Expert Systems with Applications 42 (2015) 6315–6328

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Hybrid methods for fuzzy clustering based on fuzzy c-means and improved particle swarm optimization

Telmo M. Silva Filho *, Bruno A. Pimentel, Renata M.C.R. Souza [1], Adriano L.I. Oliveira

Universidade Federal de Pernambuco, Centro de Informática, Av. Jornalista Aníbal Fernandes, s/n, 50.740-560 Recife (PE), Brazil

ARTICLE INFO

Article history:
Available online 27 April 2015

Keywords:
Fuzzy clustering
Fuzzy c-means
Improved particle swarm optimization
Adaptive weights

ABSTRACT

Fuzzy clustering has become an important research field with many applications to real world problems. Among fuzzy clustering methods, fuzzy c-means (FCM) is one of the best known for its simplicity and efficiency, although it shows some weaknesses, particularly its tendency to fall into local minima. To tackle this shortcoming, many optimization-based fuzzy clustering methods have been proposed in the literature. Some of these methods are based solely on a metaheuristic optimization, such as particle swarm optimization (PSO) whereas others are hybrid methods that combine a metaheuristic with a traditional partitional clustering method such as FCM. It is demonstrated in the literature that methods that hybridize PSO and FCM for clustering have an improved accuracy over traditional partitional clustering approaches. On the other hand, PSO-based clustering methods have poor execution time in comparison to partitional clustering techniques. Another problem with PSO-based clustering is that the current PSO algorithms require tuning a range of parameters before they are able to find good solutions. In this paper we introduce two hybrid methods for fuzzy clustering that aim to deal with these shortcomings. The methods, referred to as FCM–IDPSO and FCM2–IDPSO, combine FCM with a recent version of PSO, the IDPSO, which adjusts PSO parameters dynamically during execution, aiming to provide better balance between exploration and exploitation, avoiding falling into local minima quickly and thereby obtaining better solutions. Experiments using two synthetic data sets and eight real-world data sets are reported and discussed. The experiments considered the proposed methods as well as some recent PSO-based fuzzy clustering methods. The results show that the methods introduced in this paper provide comparable or in many cases better solutions than the other methods considered in the comparison and were much faster than the other state of the art PSO-based methods.

1. Introduction

With the growing interest in automatically understanding, processing and summarizing data, many application domains have employed various pattern recognition methods (Xu & Wunsch, 2005). One way of identifying patterns within a dataset is by using clustering analysis. Clustering methods work by assigning objects to a group, if they show a high level of similarity and by assigning objects to different groups, if they show a high level of dissimilarity. These methods are further classified in two categories: hard and fuzzy (Xu & Wunsch, 2005). Hard clustering methods assign each object to a single group, while fuzzy methods introduce membership degrees between objects and the different groups of the dataset (Pal, Pal, Keller, & Bezdek, 2005).

Fuzzy c-means (FCM), proposed by Bezdek, Ehrlich, and Full (1984), is the most popular fuzzy clustering method. In FCM, the goal is to minimize the criterion function, taking into account the similarity of elements and cluster centers. It is more useful for data sets that have highly overlapping groups. Since FCM is easily implemented and has obtained satisfactory results in many applications, it has become an important tool for pattern recognition (Jain, Murty, & Flynn, 1999). However, FCM has some shortcomings that have motivated the proposal of alternative approaches for fuzzy clustering, many of which are extensions of FMC. For instance, Zhang, Pedrycz, Lu, Liu, and Zhang (2014) proposed an FCM which uses a genetic heuristic strategy to search for interval weights for the data attributes, to model their different importance for the clustering performance. In another effort to improve clustering quality of FCM, Sabzekar and Naghibzadeh (2013) employed relaxed constraints support vector machines to solve the problem of multiple objects being assigned to clusters with low membership values. In the area of image segmentation, Zhao, Fan, and

* Corresponding author.
E-mail addresses: tmsf@cin.ufpe.br (T.M. Silva Filho), bap@cin.ufpe.br (B.A. Pimentel), rmcrs@cin.ufpe.br (R.M.C.R. Souza), alio@cin.ufpe.br (A.L.I. Oliveira).
[1] Main corresponding author.

http://dx.doi.org/10.1016/j.eswa.2015.04.032
0957-4174/

Liu (2014) used an optimal-selection-based suppressed FCM algorithm with self-tuning non-local spatial information to improve segmentation performance on images with high noise disturbance. In order to handle the memberships based on the inherent information in each feature Pimentel and De Souza (2013) introduced multivariate memberships which are different from one variable to another and from one cluster to another. To deal with more complex data types, such as interval data, Pimentel and Souza (2014) proposed a multivariate FCM method with relevance weights for each variable that are different from one cluster to another.

Although these FCM versions aim to achieve good performance in fuzzy clustering, they have disadvantages. Some of them are (i) the initialization with randomly generated cluster centers (Bezdek et al., 1984) and (ii) the high change of getting trapped to local minima.

Metaheuristic optimization algorithms such as genetic algorithm (GA), simulated annealing (SA), ant colony optimization (ACO) and particle swarm optimization (PSO) have been successfully employed in many applications (Izakian & Abraham, 2011). They aim to solve optimization problems without being trapped into local minima. PSO has become one of the most popular metaheuristics and an important tool for many applications due to its versatility and simplicity (Alam, Dobbie, Koh, Riddle, & Rehman, 2014).

