

一种带混合进化机制的膜聚类算法*

彭宏¹, 蒋洋¹, 王军², Mario J. PÉREZ-JIMÉNEZ³

¹(西华大学 无线电管理技术研究中心, 四川 成都 610039)

²(西华大学 电气信息学院, 四川 成都 610039)

³(Research Group of Natural Computing, University of Seville, Sevilla, Spain)

通讯作者: 彭宏, E-mail: ph.xhu@hotmail.com, http://www.xhu.edu.cn

摘要: 膜计算(也称为 P 系统或膜系统)是一种新颖的分布式、并行计算模型。为了处理数据聚类问题,提出了一种采用混合进化机制的膜聚类算法。它使用了一个由 3 个细胞组成的组织 P 系统,为一个待聚类的数据集发现最优的簇中心。其对象表示候选的簇中心,并且这 3 个细胞分别使用了 3 种不同的进化机制:遗传算子、速度-位移模型和差分进化机制。然而,所使用的速度-位移模型和差分进化机制是结合了这个特殊膜结构和转运机制所提出的改进版本。这种混合进化机制能够增强系统中对象的多样性和改善收敛性能。在混合进化机制和转运机制控制下,这种膜聚类算法能够确定一个数据集的良好划分。所提出的膜聚类算法在 3 个人工数据集和 5 个真实数据集上被评估,并与 *k*-means 和几种进化聚类算法进行比较。统计显著性测试建立了所提出的膜聚类算法的优势。

关键词: 膜计算;P 系统;组织 P 系统;数据聚类;膜聚类算法;混合进化机制

中图法分类号: TP18

中文引用格式: 彭宏,蒋洋,王军,Pérez-Jiménez MJ. 一种带混合进化机制的膜聚类算法. 软件学报, 2015, 26(5): 1001-1012. <http://www.jos.org.cn/1000-9825/4657.htm>

英文引用格式: Peng H, Jiang Y, Wang J, Pérez-Jiménez MJ. Membrane clustering algorithm with hybrid evolutionary mechanisms. Ruan Jian Xue Bao/Journal of Software, 2015, 26(5): 1001-1012 (in Chinese). <http://www.jos.org.cn/1000-9825/4657.htm>

Membrane Clustering Algorithm with Hybrid Evolutionary Mechanisms

PENG Hong¹, JIANG Yang¹, WANG Jun², Mario J. PÉREZ-JIMÉNEZ³

¹(Center for Radio Administration and Technology Development, Xihua University, Chengdu 610039, China)

²(School of Electrical and Information Engineering, Xihua University, Chengdu 610039, China)

³(Research Group of Natural Computing, University of Seville, Sevilla, Spain)

Abstract: Membrane computing, known as P systems or membrane systems, is a novel class of distributed and parallel computing models. This paper proposes a membrane clustering algorithm using hybrid evolutionary mechanisms to address data clustering problem. It uses a tissue P system consisting of three cells to find the optimal cluster centers for a data set to be clustered. Its object is used to express candidate cluster centers, and the three cells use three different evolutionary mechanisms: genetic operators, velocity-position model and differential evolution mechanism. Particularly, the velocity-position model and differential evolution mechanism used in the process are the improved versions proposed in this paper according to the special membrane structure and communication mechanism. The hybrid evolutionary mechanisms can enhance the diversity of objects in the system and improve the convergence performance. Under the control of the hybrid evolutionary mechanisms and communication mechanism, the membrane clustering algorithm can determine a good partition for a data set. The proposed membrane clustering algorithm is evaluated on three artificial data sets and five real-life data

* 基金项目: 国家自然科学基金(61170030); 教育部春晖计划(Z2012031); 四川省科技支撑计划(2013GZX0155)

收稿时间: 2013-10-28; 修改时间: 2014-02-27; 定稿时间: 2014-05-21; jos 在线出版时间: 2014-08-19

CNKI 网络优先出版: 2014-08-19 14:33, <http://www.cnki.net/kcms/doi/10.13328/j.cnki.jos.004657.html>

sets and compared with k -means and several evolutionary clustering algorithms. Statistical significance tests have been performed to establish the superiority of the proposed membrane clustering algorithm.

Key words: membrane computing; P system; tissue P system; data clustering; membrane clustering algorithm; hybrid evolutionary mechanism

聚类分析是数据挖掘中的一个核心问题.在很多场合中所获取的数据仅有输入而没有对应的输出,所以这些数据是未标记的或无监督的.数据聚类是用于处理这种数据的无监督学习过程,它依据某个相似性度量将对象进行分组,使得同组中的样本是相似的但不同组的样本是不相似的^[1].聚类方法能够大体上分为 3 类:层次方法、划分方法和重叠方法^[2,3].层次聚类通过连续合并较小的簇为较大的簇或者分裂较大的簇来实现数据聚类,层次聚类包括凝聚和分裂两种子类型;划分方法依据某个准则将数据集分解为几个不相交的簇;而重叠方法则以某种方式放宽相互不重叠的约束实现软的或模糊划分.数据聚类已成功地应用于诸多领域,如数据挖掘、机器学习、图像处理和生物学等^[4-7].

对于划分聚类,普遍采用的准则是最小化每个簇中样本的不相似性和最大化不同簇间相异性. K -means 是最流行的划分聚类算法之一,它之所以受到广泛关注,是由于其算法的简单性和伸缩性以及对于问题中任何变量具有线性渐进时间^[8,9].然而, k -means 对初始簇中心的选取非常敏感,以至于当未正确选取初始簇中心时,它可能会陷入局部最优解.近年来,进化算法的全局搜索能力已被采用,以克服 k -means 的缺陷.首先,几种基于遗传算法(GA)的聚类算法相继被开发,它们采用了不同形式表达聚类解.第 1 种算法直接采用染色体编码每个数据点所属的簇号^[10,11],然而当数据点激增时,这种算法会遭遇巨大的搜索空间和高的搜索代价.为了改善搜索性能,第 2 种算法使用一种间接表示法,即,染色体编码一个数据集的全部簇中心^[12-15],间接表示法随后被多数进化聚类算法所采用.此外,也开发了混合 GA 和 k -means 的方法,其中,GA 用于聚类的特征选取^[16].然而,GA 可能遭遇退化是这类聚类算法的一个严重缺陷.为此,几个基于粒子群优化(PSO)、蚁群优化(ACO)的聚类算法已相继被提出.Kao 等人提出了一种组合 PSO 和 k -means 的聚类方法^[17],而 Shelokar 等人引入 ACO 去处理聚类问题^[18].另外,Niknam 等人提出了一种组合 PSO,ACO 和 k -means 的聚类方法^[19].

