Information Sciences xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Information Sciences

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

Interval Type-2 Relative Entropy Fuzzy C-Means clustering

M. Zarinbal [a], M.H. Fazel Zarandi [a,c,*], 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] Knowledge Intelligent Systems Laboratory, University of Toronto, Toronto, Canada

ARTICLE INFO

Article history:
Received 21 April 2013
Received in revised form 31 October 2013
Accepted 9 February 2014
Available online xxxx

Keywords:
Interval Type-2 fuzzy set theory
Interval arithmetic
Relative entropy
Fuzzy c-means clustering
Interval Type-2 Relative Entropy Fuzzy C-Means clustering

ABSTRACT

Fuzzy set theory especially Type-2 fuzzy set theory provides an efficient tool for handling uncertainties and vagueness in real world observations. Among various clustering techniques, Type-2 fuzzy clustering methods are the most effective methods in the case of having no prior knowledge about observations. While uncertainties in Type-2 fuzzy clustering parameters are investigated by researchers, uncertainties associated with membership degrees are not very well discussed in the literature. In this paper, investigating the latter uncertainties is our concern and Interval Type-2 Relative Entropy Fuzzy C-Means (IT2 REFCM) clustering method is proposed. The computational complexity of the proposed method is discussed and its performance is examined using several experiments. The obtained results show that the proposed method has a very good ability in detecting noises and assignment of suitable membership degrees to observations.

1. Introduction

Pattern recognition methods are types of data analysis methods, in which observations are investigated without assuming to have any mathematical model. The main concern of these methods is to discover the regularities in observations and perhaps classifying them into different categories, automatically. The motivation is to perform these tasks more accurate and more economical than humans do [14].

Supervised and unsupervised classification methods are two main approaches in pattern recognition methods. In supervised classification methods, the labeled observations are used to provide basis for learning, whereas, if there were no prior information, unsupervised classification or clustering methods would be the most effective ones. In this case, the goal is to find the natural grouping exists in observations. High dimensional observations along with unknown number of clusters have resulted in thousands of automated clustering methods such as K-means, Gaussian mixture models, density-based spatial clustering of applications with noise (DBSCAN) and Canopy methods. In all these methods, it is assumed that the observations have to be belong to just one cluster, whereas in practice the observations could belong to more than one cluster with some degree of belonging. This gap between theory and practice could be effectively handled by fuzzy set theory.

However, different meanings of words to different people, not agreeing knowledge extracted from experts, noisy and uncertain measurements, and noisy data used to tune parameters, may cause uncertainties in fuzzy sets' parameters [23]. Type-2 fuzzy sets in contrast with Type-1 fuzzy sets (classical fuzzy sets) are able to model such uncertainties as their

* Corresponding author at: Department of Industrial Engineering, Amirkabir University of Technology, 424 Hafez Ave, P.O. Box 15875-4413, Tehran, Iran. 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).

http://dx.doi.org/10.1016/j.ins.2014.02.066
0020-0255/

membership functions are themselves fuzzy. These secondary membership functions could be Type-1 fuzzy sets (General Type-2 fuzzy sets) or crisp intervals in $[0,1]$ (Interval Type-2 fuzzy sets). Obviously, Interval Type-2 fuzzy sets have less computational complexity than General Type-2 fuzzy sets [23] and have been applied in many application areas such as granular computing [31,32] pattern recognition [21,9], control [22,2], prediction [27,20], and decision making [37,13].

The combination of Interval Type-2 fuzzy set theory and clustering methods gives more flexibility to handle uncertainties in real observations and it has resulted in many clustering methods, including Interval Type-2 fuzzy c-means clustering [15], Interval Type-2 fuzzy c-regression clustering [8], Interval Type-2 approach to kernelized fuzzy c-means clustering [16], Interval Type-2 approach to kernel possibilistic c-means clustering [34], Interval Type-2 possibilistic c-means clustering [10], and Interval-valued fuzzy relation-based clustering [12] methods. In addition, by combining General Type-2 fuzzy sets and fuzzy c-means (FCM), Linda and Manic [19] proposed General Type-2 fuzzy c-mean clustering method. These clustering methods have been applied in many areas such as image processing [9,16,30,33], performance evaluation [12], stock market prediction [7], and detection [6].

