Neurocomputing 83 (2012) 98–109

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

A novel chaotic particle swarm optimization based fuzzy clustering algorithm

Chaoshun Li*, Jianzhong Zhou, Pangao Kou, Jian Xiao

School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

A R T I C L E I N F O

Article history:
Received 15 July 2011
Received in revised form
18 October 2011
Accepted 1 December 2011
Communicated by M. Sato-Ilic
Available online 5 January 2012

Keywords:
Fuzzy clustering model
Fuzzy *c*-means
Chaotic particle swarm optimization
Gradient method
Iterative chaotic map with infinite collapses

A B S T R A C T

Clustering is a popular data analysis and data mining technique. In this paper, a novel chaotic particle swarm fuzzy clustering (CPSFC) algorithm based on chaotic particle swarm (CPSO) and gradient method is proposed. Fuzzy clustering model optimization is challenging, in order to solve this problem, adaptive inertia weight factor (AIWF) and iterative chaotic map with infinite collapses (ICMIC) are introduced, and a new CPSO algorithm combined AIWF and ICMIC based chaotic local search is studied. The CPSFC algorithm utilizes CPSO to search the fuzzy clustering model, exploiting the searching capability of fuzzy *c*-means (FCM) and avoiding its major limitation of getting stuck at locally optimal values. Meanwhile, gradient operator is adopted to accelerate convergence of the proposed algorithm. Its superiority over the FCM algorithm and another two global optimization algorithm-based clustering methods is extensively demonstrated for several artificial and real life data sets in comparative experiments.

1. Introduction

Clustering is a classification technique for data analysis, which is used to classify a set of data or patterns usually multidimensional in nature into different groups according to a predefined criterion, so that objects in the same group are more similar than those in different groups. More specifically, the patterns that are usually s dimensional vectors are distributed to c classes while certain kind of optimization criterion is minimized, and the patterns in the same class are more similar than those in different classes in the end. In recent decades, clustering has played central role in diverse domains of science and engineering applications, such as, data mining, pattern recognition, machine learning, image segmentation and fault diagnosis [1–3].

Many clustering algorithms have been developed, which can be performed in two different modes, namely, crisp and fuzzy. Crisp clustering algorithms have the advantages that they are easy to implement and efficient. *K*-means [4], ISODATA [5], LVQ [6] are typical crisp clustering algorithms. Although crisp clustering is widely used and well developed, it is not suitable to deal with fuzzy data sets, because it assumes classes of data set to be non-overlapped and strictly separated, so the membership value to clusters is zero or one. Bezdek developed a fuzzy clustering algorithm, the well-known fuzzy *c*-means (FCM) [7], to overcome the limitation of crisp clustering by distributing fuzzy membership value to clusters for patterns. In fact, crisp clustering may be viewed as a special case of fuzzy clustering, where a pattern gets a membership of 1 in the cluster to which it belongs and a membership value of 0 to all the other clusters. Once a fuzzy cluster structure is determined, one can determine a crisp cluster structure by replacing the highest membership value of a pattern by one and all other membership values by zero. However, when the clusters present in the data are overlapping in nature, the fuzzy clustering may provide more information to the higher level processes using it [8].

For the original clustering algorithms, for example, FCM or *K*-means, the number of clusters, c, is known a priori. In such situations clustering may be formulated as distribution of n patterns in N dimensional metric space among c groups. This involves minimization of some extrinsic optimization criterion. The major drawback of these algorithms is that they often get stuck at local minima and the result is largely dependent on the choice of the initial cluster centers [9]. Researchers have tried to improve FCM by introducing excellent optimization methods to optimize the objective function of FCM, trying to avoid trapping into local minima. Global optimization algorithms, like Genetic algorithms (GA) [10–12], ant colony optimization [13], Chaos optimization [14], have already been adopted to improve FCM. Particle swarm optimization (PSO) algorithms are randomized search and optimization techniques based on the concept of swarm. They are efficient, adaptive and robust search processes, performing multi-dimensional search in order to provide near

