Evaluating the performance of Fuzzy Clustering using different distance metrics for Image Segmentation

Conference Paper · January 2022

DOI: 10.1109/ICONAT53423.2022.9725950

3 authors, including:

Prabhjot Kaur
Maharaja Surajmal Institute Of Technology

2022 International Conference for Advancement in Technology (ICONAT)
Goa, India. Jan 21-22, 2022

Evaluating the performance of Fuzzy Clustering using different distance metrics for Image Segmentation

Jyotsna Rathee
Research Scholar,CSE,DCRUST, Murthal, Sonipat
Assistant professor, MSIT, GGSIP University
Delhi, India
jyotsnarathe@gmail.com

Prabhjot Kaur
IT Department
Associate Professor, MSIT, GGSIP University
Delhi, India
prabhjot.kaur@msit.in

Ajmer Singh
CSE Department
Assistant Professor, DCRUST, Murthal
Sonipat, India
ajmersingh.cse@dcrustm.org

2022 International Conference for Advancement in Technology (ICONAT) | 978-1-6654-2577-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICONAT53423.2022.9725950

***Abstract*—Segmentation in image processing is an important part to analyze an image automatically. Object detection and recognition in images are done with the help of segmentation process. This paper evaluates the performance of Fuzzy Clustering method for Image Segmentation using different distance metrics namely Euclidean, Canberra, Chebyshev. The performance is tested using two digital images and is quantitatively accessed using four metrics namely Partition Entropy ($V_{par.entr.}$), Partition Coefficient ($V_{par.coef.}$), Fukuyama-Sugeno ($V_{fuku.sugn.}$) and XieBeni function ($V_{xie.ben.}$).**

***Index Terms*—Image Segmentation,Fuzzy Clustering,Fuzzy C-means,Distance metrics, Euclidean, Canberra, Chebyshev**

I. Introduction

In this running world, huge amount of digital data is produced in daily scenario's like social networking websites (facebook, instagram etc.), medical images. This data need to be processed, so that valuable information can be gathered and further analyzed and can be used for different applications. An image is broken down into several segments so that analysis can be done properly.

Segmentation process of an image (digital) wherein the image is divided, based upon similar characteristics, into various segments (i.e. set of pixels or image objects) is known as Image Segmentation. The segmentation process is important to make an image meaningful so that it can be analyzed more efficiently. Image segmentation is basically used to locate objects and boundaries of objects (such as lines, curves, etc.) in an images. Image segmentation has been used in various real life applications like, recognition tasks,content bases image retrieval, object detection, medical imaging , pedestrian detection etc. [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] Various methods for Image Segmentation techniques are namely threshold, region based, edge based, watershed and clustering.

Clustering algorithms are used to identify different clusters from the data based upon some common characteristics of data so that same points within one class must have strong association towards that cluster and weak connection with other clusters. Clustering is divided into two categories i.e hard clustering and soft clustering.In hard clustering data is divided into distinct clusters, where each data element belongs to exactly one cluster whereas in Fuzzy clustering or soft clustering is based upon fuzzy logic using which one data point in the space can be associated to multiple clusters based upon their membership values. This paper accesses fuzzy clustering using three distance metrics namely Euclidean, Canberra and Chebyshev for Image segmentation.

II. Background

This section briefly describes Fuzzy clustering, Fuzzy c-means [11], distance metrics referred in this paper and four cluster validity metrices used to access digital images.

A. Fuzzy C-Means

Fuzzy clustering is the process of assigning membership levels to the data points and then using them to assign data elements to one or more clusters [12]. Fuzzy clustering uses distance measures to find similarity or dissimilarity between any pair of objects. The fuzzy c-means clustering algorithm developed by Dunn [13] and later improved by [11] uses Euclidean distance metric as a similarity measure to assign points to different clusters.

Algorithm

- *Number of clusters are chosen.*
- *Coefficients are randomly assigned to each data point in the cluster.*
- *Repeat until the algorithm is converged.*
- *Centroid for each cluster is computed.*

- *Coefficients of data points for being in clusters are computed.*

FCM algorithm tries to segment n (finite) elements say q=(q_1, q_2, q_3......., q_n) into c fuzzy clusters with certain characteristics.