膜计算(membrane computing)是一类新颖的分布式、并行计算模型,由 Gheorghe Păun 于 2000 年正式提出^[20].作为自然计算的一个新分支,它旨在从生命细胞的结构与功能中以及从组织和器官等细胞群的协作中抽象出计算模型,这类计算模型也常称为 P 系统(P systems).P 系统包含 3 个核心要素:膜结构、对象和进化规则^[21].P 系统通过细胞(膜)中对象的进化实现信息的计算和处理.此外,源自于细胞生物学的不同机制已融入到 P 系统中,提出了众多的变体^[22-28].这些多种多样的机制为处理实际问题提供了新颖的思路.

为了弥补上述缺陷,本文引入膜计算处理数据聚类问题,提出了一种在膜计算框架下的聚类算法,称为膜聚类算法(membrane clustering algorithm),简称 MC 算法.为此开发了一个组织 P 系统.它由 3 个细胞构成,每个细胞包含若干对象,并且每个对象表达一组候选的簇中心.分别引入 3 种不同的进化机制以实现对象进化,即遗传算子^[29]、速度-位移模型^[30]和差分进化机制^[31].而且,通过融入细胞膜间的转运机制提出了原始的速度-位移模型和差分进化机制的改进版本,其中,两种不同源的最好对象在这些改进进化机制中将指导对象进化.在这些进化机制和转运机制的控制下,这个组织 P 系统能够充分地开采整个搜索空间并发现出最优簇中心.

近年来,膜计算由于并行计算、非确定性、可编程性等特征已经受到越来越多的关注.膜计算的研究主要涵盖 3 个方面:(1) 建模与理论分析,即,各种 P 系统的建立及其计算能力和有效性分析;(2) 面向实际系统的 P 系统构建与应用;(3) P 系统仿真平台的开发.已有理论成果揭示了 P 系统是强大的(在大多数情况下具有图灵完备性)和高效的(特别是其能以线性或多项式时间求解众多的 NP 完备问题);同时,多个 P 系统软、硬件仿真平台相继开发成功^[24].近年来,膜计算的应用研究逐步受到重视,学者们陆续采用膜计算求解一些实际问题,诸如模糊推理^[25,27]、故障诊断^[26]、图像处理^[28]、优化问题^[32]、生态系统仿真^[33]和合成生物学^[34]等.不同于已有的工作,本文的动机是采用膜计算求解数据聚类问题,并且成功开发出一种在膜计算框架下的聚类算法.

1 数据聚类问题

数据聚类是诸如数据挖掘、机器学习和模式识别的基础工具之一.众所周知,通过最小化某个相异性度量在非均匀数据中发现数据分组是一个 NP 难问题.一个 d 维欧式空间中的数据聚类可被描述为这样一个过程,它根据某种相似性(距离)度量将一个含 n 个模式或数据点的数据集划分到几个组(或簇)中.记 $X=\{X_1, X_2, \dots, X_n\}$ 是一个由 n 个未标记数据点组成的数据集,其中, $X_i=(x_{i1}, x_{i2}, \dots, x_{id})$ 是一个 d 维向量, $i=1, 2, \dots, n$, 且 x_{ij} 表示 X_i 的第 j 个实值特征.对于数据集 X ,划分聚类算法尝试发现 K 个簇的一个划分 $\{C_1, C_2, \dots, C_K\}$,使得相同簇中的模式有最大的相似性,而不同簇中的模式有尽可能高的相异性.这个划分应满足如下的性质:(1) $C_i \neq \emptyset$, 对任意 $i \in \{1, 2, \dots, K\}$; (2) $C_i \cap C_j = \emptyset$, 对任意 $i \neq j$ 且 $i, j \in \{1, 2, \dots, K\}$; (3) $\bigcup_{i=1}^K C_i = S$.

在评估两个模式的相似性时,多数方法使用了欧式距离.任何两个 d 维模式 X_i 和 X_j 的欧式距离由下式给出:

$$d(X_i, X_j) = \sqrt{\sum_{p=1}^d (x_{ip} - x_{jp})^2} = \|X_i - X_j\| \tag{1}$$

为了发现 K 个簇的最优划分,聚类问题常定义为一个性能函数的最优(最小)问题.对于这 K 个簇的优良性度量,一种较流行的性能函数是整体簇内方差或整体均方差(MSE),即

$$J_m(C_1, C_2, \dots, C_K) = \sum_{i=1}^K \sum_{X_j \in C_i} \|X_j - z_i\| \tag{2}$$

其中, z_1, z_2, \dots, z_K 分别是划分 C_1, C_2, \dots, C_K 的簇中心,且 $z_i \in R^d, i=1, 2, \dots, K$.

2 带混合进化机制的膜聚类算法

组织 P 系统是膜计算模型的主要类型之一,它抽象于在一个公共环境中进化的多个单膜细胞行为,其结构为网状结构,其中每个细胞视为一个信息处理器,它处理对象并且沿细胞之间预先指定的通道转运这些对象^[22].对象的处理由进化规则完成,而转运通过转运规则所实现.所探讨的膜聚类算法使用一个度为 3 的组织 P 系统,这里,我们仅讨论这个特殊设计的组织 P 系统,而有关更一般的组织 P 系统的知识请参阅文献[21,22].

2.1 一个度为3的组织P系统

我们所考虑的度为 3 的组织 P 系统(由 3 个细胞组成)形式地定义为如下的一个元组:

$$IF=(O_1, O_2, O_3, R_1, R_2, R_3, R', i_o).$$

其中,(1) O_1, O_2 和 O_3 分别为这 3 个细胞的对象集;(2) R_1, R_2 和 R_3 分别为这 3 个细胞的进化规则集;(3) R' 为细胞膜之间转运规则集;(4) $i_o=0$, 指示环境为系统的输出区域.

图 1 给出了这个组织 P 系统的膜结构,其中,3 个细胞分别由 3 个基本膜所围成,分别用 1,2 和 3 标记.环境由 0 标记.这 3 个基本膜称为进化膜,因为其作用是进化系统中的对象.将引入混合进化机制实现对象的进化,具体设计如下:细胞 1 采用 3 种遗传算子(选择、交叉和变异)作为对象进化机制,细胞 2 使用速度-位移模型,而细胞 3 则应用差分进化机制实现对象的进化.混合进化机制带来的益处是增强系统中对象的多样性.图 1 中的有向线表示进化膜之间对象的转运通道,对象转运将实现进化膜之间对象交换与共享.对象的转运是由转运规则完成.这种转运机制能够实现在这 3 个进化膜之间对象的协同进化,并加速系统的收敛.约定环境为系统的输出膜.

