Information Sciences 304 (2015) 80–91

Contents lists available at ScienceDirect

Information Sciences

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

An unsupervised learning algorithm for membrane computing

Hong Peng [a,*], Jun Wang [b], Mario J. Pérez-Jiménez [c], Agustín Riscos-Núñez [c]

[a] Center for Radio Administration and Technology Development, Xihua University, Chengdu, Sichuan 610039, China
[b] School of Electrical and Information Engineering, Xihua University, Chengdu, Sichuan 610039, China
[c] Research Group of Natural Computing, Department of Computer Science and Artificial Intelligence, University of Seville, Sevilla 41012, Spain

ARTICLE INFO

Article history:
Received 29 December 2013
Received in revised form 10 January 2015
Accepted 27 January 2015
Available online 3 February 2015

Keywords:
Membrane computing
P system
Evolution–communication P system
Unsupervised learning
Data clustering
Fuzzy clustering

ABSTRACT

This paper focuses on the unsupervised learning problem within membrane computing, and proposes an innovative solution inspired by membrane computing techniques, the fuzzy membrane clustering algorithm. An evolution–communication P system with nested membrane structure is the core component of the algorithm. The feasible cluster centers are represented by means of objects, and three types of membranes are considered: evolution, local store, and global store. Based on the designed membrane structure and the inherent communication mechanism, a modified differential evolution mechanism is developed to evolve the objects in the system. Under the control of the evolution–communication mechanism of the P system, the proposed fuzzy clustering algorithm achieves good fuzzy partitioning for a data set. The proposed fuzzy clustering algorithm is compared to three recently-developed and two classical clustering algorithms for five artificial and five real-life data sets.

1. Introduction

The learning problem aims at finding hidden patterns in data or estimating the unknown model parameters from the given data. From the viewpoint of machine learning, learning methods can be categorized into three main variants: supervised learning, unsupervised learning and reinforcement learning. Unsupervised learning, also known as clustering, is the process of finding natural groupings from unlabeled data [8], that is, finding k clusters from a set of n data points according to some similarity measure such that patterns within the same cluster are more similar than those from different clusters. Over the past years, a large number of unsupervised learning methods have been introduced [13,47]. These methods can be fall into three categories: hierarchical, partitioned, and overlapping methods. Hierarchical methods can be either agglomerative, which begin with each element as a separate cluster and merge them into larger clusters, or divisive, which begin with the whole set and successively divide it into smaller clusters. Partitioned methods attempt to directly decompose the data set into several disjoint clusters without the hierarchical structure, while overlapping methods search soft or fuzzy partitioning by relaxing the mutually disjoint constraint.

K-means is one of most popular unsupervised learning methods due to its simplicity and effectiveness, and has been used in a wide variety of areas such as pattern recognition, data mining, and bioinformatics. However, *k*-means has several drawbacks: it is problematic to avoid local minima, it is sensitive to initial cluster centers, and it takes significant time to search the global optimal solution when the number of data points is large. To overcome these, evolutionary clustering algorithms have been considered in recent years based on genetic algorithms (GA), particle swarm optimization (PSO), ant colony optimization (ACO), differential evolution (DE), artificial bee colony (ABC), and black hole (BH) algorithm.

* Corresponding author.
E-mail address: ph.xhu@hotmail.com (H. Peng).

http://dx.doi.org/10.1016/j.ins.2015.01.019

GA-based clustering methods were first reported in the literature [5,17], and have two approaches to express the solution of a clustering problem: point-based schemes [18,23] and center-based schemes [2]. Point-based schemes can suffer from large search space and high computing cost when data points proliferate, so center-based schemes are used by most evolutionary clustering methods developed in recent years. Kao et al. [14] proposed a clustering method that uses PSO to optimize a set of cluster centers. Shelokar et al. [40] introduced an ACO-based method to find appropriate cluster centers. Das et al. [6] used DE in image pixel clustering, while Karaboga et al. [15] presented an ABC-based clustering method. Hatamlou [11] proposed a new optimization approach based on black hole for data clustering. In addition, several evolutionary clustering algorithms with hybrid mechanisms have been developed in recent years [16,24,25]. Fuzzy c-means (FCM) is a well-known overlapping clustering technique that uses the principles of fuzzy sets to evolve a partitioning matrix $U(X)$ [3]. However, FCM still has the limitation of finding suboptimal solutions. To overcome this, Maulik et al. [19] proposed a GA-based fuzzy clustering technique, Fuzzy-VGA, which automatically evolves the appropriate fuzzy partitioning for a data set by optimizing the well-known XB-index. Saha et al. [36] developed a fuzzy variable string length genetic point symmetry (Fuzzy-VGAPS) based clustering technique, where membership values of points to different clusters are computed based on a point symmetry-based distance rather than the Euclidean distance. Maulik et al. [20] presented a modified DE based fuzzy clustering (MoDEFC) algorithm and applied it to deal with pixel classification of remote sensing imagery. Sanchez et al. [39] proposed a fuzzy granular gravitational clustering algorithm for multivariate data. Saha et al. [37] developed a multiobjective modified differential evolution based fuzzy clustering algorithm.

