Gene Expression Data Analysis using Fuzzy C-means Clustering Technique

Abstract

The challenging issue in microarray technique is to analyze and interpret the large volume of data. This can be achieved by clustering techniques in data mining. In hard clustering like hierarchical and k-means clustering techniques, data is divided into distinct clusters, where each data element belongs to exactly one cluster so that the outcome of the clustering may not be correct in many times. The problems addressed in hard clustering could be solved in fuzzy clustering technique. Among fuzzy based clustering, fuzzy c means (FCM) is the most suitable for microarray gene expression data. The problem associated with fuzzy c-means is the number of clusters to be generated for the given dataset needs to be specified in prior. The main objective of this proposed Possibilistic fuzzy c-means method is to determine the precise number of clusters and interpret the same efficiently. The PFCM is a good clustering algorithm to perform classification tests because it possesses capabilities to give more importance to topicalities or membership values. PFCM is a hybridization of PCM and FCM that often avoids various problems of PCM, FCM and FPCM. Based on the sample dataset ‘lung’ the entire research has been developed. The available research works already developed in this area are
not exclusively working with cancer genes. At this juncture, using of the Modified Possibilistic fuzzy c-means algorithm could be found matching with cancer genes in a better fashion. “Matlab” is used for the algorithm. The accuracy of the dataset may be identified with the usage of different training sets. Possibilistic fuzzy c-means algorithm has provided better results while identifying the cancer gene. For evaluating the feasibility of the Possibilistic Fuzzy C-Means (PFCM) clustering approach, the researcher has carried out the experimental analysis.

References

5. Seo Young Kim, Tai Myong Choi, ”Fuzzy types clustering for microarray data”, PWASET Volume 4 February 2005 ISSN 1307-6884.

Index Terms

Computer Science

Fuzzy Systems
Keywords

Clustering, Microarray, Gene Expression, Fuzzy Clustering, FCM, PFCM