Cluster centres (c=(c_1, c_2,...., c_c)) and partition matrix are returned by FCM algorithm, when finite set of data is given. M_{ij} tells degree to which element q_1, belongs to cluster c_j. Fuzzy c-means aims to minimize the following objective function:

$$ObjectiveFunction = \sum_{i=1}^{n} \sum_{j=1}^{c} M_{ij}^{y} \left\| q_i - c_j \right\|^2 \tag{1}$$

where M_{ij}, is

$$M_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|q_i - c_j\|}{\|q_i - c_a\|} \right)^{\frac{2}{y-1}}} \tag{2}$$

Centroids are calculated as per the following equation:

$$c_k = \frac{\sum_x m_k(x)^h x}{\sum_x m_k(x)^h} \tag{3}$$

where h is the hyper-parameter that controls how fuzziness of cluster. Higher value for h signifies higher fuzzier the cluster will be.

B. Distance Metrics

Distance metric plays a very important role in clustering process. The points in the data during clustering process are allocated to the different clusters using distance metric. The actual shape of clusters detected by any clustering algorithm depends upon the distance metric used by that algorithm. Researchers have used many distance metrics to implement clustering approach [14] [5] [12]. This section discusses Euclidean, Canberra, Chebyshev distance metrics.

1) Euclidean Distance Metric: Euclidean distance has been widely used in various image processing applications [11] [6] [7]. Euclidean distance calculates distance between two points as per following equation :-

$$Distance(q_i, q_j) = \left(\sum_{p=1}^{k} |q_{ip} - q_{jp}|^2\right)^{\frac{1}{2}} \tag{4}$$

Where q_i and q_j represents two points between which distance is to be calculated, $Distance(q_i, q_j)$ represents distance between two points. One limitation of Euclidean distance metric is that it favours hyper spherical clusters [15]. It is a special case of Minkowski distance when $m = 2$. Euclidean distance is a straight distance between two points, so as number of dimensions increases Euclidean distance metrics become inconvenient [16].

2) Canberra Distance Metric: The distance metric calculates the sum of absolute fractional differences of the data point characteristics and formulated as below [17]

$$Distance(q_i, q_j) = \left(\sum_{p=1}^{k} \frac{|q_{ip} - q_{jp}|}{|q_{ip}| - |q_{jp}|}\right) \tag{5}$$

where k, is the number of characteristics of a data point. Canberra distance is the weighted version of the Manhattan distance and calculates distance in vector space proposed by Lance and Williams.

3) Chebyshev Distance Metric : It is used to calculate the maximum value of the absolute differences between the characteristics of two data points [15]. Another name for this distance metrics are tchebyschev distance,chessboard distance or maximum metric proposed by Panfnuty Chebyshev.

$$Distance(q_i, q_j) = max_{1 \leq p < k}(|q_{ip} - q_{jp}|) \tag{6}$$

where p ranges from $1tok$, and k are the number of characteristics of data point.

C. Performance Evaluation

Performance of clustering can be accessed with the compactness and separation property of the clusters.Here compactness is the measure of how the data is represented within one cluster and separation measures how well the two clusters are separated. To check the performance of three distance metrics, this paper used four cluster validity functions: [18] [19].

1) Partition Coefficient function $(V_{par.coef.})$ *:* Partition coefficient function accesses clustering based upon fuzziness of the segment. It measures data heuristically because it is not connected to any property of data. In fuzzy clustering good segmentation is implied by the maximum value of $(V_{par.coef.})$ [9].

$$V_{par.coef.} = \sum_{p=1}^{n} \sum_{k=1}^{c} f^{2}_{kp} \tag{7}$$

Here f is the membership function, n is the number of pixels in data and c is the number of clusters.