Uncertainty in clustering parameters, such as degree of fuzziness, has been discussed in most of Type-2 fuzzy clustering investigations. Although the uncertainties in these parameters would result in uncertain shape of membership functions, uncertainty associated with membership functions themselves are not clearly investigated. In addition, in some investigations, using kernel functions has been discussed, but the choice of kernel functions and kernel width depends on the problem at hand and thus these clustering methods could not be easily generalized. Moreover, in some of these methods, the Type-2 fuzzy membership functions are defuzzified into Type-1 fuzzy membership functions during each iteration, and therefore some information would be lost. On the other hand, almost all of the developed Type-2 fuzzy clustering methods are based either on FCM or on possibilistic c-means (PCM) clustering methods. FCM is a kind of partitioning algorithm that may cause serious problems in the case of having noisy observations. Moreover, there is no interaction between clusters in PCM, which may cause closely located or even coincide clusters [28]. Hence, introducing a new generalized clustering method based on Interval Type-2 fuzzy sets is essential for real world applications.

Therefore, in this paper, we proposed a novel Interval Type-2 fuzzy clustering method, in which the uncertainty associated with membership functions is the main concern. That is, in the proposed clustering method, Interval Type-2 fuzzy membership functions are directly modeled using interval arithmetic[1] theorems. Moreover, in order to overcome the FCM and PCM's shortcomings, the proposed clustering method is modeled using Relative Entropy Fuzzy C-Means (REFCM) method proposed by Zarinbal et al. [36]. The performance of this new clustering method, Interval Type-2 Relative Entropy Fuzzy C-Means (IT2 REFCM), is investigated using several experiments. In the first four experiments, the dimensions of the datasets are increased while the number of clusters is kept unchanged. In the fifth experiment, eight two-dimensional datasets with different cluster numbers are provided. Finally, the performance of the proposed method in real applications is evaluated using four medical images. The obtained results show that the proposed method has a very good ability in detecting noises and assignment of suitable membership degrees to observations.

Based on above discussions, the rest of this paper is organized as follows: Section 2, related works, provides some discussions on Type-2 fuzzy clustering methods in literature. The Interval Type-2 Relative Entropy Fuzzy C-Means clustering method and its properties are presented in Section 3. Computational complexity and the performance of the proposed method are addressed in Section 4. The conclusions are stated in Section 5. Finally, Appendix A provides some properties of interval arithmetic and the proofs needed for property and theorem of the proposed method.

2. Related works

As mentioned in previous section, the membership functions of Type-2 fuzzy set are themselves fuzzy. That is, the Type-2 fuzzy set, $\tilde{F}$, is comprised of membership function $U_{\tilde{F}}(x,u)$ as [23]:

$$\widetilde{F} = \{((x,u), U_{\tilde{F}}(x,u)) | \; \forall x \in X, \; \forall u \in U_x \subseteq [0,1]\} \tag{1}$$

where X is universe of discourse, and $0 \leqslant u \leqslant 1$ and $0 \leqslant U_{\tilde{F}}(x,u) \leqslant 1$ denotes the primary and the secondary membership functions, respectively. In fact, $U_{\tilde{F}}(x,u)$ could be Type-1 fuzzy set (General Type-2 fuzzy set) or crisp interval in $[0,1]$ (Interval Type-2 fuzzy set). $\widetilde{F}$ can also be expressed as [23]:

$$\widetilde{F} = \int_{x \in X} \int_{u \in J_x} U_{\tilde{F}}(x,u)/(x,u) \quad J_x \subseteq [0,1] \tag{2}$$

where $\int\int_{(\cdot)}$ denotes union over all admissible x and u [23].

Combination of Type-2 fuzzy set theory and clustering methods would give more flexibility to these methods in handling uncertainties and it has resulted in many clustering methods. Interval Type-2 fuzzy c-means clustering [15], Interval Type-2 fuzzy c-regression clustering [8], Interval Type-2 approach to kernelized fuzzy c-means clustering [16], Interval Type-2 approach to kernel possibilistic c-means clustering [34], Interval Type-2 possibilistic c-means clustering [10], and Interval-valued fuzzy relation-based clustering [12] are fuzzy clustering methods that utilize Interval Type-2 fuzzy set theory.

[1] Interval arithmetic performs arithmetic operations on closed intervals and represents real numbers between the lower and upper endpoints such that the true result certainly lies within this interval [5].