基于免疫优势的克隆选择聚类算法

刘若辰[1],沈正春[1],贾 建[2],焦李成[1]
(1.西安电子科技大学智能信息处理研究所 智能感知与图像理解教育部重点实验室,陕西西安 710071;
2.西北大学数学系,陕西西安 710069)

摘 要: 基于克隆选择原理和免疫优势理论,本文提出一种新的基于免疫优势的克隆选择聚类算法(Immunodomaince based Clonal Selection Clustering Algorithm,IDCSCA),该算法通过在经典的克隆选择算法框架中,引入基于免疫优势理论的免疫优势算子实现了在线自适应动态获得先验知识和个体间的信息共享.新算法首先通过对群体中若干最优抗体的分析,提取免疫优势,然后将其推广到整个抗体群,通过在进化过程中利用积累的先验知识,在保证抗体种群多样性的基础上加快收敛速度.采用个5个数据集对算法性能进行了测试,与模糊C均值算法(Fuzzy C-means,FCM)、基于遗传算法的模糊聚类算法(Genetic Algorithm based Fuzzy C-means,GAFCM)以及基于克隆选择的模糊聚类算法(Clonal Selection Algorithm based Fuzzy C-means,CSAFCM)比较,结果表明IDCSCA能有效避免聚类中心迭代过程中陷入局部最优点的问题,而且聚类性能更稳定.

关键词: 免疫克隆;遗传算法;免疫优势;聚类
中图分类号: TP18 **文献标识码:** A **文章编号:** 0372-2112 (2010) 04-0960-06

Immunodomaince Based Clonal Selection Clustering Algorithm

LIU Ruo-chen[1], SHEN Zheng-chun[1], JIA Jian[2], JIAO Li-cheng[1]
(1. *Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, institute of Intelligent Information Processing, Xidian University, Xi'an, Shaanxi* 710071, *China*;
2. *Department of Mathematics, North West University, Xi'an, Shaanxi* 710069, *China*)

Abstract: Based on clonal selection principle and the immunodominance theory, a new immune clustering algorithm, Immunodomaince based Clonal Selection Clustering Algorithm (IDCSCA) is proposed in this paper. An immunodomaince operator is introduced to the clonal selection algorithm, which can realize on-line gaining priori knowledge and sharing information among different individuals. Firstly, the gene of elites in antibody population can be extracted and generalized to ordinary antibodies, by the interaction, the whole antibody population evolves. The proposed method has been extensively compared with Fuzzy C-means (FCM), Genetic Algorithm based FCM (GAFCM) and Clonal Selection Algorithm based FCM (CSAFCM) over a test suit of several real life data sets and synthetic data sets. The result of experiment indicates the superiority of the IDCSCA over FCM, GAFCM and CSAFCM on stability and reliability for its ability to avoid trapping in local optimum.

Key words: clone selection; genetic algorithm; immunodominance; fuzzy clustering

1 引言

聚类分析是多元统计分析的一种,也是非监督模式识别的一个重要分支.它把一个没有类别标记的样本集按某种准则划分成若干个子集,使相似的样本尽可能归为一类,而不相似的样本尽量划分到不同的类中.聚类分析的研究过程中,出现了一批经典的聚类算法,如k-means聚类算法、FCM聚类算法等.但是这些传统经典聚类算法在获得成功应用的同时,也存在一些问题,如对初始化敏感、仅能收敛到局部最优点、算法伸缩性较差、无法处理大规模数据等.为解决这些问题,人们将自然计算引入了聚类研究领域,提出了一系列基于自然计算的聚类算法[1~10],如基于遗传算法的聚类方法[1~5],基于免疫进化的聚类算法[6,7]、基于克隆选择的聚类算法[8]、基于免疫网络的聚类算法[9,10]、基于基因表达式编程的聚类算法[13]等.

