Expert Systems With Applications 91 (2018) 27–35

Contents lists available at ScienceDirect

Expert Systems With Applications

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

A histogram approach for determining fuzzifier values of interval type-2 fuzzy *c*-means

Kashif Hussain Memon*

Department of Electronics and Communication Engineering, Hanyang University ERICA Campus, Ansan, Korea

ARTICLE INFO

Article history:
Received 9 May 2017
Revised 30 July 2017
Accepted 22 August 2017
Available online 24 August 2017

Keywords:
Fuzzy *c*-means
Interval type-2 FCM
Fuzzifier

ABSTRACT

Type-2 fuzzy sets are preferred over type-1 sets as they are capable of addressing uncertainty more efficiently. Fuzzifier values play a pivotal role in managing these uncertainties; still selecting an appropriate value of fuzzifier has been a tedious task. Generally, based on observation, a particular value of fuzzifier is chosen from a given range of values for a given dataset. In this paper, I have tried to adaptively compute suitable fuzzifier values of interval type-2 fuzzy *c*-means for a given pattern. Information is extracted from individual data points using histogram approach and this information is further processed to give us the two fuzzifier values m_1 and m_2. These obtained values are bounded within some upper and lower bounds based on existing methods.

1. Introduction

Fuzzy *c*-means (FCM) is an unsupervised form of a clustering algorithm in which unlabeled data $X=\{x_1,x_2,..., x_N\}$ is grouped together according to their fuzzy membership values (Bezdek, Ehrlich, & Full, 1984; Cannon, Dave, & Bezdek, 1986). Since, in data analysis and computer vision problems, analyzing and dealing the uncertainties is a very important issue, FCM is being vastly used in these fields. Type-1 fuzzy sets cannot deal uncertainties therefore, type-2 fuzzy sets were defined to represent the uncertainties associated with type-1 fuzzy sets (Mendel & John, 2001; Mendel & John, 2002; Mendel, 2001; Mendel, 2004; Zadeh, 1975). Though the computational complexity of typ-2 fuzzy sets is higher than that of type-1 fuzzy sets, but the results obtained through the type-2 fuzzy sets are much better than the results obtained through the type-1 fuzzy sets. Therefore, if type-2 fuzzy sets can provide significant improvement on performance (depending on the application), the increase of computational complexity due to type-2 fuzzy sets may be a small price to pay (Mendel, 2004). Interval type-2 fuzzy sets has lower computational complexity compared to the general type- fuzzy sets.

The degree of fuzziness is one type of uncertainty which is dealt in interval type-2 fuzzy *c*-means. The fuzzifier m used in type-1, type-2 FCM, single and multiple kernel FCM, ranges from $[1, +\infty)$, and its value significantly affects the formation of clusters (Bezdek, 1981; Huang, Chuang, & Chen, 2012; Linda & Manic, 2012; Melin & Castillo, 2014; Nguyen & Ngo, 2013; Zhang, Zhang, & Chen, 2003). Generally, $m=2$ is selected as the fuzzifier value in FCM algorithm (Bezdek, 1976). Several theories have been put forward on the possible range of values of m.

Based on observation, initially upper and lower bound was given to be in between 1.1 and 5 (Bezdek, 1981). Again, based on practical observation, for a given number of data points (n), it was suggested that fuzzifier should be chosen such that $m \geq$ n/(n-2) (Bezdek, Hathaway, Sabin, & Tucker, 1987). Later on, some researchers showed that range of m is between 1.25 and 1.75 (Chan & Cheung, 1992), while others showed it to be in between 1.5 and 2.5 (Pal & Bezdek, 1995).

The degree of fuzziness is dependent on data, so I cannot use the same bound for all data (Yu, Cheng, & Huang, 2004). Two points which do not depend on m are the mass center and the cluster center. Exploiting this information, some researchers tried to provide upper and lower bound for fuzzifiers according to the given dataset (Huang, Xia, Wang, Zeng, & Wang, 2012; Ozkan & Turksen, 2007). It can be seen that in FCM, a fixed value m is selected and all the above research works give us a range from which I need to manually select one value for a particular pattern. Recently work was done to fix this issue and researchers tried to automatically tune the fuzziness control parameter (Das, Sinha, Chakravarty, & Konar, 2013). Apart from this, not much work has been done in this regard and none for automatic tuning of fuzzifier value of interval type-2 FCM. Therefore, I propose adaptive computation of the value of m for interval type-2 FCM.

In my proposed method, I have tried to generate footprint of uncertainty (FOU) using histogram approach and then extract

* Corresponding author at: Computer Systems Engineering, University College of Engineering Technology, Bahawalpur, Pakistan.