This has motivated the proposal of many PSO-based methods for hard clustering (Alam et al., 2014) and some PSO-based methods for fuzzy clustering Li, Zhou, Kou, and Xiao (2012) and Chen, Xu, and Tang (2014). Clustering methods may use PSO in stand alone form or in combination with FCM (Alam et al., 2014). Many of the PSO-based hard or fuzzy clustering methods were shown to have improved accuracy compared to traditional partitional clustering approaches such as K-means, K-harmonic and FCM (Alam et al., 2014; Li et al., 2012; Chen et al., 2014). However, these PSO-based methods are much slower compared to the traditional methods which may limit their practical applications. Another problem with PSO-based clustering methods, according to Alam et al. (2014), is the need to tune a range of parameters before they are able to find a better solution. Many of the PSO versions have three parameters which may influence performance and thus may have to be tuned.

Many changes to the original PSO have been proposed, focusing on improving its results and/or convergence time. Zhang, Xiong, and Zhang (2013) proposed a version of the PSO algorithm which automatically adjusts its parameters to try and achieve a better performance called improved self-adaptive particle swarm optimization (IDPSO). Other researches apply PSO to several kinds of applications. Pang, Wang, Zhou, and Dong (2004) developed a PSO version applied to the traveling salesman problem, called fuzzy discrete particle swarm optimization (FPSO). Since FCM may be interpreted as an optimization problem, it could be integrated with PSO. Concerning this combination, Izakian and Abraham (2011) proposed a hybrid fuzzy clustering algorithm based on FCM and FPSO, called FCM–PSO. Their experiments showed better results than FPSO and FCM.

This work introduces two new methods for the fuzzy clustering problem. Both methods combine IDPSO with FCM. The reason for choosing IDPSO is that this variant of PSO has outperformed other variants of PSO in function optimization and has two advantages over other versions of PSO, namely, (i) the more effective exploration of the search space, leading to the global optimum and effectively avoiding premature convergence, and (ii) the dynamic adjustment of parameters during training (Zhang et al., 2013). Thus, combined to FCM, IDPSO may lead to clustering methods that tackle the two main problems of PSO-based clustering methods mentioned by Alam et al. (2014), namely, the low speed and the need to tune a range of parameters. The first method proposed in this work is a hybrid between FCM and IDPSO along the lines of the FCM–PSO proposed by Izakian and Abraham (2011), called FCM–IDPSO. The second method is called FCM2–IDPSO and uses FCM to generate one of the initial solutions in the population of FCM–IDPSO. This is done to reduce the randomness of the initial solutions of FCM–IDPSO. Experiments using both synthetic and real-world data sets are reported and the results are compared to recent PSO-based fuzzy clustering methods (Chen et al., 2014; Izakian & Abraham, 2011; Li et al., 2012).

The remainder of this work is structured as follows: Section 2 reviews some related works and presents the contributions of this paper. Section 3 presents the existing algorithms on which this work is based and Section 4 introduces the proposed methods; synthetic and real datasets were used in experiments to compare the methods according to four metrics and their results are shown in Section 5; finally, Section 6 presents the final remarks as well as suggestions for further research.

2. Related works and contributions

There are many research directions aiming at extending or improving the FCM algorithm; some of them were discussed in Section 1 (Pimentel & De Souza, 2013; Pimentel & Souza, 2014; Sabzekar & Naghibzadeh, 2013; Zhang et al., 2014; Zhao et al., 2014). Some of the methods extend FCM by combining it with metaheuristic optimization algorithms, such as GA and PSO aiming to avoid being trapped in local minima (Alam et al., 2014). This section discusses the most relevant papers related to the methods introduced in this paper as well as the contributions of this paper. A more detailed review on nature inspired metaheuristic algorithms for partitional clustering is provided in an up-to-date paper by Nanda and Panda (2014). PSO-based clustering methods are extensively reviewed on a recent paper by Alam et al. (2014). There are many PSO-based hard clustering methods, yet there are fewer proposals for fuzzy clustering with PSO.

Concerning fuzzy clustering and metaheuristic algorithms, many works have been made by several authors. Niu and Huang (2011) proposed a fuzzy c-means clustering algorithm based on an enhanced particle swarm optimization which avoids premature convergence. Szabo, de Castro, and Delgado (2011) presented an extension of the crisp data clustering algorithm Particle Swarm Clustering (PSC) particularly tailored to deal with fuzzy clusters. Ma, Niu, Zhao, and Ma (2011) introduced a new fuzzy clustering algorithm based on Adaptive Particle Swarm Optimization Algorithm (APSO) in order to overcome the disadvantages of the FCM algorithm such as the sensitivity to initial values and the easy involvement in partial optimum, and enhance the abilities of APSO algorithm such as global search and escape from partial optimum.

A novel chaotic particle swarm fuzzy clustering (CPSFC) algorithm based on chaotic particle swarm (CPSO) and the gradient method was proposed by Li et al. (2012). An adaptive inertia weight factor (AIWF) and iterative chaotic map with infinite collapses (ICMIC) were introduced, and a new CPSO algorithm combined AIWF and ICMIC based chaotic local search was studied. Chen et al. (2014) proposed a hybrid clustering algorithm based on two improved versions of PSO (HPSOFCM), which combines the merits of both algorithms and showed that the proposed method is able to escape local optima.

These methods have made efforts to improve the quality of the clustering, but they do not consider that PSO parameters have a significant impact on performance of the algorithm. These parameters are related to the behavior of the convergence and how fast is