2) Partition Entropy function: $(V_{par.entr.})$ It is a metric of impurity and uncertainty within partition of finite set without considering data [9]. Best clustering results are depicted by the minimum value of $(V_{par.entr.})$.

$$V_{par.entr.} = -\frac{1}{n} \sum_{p=1}^{n} \sum_{k=1}^{c} [f_{kp} log f_{kp}] \tag{8}$$

3) Fukuyama-Sugeno function : $(V_{fuku.sugn.})$This function calculates difference between a) combining fuzziness in membership matrix with geometric compactness of data set by prototypes, b) by fuzziness in row of the partition matrix with the distance between prototype i to the mean of the data. Good partition is supported by minimum value of the metric [20].

$$V_{fuku.sugn.} = -\frac{1}{n} \sum_{k=1}^{c} \sum_{p=1}^{n} f_{kp}^{m} (\|q_p - Centre_k\|^2 - \left\|Centre_k - \overline{Centre_j}\right\|^2) \tag{9}$$

Where $\overline{Centre}$ is,

$$\overline{Centre} = \frac{1}{n} \sum_{k=1}^{c} [Centre_k] \tag{10}$$

Here, q represents data point, $Centre$ represents centre of the cluster and f is the membership function.

4) XieBeni function : $(V_{xie.ben.})$ Cluster separateness and compactness in XieBeni are measured by using inter-cluster distance (distances between cluster centers) and intra-cluster deviations [21]. XieBeni is a function of the centroids of the clusters and data set. It is a ratio of total variation of the centroid and segment and centroids vectors separation. Under comparison the best segment is the one supported by minimum value.

$$V_{xie.ben.} = -\frac{(\sum_{k=1}^{c}\sum_{p=1}^{n} f_{pk}^{2} \left\| q_p - Centre_k \right\|^2)}{n * (min_{k\neq 0}(\|Centre_k - Centre_z\|)^2)} \quad (11)$$

where c is the number of clusters and n is number of data points. A good fuzzy partition is expected to have a low degree of overlap and a larger separation distance [18].

III. Results and Simulations

This section evaluates the performance of Fuzzy C-means using different distance metrics for digital image segmentation. Two digital images as used as test images.

- Digital Image 1 (Bacteria Image) with size 126x140x3.
- Digital Image 2 (Hand Image) with size 243x304x3.

All the simulations are done using MATLAB version R2017b with system configuration as Intel(R) Core (TM) i5-8250U CPU @ 1.60GHz 1.80 GHz. Number of clusters in case of Digital Image 1 are 2 and in case of Digital Image 2 are 3.

Fig.1 shows original images. Fig. 2 and Fig. 3 show segmentation results of Fuzzy c Means using Euclidean, Canberra and Chebyshev distance metrics respectively. Table 1 lists the execution time and readings of the four quantitative measures $V_{par.entr.}$, $V_{par.coef.}$, $V_{fuku.sugn.}$ and $V_{xie.ben.}$.

For both digital images total time taken by Euclidean distance metric was minimum followed by Chebyshev distance.

$V_{par.coef.}$ for digital image 1 was best achieved in the case of Euclidean distance metric followed by Chebyshev. Where $V_{par.coef.}$ for digital image 2 was best achieved in the case of Chebyshev distance metric followed by Euclidean distance metric. There is a minute difference between the performance of Fuzzy c means with Euclidean and Chebyshev distance metrics, whereas the performance with Canberra is worst in both the cases.

$V_{par.entr.}$ for Digital image 1 is minimum in the case of Euclidean distance metric followed by Chebyshev and maximum in case of Canberra distance which makes Euclidean distance metric to be the best among three in this case. $V_{par.entr.}$ for Digital image 2 is minimum in the case of Chebyshev distance metric followed by Euclidean and maximum in the case of Canberra which makes Chebyshev distance metric to be the best among three. In both the cases, Canberra has the worst performance.

The value of $V_{fuku.sugn.}$ and $V_{xie.ben.}$ in both the images is minimum in case of Euclidean Distance metric followed by Chebyshev, whereas the performance of fuzzy clustering using Canberra distance metric is worst.

As per the quantitative analysis, Fuzzy clustering using Euclidean distance metric and Chebyshev distance metric provide good segmentation results than Canberra distance metric.

As per qualitative analysis (Fig. 2 & Fig. 3), performance with Chebyshev distance metric is better than Euclidean distance metric. The noise in the image is clear in case of Chebyshev than by using Euclidean distance norm. Fig. 2(d) shows better segmentation visually than Fig. 2(b) & 2(c) which concludes the performance of Chebyshev is better than Euclidean and Canberra. Digital Images in Fig. 2 ((b) & (d)) and Fig. 3 ((b) & (d)) seems to be quiet similar in detecting size of the bacteria and hand present in the digital images.

