Pattern Recognition Letters 33 (2012) 2280–2284

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

The range of the value for the fuzzifier of the fuzzy c-means algorithm

Ming Huang [a,b,*], Zhixun Xia [a], Hongbo Wang [a], Qinghua Zeng [a], Qian Wang [c]

[a] Science and Technology on Scramjet Laboratory, National University of Defense Technology, Hunan, Changsha 410073, China
[b] Xi'an Satellite Control Center, Shaanxi, Xi'an 710043, China
[c] Northwest Institute of Nuclear Technology, Shaanxi, Xi'an 710024, China

ARTICLE INFO

Article history:
Received 21 August 2011
Available online 6 September 2012

Communicated by G. Borgefors

Keywords:
Fuzzy c-means algorithm
Fuzzifier
The range of the value
The behavior of membership function

ABSTRACT

The fuzzy c-means algorithm (FCM) is a widely used clustering algorithm. It is well known that the fuzzifier, m, which is also called fuzzy weighting exponent, has a significant impact on the performance of the FCM. Most of the researches have shown that there exists an effective range of the value for m. However, since the method adopted by researchers is mainly experimental or empirical, it is still an open problem how to select an appropriate fuzzifier m in theory when implementing the FCM. In this paper, we propose a theoretical approach to determine the range of the value of m. This approach utilizes the behavior of membership function on two data points, based on which we reveal the partial relationship between the fuzzifier m and the dataset structure.

1. Introduction

The fuzzifier m, where $m \in [1, +\infty)$, is a parameter introduced into the clustering function WGSS by Bezdek (1981). m plays an important role in the FCM algorithm. A proper value of m can suppress the noise and smooth the membership function. However, there has been little theoretical basis for an optimal choice of m in the FCM.

Many heuristic strategies are recommended in the literature. There is an experiential scope of m given by Bezdek (1981), where the lower and the upper bounds are 1.1 and 5, respectively. In 1976, a physical interpretation of the FCM algorithm when $m = 2$ was also given by Bezdek (1976). From the aspect of word recognition, Chan and Cheung (1992) proposed that the value of m should be between 1.25 and 1.75. Considering the convergence of the algorithm, Bezdek and Hathaway (1987) indicated that $m > n/(n-2)$. Based on the performance of some cluster validity indices, Pal and Bezdek (1995) suggested that the value of m is probably in the interval [1.5,2.5]. Most researchers adopt $m = 2$ when performing the FCM algorithm. Some researchers (Hwang and Rhee, 2007; Yu, 2003; Yu et al., 2004) believe that the structure of dataset influences the value of m. Ozkan and Turksen (2004) focused on the uncertainty contained in m and proposed an entropy assessment for m. They (Ozkan and Turksen, 2007) also proved that the range value of m that captures the uncertainty generated by m itself is [1.4,2.6]. Gao (2004) proposed two methods to find the proper value of m. The first one is based on fuzzy decision theory, but it needs to define two membership functions which lack of theoretical basis. The second one is based on the concavo-convex property of clustering function, whose physical interpretation is of ambiguity.

In this paper, a theoretical approach used to determine the range of the value of m is proposed. This approach utilized the behavior of membership function on two special data points. The rest of the paper is organized in four sections. In Section 2, we analyze the connotation of m. In Section 3, two special data points and the behavior of their membership function are explained. In Section 4, we propose the approach to determine the range of the value of m. And the conclusions are drawn in Section 5.

2. The connotation of m

The parameters used in the FCM are number of clusters, cluster centers, level of fuzziness (fuzzifier) and similarity measure. Theory 1 describes what an important role the m plays in the FCM algorithm.

Theory 1

(1) if $m = 1$, then the FCM algorithm reduces to HCM algorithm.
(2) if $m \to 1^+$, then the FCM algorithm reduces to HCM algorithm with probability equal to 1, i.e. the FCM algorithm cannot do any fuzzy partition.

* Corresponding author at: Science and Technology on Scramjet Laboratory, National University of Defense Technology, Hunan, Changsha 410073, China. Tel.: +86 13975119796.

E-mail addresses: hmbob1@126.com (M. Huang), zxxia@nudt.edu.cn (Z. Xia), whbwatch@yahoo.com.cn (H. Wang), zqhkd@yahoo.com.cn (Q. Zeng), wangqian_gudu@163.com (Q. Wang).

0167-8655/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2012.08.014

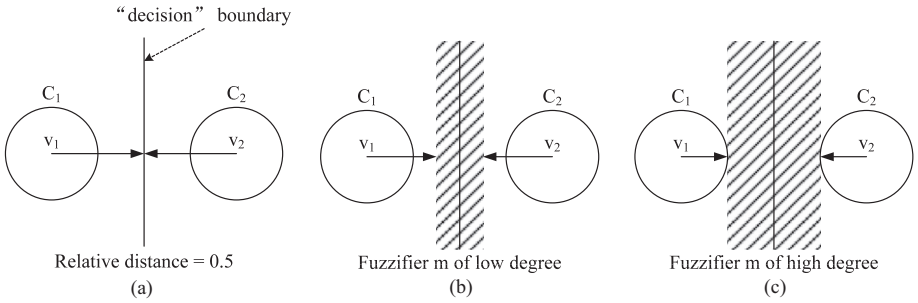

Fig. 1. An intuitionistic illustration of m.

(3) if $m \to +\infty$, then the membership matrix $U = [\mu_{ik}] = [1/c]$, i.e. the centers of various groups in the FCM are degraded into almost the center of all the data.