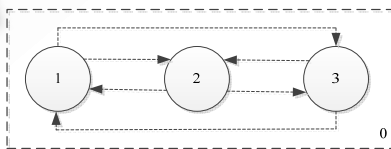


Fig.1 A tissue P system of degree 3

图 1 一个度为 3 的组织 P 系统

2.1.1 对象

组织 P 系统在膜聚类算法中的主要职责是为一个数据集搜索最优的簇中心,为此,采用对象来表达一组(候选的)簇中心.假定数据集 X 有 K 个簇 C_1, C_2, \dots, C_K , 对应的簇中心分别为 z_1, z_2, \dots, z_K . 和数据点一样,每个簇中心是一个 d 维向量,即 $z_i = (z_{i1}, z_{i2}, \dots, z_{id}), i=1, 2, \dots, K$, 于是,细胞中对象可设计为如下形式的 $K \times d$ 维向量:

$$Z = (z_{11}, z_{12}, \dots, z_{1d}, \dots, z_{i1}, z_{i2}, \dots, z_{id}, \dots, z_{K1}, z_{K2}, \dots, z_{Kd}).$$

其中, $z_{1d}, \dots, z_{i1}, z_{i2}, \dots, z_{id}$ 对应于第 i 个簇中心的 d 个分量, $i=1, 2, \dots, K$.

O_i 是第 i 个进化膜的对象集,通常包含一定数量的对象,这些对象将由相应的进化规则所进化.为简单起见,假定 3 个进化膜有相同个数的对象,并记 m 为每个进化膜中对象个数.在计算过程中,环境中始终保存着整个系统迄今为止发现的最好对象,它称为全局最优对象,记为 Z_{best} . 当系统停机后, Z_{best} 即为所求的最终解.

初始时,系统将为每个进化膜生成 m 个初始对象,组成其对象集 $O_i (i=1, 2, 3)$. 当生成一个初始对象时,系统将重复地产生 $K \times d$ 个随机实数并满足数据集中样本点相应维度的取值范围.

在对象进化过程中,系统需要一个度量来评估对象的优劣.本文直接使用聚类问题的性能函数 J_m 作为对象的度量或适应度. J_m 值越小,对象越好;否则越差.

2.1.2 混合进化机制

在组织 P 系统中,每个细胞包含若干进化规则,其作用是进化其对象.进化规则的一般形式为 $u \rightarrow v$, 其含义是对象 u 被进化为对象 v . 在所设计的组织 P 系统中,3 个进化膜分别引入了 3 种不同进化技术来实现对象的进化,它们分别是遗传操作(选择、交叉和变异)^[29]、速度-位移模型^[30]和差分进化机制^[31]. 这 3 种不同进化技术分别作为 3 个进化膜的进化规则.进化膜的混合进化机制必将增强这个组织 P 系统在进化过程中对象的多样性.为了更好地融入这 3 种进化技术,依据所使用的膜结构与转运机制,开发了这 3 种进化技术的改进版本,特别是改进的速度-位移模型和改进的差分进化机制.具体描述如下:

• 膜 1 的进化规则

进化膜 1 采用 GA^[29] 的 3 种遗传操作(选择、交叉和变异)作为进化规则,实现其对象的进化.在每个计算步中,进化膜 1 将自身的 m 个对象和转运自其他两个进化膜的 $2r$ 个对象构成一个进化池.对进化池中对象执行选择、交叉和变异操作.为了保持膜中对象规模不变,将执行截断操作.截断操作依据对象的适应度进行,以保留最好的 m 个对象.

在具体实现中,选择操作采用轮赌法而交叉操作使用实数形式的单点交叉,其中,交叉位置(即,对象的某个分量)是根据交叉概率 p_c 所确定.对象的变异采用了如下实数形式的单点变异:如果 z 是根据变异概率 p_m 所确定的对象 Z 的变异点(即某个分量),则变异后的值为

$$z' = \begin{cases} z \pm 2\delta z, & z \neq 0 \\ z \pm 2\delta, & z = 0 \end{cases} \quad (3)$$

其中,“+”和“-”以等概率出现, δ 是依据均匀分布生成的 $[0, 1]$ 中的一个随机实常数.

• 膜 2 的进化规则

PSO 的速度-位移模型^[30]作为进化膜 2 的进化规则被考虑.然而,结合组织 P 系统的相关机制,提出了一种改进的速度-位移模型.改进是源自于这样一种直觉:进化膜 2 有两种最优对象,一种是它迄今为止所发现的最优对象(称为局部最优对象,记为 Z_{lbest}),另一种是转运自其他两个进化膜的 $2r$ 个最优对象(称为外部最优对象).从这 $2r$ 个外部最优对象中随机地选取一个,记为 Z_{ebest} . 这两种最优对象将参与或指导进化膜 2 中对象的进化.这两种由不同进化技术所产生的最优对象势必增强系统对象的多样性和全局探测能力.

假设 Z_i 为进化膜 2 中第 i 个对象,这个改进的速度-位移模型可被描述如下:

$$\begin{cases} V_i = w \cdot V_i + c_1 r_1 (P_i - Z_i) + c_2 r_2 (Z_{lbest} - Z_i) + c_3 r_3 (Z_{ebest} - Z_i) \\ Z'_i = Z_i + V_i \end{cases} \quad (4)$$

其中, V_i 为 Z_i 对应的速度, Z'_i 为 Z_i 进化后的值, P_i 表示对象 Z_i 迄今为止所发现的最好位置, w 为惯性权重, c_1, c_2, c_3 为学习因子, r_1, r_2, r_3 为 3 个 $[0, 1]$ 中的随机数.在实现时,惯性权重将采用线性递减的策略.

- 膜3的进化规则

进化膜3将采用差分进化算法^[31]的差分进化机制作为进化规则。众所周知,原始的差分进化机制主要包括3种进化算子:变异、交叉和选择。基于同样原因,提出了一种改进的差分进化机制,主要体现在变异算子的设计上,然而交叉和选择算子完全同原始的差分进化机制,这里不再赘述,仅描述改进的变异算子。

假设 Z_i 为进化膜3中第 i 个对象,则变异向量 Y_i 由下式生成:

$$Y_i = \begin{cases} Z_i + \alpha(Z_{lbest} - Z_i) + \alpha(Z_{ebest} - Z_i) + \alpha(Z_{r_1} - Z_{r_2}), & \text{if } rand(0,1) < \alpha \\ Z_i + F(Z_{r_3} - Z_{r_4}), & \text{otherwise} \end{cases} \quad (5)$$

其中, Z_{lbest} 为膜3中的局部最优对象, Z_{ebest} 是外部最优对象(它是从 $2r$ 个外部最优对象中随机地选取一个), Z_{r_1} , Z_{r_2} , Z_{r_3} , Z_{r_4} 是膜3中随机挑选的4个对象。 α 和 F 是两个尺度因子,分别计算如下:

$$\alpha = \frac{1}{1 + \exp\left[-\frac{1}{t_{max}}\right]}, F = 0.5 \times (1 + rand(0,1)) \quad (6)$$

其中, t_{max} 为最大计算步数(或最大迭代次数)。在公式(5)和公式(6)中, $rand(0,1)$ 表示区间(0,1)的一个随机实数生成函数。在得到变异向量 Y_i 后,再经过通常的交叉和选择规则产生 Z_i 进化后的值。

2.1.3 转运机制

细胞膜之间可能存在转运通道,用于它们之间对象的交换与共享。这种转运机制是通过细胞间的转运规则来提供。在所设计的组织P系统中,有如下两类转运规则:

- $(i, Z_1 Z_2 \dots Z_r / Z'_1 Z'_2 \dots Z'_r, j), i \neq j, i, j = 1, 2, 3$.

