

一种融合人工免疫系统与 AP 算法的分类器设计

储岳中^{1,2} 徐 波³ 高有涛¹

(1. 南京航空航天大学航天学院, 南京, 210016; 2. 安徽工业大学计算机学院, 马鞍山, 243002;

3. 南京大学天文与空间科学学院, 南京, 210093)

摘要:为提高复杂数据分类器的分类性能, 结合人工免疫系统(Artificial immune system, AIS)的自适应识别能力与全局搜索能力, 以及近邻传播(Affinity propagation, AP)算法自动确定最佳数据类数的能力, 提出了一种基于人工免疫系统与近邻传播相结合的分类算法。通过自适应免疫算法, 获得反映数据集模式特征的抗体记忆集, 然后再利用基于聚类有效性指标的 AP 算法确定抗体记忆集的最佳聚类数, 以此构造分类器。最后, 通过人工数据集和 UCI 基准数据集来测试该分类器。实验结果表明, 与直接采用免疫算法和 AP 算法相比, 该算法在分类正确率和识别性能方面均有良好的表现。在与一些经典分类算法的对比实验中, 本文算法也表现出较好的竞争力。

关键词:人工免疫系统; 近邻传播算法; K 近邻算法; 分类

中图分类号: TP18 **文献标志码:** A **文章编号:** 1005-2615(2013)02-0232-07

Design of Classifier Based on Combination of Artificial Immune System and AP Algorithm

Chu Yuezong^{1,2}, Xu Bo³, Gao Youtao¹

(1. College of Astronautics, Nanjing University of Aeronautics & Astronautics, Nanjing, 210016, China;

2. School of Computer Science, Anhui University of Technology, Maanshan, 243002, China;

3. School of Astronomy & Space Science, Nanjing University, Nanjing, 210093, China)

Abstract: To improve the classification performance of complex data classifier, a classification algorithm based on the combination of artificial immune system (AIS) and affinity propagation(AP) algorithm is proposed which combines the global search capability and adaptive recognition of AIS with the capability of automatically determining the optimal number of data classes of AP. The antibody memory set reflecting the characteristics of data set is obtained by adaptive immune algorithm. Then the optimal number of clusters of the antibody memory set is obtained using AP algorithm based on cluster validity index, and a classifier is constructed. Finally, the proposed algorithm is tested on artificial data set and UCI data set. Compared with artificial immune algorithm and AP algorithm, it has a great competitiveness on recognition rate and recognition capability. The algorithm also shows good competitiveness compared with some of the classic classification algorithms.

Key words: artificial immune system; affinity propagation algorithm; K-nearest neighbor algorithm; classification

分类是找出描述并区分数据类或概念的模式, 以便能够使用模型预测类并标记未知的对象类。

基金项目:国家自然科学基金(11078001)资助项目; 国家高技术研究发展计划(“八六三”计划)(2012AA121602)资助项目。

收稿日期: 2012-04-23; **修订日期:** 2012-10-18

通信作者: 储岳中, 男, 博士, 副教授, 1971 年生, E-mail: mychu@126.com。

分类技术从机器学习的角度看,属于监督学习技术。人工免疫识别系统在监督学习领域已有许多应用。Carter^[1]提出了一种称为 Immunos-81 的监督学习模型,为了识别未知数据,利用 T 细胞、B 细胞,抗体和它们之间的相互作用等多种机制进行竞争,当产生与数据有最高亲和力的克隆时发生识别。Castro 等^[2]设计了一种免疫启发的模糊分类集成系统,用基于克隆选择、超变异和免疫网络原理的免疫算法进化模糊规则,一旦免疫算法找到问题的多个良好解,则集成这些模糊系统,以得到更高的分类准确率。Sahan 等^[3]提出了基于模糊人工免疫系统和 K 近邻混合方法的分类系统,用于癌症诊断,通过十重交叉验证方法,获得了较高的诊断准确率。莫宏伟等^[4]利用免疫网络产生的记忆细胞,得到基于人工免疫网络记忆分类器,与传统算法相比,表现出一定的竞争力。由此可看出,人工免疫识别系统已获得较为成功的应用,但它仍存在记忆细胞数目多、不易确定类数及分类性能不是很高等缺陷。为了克服这些缺陷,针对无监督模式识别问题,本文在人工免疫识别系统的基础上,利用近邻传播算法自动确定数据集的最佳聚类数,以此来设计分类器。通过人工数据集和 UCI 数据集的仿真实验,并与标准 AP 算法、人工免疫算法及一些经典改进算法进行了比较,结果表明这种基于混合算法的分类器是有效的。