IV. Conclusion

This paper shows the performance of fuzzy clustering using Euclidean, Canberra and Chebyshev distance metric. It is tested with two digital images. From the qualitative and quantitative analysis, it is reported that the performance of Chebyshev is best followed by Euclidean distance metric. The performance of fuzzy clustering using Canberra distance metric is worst in both the cases.

References

[1] F. Mei, J. Mei, J. Zheng, S. Zhang, and K. Zhu, "Application of particle swarm fused kfcm and classification model of svm for fault diagnosis of circuit breaker," *Proceedings of the CSEE*, vol. 33, no. 36, pp. 134–141, 2013.

[2] Y. Wu, W. Hou, and S. Wu, "Brain mri segmentation using kfcm and chan-vese model," *Transactions of Tianjin University*, vol. 17, no. 3, pp. 215–219, 2011.

[3] D. Hu, A. Sarosh, and Y.-F. Dong, "A novel kfcm based fault diagnosis method for unknown faults in satellite reaction wheels," *ISA transactions*, vol. 51, no. 2, pp. 309–316, 2012.

[4] P. Kaur, A. Jain, and S. Mittal, "Touch-less fingerprint analysis–a review and comparison," *International Journal of Intelligent Systems and Applications*, vol. 4, no. 6, p. 46, 2012.

[5] P. Kaur, A. Soni, A. Gosain, and I. I. India, "Novel intuitionistic fuzzy c-means clustering for linearly and nonlinearly separable data," *WSEAS Trans Comput*, vol. 11, no. 3, pp. 65–76, 2012.

[6] T. Chaira, "A novel intuitionistic fuzzy c means clustering algorithm and its application to medical images," *Applied Soft Computing*, vol. 11, no. 2, pp. 1711–1717, 2011.

[7] P. Kaur, "Intuitionistic fuzzy sets based credibilistic fuzzy c-means clustering for medical image segmentation," *International journal of information technology*, vol. 9, no. 4, pp. 345–351, 2017.

[8] N. Garg, V. Sharma, and P. Kaur, "Melanoma skin cancer detection using image processing," in *Sensors and Image Processing*. Springer, 2018, pp. 111–119.

[9] P. Kaur, I. Lamba, and A. Gosain, "A robust method for image segmentation of noisy digital images," in *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*. IEEE, 2011, pp. 1656–1663.

[10] P. Kaur and A. Gosain, "A density oriented fuzzy c-means clustering algorithm for recognising original cluster shapes from noisy data," *International Journal of Innovative Computing and Applications*, vol. 3, no. 2, pp. 77–87, 2011.

[11] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm," *Computers & geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984.

[12] J. Arora, K. Khatter, and M. Tushir, "Fuzzy c-means clustering strategies: A review of distance measures," *Software Engineering*, pp. 153–162, 2019.

[13] J. C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.

[14] P. Kaur, P. Gupta, and P. Sharma, "Review and comparison of kernel based fuzzy image segmentation techniques," *International Journal of Intelligent Systems and Applications*, vol. 4, no. 7, p. 50, 2012.

TABLE I
QUANTITATIVE ANALYSIS USING FOUR METRICS

Image	Distance Metric	Time (Sec)	$(V_{par.entr.})$	$(V_{par.coef.})$	$(V_{xie.ben.})$	$(V_{fuku.sugn.})$
Digital Image 1	Euclidean	0.0492	0.0941	0.9495	0.0254	-3269.98
	Canberra	0.0696	0.1572	0.9167	0.0311	-2711.9960
	Chebyshev	0.0535	0.1031	0.9442	0.0272	-3054.6520
Digital Image 2	Euclidean	0.6975	0.3728	0.7983	0.0822	-6612.2802
	Canberra	1.8449	0.4685	0.7395	0.0918	-5372.4916
	Chebyshev	0.9565	0.3557	0.8058	0.0848	-6578.1311

