Information Sciences 260 (2014) 74–97

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Relative entropy fuzzy c-means clustering

M. Zarinbal [a], M.H. Fazel Zarandi [a,*], I.B. Turksen [b,c]

[a] Department of Industrial Engineering, Amirkabir University of Technology, 424 Hafez Ave, P.O. Box 15875-4413, Tehran, Iran
[b] TOBB Economics and Technology University, Ankara, Turkey
[c] Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada

ARTICLE INFO

Article history:
Received 17 October 2012
Received in revised form 24 September 2013
Accepted 7 November 2013
Available online 14 November 2013

Keywords:
Fuzzy clustering
Relative entropy
Fuzzy c-means
Relative entropy fuzzy c-means clustering

ABSTRACT

Pattern recognition is a collection of computer techniques to classify various observations into different clusters of similar attributes in either supervised or unsupervised manner. Application of fuzzy logic to unsupervised classification or clustering methods has resulted in many wildly used techniques such as fuzzy c-means (FCM) method. However, when the observations are too noisy, the performance of such methods might be reduced. Thus, in this paper, a new fuzzy clustering method based on FCM is presented and the relative entropy is added to its objective function as a regularization function to maximize the dissimilarity between clusters. Several examples are provided to examine the performance of the proposed clustering method. The obtained results show that the proposed method has a very good ability in detecting noises and assignment of suitable membership degrees to the observations.

1. Introduction

The main task of pattern recognition is to impose identity on observations such as attributes of an object, symptoms of patients, speech, or images. In other words, pattern recognition is a collection of computer techniques that aims at finding regularities in observations. This technique contains two different approaches, supervised classification and unsupervised classification or clustering. Supervised classification methods use the pre-labeled observations to learn the existing regularities, while unsupervised methods or clustering methods try to find the natural grouping exists in observations. This is done in such a way that objects belonging to the same cluster are as similar as possible and objects belonging to different clusters are as dissimilar as possible; that is, finding c groups in N observations based on similarity. In fact, an ideal cluster is defined as a set of compact and isolated observations. In reality, however, a cluster is a subjective entity that is in the eye of the beholder and whose significance and interpretation requires domain knowledge. While humans are excellent in finding clusters in two and possibly in three dimensions, an automated algorithm is needed for high-dimensional observations. This challenge along with the unknown number of clusters for the given data set has resulted in thousands of clustering algorithms [23]. K-means, Gaussian mixture models, density-based spatial clustering of applications with noise (DBSCAN) and Canopy are examples of these algorithms.

Fuzzy logic was introduced by Zadeh to handle uncertainties and imprecision in real world problems. In fact, fuzzy logic may be viewed as an attempt to formalize/mechanize two human capabilities: the capability to reason and make rational decisions in an environment of imprecision; and the capability to perform wide variety of physical and mental tasks without

* Corresponding author. Tel.: +98 2164545378.
E-mail addresses: mzarinbal@aut.ac.ir (M. Zarinbal), zarandi@aut.ac.ir (M.H. Fazel Zarandi), bturksen@etu.edu.tr (I.B. Turksen).

0020-0255/$ - see front matter
http://dx.doi.org/10.1016/j.ins.2013.11.004

any measurements and any computations [59]. By applying fuzzy logic to clustering methods, observations could belong to more than just one cluster. This combination would give more flexibility to handle the uncertainties in real world observations and has resulted in many fuzzy clustering methods including fuzzy c-means (FCM) [3], fuzzy c-regression model (FCRM) [17], possibilistic c-means (PCM) [30], possibilistic fuzzy c-means (PFCM) [46], cluster repulsion [55], fuzzy c-spherical shells (FCSS) [31], maximum entropy clustering [26,56], and knowledge based clustering [47,48]. These fuzzy clustering methods and many other methods have been applied in various areas including system evaluation [50,53,19], decision-making [27,39], financial forecasting [34,52,10,5], time series [24,35,12,1], image processing [45,44,13,14,54,18], speech recognition [28,32], etc.