Therefore, the fuzzifier m controls the amount of fuzziness of the final C-partition in the FCM algorithm.

According to literature 6, an intuitionistic illustration of theory 1 is shown in Fig. 1.

Suppose that there are two clusters with the same structure and density in 2-D dataset, denoted as C_1 and C_2. The cluster centers are υ_1 and υ_2 respectively. The vertical line in Fig. 1(a) can be considered as a "decision" boundary where patterns are equally distant from the two cluster center, that is, the relative distance between a pattern and each cluster center equaling 0.5. The case shown in Fig. 1(a) is a crisp membership assignment in the FCM. That means the patterns located to the left (right) of the boundary belongs to cluster C_1 (C_2). This boundary can be expanded by a fuzzifier m as shown in Fig. 1(b) and (c). If the value of fuzzifier m is increased then the maximum fuzzy boundary becomes wider.

3. Membership grade

3.1. The membership function

The FCM membership function is calculated as:

$$\mu_{i,k} = \left[\sum_{t=1}^{c}\left(\frac{\|x_k - v_i\|_A}{\|x_k - v_t\|_A}\right)^{\frac{2}{m-1}}\right]^{-1} \tag{1}$$

where $\mu_{i,k}$ is the membership value of kth sample in ith cluster such that $\mu_{i,k} \in [0,1]$, c is the number of clusters, x_k is the kth sample, v_i is the cluster center of ith cluster, $\|\cdot\|_A$ is the norm function, and $\sum_{t=1}^{c}\mu_{t,k} = 1$ for a given $m > 1$. This means that the sum of the degrees of membership values of any data is one, or in other words, any data should be a member of at least one of the clusters with a membership value greater than zero.

3.2. Two rules

The expression indicates that the membership value is controlled by fuzzifier m. However, there are two points where the membership values do not depend on m. One of the points is the mass center that has equal distance to all cluster centers and thus has a membership value $1/c$ to all cluster centers. It is identified by cluster centers and continuous membership function such that it has equal distance from all the cluster centers. In addition, when m goes to infinity, cluster centers collapse to this point. Hence this value clearly does not depend on m. The other points are the cluster center values which have a membership value 1 in its cluster and 0 to all others. Hence these values also do not depend on m. According to the definition of membership, we can obtain two tenable rules which can assist us to find the reasonable range of the value of m.

Rule 1. The membership value of sample p is $\mu_{i,p} \to 1/c$, if p is located in the neighborhood of the mass center.

Rule 2. The membership value of sample q is $\mu_{i,q} \to 1$, if q is located in the neighborhood of the cluster center v_i.

3.3. Calculation of $\mu_{i,p}$ and $\mu_{i,q}$

In order to calculate the $\mu_{i,p}$, we expand the function around the mass center by using Taylor series expansion.

One-dimensional Taylor series of a real function $f(x)$ about a point $x = x_0$ is given by

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!}f''(x_0)(x - x_0)^2 + \cdots \\ &\quad + \frac{1}{n!}f^{(n)}(x_0)(x - x_0)^n + \cdots \\ &= f(x_0) + f'(x_0)(x - x_0) + R \end{aligned} \tag{2}$$

where R is the remainder. Let d^* denote the distance measure to all cluster centers from the mass center, d_i denotes the distance to ith cluster center of the point located in the neighborhood of the mass center. Then, Taylor series of the $\mu_{i,p}$ can be written as

$$\mu_{i,p} = \mu_{i,p}\Big|_{d_i^*} + \frac{\partial}{\partial d_i}\mu_{i,p}\Big|_{d_i^*}(d_i - d_i^*) + R \tag{3}$$

wherein,

$$\mu_{i,p}|_{d_i^*} = \frac{1}{c}, \quad \frac{\partial \mu_{i,p}}{\partial d_i} = -\frac{\sum_{t=1,t\neq i}^{c}\left(\frac{2}{m-1}\right)\left(\frac{d_i}{d_t}\right)^{\frac{2}{m-1}}d_i^{-1}}{\left[\sum_{t=1}^{c}\left(\frac{d_i}{d_t}\right)^{\frac{2}{m-1}}\right]^2}$$

Since the derivative $\frac{\partial \mu_{i,p}}{\partial d_i}$ should be evaluated at the mass center where $d_j = d^*$, for $j = 1, \ldots, c$, we obtain

$$\frac{\partial}{\partial d_i}\mu_{i,p}|_{d_i^*} = -\frac{(c-1)\left(\frac{2}{m-1}\right)}{\left[\sum_{t=1}^{c}1\right]^2}\frac{1}{d_i^*} = -\frac{(c-1)\left(\frac{2}{m-1}\right)}{c^2}\frac{1}{d_i^*} \tag{4}$$

Neglect the remainder R, then

$$\mu_{i,p} \cong \frac{1}{c} - \frac{(c-1)\left(\frac{2}{m-1}\right)}{c^2}\left(\frac{d_i - d_i^*}{d_i^*}\right) = \frac{1}{c} - \frac{(c-1)\left(\frac{2}{m-1}\right)}{c^2}(\Delta) \tag{5}$$

We can use the membership function directly to calculate $\mu_{i,q}$. Let d_i denote the distance to the ith cluster center from a point located in the neighborhood of v_i. d_i is very small in value compared to the distance to all the other cluster centers from this point. Thus in general, let $\frac{d_i}{d_{j,j\neq i}} = d$, then

$$\mu_{i,q} = \left[1 + (c-1)d^{\frac{2}{m-1}}\right]^{-1} \tag{6}$$