这是一条细胞 i 和细胞 j 之间的双向转运规则,其中, Z_1, Z_2, \dots, Z_r 为细胞 i 的 r 个对象,而 Z'_1, Z'_2, \dots, Z'_r 为细胞 j 的 r 个对象。该规则执行意味着:细胞 i 中 r 个对象 Z_1, Z_2, \dots, Z_r 被转运到细胞 j 中;同时,细胞 j 中 r 个对象 Z'_1, Z'_2, \dots, Z'_r 被转运到膜 i 中。

- $(i, Z_{lbest} / \lambda, 0), i = 1, 2, 3$.

这是一条细胞 i 和环境之间的单向转运规则,其中, Z_{lbest} 为当前计算步中细胞 i 的(局部)最优对象,且 λ 表示空对象。该规则的执行意味着:细胞 i 中对象 Z_{lbest} 被转运到环境中,并更新环境中的全局最优对象 Z_{best} 。

在所设计的组织P系统中,3个进化膜利用第1类转运规则建立起它们之间的对象转运关系,这种转运关系逻辑上为任意两个细胞提供了两条虚拟有向弧,如图2所示。每个进化膜通过第1类转运规则将其 r 个最好对象转运到其他两个进化膜中。于是,每个进化膜也将接收到来自于其他两个进化膜的最好对象。所接收到的这 $2r$ 个对象构成该进化膜的一个外部最优对象子集,将在下一计算步中参与对象进化。

在每个计算步中,进化膜使用第2类单向转运规则将其最好对象转运到环境中,并更新全局最优对象 Z_{best} 。更新策略如下:如果转运的对象优于这个全局最优对象 Z_{best} ,则这对象将替换 Z_{best} ;否则,它被丢弃。

从上面的描述中可以看出,细胞间的这种转运关系隐含地揭示了组织P系统的网状结构。这种内在转运关系能够为采用混合进化技术的这3个进化膜提供一种协同进化机制,有助于加速整个系统的收敛。

2.1.4 停机与输出

组织P系统的各个细胞作为执行单元并行地运行(假定存在一个全局时钟,用于标定时间),所以,系统的功能是分布式和并行的。定义组织P系统的一个计算为一系列的计算步,它们从包含初始对象集的3个细胞开始,并且在每个计算步中,一个或多个规则作用于当前对象集上。如果系统达到停机条件,则系统停机。这时,计算结果位于输出膜(即环境)中。

为简单起见,采用了一种简单的停机条件:设置最大执行步数。这个组织P系统会持续运行直到其达到最大执行步数为止,系统停机。当系统停机时,环境中所保存的最优对象为最终输出结果,即所发现的最优簇中心。

2.2 膜聚类算法

本文所提出的膜聚类算法是一种在膜计算框架下的划分聚类算法,其核心是一个度为3、具有混合进化机制的组织P系统,其作用是自动地为一个待聚类的数据集搜索最优的簇中心。

对于一个具有 n 个数据点的数据集 X , 假定将这些数据划分为 K 个簇. 组织 P 系统中的每个对象表达一组候选的簇中心. 膜聚类算法首先为每个进化膜生成 m 个初始对象, 然后执行这个组织 P 系统. 在混合进化机制和转运机制的控制下, 这个组织 P 系统不断地进化系统中的对象并更新全局最优对象 Z_{best} , 直到系统满足停机条件为止. 当系统停机时, 这个全局最优对象 Z_{best} 即为最优簇中心. 最后, 膜聚类算法根据这个最优簇中心重新划分 n 个数据点到相应的 K 个簇中完成数据聚类.

图 2 给出了这种膜聚类算法的流程图.

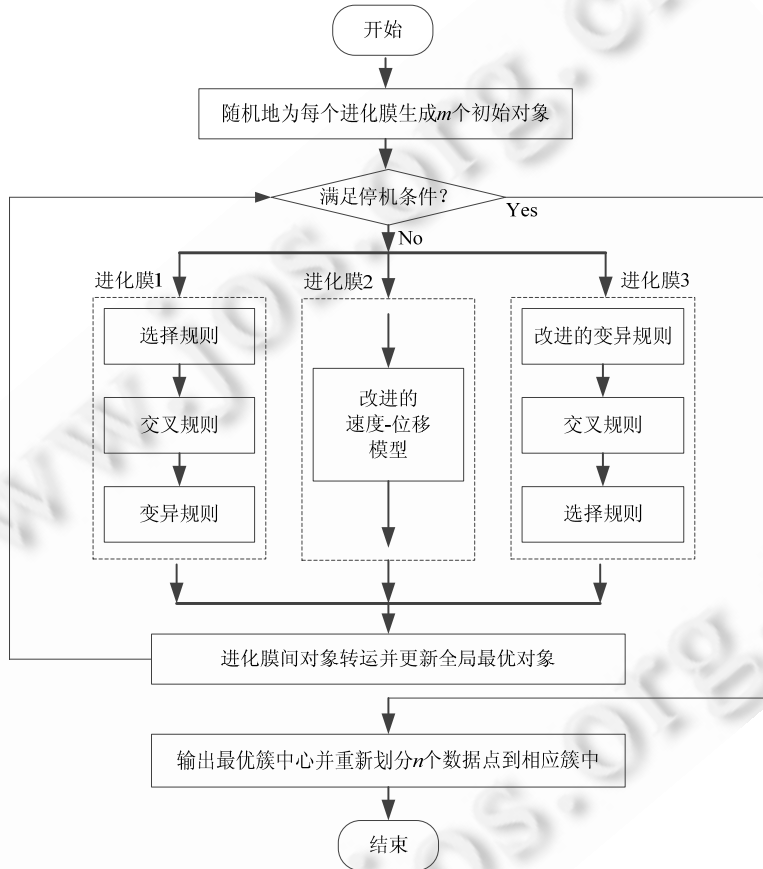


Fig.2 Flow chart of the membrane clustering algorithm

图 2 膜聚类算法的流程图

3 仿真实验

3.1 实验数据集

为了评估膜聚类算法, 实验中使用了 8 个数据集, 其中 3 个是 2 维的人工数据集(AD_5_2, Square_4 和 Sym_3_22), 另外 5 个是来自于 UCI 的真实数据集(Iris, LungCancer, Newthyroid, Wine 和 Livedisorder)^[35]. 图 3 显示了这 3 个人工数据集. 表 1 给出了这 8 个数据集的简单描述.