1 自适应免疫网络

人工免疫系统是基于免疫机制和理论免疫学而发展的各种人工范例的统称。2000 年,de Castro 等^[5]提出了人工免疫网络的概念,其主要功能是用一个小规模的内镜像记忆网络映射源输入数据集,从而达到将数据压缩的目的,同时从数据集合抽取相关特征信息。人工免疫网络一般由抗原、抗体和记忆细胞组成。其中:抗原表示待分析的数据集合;抗体为各模式类的特征值,在网络中表现为一个数据项;而记忆细胞是算法产生的逼近原始数据集并带有模式特征的数据,由它构成记忆网络。本文将待处理的数据集定义为抗原集合 $AG = \{ag_1, ag_2, \dots, ag_l\}$, l 为抗原空间规模。通过免疫应答识别过程,获得的记忆细胞集合 M 即为需要求解的模式。设 M 的规模为 I' , 则免疫应答结束后,应有 $I > I'$ 。设系统中的抗体表示为 $AB = \{ab_1, ab_2, \dots, ab_k\}$, k 为初始抗体空间的规

模,显然 $AB \subset AG$ 。

免疫空间中样本间的相互作用称为免疫亲和力,大多文献采用欧氏距离的倒数计算亲和力,这种方式只有当样本的自然分布接近球状时,即每个类的各个分量的方差接近相等时,才可能有较好的效果。本文亲和力的计算在欧氏距离的基础上改进如下^[6]:设 $V_i = \{v_{i1}, v_{i2}, \dots, v_{iL}\}$, $V_j = \{v_{j1}, v_{j2}, \dots, v_{jL}\}$, 分别表示免疫空间中的两个个体 (L 为向量维数), 则其亲和力为 $f_{ij} = \left(1 + \sum_{k=1}^L \frac{(v_{ik} - v_{jk})^2}{v_{ik}^2 + v_{jk}^2}\right)^{-\frac{1}{2}}$ 。对于这种亲和力计算方式,当样本不接近球形分布时,由于 $\frac{1}{v_{ik}^2 + v_{jk}^2}$ 的作用,那些 $(v_{ik} - v_{jk})^2$ 较大的累加项在公式中的权重有所减少,相对使用普通欧氏距离的亲和力计算方式,这会提高算法的适用性。

免疫网络通过迭代循环,对所处理的数据进行学习,直到满足学习终止条件(本文采用设定迭代次数),最后的抗体记忆集 M 为获得的学习模式,待分析数据集的模式特征完全包含于抗体记忆集。常规免疫算法可利用抗体记忆集对测试数据(新抗原)进行二次应答,以识别测试数据的模式。当这种免疫学习的结果用于分类问题时,免疫网络并没有确定抗体记忆集中抗体的类别,因此还达不到自动分类的目的,为此引入基于有效性评估指标的近邻传播算法来自动确定抗体记忆集的最佳类别。

2 近邻传播聚类算法

近邻传播(Affinity propagation, AP)聚类算法是 Frey 等^[7]在 Science 杂志上提出的一种新的聚类算法。它是根据 N 个数据点之间的相似度矩阵 S 进行聚类。AP 算法不需要事先指定聚类数目,相反它将所有数据点都作为潜在的聚类中心,称之为 exemplar。相比一些传统聚类算法,它能更快地处理大规模数据,得到较好的聚类结果。

