Applied Soft Computing 12 (2012) 3442–3451

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

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

A new ECG beat clustering method based on kernelized fuzzy c-means and hybrid ant colony optimization for continuous domains

Berat Doğan*, Mehmet Korürek

Istanbul Technical University, Department of Electronics and Communication Engineering, Turkey

ARTICLE INFO

Article history:
Received 3 December 2011
Received in revised form 17 April 2012
Accepted 5 July 2012
Available online 20 July 2012

Keywords:
Fuzzy c-means
Kernelized fuzzy c-means
ECG classification
Ant colony optimization
Arrhythmia classification
Swarm
ECG clustering

ABSTRACT

The kernelized fuzzy c-means algorithm uses kernel methods to improve the clustering performance of the well known fuzzy c-means algorithm by mapping a given dataset into a higher dimensional space non-linearly. Thus, the newly obtained dataset is more likely to be linearly seprable. However, to further improve the clustering performance, an optimization method is required to overcome the drawbacks of the traditional algorithms such as, sensitivity to initialization, trapping into local minima and lack of prior knowledge for optimum paramaters of the kernel functions. In this paper, to overcome these drawbacks, a new clustering method based on kernelized fuzzy c-means algorithm and a recently proposed ant based optimization algorithm, hybrid ant colony optimization for continuous domains, is proposed. The proposed method is applied to a dataset which is obtained from MIT–BIH arrhythmia database. The dataset consists of six types of ECG beats including, Normal Beat (N), Premature Ventricular Contraction (PVC), Fusion of Ventricular and Normal Beat (F), Artrial Premature Beat (A), Right Bundle Branch Block Beat (R) and Fusion of Paced and Normal Beat (f). Four time domain features are extracted for each beat type and training and test sets are formed. After several experiments it is observed that the proposed method outperforms the traditional fuzzy c-means and kernelized fuzzy c-means algorithms.

1. Introduction

Electrical activity of the heart is measured by a noninvasive method called the electrocardiogram (ECG). ECG is a very important tool for the diagnosis of the heart beat abnormalities. If there exists an abnormality in usual behavior of the heart, the measured ECG signal deviates from its normal shape which helps the physician to diagnose this abnormality. However, it is really difficult to diagnose an abnormality in a long term ECG record which consists of thousands of ECG beats. For this reason, supplementary tools for automatically diagnosis of the heart beat abnormalities are proposed. These tools use both the signal processing and pattern recognition methods.

Pattern recognition methods can be categorized into two groups according to their learning procedure. Supervised learning, requires prior labeling of the training data to create a model of the given dataset. A supervised learning algorithm analyses the given training dataset and creates an output. This output is then compared with the desired output (label) and an error or feedback signal is created. Algorithm then updates itself according to this feedback signal in order to create a model of the given dataset. Once the algorithm is terminated the obtained model should generalize the training data such that when an unknown input pattern is given to the model it should be classified correctly. However, unsupervised learning does not need a prior labeling. It creates clusters from a given dataset according to a similarity measure which is usually a distance function. After the clustering process, similar patterns are grouped in the same cluster and dissimilar patterns are grouped in different clusters.

Probably the most popular clustering algorithm is the hard c-means (or k-means) algorithm which assigns a data point in a given dataset to exactly one cluster. Such an assignment can be inadequate because some data points can be in a location which is almost equally distant from two or more cluster centers. By forcing such a point to exactly one cluster, the similarity of this point to other clusters is totally ignored. For this reason, fuzzy clustering methods are proposed. In fuzzy clustering methods a data point can belong to more than one cluster with different degrees of membership which is useful especially when the clusters overlap each other. Actually, this is the case for most of the real world problems, as in the ECG beat clustering problem. In real world problems, there is often no sharp boundary between clusters so that, the fuzzy clustering methods are usually better suited for these kind of problems when compared to the crisp methods.

Generally speaking, in ECG classification algorithms, a similarity measure is used to measure the distances between the query

* Corresponding author.
E-mail address: bdogan@itu.edu.tr (B. Doğan).

1568-4946/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2012.07.007

beat and the templates in the database. The smaller the distance is, the more similar the template to the query [1]. However, as stated above, in many cases a data point can be in a location which is almost equally distant to more than one cluster center. In such a situation, fuzzy clustering methods could prevent the misclassification of a query beat by utilizing the membership values of the datapoints to each cluster.

Based on the above considerations, recently a number of studies which use fuzzy clustering algorithms for ECG beat classification are proposed. One of these fuzzy clustering algorithms is the fuzzy c-means (FCM) algorithm [2]. After a clustering process, the FCM algorithm gives two outputs namely, the cluster centers and a fuzzy partition matrix which contains the membership values of each datapoint to these clusters. The obtained cluster centers or the membership values are then utilized for classification. Özbay et al. [3] utilizes the membership values obtained by using FCM algorithm to train a multilayer perceptron (MLP) for classification of ten types of ECG beats. In more recent studies, Ceylan et al. [4] and Özbay [5] use type-2 fuzzy c-means algorithm in a similar manner. Another study which utilizes membership values for ECG beat classification is proposed by Yeh et al. [6]. In this study, four morphological features out of nine are selected with a method called Range-Overlap Method (ROM) and then clustered by fuzzy c-means algorithm. Different from the above studies, Haseena et al. [7] use cluster centers as input to a neural network for classification of eight types of ECG beats.

