Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 79 (2016) 100 – 111

Procedia Computer Science

7th International Conference on Communication, Computing and Virtualization 2016

Performance Analysis of Various Fuzzy Clustering Algorithms: A Review

Anjana Gosain[a], Sonika Dahiya[b]

[a]*Professor, USICT, GGS Indraprastha University, Sector 16C, Dwarka, Delhi-110075, India*
[b]Assistant *Professor, USICT, GGS Indraprastha University, Sector 16C, Dwarka, Delhi-110075, India*

Abstract

Fuzzy clustering is useful clustering technique which partitions the data set in fuzzy partitions and this technique is applicable in many technical applications like crime hot spot detection, tissue differentiation in medical images, software quality prediction etc. In this review paper, we have done a comprehensive study and experimental analysis of the performance of all major fuzzy clustering algorithms named: FCM, PCM, PFCM, FCM-σ, T2FCM, KT2FCM, IFCM, KIFCM, IFCM-σ, KIFCM-σ, NC, CFCM, DOFCM. To better analysis their performance we experimented with standard data points in the presents of noise and outlier. This paper will act as a catalyst in the initial study for all those researchers who directly or indirectly deal with fuzzy clustering in their research work and ease them to pick a specific method as per the suitability to their working environment.

Peer-review under responsibility of the Organizing Committee of ICCCV 2016

Keywords: Fuzzy Clustering; FCM; PCM; FCM-σ; IFCM-σ; KFCM-σ; Type 2; Intuitionisitic; IFCM; DOFCM; CFCM; NC;

1. Introduction

Clustering is an important unsupervised form of classification technique of data mining [1,3]. It divides the data elements in a number of groups such that elements within a group possess high similarity while they differ from the elements of other groups. Clustering can of two types: Hard Clustering and Fuzzy Clustering[1,3]. When each element is solely dedicated to one group, that type of clustering is called Hard clustering. In hard clustering, clusters have crisp sets for representing element's membership, i.e. the membership of elements in a cluster is assessed in binary terms according to a bivalent condition that an element either belongs or does not belong to the set. In contrast, when the elements are not solely belonging to any one group, instead they share some fraction of

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: sonika.dahiya11@gmail.com.

Peer-review under responsibility of the Organizing Committee of ICCCV 2016
doi:10.1016/j.procs.2016.03.014

membership in a number of groups, that type of clustering is called Fuzzy clustering. So, fuzzy clustering permits the gradual assessment of the membership of elements in a set which is described by a membership function valued in the real unit interval [0, 1]. Thus membership functions are represented as a fuzzy set which can be either Type-I, Type-II or Intuitionistic. Algorithms proposed in the literature are based one of the three fuzzy set theory as shown in figure 1.

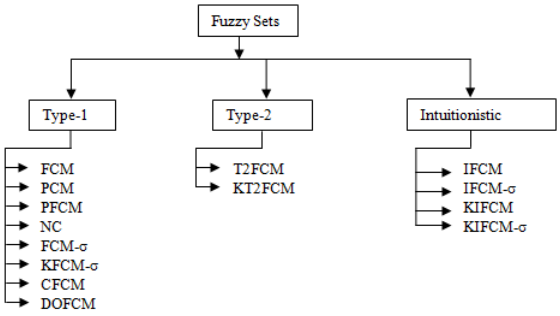

Fig. 1. Classification of Fuzzy algorithms discussed.

Bezdek's famous fuzzy clustering algorithm named as Fuzzy C-Means Algorithm and other algorithms such as PCM, PFCM, FCM-σ which are derived from FCM algorithm have been useful in many technical areas like image analysis, pattern recognition [3, 4, 5, 6,7]. FCM performs well with noise free data but are highly sensitized to noise and outliers. As they fail to distinguish between data points and noise or outliers, so centroid is attracted towards outliers instead of the center of the cluster. Though PCM and PFCM perform better in presence of noise in comparison to FCM, but PCM fails to find optimal clusters in presence of noise[4]. PFCM fails to give appreciable results when data set consists of two clusters which are highly unalike in size and outliers are present[5].

T2FCM and KT2FCM algorithms are based on Type-2 fuzzy sets, so some data elements contribute more in computing appropriate cluster centroids [8,9]. Though they outperform earlier proposed algorithms but T2FCM fails when the data structure of input patterns is non-spherical and complex[8]. KT2FCM overcome the problem of T2FCM by introducing kernel, tangent function and lagrangian methods, with basic objective function of the T2FCM algorithm and gives better segmentation over noisy data [9].

To further improve the effectiveness of clustering algorithms new concept of hesitation degree was merged with membership degree, based on which IFCM, IFCM-σ, KIFCM, KIFCM-σ were proposed. IFCM extends conventional FCM by adding intuitionistic feature to membership and objective functions. IFCM improved cluster computation over existing algorithms but failed to efficiently cluster non-spherically separable data. Then next Robust Intuitionistic Fuzzy C-means (IFCM-σ) was proposed that takes into account new distance metric to work well with non-spherically separable data. KIFCM was proposed next by introducing Radial Basis kernel function to improve the accuracy of the intuitionistic fuzzy c-means. KIFCM-σ is a hybridization of IFCM, kernel function, and new distance metric in the feature space and in the data space and thus overcomes various problems of IFCM and FCM-σ.

The main focus of all these algorithms is on improving the clustering or centroid computation without considering the noise and outliers. NC and CFCM are proposed specifically to work efficiently with noisy data. They emphasise on reducing the effect of outliers on resulting clusters rather than exactly identifying them. CFCM does this by introducing a new variable named credibility. As these algorithms emphasis on reducing the effect of outliers on resulting clusters. Therefore, the problem with these algorithms is that they fail to find accurate clusters. To overcome this problem, DOFCM works on identifying outliers based on density of data set and then performs clustering considering actual data points. Thus, it improved the accuracy of clusters.

In this paper we discussed all these algorithms in detail and experimentally showed a performance comparison of all these algorithms.