Knowledge-Based Systems 96 (2016) 76–83

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Exploring the uniform effect of FCM clustering: A data distribution perspective

Kaile Zhou[a,b,*], Shanlin Yang[a,b]

[a] Key Laboratory of Process Optimization and Intelligent Decision-making of Ministry of Education, Hefei University of Technology, Hefei 230009, China
[b] School of Management, Hefei University of Technology, Hefei 230009, China

ARTICLE INFO

Article history:
Received 6 August 2015
Revised 25 November 2015
Accepted 2 January 2016
Available online 8 January 2016

Keywords:
Fuzzy c-means (FCM)
Data distribution
Uniform effect
Coefficient of variation (CV)
Clustering

ABSTRACT

Fuzzy c-means (FCM) is a well-known and widely used fuzzy clustering method. Though there have been considerable studies that focused on the improvement of FCM algorithm or its applications, it is still necessary to understand the effect of data distributions on the performance of FCM. In this paper, we present an organized study of FCM clustering from the perspective of data distribution. We first analyze the structure of the objective function of FCM and find that FCM has the same uniform effect as K-means. Namely, FCM also tends to produce clusters of relatively uniform sizes. The coefficient of variation (CV) is introduced to measure the variation of cluster sizes in a given data set. Then based on the change of CV values between the original "true" cluster sizes and the cluster sizes partitioned by FCM clustering, a necessary but not sufficient criterion for the validation of FCM clustering is proposed from the data distribution perspective. Finally, our experiments on six synthetic data sets and ten real-world data sets further demonstrate the uniform effect of FCM. It tends to reduce the variation in cluster sizes when the CV value of the original data distribution is larger than 0.88, and increase the variation when the variation of original "true" cluster sizes is low.

1. Introduction

Clustering [1] is an unsupervised learning process to discover significant patterns in a given data set by partitioning the data set into groups (i.e., clusters) such that the data objects assigned to the same group are as similar as possible while those in different groups are dissimilar to the greatest extent. Clustering has been widely applied in many fields [2–5]. Fuzzy c-means (FCM) clustering introduced the concept of membership degree to measure the extent to which a data object belongs to different groups. Since it was first proposed by Dunn [6] and generalized by Bezdek [7], FCM has become a well-known and widely used clustering method [8–11].

Currently, there have been considerable research works that focused on the algorithm improvement and application of FCM. Some researchers focused on the improvement of FCM so that it can be applied to different types of data, such as time series data [12], relational data [13], categorical data [14], etc. The improvement of similarity/dissimilarity function in fuzzy clustering is also an important research topic [15–17]. In addition, to expand the application scope of FCM clustering, some supervised knowledge was integrated to form the semi-supervised fuzzy clustering methods [18–20]. The fuzzifier, an important parameter in FCM to measure the fuzziness of partitions, also has attracted wide attention in literature [21–23].

From the perspectives of data distributions, K-means algorithm [24] and hierarchical clustering [25] have been studied. Xiong et al. [24] found that K-means tends to produce clusters of relatively uniform sizes, even the input "true" cluster sizes are varied. The research of Wu et al. [25] revealed that hierarchical clustering tends to produce clusters with high variation on cluster sizes regardless of the "true" cluster distributions. However, FCM is different from these two kinds of clustering methods, due to the fact that both K-means and hierarchical clustering are hard (crisp) clustering methods while FCM is a fuzzy clustering method. Both the membership degree and fuzzifier parameter were introduced in FCM algorithm. Therefore, along this line, we aim to investigate the FCM clustering from a data distribution perspective in the current study.

Although people have identified that the data characteristics may influence the performance of FCM clustering in existing studies [26,27], an organized and systematic study is still needed to investigate how data distributions can affect the performance of FCM clustering, particularly the interaction between fuzzifier parameter

* Corresponding author at: School of Management, Hefei University of Technology, Hefei 230009, China. Tel.: +86 15955112340.
E-mail address: zhoukaile@hfut.edu.cn, kailezhou@gmail.com (K. Zhou).

http://dx.doi.org/10.1016/j.knosys.2016.01.001
0950-7051/

in FCM and the data distributions partitioned by FCM clustering. Therefore, to further enhance our understanding and guide us for the better use of FCM clustering in practical applications, it is necessary to study the effect of data distribution on the performance of FCM clustering.

In this paper, we first present a theoretical analysis of the objective function of FCM clustering. Based on the analysis of the structure of its object function, we find that there are three factors, namely the cluster sizes, the distances among clusters and the fuzzifier parameter, which can affect the performance of FCM. To further demonstrate the effect of the original data distribution on the clustering result of FCM, we then conducted extensive experiments on both synthetic data sets and real-world data sets from the UCI machine learning repository [28], with different data distributions in cluster sizes. We also introduced the coefficient of variation (CV) [29] to measure the dispersion of a given data set. The CV is a dimensionless number that can be used to compare the variations of populations with significantly different mean values. Generally, a larger value of CV means a greater variability in cluster sizes of a data set. Therefore, the change of CV values, represented by DCV, before and after clustering can serve as a necessary but not sufficient criterion to validate the performance of FCM clustering. In addition, since the fuzzifier in FCM also has an impact on the clustering results of FCM, the DCV values can also help us to better understand the effect of fuzzifier.