Membrane computing was initiated by Gh. Păun [28], as a new branch of natural computing, aiming to abstract computing models from the structure and functioning of living cells, as well as from the cooperation of cells in tissues, organs, and cell populations [10,32]. Following this inspiration, classes of distributed parallel computing models have been defined, usually known as P systems or membrane systems. P systems have several interesting features: non-determinism, programmability, extensibility, readability, they are easy to communicate, etc., and many variants have been proposed [9,12,27,33,45,48,51]. Most P systems variants have proved to be powerful (in the sense of Turing completeness) and effective (since they have successfully solved a large number of NP-hard problems in a linear or polynomial time [31]). In recent years, the potentiality and characteristics of membrane computing have attracted much attention in relation to real-life applications, such as membrane algorithms for solving optimization problems [26,34,49,50,52], and fuzzy spiking neural P systems for dealing with knowledge representation and fault diagnosis [35,42,44].

Learning capability is an important and useful characteristic of natural computing methods, such as neural computing and evolutionary computing [21,22,38]. Unfortunately, existing P systems variants generally lack learning capability, due to their orientation towards a formal language framework. Therefore, Gh. Păun has listed providing learning capability to P systems as an interesting open problem [29,30].

The main motivation behind the work presented in this paper is to focus on the problem of unsupervised learning in membrane computing, and to propose a novel clustering method in this framework to solve fuzzy clustering problems: the fuzzy membrane clustering algorithm. An evolution–communication P system with a nested membrane structure is considered as the computing framework, and both its evolution–communication mechanism and the differential evolution mechanism of DE are integrated into a new P system based learning mechanism. The scientific contribution of this paper has two aspects: (i) this is the first attempt to use P systems for solving fuzzy clustering problems and (ii) a novel clustering method for data mining and/or machine learning is developed.

Section 2 describes the fuzzy clustering problem and several clustering validity indexes. In Section 3, we briefly review evolution–communication P systems and classical DE algorithms. Section 4 describes the proposed evolution–communication P system, and details the proposed fuzzy membrane clustering algorithm. Experiments and results are provided in Section 5. Finally, Section 6 draws the conclusions.

2. Problem statement

2.1. Fuzzy clustering problem

Data clustering can be seen as the task of distributing (partitioning) n data points into several groups according to some similarity measure. Let us consider $X = \{X_1, X_2, \ldots, X_n\} \subseteq R^{n\times d}$, a set of n unlabeled data points in a d-dimensional Euclidean space, where $X_i = (x_{i1}, x_{i2}, \ldots, x_{id})$, that is, x_{ij} is the j-th real-value feature of the i-th data point. The goal of a fuzzy clustering algorithm is to find a fuzzy partition of the data set, X, into k clusters, $C_1, C_2, \ldots, C_K$, in such a way that the similarity of data points in the same cluster is very high, and at the same time the similarity of data points in different clusters is very low. In the context of fuzzy clustering, a data point may belong to each of the clusters with a certain fuzzy membership degree, which can be captured by a fuzzy partitioning matrix, $U_{K\times n}$, where each u_{ij} is the fuzzy membership degree of X_j to C_i. The elements of such matrix should satisfy the following properties:

$$\begin{cases} 0 < \sum_{j=1}^{n} u_{ij} < n & \text{for } i = 1, 2, \ldots, K \\ \sum_{i=1}^{K} u_{ij} = 1 & \text{for } j = 1, 2, \ldots, n \\ \sum_{i=1}^{K}\sum_{j=1}^{n} u_{ij} = n & \end{cases} \quad (1)$$