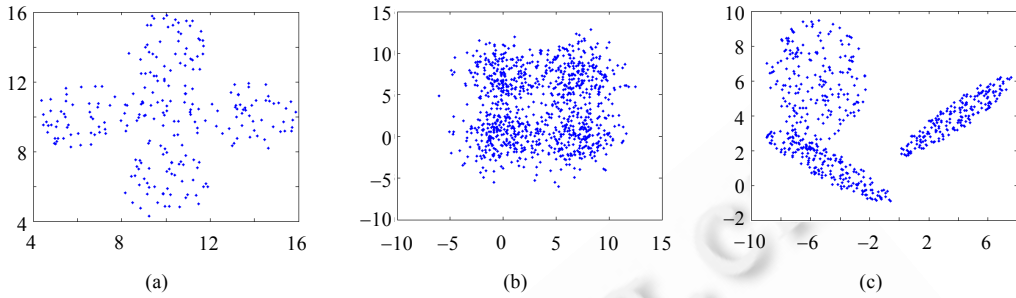


Fig.3 Three artificial data sets
图 3 3 个人工数据集

Table 1 Data sets used in experiments
表 1 实验中所使用的数据集

Data sets	No. of data	Dimension	No. of clusters
AD_5_2	250	2	5
Square_4	1 000	2	4
Sym_3_22	600	2	3
Iris	150	4	3
LungCancer	32	56	3
Newthyroid	215	5	3
Wine	178	13	3
Livedisorder	345	6	2

3.2 输入参数

在实验中,所提出的膜聚类算法将与 k -means 算法和几种典型的进化算法进行对比,包括 GA^[29],PSO^[17],ACO^[18]和 PSO-ACO^[19].同时,考虑到所设计的组织 P 系统中采用了改进的速度-位移模型和改进的差分进化机制,为了研究其对聚类性能的提升效果,我们实现了这个膜聚类算法的另外两个副本,它们分别使用标准的速度-位移模型和标准的差分进化机制(“rand/1”型的变异规则).记采用混合进化机制的系统为 Hybrid-P systems,而另外两个副本分别为 PSO-P systems 和 DE-P systems.通过在上述 8 个数据集上的仿真结果来评估这些聚类算法的聚类性能.

Hybrid-P systems 的参数设置如下:每个细胞(膜)中对象的个数 m 为 100,最大执行步数为 100;(细胞 1 中)遗传操作采用自适应的交叉概率和变异概率;(细胞 2 中)改进的速度-位移模型采用线性递减的惯性权重,且 $c_1=c_2=c_3=1.0$;(细胞 3 中)改进的差分进化机制所使用的交叉概率为 0.9.为了对比的需要,PSO-P systems 和 DE-P systems 分别采用了与 Hybrid-P systems 的参数.

在实现其他 4 种进化算法(GA,PSO,ACO 和 PSO-ACO)时,通过几组典型参数的结果比较来确定最佳的参数值.这些参数设置如下:(1) PSO 参数为 $c_1=c_2=2.0$;(2) ACO 参数为 $\gamma_1=\gamma_2=0.5, \rho=0.99$;(3) 除了与 PSO 和 ACO 相同的参数以外,PSO-ACO 的其他参数为 $D_0=10, a=15, r=0.5$;(4) GA 的交叉概率和变异概率分别为 0.8 和 0.001.另外,这些算法的最大迭代步数均设置为 100.

3.3 实验结果

为了评估这些聚类算法的聚类质量,实验采用如下 3 种度量:

(i) 整体均方差,见公式(2).一般地,其值越小,聚类质量越好.

(ii) Silhouette Index (SI).假设 a 表示一个点与来自于这个簇的其他点的平均距离, b 表示这个点与其他簇中点的平均距离的最小值,于是这个点的 Silhouette 宽度定义为

$$s = \frac{b - a}{\max\{a, b\}} \tag{7}$$

SI 值是所有数据点的 Silhouette 宽度的平均值,它介于[-1,1]之间.一般地,较高的 SI 值意味着较好的聚类质量.

(iii) CS 度量.该度量是由 Chou 等人^[36]提出的.研究表明,CS 度量能够有效地处理不同密度和规模的簇.CS 度量定义如下:

$$CS(K) = \frac{\sum_{i=1}^K \left[\frac{1}{N_i} \sum_{X_i \in C_i} \max_{X_q \in C_q} \|X_i - X_q\| \right]}{\sum_{i=1}^K \min_{j \in K, j \neq i} \|m_i - m_j\|} \quad (8)$$

其中, m_i 由下式计算:

$$m_i = \sum_{X_j \in C_i} X_j / N_i.$$

一般地,CS 度量值越小则意味着所得到的划分越好.即,聚类算法获得了良好的聚类性能,反之亦然.

考虑到这些聚类算法中包含随机因素,并且每次运行的结果可能不一样,所以在每一个数据集上对每种算法都独立地运行 100 次,分别计算上述 3 个度量对这 100 次运行的平均值和标准差.平均值将反映算法的平均性能,而标准差则指示算法的鲁棒性.表 2~表 9 分别给出了这些聚类算法在 3 个人工数据集和 5 个真实数据集上的仿真结果.每一个表分别提供了这 8 个聚类算法的 J_m ,SI 和 CS 度量的平均值和标准差,它们是在对应数据集上独立地运行 100 次的结果.

Table 2 Experimental results on AD_5_2

表 2 在 AD_5_2 上的实验结果

Methods	J_m		SI		CS	
	Average	Std.	Average	Std.	Average	Std.
Hybrid-P systems	326.438 4	0.0	0.512 1	0.0	0.603 5	0.046
PSO-P systems	326.440 8	0.001 2	0.511 2	0.003	0.605 1	0.053
DE-P systems	326.439 7	0.000 1	0.511 5	0.001	0.604 8	0.052
<i>k</i> -means	332.478 1	3.128 6	0.411 5	0.093	0.657 3	0.231
GA	328.316 5	0.479 2	0.492 1	0.038	0.615 2	0.093
PSO	326.448 6	0.012 8	0.508 3	0.017	0.609 8	0.065
ACO	326.452 3	0.034 4	0.505 9	0.019	0.608 2	0.068
PSO-ACO	326.441 1	0.007 5	0.510 8	0.003	0.606 5	0.055