Our experimental results demonstrated that FCM tends to reduce the variation of a data set in cluster sizes when the CV value of original data distribution is larger than 0.88, but increase the variation when the variation of original "true" cluster sizes is low. Namely, FCM has the similar uniform effect as K-means clustering [24]. However, the uniform effect of FCM is also influenced by the value of fuzzifier.

The remainder of this paper is organized as follows. A brief introduction of FCM clustering and the selection of fuzzifier are presented in Section 2. Then in Section 3, the objective function of FCM is analyzed and discussed theoretically, followed by the introduced CV measure and an illustrative example. Section 4 provides the results of the extensive experiments and some discussions. Finally, conclusions are drawn in Section 5.

2. Related works

2.1. Clustering and FCM clustering

For a given data set $X=\{x_1,x_2,\ldots,x_n\}$, clustering algorithms [30] partition the n data objects in X into c groups $C=\{C_i,C_2,\ldots,C_c\}$ based on similarity/dissimilarity metric, and a partition matrix $U(X)$ is obtained. The partition matrix is expressed as $U=[\mu_{ij}]_{c\times n}$ $(i=1,\ldots,c,\ j=1,\ldots,n)$, where μ_{ij} is the membership degree of data object x_j to cluster C_i. If μ_{ij} satisfies

$$\mu_{ij}=\begin{cases}1 & \text{if } x_j\in C_i\\ 0 & \text{otherwise}\end{cases} \tag{1}$$

and the partition results satisfy: $C_i\neq\emptyset$ $(i=1,\ldots,c)$, $C_i\cap C_j=\emptyset$ $(i=1,\ldots,c,\ j=1,\ldots,n)$, and $\cup_{i=1}^{c}C_i=X$, then this kind of clustering partition is hard clustering, also called as crisp clustering. K-means clustering [31] and hierarchical clustering [32] are two typical hard clustering methods.

While in fuzzy clustering, the degrees of each data object to different clusters are represented by the values of membership degree. Generally, a data object is grouped into the cluster to which it has the maximum value of membership degree. The membership degree μ_{ij} in fuzzy clustering satisfies $\mu_{ij}\in[0,1]$, and

$$0<\sum_{j=1}^{n}\mu_{ij}<n,\quad \forall i=1,\ldots,c \tag{2}$$

$$\sum_{i=1}^{c}\mu_{ij}=1,\quad \forall j=1,\ldots,n \tag{3}$$

$$\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}=n \tag{4}$$

Fuzzy c-means (FCM) clustering is a popular fuzzy clustering method [33–35]. It starts with determining the number of groups followed by the random selection of initial cluster centers. Then, each data object is assigned a membership degree to each cluster. Each cluster center point and corresponding membership degrees are updated iteratively by minimizing the objective function until the positions of the cluster centers will not change or the difference of objective function values between two iterations is within a threshold.

The objective function of FCM is defined as

$$J_c=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}||x_j-v_i||^2 \tag{5}$$

where v_i is the cluster center of cluster C_i, and$v_i=(1/n_i)\sum_{x_j\in C_i}x_j$. $n_i$is the number of data objects in cluster C_i. m is the fuzzifier, also referred to as the weighting exponent or fuzziness parameter in FCM. $\|\cdot\|$ represents the Euclidean distance.

In the iterations, the membership degree $\mu_{ij}$and the cluster centers v_i are updated as

$$\mu_{ij}=1/\sum_{r=1}^{c}(d_{ij}/d_{rj})^{\frac{2}{m-1}} \tag{6}$$

$$v_i=\sum_{j=1}^{n}\mu_{ij}^{m}x_j/\sum_{j=1}^{n}\mu_{ij}^{m} \tag{7}$$

2.2. Fuzzifier selection of FCM algorithm

The fuzzifier, denoted as m, is an important parameter and has a significant impact on the clustering result of FCM [36]. There have been considerable research efforts that focused on the selection of this important parameter in FCM. Pal and Bezdek [37] presented a heuristic rule for the optimal selection of m, and m was limited to [1.5, 2.5]. The similar suggestion was provided in [38]. In addition, Bezdek [39] studied the physical interpretation of FCM when $m=2$, and suggested that $m=2$ is the optimal selection. Based on the study of word recognition, Chan and Cheung [40] suggested that the value of m should range in [1.25, 1.75]. However, Choe and Jordan [41] pointed out that the performance of FCM was not sensitive to the value of m based on fuzzy decision theory. To obtain the uncertainty brought by m in FCM, Ozkan and Turksen [22] identified that the upper and lower values of m were 1.4 and 2.6 respectively. Wu [21] proposed a new guideline for the selection of m based on a robust analysis of FCM, and suggested to implement FCM with $m\in[1.5,4]$. Additionally, some researchers [23,42,43] studied the influence of data set structure on the selection of m.

Currently, there is still little theoretical guidance and widely accepted criterion to support the selection of m in FCM [23,37]. This parameter is always selected subjectively in practical applications. The general and most widely used fuzzifier value in FCM applications is $m=2$ [37–39].

In this paper, the effect of fuzzifier m on the clustering results of FCM from a data distribution perspective is also discussed, which can provide some support for us to select the optimal value of m.