



Using possibilistic clustering method in microarray dataset of Alzheimer



The 4th Iranian Conference on
Systems Biology

Zohre Moattar-Husseini¹, Mohammad Hossein Fazel Zarandi^{1*}, Abbas Ahmadi¹

Amirkabir University of Technology
(Tehran Polytechnic)

¹ Department of Industrial Engineering, Amirkabir University of Technology, Tehran, Iran

*Corresponding Author's E-mail: (zarandi@aut.ac.ir)

Abstract

Microarray technology enables the measurement of a large number of genes or proteins under a specific condition in a short period of time. To make better use of the large amount of information provided by microarray data, biologically clustering of data has been recommended. Due to the high dimensions, noise and redundancy in the microarray data set, it is necessary to use robust clustering approaches. Accordingly, in this paper, to overcome these characteristics, Fuzzy C-Mean (FCM) and Possibilistic C-Mean (PCM) clustering approaches have been used to deal with noise, uncertainty and outliers. These approaches are then evaluated, using Alzheimer's microarray datasets and validation indices, which indicates the effectiveness of the approaches.

Keywords

Soft clustering; Fuzzy c-means clustering; Possibilistic c-means clustering; Microarray data; Alzheimer dataset.

Introduction

- In the field of gene expression, the development of clustering algorithms is generally based on (Vijverberg, 2007; Xu and Wunsch, 2010):
 - **Managing the characteristics of gene expression data (high dimensions, uncertainty, redundancy and noise) by developing robust clustering methods,**
 - **The possibility of assigning cluster elements to more than one cluster**
 - **Minimizing the number of parameters of the clustering approach for ease of use by biologists**
- According to the mentioned issues, soft clustering approaches are effective in dealing with the complex relationships between gene expression data (Shukla and Muhuri, 2019).
- The most popular soft clustering approaches are Fuzzy C-Means (FCM) and Possibilistic C-Means (PCM), in which cluster elements can simultaneously be members of other clusters with specific membership functions (Bezdek et al., 1984; Krishnapuram and Keller, 1993).
- Also, the fuzzy set can manage uncertainty in the modeling (Mendel, 2017). In some sources (Cano et al., 2007; Truong et al., 2017), possibilistic clustering approaches have been used to deal with the negative effects of noise and outliers.
- Therefore, these issues express the need for research in the field of microarray data and the use of fuzzy clustering and possibilistic clustering methods.

Materials and Methods

- **Datasets:** Due to the polygenic nature of Alzheimer's disease (AD), this study focuses on the clustering of microarray AD data generated by Liang et al. (2008). The dataset includes 54675 genes of 161 samples. In our study, samples of four regions are considered such as: Middle Temporal Gyrus (MTG), Entorhinal Cortex (EC), Hippocampus (HIP), Posterior Cingulate Cortex (PC).
- **Methods:** FCM and PCM clustering approaches
- **Programming language:** MATLAB is applied for microarray dataset of AD
- **Cluster Validity Index (CVI):** Two cluster validity indices including number of Incorrect Clustering (IC) and Random Index (RI), are considered for the soft clustering approaches.

Results

- The following table presents the clustering results of four Alzheimer's datasets (MTG, EC, HIP and PC) based on RI and IC evaluation indicators:

Table 1: A comparative analysis of soft clustering for Alzheimer dataset

#	Data	#AD/#Samples	IC		RI	
			FCM	PCM	FCM	PCM
1	MTG	16/28	0	0	1.00	1.00
2	EC	10/23	10	4	0.56	0.83
3	HIP	10/23	6	5	0.75	0.79
4	PC	9/22	0	5	1.00	0.77

- As shown in the table, clustering results depend on data structure as well as method performance. The rand index for the MTG dataset is at its best for both methods. The lowest and highest values of RI for PCM approach are 0.77 and 1, respectively, while the lowest and highest RI for FCM are 0.56 and 1, respectively. Therefore, in general, the results of this study indicate the effectiveness of the soft clustering approaches in the Alzheimer's microarray data set. Moreover, in all data except PC, the PCM approach is superior to the FCM.

Discussion, Conclusion and

- Clustering as a methods of microarray data analysis needs to have sufficient robust capability. Soft clustering approaches in the field of biology are more effective due to their ability to assign cluster elements to multiple clusters. Fuzzy and possibilistic clustering approaches can also be appropriate through managing the negative effects of uncertainty, noise, and outliers.
- Accordingly, this study uses clustering approaches including fuzzy c-mean and possibilistic c-mean for the Alzheimer's microarray dataset.
- In general, it is concluded that the validity indices are influenced by the capabilities of clustering methods and data structure. The experiments also indicated that, all the datasets of Alzheimer's except the PC, the PCM approach were superior to the FCM. Therefore, in general, the results of this study indicate the effectiveness of the soft clustering approach, particularly PCM approach, in the Alzheimer's microarray data set.

References

- Benaichouche, A.N., Oulhadj, H., Siarry, P., 2013. Improved spatial fuzzy c-means clustering for image segmentation using PSO initialization, Mahalanobis distance and post-segmentation correction. Digit. Signal Process. A Rev. J. 23, 1390–1400.
- Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. Comput. Geosci. 10, 191–203.
- Cano, C., Adarve, L., López, J., Blanco, A., 2007. Possibilistic approach for biclustering microarray data. Comput. Biol. Med. 37, 1426–1436.
- Krishnapuram, R., Keller, J.M., 1993. A Possibilistic Approach to Clustering. IEEE Trans. Fuzzy Syst. 1, 98–110.
- Liang, W.S., Dunckley, T., et al., 2008. Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set. Physiol. Genomics 33, 240–256.
- Mendel, J.M., 2017. Uncertain rule-based fuzzy systems. Springer.
- Shukla, A.K., Muhuri, P.K., 2019. Big-data clustering with interval type-2 fuzzy uncertainty modeling in gene expression datasets. Eng. Appl. Artif. Intell. 77, 268–282.
- Truong, H.Q., Ngo, L.T., Pedrycz, W., 2017. Granular Fuzzy Possibilistic C-Means Clustering approach to DNA microarray problem. Knowledge-Based Syst. 133, 53–65.
- Xu, R., Wunsch, D.C., 2010. Clustering algorithms in biomedical research: A review. IEEE Rev. Biomed. Eng. 3, 120–154.