对于大规模数据集,一方面全体样本的总数很大,这是对聚类算法尤其是经典聚类算法可扩展性的一个挑战;另一方面,数据集类别数的增加,将导致某些类别由于类内样本数较少,或类内分布相对密集,而以近似于孤立点的形式存在于样本空间中.这种数据集对聚类算法尤其是基于自然计算的聚类算法,所采用的聚类中

收稿日期:2009-01-16;修回日期:2009-08-31
基金项目:国家自然科学基金(No.60803098,60703108);国家教育部博士点基金(No.20070701022);中国博士后科学基金(No.20080431228,20090451369)

心搜索策略提出了更高的要求.现有的聚类算法对此类数据进行聚类时,由于聚类中心在迭代的过程中容易陷入局部最优值,而无法发现该类型数据集的正确聚类.因此,在大规模多类别数据的聚类分析中,聚类算法在可靠性,稳定性方面亟待改进.

作为一种人工免疫系统算法,克隆选择算法已经被成功地应用于聚类分析的研究[8~11],为进一步改善基于人工免疫系统的聚类算法的性能,并使其适应于大规模、多类别、分布复杂的数据集的聚类,本文提出了一种新的聚类算法,该算法在克隆选择算法的框架内引入免疫优势算子,使算法能够利用聚类中心搜索过程中积累的先验知识,加快收敛过程,防止陷入局部最优点,提高算法的稳定性.

2　克隆选择理论以及免疫优势定义

2.1　克隆选择理论

1958 年,Burnet 等人提出了著名的免疫克隆选择学说[12].克隆选择原理认为,淋巴细胞除了扩增或者分化成浆细胞以外,也能分化成生命期较长的 B 记忆细胞.当再次遇到相应的抗原时,记忆细胞将预先被免疫系统选择出来,并迅速活化、增殖、分化为抗体生成细胞,执行高效而持久的免疫功能.

受克隆选择学说启发,De Castro 等人提出了一种克隆选择算法(Clonal Selection Algorithm, CSA)[13],它借助于免疫系统的抗体克隆选择机理,构造适用于人工智能的克隆算子.基于克隆算子的克隆选择算法是一种群体搜索策略,具有并行性和搜索变化的随机性,在搜索中不易陷入局部最优值,能以较大的概率获得问题的全局最优解,且具有较快的收敛速度.基于以上优点,本文将以其作为主框架构造新的聚类方法.

2.2　免疫优势定义

免疫学认为,虽然一个抗原分子上可以有多个表位,但在诱导宿主免疫应答时可能只有一种或一个表位起主要作用,使宿主产生以该特异性为主的免疫应答;这种现象称为免疫显性或免疫优势 (immunodominance),起关键作用的表位称为显性表位.免疫优势是在抗体与抗原相互作用中产生的,其产生和作用都是一个动态的过程.免疫优势位点决定了在自然选择中哪一种抗原将面临更大的压力.

在人工免疫系统中,用抗体免疫优势的概念来表示抗体编码每一位对抗原的不同重要程度[14].需要指出的是,这里没有完全遵循免疫优势的生物学定义,一方面是因为依据独特型调节网络理论,抗体也具有抗原特性,因此将这一概念用于抗体也是有生物学基础的;另一方面,在人工免疫系统主要操作是针对抗体,而非抗原.为便于说明,有如下定义:

对于抗体 $\boldsymbol{a}=(a_1,a_2,\cdots,a_l)$,抗体 $\boldsymbol{a}$ 的解码为 $e^{-1}(\boldsymbol{a})$,如果对于任意的 $a_k\in\{0,1\}$,$k\neq i$ 有

$$f(\boldsymbol{a})=f(\boldsymbol{e}^{-1}(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_{i-1},1,\boldsymbol{a}_{i+1},\boldsymbol{a}_l))\geqslant\boldsymbol{f}(\boldsymbol{e}^{-1}(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_{i-1},0,\boldsymbol{a}_{i+1},\boldsymbol{a}_l)) \quad (1)$$

或

$$\boldsymbol{f}(\boldsymbol{a})=f(e^{-1}(a_1,a_2,\cdots,a_{i-1},0,a_{i+1},a_l))\geqslant f(e^{-1}(a_1,a_2,\cdots,a_{i-1},1,a_{i+1},a_l)) \quad (2)$$

则称抗体编码的第 i 位具有免疫优势或免疫优势点(显性表位),该位的取值称为优势值.抗原或抗体具有免疫优势的过程统称为获得免疫优势.