* Corresponding author. Tel.: +86 2787543992.
E-mail address: lichaoshun@yahoo.com.cn (C. Li).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2011.12.009

optimal solutions of an evaluation (fitness) function in an optimization problem [15]. Since the problem of clustering may be viewed as searching for a number of clusters in the feature space such that a given clustering metric is optimized, application of PSO to this problem seems natural and appropriate. Such attempts can be found in [16,17] However, the performance of PSO greatly depends on its parameters, and similar to GA, it often suffers from being trapped in local optimum [18,19].The clustering problem discussed in this paper are treated as an optimization problem, so it is necessary to bring in or designed more excellent optimization method to improve FCM.

Due to the ergodic and dynamic properties of chaos variables, chaotic search is more capable of hill-climbing and escaping from local optima than random search [20], and thus has been applied to the area of optimization computation. In the last two decades, various chaos-based optimization algorithms, for example, a chaos-based simulated annealing algorithm (CSA) [21], a hybrid chaotic ant swarm optimization [22], chaotic bee colony algorithms (CABC) [23] and chaotic particle swarm optimization algorithms (CPSO) [19], have been proposed for solving complex optimization problems more effectively.

In [24], Yang et al. proposed a chaotic map particle swarm optimization with an acceleration strategy for fuzzy clustering problem, while chaotic map was used to generate sequences to substitute the random parameters r_1 and r_2 of PSO and acceleration strategy was to use one-third of the particles to accelerate the convergence rate of the particles. In this paper, chaotic map is focused and applied in chaotic search for CPSO for fuzzy clustering. Furthermore, in the searching process of fuzzy clustering objective function, gradient method is designed to accelerate the convergence rate. Chaotic map is essential for chaotic search, there are several chaotic maps frequently used for chaotic search [25], such as logistic map, tent map, Chebyshev map. In this paper, the iterative chaotic map with infinite collapses (ICMIC) map is considered in chaotic search, and a new chaotic particle swarm optimization is proposed. The new CPSO is applied in fuzzy clustering model to provide good performance in capturing the global optimal fitness, thus getting the best clustering results. Furthermore, in order to increase the efficiency of optimization process, gradient method is introduced to accelerate the convergence. Based on the new CPSO and gradient method, the fuzzy clustering (CPSFC) algorithm is proposed and compared with three other approaches, FCM, fuzzy clustering model with GA, namely the GGAFCM [26] and fuzzy clustering model with PSO, namely the PSOFCM [17].

Paper structure is as follows: Section 2 briefly discusses fuzzy clustering model, FCM algorithm and introduce several fuzzy clustering indices. Section 3 describes PSO roots and the working scheme of CPSO. Section 4 illustrates CPSO applied to fuzzy clustering model, and presents the CPSFC algorithm. Section 5 shows the databases faced, the experiments carried out, the results achieved when compared with the three typical classification techniques. Finally, in Section 6 our conclusions are drawn.

2. Fuzzy *c*-means algorithm and cluster validity indices

The core of a fuzzy clustering model is to determine the measure of similarity, by which the patterns distances can be calculated, and then, it can be estimated how similar two patterns are. Euclidean distance is chosen as the measure of similarity in FCM, so patterns can be treated as vectors in the Euclidean Space. As FCM algorithm and other clustering algorithms have been developed, a number of fuzzy validity indices are designed to measure and help to judge the clustering results. The compactness of a cluster and the separation of clusters are usually taken into consideration.

