A Kernel Fuzzy Clustering Algorithm with Generalized Entropy Based on Weighted Sample

Kai Li¹, Lijuan Cui²

Abstract

Aiming at fuzzy clustering with generalized entropy, a kernel fuzzy clustering algorithm with generalized entropy based on weighted sample is presented. By introducing weight of sample into objective function for fuzzy clustering with generalized entropy, we obtain optimization problem for fuzzy clustering with generalized entropy based on weighted sample. And we use Lagrange multiplier method to solve corresponding optimization problem and obtain degree of membership for each sample belonging to different cluster, centers of clusters and weights of samples. Following that, a kernel fuzzy clustering algorithm with generalized entropy based on weighted sample is presented. We select the representative dataset Iris from UCI repository to conduct experiments. Experimental results show the effectiveness of presented algorithm above.

Keywords

Fuzzy clustering, generalized entropy, weighted sample, kernel.

1. Introduction

Clustering is an important data analysis method and has been applied to pattern recognition, data mining, image segmentation and etc. Up to now, researchers have proposed many different clustering algorithms, where division-based cluster analysis (also called as objective function based cluster analysis) is one of the commonly used methods, such as K-means and Fuzzy C-means. Recently, to deal with clustering for difficult dataset, researchers introduce kernel into fuzzy clustering. Zhang et al [1] presented fuzzy c-means and possible c-means algorithm based kernel. Then they modified BCFCM (Bias Corrected FCM) algorithm and proposed the KFCM algorithm with spatial constraints[2].

Yang [3] further studied kernel clustering algorithm with spatial correction to propose GKFCM, and was successfully applied to image segmentation. Kannan et al [4], Swagatam et al [5] studied image segmentation based on kernel fuzzy clustering. It is seen that clustering algorithms above only consider data points or data attributes with the same importance. To solve these problems, researchers have proposed many different improved algorithms. Huang et al [6] introduced variable weights to the k-means clustering process and presented a k-means type clustering algorithm that can automatically calculate variable weights. Jing et al[7] included the weight entropy in the objective function to extend the k-means clustering process. They calculate a weight for each dimension in each cluster and use the weight values to identify the subsets of important dimensions that categorize different clusters. To reduce the FCM algorithm’s dependence on the initial cluster centers and data sets, Su et al [8] introduced weighting parameter to adjust the location of cluster centers and noise problem. To consider the particular contributions of different features, Li et al[9] presented a new feature weighted fuzzy clustering algorithm. In this paper, a kernel fuzzy clustering with generalized entropy based on weighted sample is studied. In the process of clustering, with changes of degree of membership for each sample and centers of clusters, weights of samples are updated.

This paper is organized as follows. In section 2, we study kernel fuzzy clustering with the generalized entropy based on weighted sample and obtain a kernel fuzzy clustering algorithm with generalized entropy based on weighted sample. In the section 3, we choose commonly used dataset Iris from UCI to test the presented algorithm’s performance. In the final section, conclusion is given.

2. Kernel fuzzy clustering with generalized entropy based on weighted sample

Let \( X = \{x_1, x_2, \ldots, x_n\} \) be a data set, where \( x_i \in \mathbb{R}^c \), \( c \) is a positive integer greater than one and \( m \geq 1 \) is fuzzy index, \( \mu_i = \mu_i(x_i) > 0 \) is degree of membership...
Objective \( J(U,V) = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij} m \left( 1 - K(x_j, v_i) \right) \) + \( \beta \left( \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^m - 1 \right) \) \( \lambda \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^m - 1 \) + \( \gamma \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^m - 1 \)

Here, we have introduced weights \( w_j \) into objective function (1) above. So, we obtain the following objective function on the basis of (1)

\[
J_{AWGEK}(U,V) = \sum_{j=1}^{c} \sum_{i=1}^{n} w_j^m \left( \Phi(x_j) - \Phi(v_i) \right) \left\| x_j - v_i \right\|^2 + \delta \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^m \left( 1 - K(x_j, v_i) \right) \frac{1}{\sigma^2} \]

where \( \beta \) is parameter, \( w_j \) is the weight of the \( j \)-th sample.

Optimization problem for kernel fuzzy clustering thus is as follows:

\[
\text{min } J_{AWGEK}(U,V) \]

s.t. \( \sum_{j=1}^{c} \mu_{ij} = 1, j = 1, 2, \ldots, n \) \( \sum_{j=1}^{c} w_j = 1 \)
Step 2 Compute degree of membership $\mu_{ij}$ for each sample according to (11).
Step 3 Compute center of cluster $v_i$ for each cluster according to (12).
Step 4 Calculate weights of samples $w_j$ according to (13).
Step 5 Repeat step 2 to step 4 until the center of cluster $v_i$ does not change.

3. Experimental results and analysis

In order to verify the effectiveness of the proposed algorithm AWGEKFCM, we select Iris dataset from UCI data repository. In addition, we use two indexes to evaluate performance of clustering algorithm. They are accuracy $ACC$ and mutual information MI, respectively, where $ACC$ is represented as $ACC = \frac{n - err}{n} \times 100\%$, $n$ is number of samples in dataset $X$ and $err$ is number of misclassified samples. Moreover, we initialize centers of clusters by randomly choosing method from dataset.

Firstly, we conduct the detailed experimental study for dataset Iris. In the experiment, parameters $\delta$ and $\sigma$ is fixed as value -0.01 and 15, respectively; fuzzy index $m$ is taken as 1.1, 1.5, 2, 2.5, 3, 5, 7, 9 and 11, respectively and $\beta$ is taken as 3, 0, -1, -3, -5 and -7, respectively. Experimental results are shown in Figure 1, where (a), (c), (e), (g), (i) and (k) are relation between fuzzy index $m$ and accuracy whereas (b), (d), (f), (h), (j) and (l) are relation between fuzzy index $m$ and mutual information (MI). It is seen that when $\beta>0$, performance of clustering is very poor, when $\beta=0$ or $\beta<0$, better performance of clustering is obtained.
Secondly, for dataset Iris, we fix $\delta$ and $\beta$ as value -0.01 and -5, respectively. Let fuzzy index $m$ be taken value as 1.1 and 9. Let $\sigma$ be take value as 1.5, 5, 10, 15, 25, 50 and 100, respectively. Experimental results are shown in Figure 2.

By Figure 2, we know that when $\sigma$ is in the range of 5 to 25, better clustering performance is obtained. Besides, we compare performance of AWGEKFCM with that of FCM, GEFCM and KGEFCM [10]. Experimental result is given in Figure 3. It is seen that we obtain better clustering results compared with FCM, GEFCM and KGEFCM in dataset Iris.
Different clustering method

Figure 3: Performance of clustering for different method

4. Conclusions

Aiming at fuzzy clustering with generalized entropy, we study its kernel fuzzy clustering with weights of samples. Objective function for fuzzy clustering with generalized entropy based on weighted sample is obtained. Then, kernel fuzzy clustering algorithm with generalized entropy based on weighted sample is presented. We select representative dataset Iris to conduct experimental study. Experimental results show that the presented algorithm is effective.

References