显然,抗体免疫优势本质上都是将先验知识引入算法,从而改善算法性能.本文将免疫优势算子引入克隆选择算法,构造基于免疫优势的免疫克隆聚类算法,进一步提高了人工免疫系统算法对多类别数据集聚类的效率.

3　基于免疫优势的克隆选择聚类算法

鉴于免疫优势算子的优良特性,我们将其引入聚类分析,提出了基于免疫优势的克隆选择聚类算法,用于解决多类别数据的聚类问题.本节首先介绍算法中的几个关键技术.最后给出算法的框架.

3.1　抗体编码及参考抗体的生成

考虑到算法在各阶段针对抗体所采取的不同操作及计算复杂度等问题,基于免疫优势克隆聚类算法对抗体采用了两种编码方式,以适应于相关的操作.

首先,在抗体初始化阶段,采用基于聚类中心的实数编码方式,每个抗体 A_i 由 k 个聚类中心组成,它可表示为长度为 $k\times d$ 的实数编码.

$$A_i=\{\underbrace{a_{11}a_{12}\cdots a_{1d}}_{c_1}\cdots\underbrace{a_{i1}\cdots a_{id}}_{c_i}\cdots\underbrace{a_{k1}a_{k2}\cdots a_{kd}}_{c_k}\} \quad (3)$$

其中,$c_1,c_2,\cdots,c_k$ 所对应的编码分别为各聚类中心在样本空间中的坐标.

考虑到,免疫优势算子提取最优抗体的免疫优势的特殊方式,算法中,对抗体进行免疫优势的获取与推广操作时,相关抗体均转化为对应的 0 - 1 二进制编码.具体如下:

(1)计算规模为 N 的当前抗体群 $A(t)=\{A_1,A_2,\cdots,A_N\}$ 中所有抗体的亲和度 $Aff(t)=\{aff_1,aff_2,\cdots,aff_{N-1},aff_N\}$.

(2)对抗体亲和度 $Aff(t)$ 进行排序,对占抗体规模比例 η_1 的优秀抗体,进行二进制编码,按照抗体免疫优势算子的相关操作,从抗体群中比例为 η_1 的优秀抗体中提取免疫优势,生成参考抗体 $\boldsymbol{m}$.

(3)以概率 p_{id} 使抗体群中的其它抗体获得免疫优势,即满足概率 p_{id} 的抗体转化为0-1二进制抗体编码,

并执行获取免疫优势的操作.

(4)所有参与免疫优势算子相关操作的抗体,重新编码为基于聚类中心的实数编码,进入新的抗体群.

3.2 抗体亲和度

基于人工免疫系统的聚类算法,通过优化聚类有效性评价函数,搜索与评价函数最优值所对应的聚类中心来获得数据集的最优划分.在聚类研究领域,针对不同类型的数据集,已经提出了多种类型的聚类有效性评价函数[15].本文,我们采用一种简单误差平方和准则来定义抗体的亲和度.

抗体解码为对应的 k 个聚类中心,以矩阵 $\boldsymbol{V}$ 的形式表示,则对应于抗体 A_i 的聚类划分的误差平方和准则函数:

$$J(U,V)=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{m}D^2(\boldsymbol{c}_i,\boldsymbol{x}_j) \tag{4}$$

其中 $U=[u_{ij}]_{k\times n}$ 是模糊指数为 m 隶属度矩阵 $D^2(\boldsymbol{c}_i,\boldsymbol{x}_j)$ 表示 $\boldsymbol{c}_i$ 与 $\boldsymbol{x}_j$ 的欧氏距离,最小值 $J_{\min}=\min\{J(U,V)\}$,所确定的聚类中心对应于最优的聚类划分.基于误差平方和准则的抗体亲和度函数:

$$aff\ (A_i)=\frac{1}{1+J} \tag{5}$$

3.3 抗体免疫优势的获得

一方面,对求解问题可能一时难以形成可以利用的先验知识;另一方面,按照定义确定抗体免疫优势的代价太大,计算量为 $O(2^{l-1})$,从而使寻找抗体免疫优势的工作失去意义.一个明显的事实是,抗体群落中的最优抗体与其它抗体相比,更适应抗原的刺激,广义地讲,它具有免疫优势.因此,分析每次迭代中最优抗体的组成结构,可以从中获得免疫优势,使算法更具有通用性以及应用的便利性.由于没有依赖与问题有关的先验知识,保持了算法作为弱方法适应性强的特点.