Table 3 Experimental results on Square_4

表 3 在 Square_4 上的实验结果

Methods	J_m		SI		CS	
	Average	Std.	Average	Std.	Average	Std.
Hybrid-P systems	2 379.735 7	0.0	0.478 9	0.0	0.578 2	0.025
PSO-P systems	2 379.736 9	0.003 2	0.477 8	0.003	0.579 7	0.031
DE-P systems	2 379.736 4	0.000 5	0.478 0	0.0	0.579 3	0.033
<i>k</i> -means	2 386.003 5	4.521 7	0.263 1	0.125	0.598 6	0.132
GA	2 380.332 5	0.631 9	0.384 2	0.059	0.584 5	0.076
PSO	2 379.743 7	0.022 6	0.466 2	0.027	0.582 2	0.041
ACO	2 379.792 1	0.042 8	0.461 5	0.032	0.581 3	0.044
PSO-ACO	2 379.738 6	0.008 3	0.477 5	0.005	0.579 9	0.038

Table 4 Experimental results on Sym_3_22

表 4 在 Sym_3_22 上的实验结果

Methods	J_m		SI		CS	
	Average	Std.	Average	Std.	Average	Std.
Hybrid-P systems	1 247.693 3	0.0	0.530 5	0.0	0.635 2	0.025
PSO-P systems	1 247.696 2	0.002 5	0.528 8	0.007	0.636 2	0.038
DE-P systems	1 247.695 7	0.000 9	0.529 2	0.0	0.635 9	0.036
<i>k</i> -means	1 255.457 2	3.872 5	0.327 9	0.135	0.735 2	0.137
GA	1 249.361 9	1.216 3	0.465 8	0.048	0.642 7	0.085
PSO	1 247.722 5	0.014 9	0.526 9	0.009	0.638 1	0.059
ACO	1 247.758 6	0.031 5	0.525 1	0.017	0.637 9	0.053
PSO-ACO	1 247.697 1	0.005 3	0.528 3	0.003	0.637 5	0.042

Table 5 Experimental results on Iris

表 5 在 Iris 上的实验结果

Methods	J_m		SI		CS	
	Average	Std.	Average	Std.	Average	Std.
Hybrid-P systems	96.655 5	6.3246×10^{-7}	0.550 9	0.0	0.697 3	0.015
PSO-P systems	96.686 5	0.002 6	0.543 8	0.006	0.708 6	0.038
DE-P systems	96.679 2	8.7451×10^{-5}	0.547 1	0.001	0.699 2	0.033
<i>k</i> -means	104.117 0	12.456 3	0.334 7	0.174	0.932 8	0.547
GA	99.834 7	5.523 9	0.465 5	0.089	0.886 1	0.418
PSO	97.235 6	0.351 3	0.522 3	0.016	0.789 5	0.225
ACO	97.253 8	0.415 2	0.513 8	0.021	0.758 2	0.147
PSO-ACO	96.694 3	0.008 9	0.542 1	0.007	0.731 5	0.053

Table 6 Experimental results on LungCancer

表 6 在 LungCancer 上的实验结果

Methods	J_m		SI		CS	
	Average	Std.	Average	Std.	Average	Std.
Hybrid-P systems	126.454 0	1.009 4	0.073 4	0.012	0.474 9	0.039
PSO-P systems	126.489 6	1.011 8	0.072 2	0.014	0.523 6	0.045
DE-P systems	126.493 5	1.013 4	0.072 8	0.014	0.489 9	0.042
<i>k</i> -means	139.407 2	7.313 6	0.033 7	0.063	0.803 9	0.582
GA	129.521 9	4.496 1	0.051 3	0.041	0.683 1	0.153
PSO	127.231 1	1.152 8	0.069 4	0.022	0.602 8	0.088
ACO	127.314 4	1.293 6	0.068 9	0.026	0.593 2	0.093
PSO-ACO	126.501 2	1.021 5	0.072 1	0.015	0.554 8	0.058

Table 7 Experimental results on Newthyroid

表 7 在 Newthyroid 上的实验结果

Methods	J_m		SI		CS	
	Average	Std.	Average	Std.	Average	Std.
Hybrid-P systems	1 872.357 6	11.048 5	0.335 2	0.013	0.632 5	0.231
PSO-P systems	1 872.413 5	11.057 3	0.334 9	0.014	0.648 8	0.245
DE-P systems	1 872.392 7	11.049 8	0.335 1	0.013	0.638 6	0.242
<i>k</i> -means	1 886.252 6	16.218 9	0.248 3	0.057	0.852 6	0.692
GA	1 875.114 5	13.583 4	0.328 5	0.034	0.783 4	0.458
PSO	1 872.512 3	11.092 3	0.334 1	0.015	0.719 8	0.352
ACO	1 872.562 5	11.104 5	0.333 7	0.016	0.715 5	0.356
PSO-ACO	1 872.427 3	11.066 8	0.334 9	0.014	0.689 4	0.261

Table 8 Experimental results on Wine

表 8 在 Wine 上的实验结果

Methods	J_m		SI		CS	
	Average	Std.	Average	Std.	Average	Std.
Hybrid-P systems	16 292.255 4	0.039 3	0.565 9	0.0	0.892 8	0.032
PSO-P systems	16 292.256 3	0.039 7	0.565 0	0.003	0.895 3	0.035
DE-P systems	16 292.255 9	0.039 5	0.565 6	0.0	0.893 2	0.033
<i>k</i> -means	16 312.431 7	9.426 9	0.362 3	0.062	0.997 3	0.147
GA	16 298.423 5	2.152 3	0.439 1	0.046	0.912 6	0.095
PSO	16 292.257 1	0.153 1	0.556 3	0.022	0.896 9	0.041
ACO	16 292.259 2	0.167 2	0.552 7	0.037	0.897 3	0.039
PSO-ACO	16 292.256 9	0.039 8	0.564 8	0.006	0.895 8	0.036

Table 9 Experimental results on Livedisorder

表 9 在 Livedisorder 上的实验结果

Methods	J_m		SI		CS	
	Average	Std.	Average	Std.	Average	Std.
Hybrid-P systems	9 851.721 4	1.2092×10^{-5}	0.554 7	0.0	0.723 6	0.025
PSO-P systems	9 851.728 6	0.0007	0.553 9	0.004	0.724 5	0.038
DE-P systems	9 851.724 6	5.6315×10^{-5}	0.554 3	0.001	0.724 1	0.032
<i>k</i> -means	9 868.325 7	7.927 4	0.389 1	0.056	0.823 6	0.354
GA	9 856.142 8	1.952 3	0.412 5	0.037	0.735 1	0.134
PSO	9 851.738 5	0.035 6	0.543 7	0.013	0.728 6	0.098
ACO	9 851.746 8	0.069 2	0.529 3	0.025	0.729 1	0.083
PSO-ACO	9 851.730 5	0.001 8	0.553 5	0.009	0.725 8	0.051

首先,我们观察 Hybrid-P systems 与它的两个副本 PSO-P systems 和 DE-P systems 的对比结果.在所有的数据集上,对比于两个副本 PSO-P systems 和 DE-P systems,Hybrid-P systems 有更低的 J_m 值和 CS 值以及更高的 SI 值.这说明本文采用的混合进化机制在提升膜聚类算法平均性能方面具有较明显的优势.从标准差来看,Hybrid-P systems 和 DE-P systems 均低于 PSO-P systems.这说明 Hybrid-P systems 和 DE-P systems 均比 PSO-P systems 更加鲁棒,但两者却非常接近.

