Dynamic segmentation of behavior patterns based on quantity value movement using fuzzy subtractive clustering method. *Journal of Physics: Conference Series*, 974(1), 0–7. <https://doi.org/10.1088/1742-6596/974/1/012009>
- Sharma, R., & Verma, K. (2019). Fuzzy shared nearest neighbor clustering. *International Journal of Fuzzy Systems*, 21(8), 2667–2678. <https://doi.org/10.1007/s40815-019-00699-7>
- Surono, S., & Putri, R. D. A. (2020). Optimization of fuzzy c-means clustering algorithm with combination of minkowski and chebyshev distance using principal component analysis. *International Journal of Fuzzy Systems*. <https://doi.org/10.1007/s40815-020-00997-5>
- Utomo, V., & Marutho, D. (2018). Measuring hybrid sc-fcm clustering with cluster validity index. *2018 International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2018, C*, 322–326. <https://doi.org/10.1109/ISRITI.2018.8864459>
- Zeng, S., Chen, S. M., & Teng, M. O. (2019). Fuzzy forecasting based on linear combinations of independent variables, subtractive clustering algorithm and artificial bee colony algorithm. *Information Sciences*, 484, 350–366. <https://doi.org/10.1016/j.ins.2019.01.071>

HASIL CEK_Bapak Sugiyarto_3

ORIGINALITY REPORT

16%

SIMILARITY INDEX

2%

INTERNET SOURCES

12%

PUBLICATIONS

4%

STUDENT PAPERS

PRIMARY SOURCES

- | | | |
|---|--|----|
| 1 | Annisa Eka Haryati, Sugiyarto. "Clustering with Principal Component Analysis and Fuzzy Subtractive Clustering Using Membership Function Exponential and Hamming Distance", IOP Conference Series: Materials Science and Engineering, 2021
Publication | 9% |
| 2 | Submitted to Universitas Brawijaya
Student Paper | 2% |
| 3 | docplayer.info
Internet Source | 1% |
| 4 | P Laras Sakti Cahyaningrum, Nursyiva Irsalinda. "Indonesian community welfare levels clustering using the fuzzy subtractive clustering (FCM) method", Journal of Physics: Conference Series, 2019
Publication | 1% |
| 5 | Iriansyah Sangadji, Yozika Arvio, Indrianto. "Dynamic Segmentation Of Behavior Patterns Based On Quantity Value Movement Using | 1% |

Fuzzy Subtractive Clustering Method", Journal of Physics: Conference Series, 2018

Publication

6

Ullah Saif, Zailin Guan, Weiqi Liu, Baoxi Wang, Chaoyong Zhang. "Multi-objective artificial bee colony algorithm for simultaneous sequencing and balancing of mixed model assembly line", The International Journal of Advanced Manufacturing Technology, 2014

Publication

1 %

7

Endra Putra Raharja, Yudhiakto Pramudya. "Outreach activities in anticipation of the 2016 solar eclipse in Sorong", Journal of Physics: Conference Series, 2016

Publication

<1 %

8

iopscience.iop.org

Internet Source

<1 %

9

Erol Egrioglu, Robert Fildes, Eren Baş. "Recurrent fuzzy time series functions approaches for forecasting", Granular Computing, 2021

Publication

<1 %

10

Shouzhen Zeng, Shyi-Ming Chen, Mario Orlando Teng. "Fuzzy forecasting based on linear combinations of independent variables, subtractive clustering algorithm and artificial bee colony algorithm", Information Sciences, 2019

<1 %

11

link.springer.com
Internet Source

<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On