基于这一思路,可以构造很多获得免疫优势的方法.针对采用二进制编码的抗体群落,可以采用如下的抗体免疫优势算子.

记抗体群 $\boldsymbol{A}=[\,a_1\quad a_2\quad\cdots\quad a_N\,]$ 为:

$$A=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1l}\\a_{21}&a_{22}&\cdots&a_{2l}\\\vdots&\vdots&\ddots&\vdots\\a_{N1}&a_{N2}&\cdots&a_{Nl}\end{bmatrix} \tag{6}$$

定义参考抗体 $\boldsymbol{m}=[\,m_1\quad m_2\quad\cdots\quad m_l\,]$,其中:

$$m_i=\begin{cases}1\ ,\ \dfrac{1}{N}\displaystyle\sum_{j=1}^{N}a_{ji}\geqslant 0.5\\0\ ,\ \text{others}\end{cases} \tag{7}$$

设,$a_i\in A$ 是抗体群中的最好抗体,如果 $f(m)>f(a_i)$,则互换.

以一定的概率 p_{id},使 $a_j\in\boldsymbol{A}$,$j=1,2,\cdots,l$ and $j\neq i$ 获得免疫优势.具体地,令:

$$a_j^{'}=H(a_j+a_i-m-1) \tag{8}$$

$H(\cdot)$是一门限函数:

$$H(x)=\begin{cases}1\ ,\ x>0\\0\ ,\ \text{others}\end{cases} \tag{9}$$

p_{id}是一个自适应调节的参数,如果 $f(a_j^{'})>f(a_j)$或 $f(m)>f(a_i)$,表明免疫优势的作用是有效的,则增大 p_{id},否则减小.

这种获得抗体免疫优势过程通过引入了先验知识和在线自适应动态获得先验知识的机制,实现了个体间的信息交换(不同于遗传算法的交叉操作),因此,更好地保证了抗体种群的多样性,进一步提高了算法的效率.

3.4 算法步骤及流程

每代进化结束后,根据亲和度大小,更新抗体群的当前最优抗体.若当前最优抗体连续 10 代无改进(改进小于 ε),停止迭代过程.算法步骤如图 1 所示.

Step 1 初始化抗体群 $A(0)=\{A_1,A_2,\cdots,A_i,\cdots,A_N\}$,设置各项参数及终止条件.

Step 2 计算抗体群 $A(0)$中各抗体亲和度.

Step 3 $t=t+1$,选取当前抗体群 $A(t)$中具有较大亲和度的,比例 η_1 优秀抗体作为提取免疫优势的抗体源,依据抗体源生成参考抗体 $\boldsymbol{m}$.

Step 4 进行获取免疫优势的操作,将参考抗体 $\boldsymbol{m}$ 中所具有的免疫优势依概率 p_{id}推广到其它抗体中.

Step 5 计算获取免疫优势后的抗体群 A_H 的亲和度,并对 A_H 中的抗体依次进行如下操作;

1)抗体克隆操作;

2)克隆变异操作;

3)FCM 一步迭代更新;

4)克隆选择操作;得到进化的抗体群 $A_S(t)$.

Step 6 更新当前最优抗体;若满足终止条件,算法停止迭代,解码最优抗体输出聚类划分.否则对抗体群中比例为 η_2 的具有较小亲和度的抗体,进行重新初始化,转 step 3.

图 1 基于免疫优势的克隆选择聚类算法步骤

4 仿真实验

为了测试基于免疫优势的克隆选择聚类算法的有效性,本节中,将新算法和模糊聚类算法(FCM)、基于遗传算法模糊聚类算法(GAFCM)以及及基于克隆选择的模糊聚类算法(CSAFCM)的聚类性能进行了综合比较.