1.1. Need for an optimization algorithm

Fuzzy c-means algorithm is used in several clustering problems efficiently. However, it has two major drawbacks which reduce its performance. Firstly, it is sensitive to initial values of the clusters and secondly, it can be easily trapped into local minima. Therefore, several extensions of the FCM algorithm are proposed to improve its performance. One of these algorithms is called the kernelized fuzzy c-means algorithm (KFCM) which uses kernel methods to improve the clustering performance of the fuzzy c-means algorithm. In KFCM a given dataset is mapped to a higher dimensional space non-linearly by a kernel function. Thus, the newly obtained dataset is more likely to be linearly separable. However, the KFCM algorithm also has the above-mentioned drawbacks. Additionally, there is no prior knowledge about the optimum parameters of the kernel functions.

Effective clustering generally involves to optimize clustering algorithms during the clustering process, such that the optimization should be a part of overall design. Ant-based and swarm-based techniques are proven to be powerful ways of optimization and recently used in several engineering problems efficiently. Therefore, in this study an ant-based algorithm, hybrid ant colony optimization for continuous domains (HACO), is used to overcome above-mentioned drawbacks by optimizing the objective function of the KFCM algorithm.

Hybrid ant colony optimization algorithm is recently proposed by Xiao and Li [8]. It is an extension of the ant colony optimization for continuous domains (ACO_R) algorithm, which is proposed by Dorigo and Di Caro [9]. Here, in this study HACO is used to find optimum cluster centers and the kernel function parameter, which in turn minimizes the objective function of the KFCM algorithm. The proposed method is then compared to the traditional fuzzy c-means and kernelized fuzzy c-means algorithms.

The rest of this paper is organized as follows: the next section comprehensively reviews the literature on ant-based and swarm-based clustering methods and its usage on ECG beat classification problem. In Section 3, existing algorithms which provide a theoretical perspective for the proposed improved method is presented and then the proposed method is given. In Section 4, the proposed ECG beat classification system and its building blocks are explained as a case study. Section 5 covers the experimental results and finally, Section 6 concludes the work.

2. Related works

Over the last decade, clustering with ant-based and swarm-based algorithms is proposed as an alternative to traditional clustering methods. Here, the term “ant-based clustering” is used for algorithms that are based on other types of natural “swarm” behavior and for which the more general term “swarm-based” is more appropriate. Ant based clustering stands out as the most widely used group of swarm-based clustering algorithms. Broadly speaking, there are two main types of ant-based clustering: the first group of methods directly mimics the clustering behavior observed in real ant colonies. The second group is less directly inspired by nature: the clustering task is reformulated as an optimization task and general purpose ant-based optimization heuristics are utilized to find good or near optimal clustering [10]. In [11] a detailed survey which includes recent studies in this research are can be found.

Depending upon the continuous development of the swarm-based techniques, their use for medical applications become a vast and promising area in the last few years. Several swarm-based tools are proposed for this purpose, including those deal with the classification of ECG signals. In [12] a method which based on particle swarm optimization (PSO) and support vector machine (SVM) is proposed for the diagnosis of the arrhythmia cordis. It is shown that the proposed method is superior than the back-propagation neural network (BP-NN) and radial basis function neural network (RBF-NN). Korürek and Doğan [13] evolved a RBF-NN with PSO to classify six types of ECG beats. Experimental results show that the proposed method classifies the ECG beats with a smaller size of network without making any concessions on classification performance. In another study, Doğan and Korürek [14] used ant colony optimization for continuous domains (ACO_R) for the same purpose. Melgani and Bazi [15] evolved the SVM with PSO to classify six types of ECG beats. They compared the proposed method with the k-nearest neighbor (k-NN) classifier and RBF-NN. It is shown that the proposed method outperforms the traditional classifiers. Ince et al. [16] designed a generic and patient-specific classification system for robust and accurate detection of ECG heart beat patterns. For this purpose, they used an artificial neural network classifier which is optimally designed with the multidimensional particle swarm optimization. In [17,18] six types of ECG beats are clustered with a method which is based on ant colony optimization (ACO). In both of these studies authors used the well-known ACO algorithm which is proposed by Dorigo and Di Caro [19] for combinatorial optimization problems.

Due to the successful results obtained with swarm-based methods in many engineering problems, their usage for fuzzy clustering algorithms also gets popular nowadays. However, up to our knowledge a study about ECG beat clustering problem is not reported before. Some examples of different applications are as follows. In [20] features such as energy, standard deviation, autocorrelation, mean, variance and normalized values are extracted from the non-stationary power signals and then these features are given as input to a fuzzy c-means clustering algorithm to generate a decision tree for power signal disturbance pattern classification. By using a hybrid ant colony optimization technique the cluster centers are updated and classification performance is improved. Obtained results also compared with different population based optimization algorithms like the genetic algorithm and particle swarm optimization. Note that, the algorithm used in this study is different from the algorithm which is used in this paper. Niknam et al. [32] proposed a hybrid evolutionary algorithm which is based on fuzzy adaptive