其次,我们比较 Hybrid-P systems 与其他 4 种聚类算法,Hybrid-P systems 和 PSO-ACO 在这 8 个数据集上获得了更低的 J_m 值和 CS 值以及更高的 SI 值.这说明 Hybrid-P systems 和 PSO-ACO 有较好的聚类效果.与此同时,也能看到,Hybrid-P systems 的聚类性能稍好于 PSO-ACO 算法.在标准差方面,Hybrid-P systems 对比其他 5 种聚类算法有更加显著的优势,在多数情况下,Hybrid-P systems 得到了为 0 或非常低的标准差.该结果说明,Hybrid-P systems 有更好的鲁棒性.

从实验结果以及以上分析可以得出如下结论:(1) 由组织 P 系统内在机制所启发的膜聚类算法有更好的平均聚类性能以及较高的鲁棒性;(2) 混合进化机制以及改进的速度-位移模型和差分进化机制能够较好地改进膜聚类算法的聚类性能.

Wilcoxon 秩和检验^[37]是一种用于独立样本的非参数统计显著性检验方法.实验中,这种统计显著性检验已在 5% 的显著性级上完成.对每个数据集建立 6 个组,分别对应于这 6 种聚类算法(hybrid-P systems,*k*-means,GA,PSO,ACO 和 PSO-ACO),其中,每一个组由相应聚类算法分别在这些数据集上连续执行 100 次的 SI 值组成.表 2~表 9 分别给出了这些数据集上每个聚类算法所对应的一组的 SI 的平均值.这些结果表明,Hybrid-P systems 的平均性能比其他 5 个聚类算法的平均性能要好.

为了进一步说明这种优良性是统计显著的,我们完成了一个统计显著性检验.零假设是假定两个组的平均值之间没有显著性差别,然而另一种假设是两个组的平均值之间有显著性差别.实验依次比较了 Hybrid-P systems 这组与其他 5 个组中的任何一组,计算相应的 Wilcoxon 秩和检验所生成的 P-值(接受零假设的概率).

表 10 给出了在这些数据集上 Wilcoxon 秩和检验的结果.在表 10 中,所有的 P-值是低于 0.05(5%显著性级),这表明以强的理由拒绝零假设.实验结果证实了 Hybrid-P systems 所产生的较好的 SI 平均值是统计显著的而不是偶然得到的.

Table 10 Results of P-values produced by Wilcoxon's rank sum test

表 10 由 Wilcoxon 秩和检验产生的 P-值结果

Data sets	<i>k</i> -means	GA	PSO	ACO	PSO-ACO
AD_5_2	1.1876×10^{-4}	1.3589×10^{-4}	1.7323×10^{-3}	1.0048×10^{-3}	0.009 5
Square_4	1.2035×10^{-4}	1.2461×10^{-4}	1.2348×10^{-3}	1.1253×10^{-3}	0.018 3
Sym_3_22	1.1980×10^{-4}	1.2238×10^{-4}	1.7951×10^{-3}	1.1321×10^{-3}	0.006 8
Iris	1.2246×10^{-4}	1.3043×10^{-4}	3.4128×10^{-3}	1.5871×10^{-3}	0.012 7
LungCancer	1.2073×10^{-4}	1.2936×10^{-4}	2.3526×10^{-3}	1.2148×10^{-3}	0.014 6
Newthyroid	1.2125×10^{-4}	1.2724×10^{-4}	2.1136×10^{-3}	1.2463×10^{-3}	0.020 5
Wine	1.2572×10^{-4}	1.2197×10^{-4}	2.4258×10^{-3}	1.1279×10^{-3}	0.015 8
Livedisorder	1.2319×10^{-4}	1.2468×10^{-4}	2.2541×10^{-3}	1.2246×10^{-3}	0.012 2

4 结束语

本文探讨了使用具有混合进化机制的组织 P 系统去处理数据聚类问题,提出了一种在膜计算下的划分聚类算法,即,膜聚类算法.一个由 3 个细胞构成的组织 P 系统是这种膜聚类算法的核心,并依据它的膜结构和转运机制提出了一种改进的速度位移模型和一种改进的差分进化机制.这种膜聚类算法的性能分别在 3 个人工数据集和 5 个真实数据集上进行评估.实验结果显示:与 k -means 和几种进化聚类算法相比,膜聚类算法有更好的优势.这个研究也揭示了组织 P 系统解决数据聚类问题的有效性.

组织 P 系统利用多个细胞中对象进化以及细胞间对象转运来搜索最优解,这类似于已有的多种群进化算法,然而,这里组织 P 系统的机理是源自于细胞生物学.众所周知,已有的众多 P 系统模型分别源自于细胞生物学的不同机理,诸如酶机制、催化机制、细胞分裂与溶解机制、转运通道的激活与抑制等.这些机制的引入,是我们进一步研究的方向.从另一个角度来讲,融合不同细胞生物学机理的、处理优化问题的 P 系统也可以视为多种群进化算法的一个补充.另外,本文的膜聚类算法存在一个限制,即,需要预先指定簇个数 K .在进一步的工作中,我们考虑引入细胞分裂与溶解机制来解决这个问题.

References:

- [1] Gan G, Ma C, Wu J. Data Clustering: Theory, Algorithms, and Applications. SIAM, 2007.
- [2] Xu R, Wunsch D. Survey of clustering algorithm. IEEE Trans. on Neural Networks, 2005,16(3):645–678. [doi: 10.1109/TNN.2005.845141]
- [3] Jain AK. Data clustering: 50 years beyond k -means. Pattern Recognition Letters, 2010,31(8):651–666. [doi: 10.1016/j.patrec.2009.09.011]
- [4] Everitt B, Landau S, Leese M. Cluster Analysis. London: Arnold, 2001.
- [5] Saha S, Bandyopadhyay S. A symmetry based multiobjective clustering technique for automatic evolution of clusters. Pattern Recognition, 2010,43(3):738–751. [doi: 10.1016/j.patcog.2009.07.004]
- [6] Das S, Sil S. Kernel-Induced fuzzy clustering of image pixels with an improved different evolution algorithm. Information Sciences, 2010,180(8):1237–1256. [doi: 10.1016/j.ins.2009.11.041]
- [7] Naldi MC, Campello RJGB, Hruschka ER, Carvalho ACPLF. Efficiency issues of evolutionary k -means. Applied Soft Computing, 2011,11(8):1938–1952. [doi: 10.1016/j.asoc.2010.06.010]
- [8] Kanungo T, Mount D, Netanyahu NS, Piatko C, Silverman R, Wu A. An efficient k -means clustering algorithm: Analysis and implementation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002,24(7):881–892. [doi: 10.1109/TPAMI.2002.1017616]
- [9] Wu X, Kumar V. The Top Ten Algorithms in Data Mining. Chapman and Hall/CRC, 2009.
- [10] Murthy CA, Chowdhury N. In search of optimal clusters using genetic algorithms. Pattern Recognition Letters, 1996,17(8):825–832. [doi: 10.1016/0167-8655(96)00043-8]
- [11] Maulik U, Bandyopadhyay S. Genetic algorithm based clustering technique. Pattern Recognition, 2000,33(9):1455–1465. [doi: 10.1016/S0031-3203(99)00137-5]
- [12] Bandyopadhyay S, Saha S. GAPS: A clustering method using a new point symmetry-based distance measure. Pattern Recognition, 2007,40(12):3430–3451. [doi: 10.1016/j.patcog.2007.03.026]
- [13] Bandyopadhyay S. Genetic algorithms for clustering and fuzzy clustering. Data Mining and Knowledge Discovery, 2011,1(6):524–531. [doi: 10.1002/widm.47]
- [14] Chang D, Zhang X, Zheng C. A genetic algorithm with gene rearrangement for k -means clustering. Pattern Recognition, 2009,42(7):1210–1222. [doi: 10.1016/j.patcog.2008.11.006]
- [15] Nguyen CD, Cios KJ. GAKREM: A novel hybrid clustering algorithm. Information Sciences, 2008,178(22):4205–4227. [doi: 10.1016/j.ins.2008.07.016]
- [16] Laszlo M, Mukherjee S. A genetic algorithm that exchanges neighboring centers for k -means clustering. Pattern Recognition Letters, 2007,28(16):2359–2366. [doi: 10.1016/j.patrec.2007.08.006]
- [17] Kao YT, Zahara E, Kao IW. A hybridized approach to data clustering. Expert Systems with Applications, 2008,34(3):1754–1762. [doi: 10.1016/j.eswa.2007.01.028]

- [18] Shelokar PS, Jayaraman VK, Kulkarni BD. An ant colony approach for clustering. *Analytica Chimica Acta*, 2004,509(2): 187–195. [doi: 10.1016/j.aca.2003.12.032]
- [19] Niknam T, Amiri B. An efficient hybrid approach based on PSO, ACO and k -means for cluster analysis. *Applied Soft Computing*, 2010,10(1):183–197. [doi: 10.1016/j.asoc.2009.07.001]
- [20] Păun G. Computing with membranes. *Journal of Computer and System Sciences*, 2000,61(1):108–143. [doi: 10.1006/jcss.1999.1693]
- [21] Păun G, Rozenberg G, Salomaa A. *The Oxford Handbook of Membrane Computing*. New York: Oxford University Press, 2010.
- [22] Freund R, Păun G, Pérez-Jiménez MJ. Tissue-Like P systems with channel-states. *Theoretical Computer Science*, 2005,330(1): 101–116. [doi: 10.1016/j.tcs.2004.09.013]
- [23] Ionescu M, Păun G, Yokomori T. Spiking neural P systems. *Fundamenta Informaticae*, 2006,71(2-3):279–308.
- [24] Păun G, Pérez-Jiménez MJ. Membrane computing: Brief introduction, recent results and applications. *BioSystems*, 2006,85(1): 11–22. [doi: 10.1016/j.biosystems.2006.02.001]
- [25] Wang J, Zhou L, Peng H, Zhang GX. An extended spiking neural P system for fuzzy knowledge representation. *Int'l Journal of Innovative Computing, Information and Control*, 2011,7(7A):3709–3724.
- [26] Peng H, Wang J, Pérez-Jiménez MJ, Wang H, Shao J, Wang T. Fuzzy reasoning spiking neural P system for fault diagnosis. *Information Sciences*, 2013,235:106–116. [doi: 10.1016/j.ins.2012.07.015]
- [27] Wang J, Shi P, Peng H, Pérez-Jiménez MJ, Wang T. Weighted fuzzy spiking neural P systems. *IEEE Trans. on Fuzzy Systems*, 2013,21(2):209–220. [doi: 10.1109/TFUZZ.2012.2208974]
- [28] Peng H, Wang J, Pérez-Jiménez MJ, Shi P. A novel image thresholding method based on membrane computing and fuzzy entropy. *Journal of Intelligent & Fuzzy Systems*, 2013,24(2):229–237. [doi: 10.3233/IFS-2012-0549]
- [29] Davis L. *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, 1991.
- [30] Kennedy J, Eberhart R. Particle swarm optimization. In: *Proc. of the IEEE Int'l Conf. on Neural Network*, Vol.4. Piscataway, 1995. 1942–1948. [doi: 10.1109/ICNN.1995.488968]
- [31] Price K, Storn R, Lampinen J. *Differential Evolution—A Practical Approach to Global Optimization*. Berlin: Springer-Verlag, 2005. [doi: 10.1007/3-540-31306-0]
- [32] Zhang GX, Cheng JX, Gheorghe M, Meng Q. A hybrid approach based on differential evolution and tissue membrane systems for solving constrained manufacturing parameter optimization problems. *Applied Soft Computing*, 2013,13(3):1528–1542. [doi: 10.1016/j.asoc.2012.05.032]
- [33] Colomer AM, Margalida A, Pérez-Jiménez MJ. Population dynamics P systems (PDP) models: A standardized protocol for describing and applying novel bio-inspired computing tools. *Plos One*, 2013,8(4):1–13. [doi: 10.1371/journal.pone.0060698]
- [34] Frisco P, Gheorghe M, Pérez-Jiménez MJ. *Applications of Membrane Computing in Systems and Synthetic Biology*. Springer-Verlag, 2013. [doi: 10.1007/978-3-319-03191-0]
- [35] UCI datasets. <http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [36] Chou CH, Su MC, Lai E. A new cluster validity measure and its application to image compression. *Pattern Analysis and Applications*, 2004,7(2):205–220. [doi: 10.1007/s10044-004-0218-1]
- [37] Hollander M, Wolfe DA. *Nonparametric Statistical Methods*. 2nd ed., 1999.



彭宏(1966—),男,四川犍为人,博士,教授,主要研究领域为膜计算,图像处理,模式识别。



蒋洋(1988—),男,硕士生,主要研究领域为膜计算。



王军(1966—),女,博士,教授,主要研究领域为膜计算,智能控制。



Mario J. PÉREZ-JIMÉNEZ(1948—),男,博士,教授,博士生导师,欧洲科学院院士,主要研究领域为膜计算,自然计算,生物信息学。