实验中采用两个 UCI 数据集及三个多类别的合成数据集,所有算法用 Matlab 7.0,实现各算法均运行在 Intel(R) Core(TM)21.86GHz CPU,2GB 内存,Windows XP 操作系统的 PC 机上.

为了比较算法的聚类性能,我们采用上述 5 个数据集,分别对提出的算法性能进行了测试.FCM、GAFCM、

CSAFCM 及 IDCSCA 在各数据集分别独立运行 20 次,并统计出各自聚类结果的最大、最小、平均正确率以及 20 次获得最优解的次数.

四种算法的相关参数设置如下:

(1)基于免疫优势的克隆选择聚类算法(IDCSCA)参数设置:抗体群规模 $N=50$;克隆规模系数 $Nc=80$,克隆后的种群规模在 100 左右;克隆变异概率 $p_m=0.30$;获取免疫优势的概率控制参数 $\alpha=0.12$,其值越大,个体获取免疫优势的概率越大.克隆选择多样性控制参数 $\beta=0.16$,其值越大,选择后的抗体群,多样性越好.用于提取免疫优势的优秀抗体在抗体群中的比例 $\eta_1=0.05$ 最优抗体,连续三代无改进时,抗体群中较差抗体更新比例 $\eta_2=0.20$;模糊指数 $m=2.0$;停止准则参数 $\varepsilon=1e-5$.

(2)基于克隆选择的模糊聚类算法(CSAFCM)参数设置:抗体群规模 $N=50$;克隆规模系数 $Nc=80$;克隆后的种群规模在 100 左右;克隆变异概率 $p_m=0.30$;克隆重组概率 $p_c=0.80$;模糊指数 $m=2.0$;停止准则参数 $\varepsilon=1e-5$.

(3)基于遗传算法的模糊聚类(GAFCM)参数设置:种群规模 $N=100$;交叉概率 $p_c=0.80$;变异概率 $p_m=0.30$;模糊指数 $m=2.0$;停止准则参数 $\varepsilon=1e-5$.

(4)FCM 的参数设置:模糊指数 $m=2.0$;停止准则参数 $\varepsilon=1e-5$.

4.1　正确率与稳定性比较

FCM、GAFCM、CSAFCM 及 IDCSCA 四种算法,针对 Iris、WBC、ASD_10_2、ASD_11_2、ASD_12_2 以及 ASD_14_2 的 5 个多类别数据集的聚类结果统计在表 1.

表 1　四种算法的聚类结果统计(20 次运行)

数据集	聚类算法	平均正确率(%)(标准差)	最高正确率(%)	最低正确率(%)	平均目标函数值(标准差)	获得最优结果次数
Iris	FCM	89.33(0)	89.33	89.33	0.1604(0)	20
	GAFCM	89.33(0)	89.33	89.33	0.1604(0)	20
	CSAFCM	89.33(0)	89.33	89.33	0.1604(0)	20
	IDCSCA	89.33(0)	89.33	89.33	0.1604(0)	20
WBC	FCM	92.79(0)	92.79	92.79	0.0063(0)	20
	GAFCM	92.79(0)	92.79	92.79	0.0063(0)	20
	CSAFCM	92.79(0)	92.79	92.79	0.0063(0)	20
	IDCSCA	92.79(0)	92.79	92.79	0.0063(0)	20
ASD_11_2	FCM	85.25(0.0503)	92.23	69.51	0.4026(0.0111)	0
	GAFCM	92.43(0.0438)	99.60	87.40	0.4095 (0.0035)	4
	CSAFCM	97.55(0.0423)	99.61	89.32	0.4137(8.2078e-004)	16
	IDCSCA	99.10(0.0230)	99.61	89.32	0.4140(0.0013)	19
ASD_12_2	FCM	85.67(0.0517)	92.34	70.28	0.3980(0.0171)	0
	GAFCM	91.91(0.0297)	99.63	89.00	0.4125 (0.0029)	2
	CSAFCM	97.15(0.0440)	99.63	89.72	0.4168(8.9115e-004)	14
	IDCSCA	98.64(0.0305)	99.63	89.72	0.4171(6.1559e-004)	18
ASD_14_2	FCM	87.16(0.0324)	93.72	73.87	0.3884(0.0169)	0
	GAFCM	93.30(0.0289)	99.56	8686	0.4054(0.0031)	2
	CSAFCM	97.33(0.0369)	99.56	91.97	0.4091(0.0013)	12
	IDCSCA	98.07(0.0308)	99.56	92.12	0.4094(8.2078e-004)	18