2.1. Fuzzy *c*-means algorithm

The well-known fuzzy *c*-means model [7] is described as follows, where the weighted within cluster sum of squared error function is used:

$$J_m = \sum_{i}^{c}\sum_{j}^{n}(u_{ij})^m \|\mathbf{y}_j - \mathbf{z}_i\|_A^2 \tag{1}$$

where $\boldsymbol{Y}=(y_1,y_2,\ldots,y_n)$ is the data set with s dimension, $\mathbf{y}_j\in R^s$ for $1\le j\le n$, $\boldsymbol{Z}=(\boldsymbol{z}_1,\boldsymbol{z}_2,\ldots,\boldsymbol{z}_c)$ is the cluster centers or prototypes, $\boldsymbol{z_i}\in\boldsymbol{R}^s$ for $1\le i\le c$, $\boldsymbol{U}=[u_{ij}]_{c\times n}$ is the partition matrix, $u_{ij}\in[0,1]$ is interpreted to be the grade of membership of $\boldsymbol{x}_j$ in the i-th cluster. $\|\cdot\|_A^2$ is an inner product norm induced by matrix $\boldsymbol{A}$ (i.e., $\|\mathbf{y}\|_A^2=\mathbf{y}^T A\mathbf{y}$). If $A=\boldsymbol{I}$, then $||\mathbf{y}_j-\mathbf{z}_i||_A^2$ represents the Euclidean distance from $\boldsymbol{y}_j$ to the ith center $\boldsymbol{z}_i$. It is believed the minimization of J_m will produce the best cluster structure and the optimal cluster results.

It is clearly stated in [27] that the minimization of J_m can be reached by Lagrange multiplier method while the partition matrix $\boldsymbol{U}$ and cluster centers $\boldsymbol{Z}$ have expressions as follows:

$$u_{ij} = \left[\sum_{k=1}^{c}\left(\frac{d_{ij}^2}{d_{kj}^2}\right)^{1/(m-1)}\right]^{-1}, \quad 1\le i\le c,\ 1\le j\le n \tag{2}$$

$$\mathbf{z}_i = \frac{\sum_{j=1}^{n}(u_{ij})^m \mathbf{y}_j}{\sum_{j=1}^{n}(u_{ij})^m}, \quad 1\le i\le c \tag{3}$$

By iterating Eqs. (2) and (3), $\boldsymbol{z}_i$ and u_{ij} will vary towards the direction that minimize J_m gradually, when the change of $\boldsymbol{z}_i$ or u_{ij} is within the given tolerance, stop iteration. The FCM algorithm is described concretely as

(1) Set the cluster numbers c, set initial cluster centers $\mathbf{z}_i^{(0)}$, $1\le i\le c$, and set the tolerance ε to determine when to stop the algorithm.
(2) Refresh u_{ij} and $\boldsymbol{z}_i$ by calculating Eqs. (2) and (3).
(3) Calculate and judge whether $\|\boldsymbol{Z}^{new}-\boldsymbol{Z}^{old}\|<\varepsilon$ or $\|\boldsymbol{U}^{new}-\boldsymbol{U}^{old}\|<\varepsilon$. If the condition is satisfied, stop the algorithm, else go to step (2).

FCM has achieved great success in applications, however the disadvantages of FCM are also obvious. FCM algorithm initializes c cluster centers randomly and seeks the solution by iteration of cluster centers and partition matrix, which is a kind of local search strategy based on gradient method. The gradient method used in FCM will reach the local optimum corresponding to the initial clusters, since the objective function is multimodal. FCM has been proved to be sensitive to initial values, and the different initial cluster centers will lead to different clustering results [9,14]. In the following sections, excellent global search strategy will be discussed to optimize the fuzzy clustering model, avoiding local optima.

2.2. Cluster validity measures

Cluster validity indices are used to evaluate the cluster result quantificationally and test the quality of fuzzy partition. Cluster validity is the study of the relative merits of a partitioned structure in the data set $\boldsymbol{Y}$. Clustering algorithm, no matter hard or fuzzy, essentially generates a partition matrix $\boldsymbol{U}$ and other useful information regarding the cluster structure by identifying prototypes or cluster centers $\boldsymbol{Z}$. Partition $\boldsymbol{U}$ and prototype $\boldsymbol{Z}$ jointly determine the "goodness" of a cluster structure [8].