聚类时以 S 矩阵的对角线上数值 $s(k, k)$ 作为 k 点能否成为聚类中心的评判标准,该值越大,这个点成为聚类中心的可能性也就越大,这里将此值称为偏向参数 p 。聚类开始时, p 一般取输入相似度的均值。AP 算法为找出合适的聚类中心而不断进行迭代并更新、传递两种类型的消息^[8]: $r(i, k)$ 表示从点 i 发送到候选聚类中心点 k 的数值消息,反映 k 点是否适合作为 i 点的聚类中心; $a(i, k)$ 则从候选聚类中心点 k 发送到点 i 的数值消息,反映 i 点是否选择 k 点作为其聚类中心。AP 算法

通过迭代过程不断更新每一个点的吸引度和归属度值,直到产生 m 个高质量的类代表点,同时将其余的数据点分配到相应的聚类中。另外,为了防止 AP 算法发生振荡而不能收敛,可引入阻尼因子 λ ($\lambda \in [0, 1)$),每一次循环迭代过程中, $r(i, k)$ 和 $a(i, k)$ 的更新结果均由当前迭代过程中更新的值和前一步迭代的结果加权得到,增大阻尼因子可以消除振荡但同时也会稍微延长收敛时间。本文实验中,考虑对收敛时间影响不大的情况下,设置经验值 $\lambda = 0.9$ 。

AP 算法基本过程如下^[9]:先计算 N 个数据点的相似度矩阵 S ,再选取 p 值(一般取 S 的中值)作为 S 的对角线元素。设置一个最大迭代次数,每一次迭代中,分别计算代表矩阵 R 和适选矩阵 A 的值,并考虑阻尼因子 λ 的作用,当迭代次数超过最大值时终止计算。然后根据 $R(k, k) + A(k, k)$ 的值来判断是否为聚类中心(如果采用负欧氏距离进行相似度测度,当 $R(k, k) + A(k, k) > 0$ 时,可认为是数据点 k 为类代表点,即聚类中心)。如果聚类结果不满足要求,则调整 p 值重新进行聚类。

因此,如何找到一合适的 p 值,来确定数据集的最佳聚类数,这是一般 AP 算法要面对的一个问题。目前,有很多有效性指标可用于分析聚类结果并确定最佳聚类数,其中评价性能较好的指标主要有 Krzanowski-Lai (KL) 指标^[10], Homogeneity-Separation (HS) 指标^[11] 和 In-Group Proportion (IGP) 指标^[12] 等,但这些指标由于自身的一些不足,对于聚类结构难以判别的数据集,其有效性判别往往不够理想,难以获得最佳聚类数。由于 AP 算法一般以反映样本几何结构的欧氏距离作为相似度测度指标,故本文也选择基于样本几何结构的有效性指标来评估 AP 算法的聚类结果^[13],并以此来确定样本的最佳聚类数。

令 $X = \{x_1, x_2, \dots, x_n\}$ 为待处理数据集,设 n 个样本通过算法被聚类成 c 类。定义第 j 个类的第 i 个样本的最小类间距离为 $bd(j, i)$,其值为该样本到其他每个类中样本平均距离的最小值。定义第 j 个类的第 i 个样本的类内距离为 $wd(j, i)$,其值为该样本到本类所有样本的平均距离。定义第 j 个类的第 i 个样本的聚类距离为 $bawd(j, i)$,其值为该样本的最小类间距离与类内距离之和,即 $bawd(j, i) = bd(j, i) + wd(j, i)$ 。定义第 j 个类的第 i 个样本的聚类离差距离为 $bswd(j, i)$,其值为该样本的最小类间距离与类内距离之差,即 $bswd(j, i) = bd(j, i) - wd(j, i)$ 。定义第 j 个类的