从表 1 中可以看到,20 次独立运行中,对于 Iris、WBC,四算法的结果相同,20 次运行次次都获得最优的划分.

另外,从表 1 中可以看到 FCM 聚类算法对 ASD_11_2、ASD_12_2 以及 ASD_14_2 的聚类过程,均无法获得数据的最优聚类结果.可见,经典的 FCM 聚类算法已经无法成功地处理此类数据的聚类问题.同样对于 GAFCM 在处理 ASD_11_2、ASD_12_2 以及 ASD_14_2 时,20 次实验中大部分无法获得最优的聚类划分.

对于 ASD_11_2 数据,IDCSCA 仅有 1 次陷入局部最优解,CSAFCM 有 4 次无法得到最优的聚类划分.对于数据 ASD_12_2,CSAFCM 有 6 次无法得到最优的聚类划分,IDCSCA 有 2 次陷入局部最优解;对于数据 ASD_14_2,CSAFCM 有 8 次无法得到最优的聚类划分,IDCSCA 有 2 次陷入局部最优解.

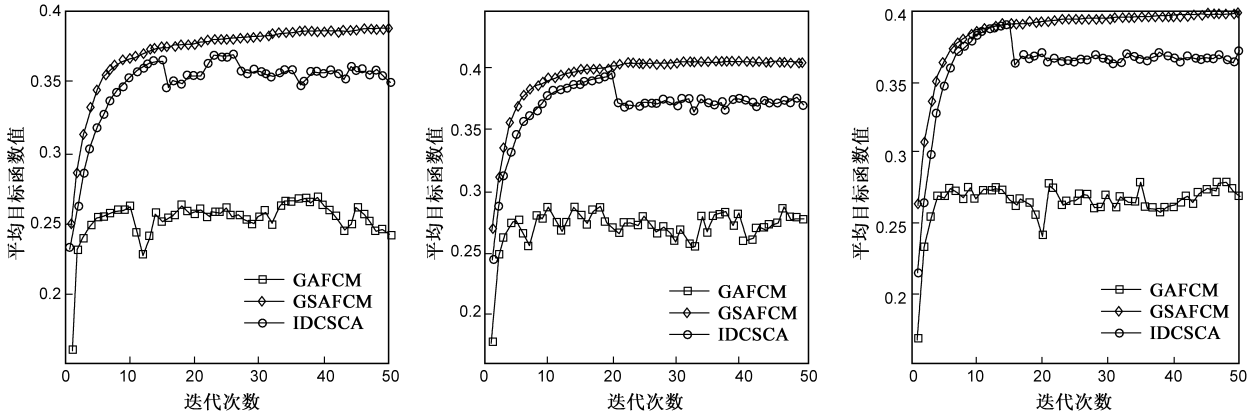
数据 ASD_11_2、ASD_12_2 和 ASD_14_2 随着数据集类别数的增加,以及数据规模的不断增大,聚类中心迭代过程中所面临的局部最优点,越来越多,在稳定性方面,两种算法都有不同程度的下降,但是,IDFCM 的聚类性能始终要优于 CSAFCM.而且,对这三个数据集,IDCSCA 在 20 次运行中无论是最大正确率、最小正确率以及平均正确率都要优于 CSAFCM,即使针对 14 类数据集 ASD_14_2 的聚类,20 次独立运行,IDCSCA 算法所得的平均聚类正确率仍优于 CSAFCM.从表 1 可以看出 IDCSCA 稳定性也较好.