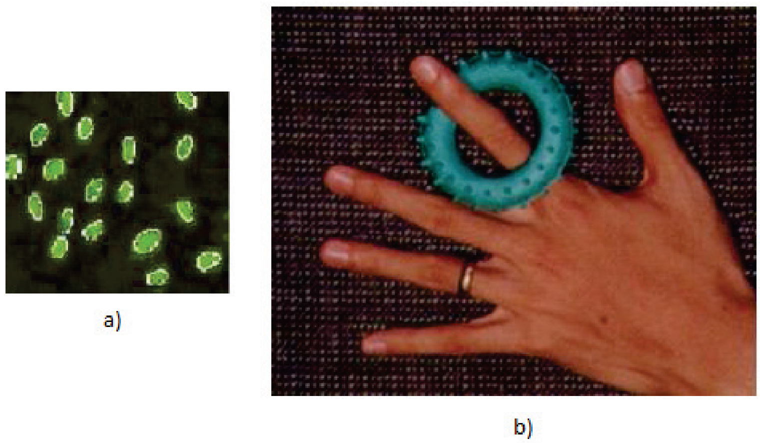

Fig. 1. a) Digital Image 1 of size 126x140x3 b) Digital Image 2 of size 243x304x3.

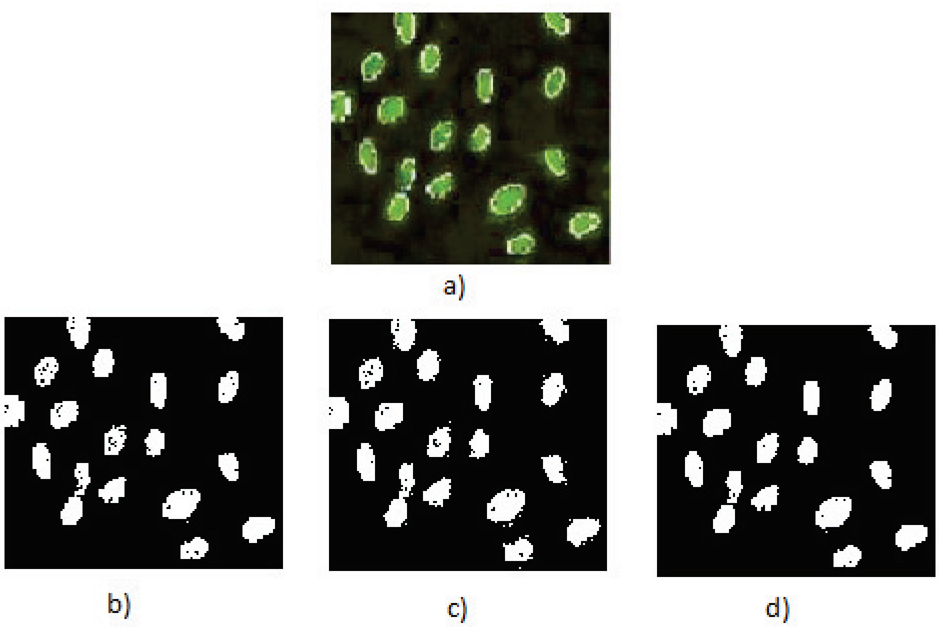

Fig. 2. a) Original Digital image 1 with no error induced, b) Digital image 1 with Euclidean Distance metrics, c) Digital image 1 with Canberra distance metrics, d) Digital image 1 with Chebyshev distance metrics.