E-mail addresses: kashi@hanyang.ac.kr, kashi.memon@gmail.com, kashif.hussain@iub.edu.pk

http://dx.doi.org/10.1016/j.eswa.2017.08.041
0957-4174/

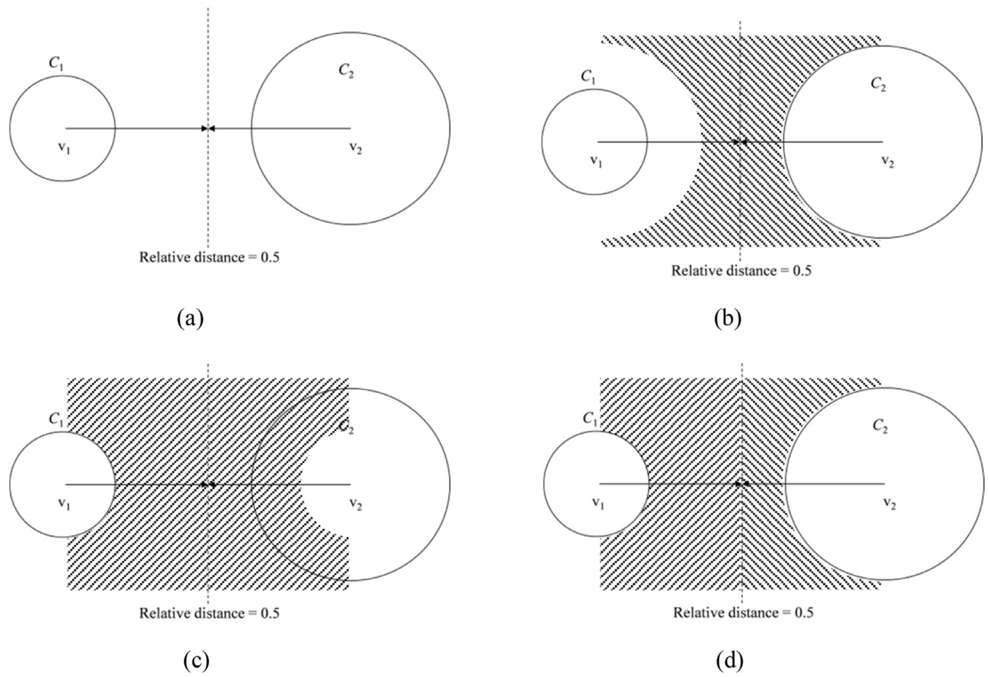

Fig. 1. (a) No fuzzy region when $m=1$ (b) low degree of fuzziness for m_1 (c) high degree of fuzziness for m_2 (d) FMR formed using both m_1 and m_2.

information from individual data points. Membership value thus obtained was further used to calculate the degree of fuzziness (m) for the given pattern set. Finally, this level of fuzziness was integrated into existing interval type-2 FCM to get the desired clusters.

The rest of the paper is organized into four sections. In Section 2 a theoretical background of some related algorithms is provided from where I have derived my own idea. After this, my proposed method is explained in detail in Section 3. Since no other algorithm exists to deal with the tuning of fuzzifier values; in Section 4 I have compared my results with results obtained while using some pre-selected fuzzifier values. Finally, in Section 5 I state my conclusion.

2. Background

This section includes discussion about fuzzifiers, their effect on memberships and cluster formation and one of the methods to determine their upper and lower bound. Later on, a brief account of fuzzy membership function generation using histogram approach method has also been given.

2.1. Effect of fuzzifier on membership generation

When the density or volume of clusters is different, then selection of the value of fuzzifier plays a crucial role in fuzzy membership generation. The maximally fuzzy membership locations (vertical line) are where the points are equidistant to all the cluster centers and their relative distance from each cluster center is same and equal to 0.5 as shown in Fig. 1(a). This vertical line can also be considered as decision boundary as $m \rightarrow 1$. At this point, FCM is nothing but hard c-means. As the value of m (degree of fuzziness) increases, the maximum fuzzy region (MFR) increases. This can be seen in Fig. 1(b) and (c). When $m \rightarrow +\infty$, the centers of various groups in FCM are degraded into almost the center of all the data. By using two fuzzifier values in interval type-2 FCM ($m_2 > m_1$), I am able to control MFR more accurately as shown in Fig. 1(d).

2.2. Deciding the range of fuzzifier values

As discussed earlier, few methods have been proposed for determining the upper and lower bounds of fuzzifier applicable for a given dataset (Huang et al., 2012; Ozkan & Turksen, 2007). The FCM membership function of x_i data point for jth cluster is given by

$$u_j(\mathbf{x}_i) = \frac{1}{\sum_{k=1}^{C} (d_{ij}/d_{ik})^{2/(m-1)}} \tag{1}$$

where C represents the number of clusters, d_{ij} is distance of i^{th} point from j^{th} cluster prototype. It is evident from this expression that membership function is independent of m at two points:

(1) The point which is equidistant to all the cluster centers (also known as the mass center). It has a membership value of 1/C (C is a number of cluster centers).
(2) The cluster centers, which have membership values of 1 for their cluster but 0 for other clusters (Huang et al., 2012).

So, to find the bounds, membership in the neighborhood is calculated and then, writing that expression in terms of m gives us the upper and the lower bound of m as given in (2).

$$1 + \frac{C-1}{C} \cdot \frac{2}{\delta} \cdot |\Delta| \le m \le \frac{2 \log d}{\log\left(\frac{\delta}{1-\delta} \cdot \frac{1}{C-1}\right)} + 1 \tag{2}$$

where d is the distance of data points from all the centers, $\Delta = \frac{d_j - d_j^*}{d_j^*}$, where d_j is the distance of data points from all jth cluster prototype, d_j^* is the distance of data points from all the cluster prototype except jth cluster prototype, and δ is the threshold.