4.2　抗体群多样性比较

本小节就数据 ASD_11_2、ASD_12_2 和 ASD_14_2 来比较 IDCSCA、CSAFCM 以及 GAFCM 的多样性,三算法

的参数设置如前所述,终止条件各自连续运行 50 代.三个算法的平均目标函数值与迭代次数间的关系如图 2.

从图 2 可以看出,GAFCM 在三个数据集的一次运行都未获得最优的划分,而 IDCSCA 和 CSAFCM 的随机一次优化聚类目标函数的结果都获得了最优,但从图 2 可以看出 IDCSCA 的平均目标函数值在迭代后期都比 CSAFCM 稍低,这也说明了采用了免疫优势算子的 IDCSCA 的多样性要比 CSAFCM 好,这一点也保证了对于样本分布不均匀的数据聚类,IDCSCA 不易陷入局部最优进而获得最优划分的稳定性.



(*a*) ASD_11_2 (*b*) ASD_12_2 (*c*) ASD_14_2

图2 算法抗体群的多样性比较

5 结论

本文将免疫优势理论运用于多类别数据的聚类问题,着重探讨了模糊聚类中抗体免疫优势的获得,并结合克隆选择机理构造了针对数据聚类的基于免疫优势的克隆选择聚类算法(IDCSCA).实验表明,由于该算法,一方面,继承了克隆选择算法的全局搜索和局部搜索相结合的特性;另一方面,通过引入免疫优势的概念,利用的搜索过程中获得的关于问题的先验知识,使新算法在解决多类别数据的聚类问题时,有效地避免了陷入局部最优的问题,提高了聚类算法的稳定性及可靠性.

参考文献:

[1] 刘健庄,谢维信.聚类分析的遗传算法方法.电子学报.1995,23(11):81 – 83.
Liu Jian-zhuang, Xie Wei-xin, Huang Jian-jun, Li Wen-hua. Clustering analysis by genetic algorithms[J]. Acta Electronica Sinica, 1995, 23(11): 81 – 83. (in Chinese)

[2] LO Hall, IB Ozyurt, JC Bezdek. Clustering with a genetically optimized approach[J]. IEEE Transactions on Evolutionary Computation. 1999, 3(7): 103 – 112.

[3] M Ujjwal, B Sanghamitra. Genetic algorithm-based clustering technique[J]. Pattern Recognition. 2000, 33(9): 1455 – 1465.

[4] L Tseng, S Yang. A genetic approach to the automatic clustering problem[J]. Pattern Recognition. 2001, 34(2): 415 – 424.

[5] M Cowgill, R Harvey, L Watson. A Genetic algorithm approach to cluster analysis[J]. Computers and Mathematics with Applications. 1999, 37(7): 99 – 108.

[6] 刘静,钟伟才.免疫进化聚类算法[J].电子学报.2001,29(12):1868 – 1872.
Liu jing, Zhong Wei-cai. A novel clustering based on the immune evolutionary algorithm[J]. Acta Electronica Sinica, 2001, 29(12): 1868 – 1872. (in Chinese)

[7] 行小帅,潘进,焦李成.基于免疫规划的 k-means 聚类算法[J].计算机学报.2003,26(5):605 – 610.
Xing Xiao-shuai, Pan Jin, Jiao Li-cheng. A novel K-means clustering based on the immune programming algorithm[J]. Chinese Journal of Compuetrs. 2003, 26(5): 605 – 610. (in Chinese)

[8] 李洁,高新波,焦李成.一种基于 CSA 的混和属性特征大数据集聚类算法[J].电子学报.2004,32(3):357 – 362.
Li Jie, Gao Xin-bao, Jiao Li-cheng. A CSA-based clustering algorithm for large data sets with mixed numeric and categorical values[J]. Acta Electronica Sinica. 2004, 32(3): 357 – 362. (in Chinese)

[9] 公茂果,焦李成,马文萍,等.基于流形距离的人工免疫无监督分类与识别算法[J].自动化学报.2008,34(3):367 – 375.
Gong Mao-guo, Jiao Li-cheng, Ma Wen-ping, Zhang Xiang-rong. Unsupervised classification and recognition using an artificial immune system based on manifold distance[J]. Acta Automatic Sinica, 2008, 34(3): 367 – 375. (in Chinese)

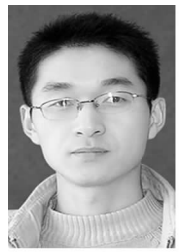
[10] C Ferreira. Gene expression programming: a new adaptive algorithm for solving problems[J]. Complex Systems. 2001, 13(2): 87 – 129.