Among all fuzzy clustering methods, FCM is the most wildly used method in which the degree of belonging of jth observation to ith cluster or its membership degree, u_{ij}, is obtained by inversion of relative distance to cluster centers, v_i $i = 1, 2, \ldots, c$. That is:

$$u_{ij} = \left(\sum_{k=1}^{c} \left(\frac{\|x_j - v_i\|^2}{\|x_j - v_k\|^2} \right)^{1/m-1} \right)^{-1} \quad 1 \leqslant i \leqslant c, \; 1 \leqslant j \leqslant N \tag{1}$$

where x_j is the jth observation, m is the degree of fuzziness or weighting coefficient, c is the total number of clusters, N is the number of observations and $\|.\|^2$ is the distance between observations and cluster centers.

This definition for membership degree may cause some serious problems in the case of having noisy observations. That is, FCM is a kind of partitioning algorithm that divides the observations into c partitions regardless of being noise or not. However, it is more natural for noise points to have very low membership degree in all clusters [11]. Possibilistic c-means clustering (PCM) [30] is formulated in such a way to overcome this problem. In this method, the degree of belonging is defined as:

$$u_{ij} = \left(1 + \left(\frac{\|x_j - v_i\|^2}{\eta_i} \right)^{1/m-1} \right)^{-1} \quad 1 \leqslant i \leqslant c, \; 1 \leqslant j \leqslant N \tag{2}$$

where η_i is a suitable positive number that determines the distance at which the membership value of a point in a cluster becomes 0.5 and other parameters have the same definition as before [30].

Clearly, in this method, the clusters have no interactions and they can be located very close to each other or even coincide. This leads to solutions in which one actual cluster in a data set is represented by two coincident clusters [11]. Moreover, the clusters do not have lots of mobility and reasonably good initialization is required for the algorithm to converge to the global minimum [30]. There are many other approaches in literature of fuzzy clustering that try to relieve the weakness of FCM by adding a regularizing function to FCM's objective function. Quadratic function [43], Pham's robust FCM [49], and entropy [56,38,58,6] are some of well-known regularizing functions.

Relative entropy is the general case of entropy that measures the distance between two distributions. This function has been applied in FCM as regularization function by many researchers including Ichihashi et al. [21], Ichihashi et al. [22] and Miyagishi et al. [41]. In these clustering methods, however, the degree of fuzziness, m is not considered, so, they reduce to hard c-means clustering algorithms [20]. To overcome this weakness, we add the relative entropy to FCM's objective function by considering the degree of fuzziness, m. The proposed clustering method, relative entropy fuzzy c-means (REFCM), is then examined using several experiments and the obtained results are compared with five well-known clustering methods, FCM [3], PCM [30] Noise clustering (NC) [9], regularized FCM [36] and regularized FCM [51].

The rest of this paper is organized as follows: Section 2 explains the fuzzy entropy clustering methods and addresses the concepts of relative entropy shortly. Section 3 is dedicated to the proposed relative entropy fuzzy c-means clustering method. The computational complexity and the performance of the proposed method are discussed in Section 4. Finally, conclusions are presented in Section 5.

2. Fuzzy entropy clustering

One of the most important aspects of real world phenomena is uncertainty and imprecision, which could be measured by entropy. In fact, entropy is the minimum descriptive complexity of a random variable [8]. Assume x as a discrete random variable with probability mass function $p(x)$. The entropy of x, $H[x]$, is defined as:

$$H[x] = -\sum_{x} p(x) \log p(x) \tag{3}$$

As $H[x]$ is strictly convex function [8,42], it could be considered as regularizing function to FCM's objective function, which is investigated by many researchers [56,38,58,25,51,36,37]. The first formulation of this idea can be found in Li and Mukaidono [36], in which the objective function is defined as:

$$\min J(U, V, c) = \sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij} d_{ij}^2 + \alpha \sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij} \log u_{ij} \quad \alpha > 0 \tag{4}$$