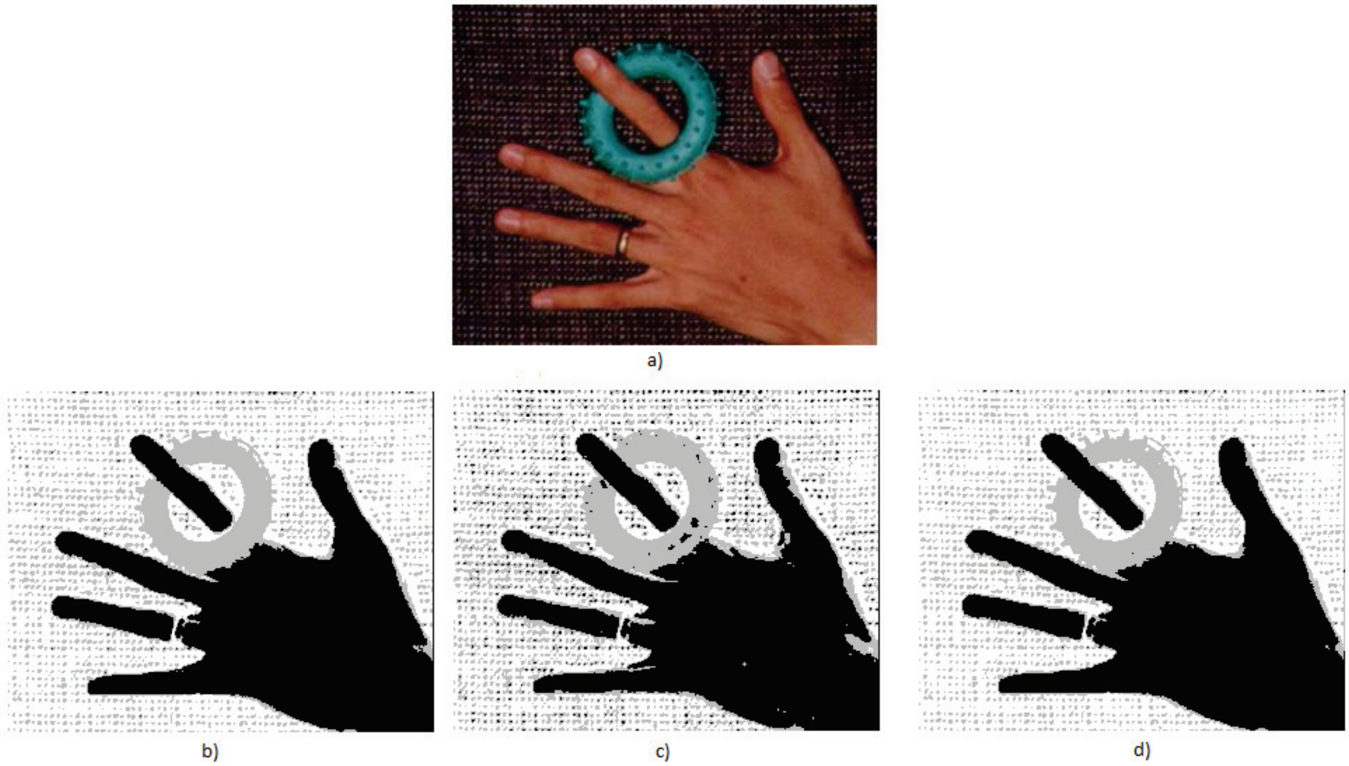

Fig. 3. a) Original Digital Image 2 with no error induced, b) Digital image 2 with Euclidean Distance metrics, c) Digital image 2 with Canberra distance metrics, d) Digital image 2 with Chebyshev distance metrics.

[15] V. Kumar, J. K. Chhabra, and D. Kumar, "Performance evaluation of distance metrics in the clustering algorithms," *INFOCOMP Journal of Computer Science*, vol. 13, no. 1, pp. 38–52, 2014.

[16] P. Kaur and T. Chaira, "A novel fuzzy approach for segmenting medical images," *Soft Computing*, vol. 25, no. 5, pp. 3565–3575, 2021.

[17] Y. Thakare and S. Bagal, "Performance evaluation of k-means clustering algorithm with various distance metrics," *International Journal of Computer Applications*, vol. 110, no. 11, pp. 12–16, 2015.

[18] D.-W. Kim, K. H. Lee, and D. Lee, "On cluster validity index for estimation of the optimal number of fuzzy clusters," *Pattern Recognition*, vol. 37, no. 10, pp. 2009–2025, 2004.

[19] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Transactions on Fuzzy systems*, vol. 3, no. 3, pp. 370–379, 1995.

[20] Y. Fukuyama, "A new method of choosing the number of clusters for the fuzzy c-mean method," in *Proc. 5th Fuzzy Syst. Symp., 1989*, 1989, pp. 247–250.

[21] M. Muranishi, K. Honda, and A. Notsu, "Xie-beni-type fuzzy cluster validation in fuzzy co-clustering of documents and keywords," in *Soft Computing in Artificial Intelligence*. Springer, 2014, pp. 29–38.