[11] 李洁,高新波,焦李成.一种基于 CSA 的模糊聚类新算法[J],电子与信息学报. 2005,27(2):302 – 305.
Li Jie, Gao Xin-bo, Jiao Li-cheng, A CSA-based new fuzzy

clustering algorithm[J]. Journal of Electronics & Information Technology. 2005, 27(2): 302 – 305. (in Chinese)
[12] M F Burnet. A modification of Jerne's theory of antibody production using the concept of clonal selection[J]. Austrian Journal of Science. 1957, 20(1): 67 – 76.
[13] L N De Castro, F J Von Zuben. Artificial Immune Systems: Part I-Basic Theory and Applications[M]. RT DCA, Brazil: 95, 1999
[14] 杜海峰,焦李成,刘若辰. 免疫优势克隆算法[J]. 电子与信息学报. 2004, 26(12): 1918 – 1924.
Du Hai-feng, Jiao Li-cheng, Liu Ruo-chen. Immunodomaince clone algorithm[J]. Journal of Electronics & Information Technology. 2004, 26(12): 1918 – 1924. (in Chinese)
[15] U Bandyopadhyay, U Maulik. Nonparametric genetic clustering: comparison of validity indices[J]. IEEE Trans. Syst, Man, Cybern. C. 2001, 31(1): 120 – 125.

作者简介:

刘若辰　女, 1974 年出生于陕西武功, 博士, 副教授, 电子学会高级会员, 主要研究领域为免疫计算和数据挖掘.
E-Mail: ruocheenliu@yahoo.com.cn



沈正春　男, 1981 年生于河南信阳, 硕士, 主要研究领域为免疫计算和数据挖掘.
E-Mail: zhchshen@yahoo.com.cn

贾　建　男, 1977 年出生于陕西西安, 博士, 讲师, 主要从事模式识别, 图像处理等方面的研究.　E-mail: jiajianbb@126.com

焦李成　男, 1959 年出生于陕西白水, 教授, 博士导师, 电子学会高级会员, 现主要从事进化计算, 模式识别, 数据挖掘智能信息处理方面的研究.　E-Mail: lchjiao@mail.xidan.edu.cn

(上接第 982 页)

[7] 薛建中,闫相国,郑崇勋. 用核学习算法的意识任务特征提取与分类[J]. 电子学报, 2004, 32(10): 1749 – 1751.
Xue Jianzhong, Yan Xiangguo, Zheng Chongxu. Classifications of EEG during Mental Tasks by Kernel Learning Algorithms[J]. Acta Electronica Sinica, 2004, 32(10): 1749 – 1751. (in Chinese)
[8] 庞彦伟,俞能海,沈道义,刘政凯. 基于核邻域保持投影的人脸识别[J]. 电子学报, 2006, 34(8): 1542 – 1544.
Pang Yanwei, Yu Nenghai, Shen Daoyi. Kernel neighborhood preserving projections for face recognition[J]. Acta Electronica Sinica, 2006, 34(8): 1542 – 1544. (in Chinese)
[9] Turk M, Pentland A. Eigenfaces for recognition[J]. Journal of cognitive Neuroscience, 1991, 3(1): 71 – 86.
[10] Swets D L, Weng J Y. Using discriminant eigenfeatures for image retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(8): 831 – 836.
[11] Zheng W M, Zhao L. Foley-Sammon optimal discriminant vectors using kernel approach[J]. IEEE Transactions on neural networks, 2005, 16(1): 1 – 9.
[12] Phillips P J, Moon H, Rizvi S A, et al. The FERET Evaluation Methodology for face-Recognition Algorithms[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(10): 1090 – 1104.