第 i 个样本的类间类内划分指标为 $BWP(j, i)$,其值为该样本的聚类离差距离与聚类距离的比值,即

$$BWP(j, i) = \frac{bswd(j, i)}{bawd(j, i)} = \frac{bd(j, i) - wd(j, i)}{bd(j, i) + wd(j, i)} \tag{1}$$

BWP 指标描述的是单个样本的聚类有效性情况,其值越大,说明这个样本的聚类效果越好。为了综合考虑数据集中所有样本的情况,一般以所有样本的 BWP 指标的平均值作为评价依据。显然,平均值越大,说明算法对该数据集的聚类效果越好,相应的,最大值所对应的聚类数也即为最佳聚类数。综上所述,可以利用式(2, 3)来确定最佳聚类数。

$$\text{AverageBWP}(k) = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} BWP(j, i) \tag{2}$$

$$k_{\text{opt}} = \arg \max_{2 \leq k \leq \sqrt{n}} \{ \text{AverageBWP}(k) \} \tag{3}$$

式中: n 为数据集的样本总数; $\text{AverageBWP}(k)$ 表示数据集被算法聚类成 k 类时的平均 BWP 指标值; k_{opt} 表示最佳聚类数。需要说明的是,AP 算法不能指定类数 k ,如果需要获得 k 个类的聚类结果,一般采用搜索的方法,算法中可通过改变偏向参数 p 的值来实现。

3 分类器算法

3.1 算法设计

本文融合人工免疫算法与 AP 算法来设计分类器,通过自适应免疫算法,获得反映数据集的模式特征的抗体记忆集 M ,然后再利用基于聚类有效性指标的 AP 算法来确定抗体记忆集的最佳聚类数,以此构造分类器。为便于描述,将本方法简称为 AIS-AP。算法描述如下:

- 步骤 1 初始化,将待处理数据集 AG 归一化并随机产生初始一定数量抗体集合 AB 。
- 步骤 2 判断免疫学习终止条件是否满足。若满足则转步骤 7,不满足转步骤 3。
- 步骤 3 对每个样本 $Ag_i \in AG$,计算亲和力,对抗体集合,计算刺激度。
- 步骤 4 对所有抗体进行克隆增殖与变异进化。
- 步骤 5 对抗体网络进行克隆抑制,然后从抗体集合中按比例 ξ 选择与抗原亲和力最高的抗体形成抗体记忆集合 M 。
- 步骤 6 完成抗体亲和力成熟过程,转步骤 2。
- 步骤 7 设抗体记忆集合 M 的规模为 n ,确定

聚类类别数的范围为 $2 \leq k < \sqrt{n}$, 利用式(2,3)来搜索最大 BWP 平均指标值时的 k 值, 并以此标注抗体记忆集合中各抗体的类别。

步骤 8 对测试数据集, 由已标注类别的抗体记忆集合表示分类器, 采用 K 近邻算法, 由预测模式匹配最好的 K 个记忆细胞确定, 预测的类根据多数表决定。

步骤 9 对测试数据集进行分类正确率评估。

3.2 算法的计算复杂度分析

算法的时间开销由自适应免疫算法和近邻传播算法先后占用。设数据集的训练样本数为 m , 免疫算法的迭代次数为 k , 则自适应免疫算法的时间复杂度为 $O(km^2)$ 。设抗体记忆集合 M 的规模为 n , 通过算法被聚类成 c 个类, 则每个类的平均样本数为 n/c , 故类数确定时, 计算平均 BWP 指标值的时间复杂度为 $O(\frac{n^2}{c^2})$, 而类数的变化范围 $2 \leq c < \sqrt{n}$, 同时 $n < m$ 。综合起来, 对类数的寻优时间不会超过免疫计算部分的时间, 故整个算法的时间复杂度应为 $O(km^2)$ 。由于免疫算法是基于抗原与抗体间的亲和度计算, 其存储矩阵由数据集的规模决定, 故算法的空间复杂度为 $O(m^2)$ 。

4 仿真实验与结果分析

为了检验本文所提算法的有效性与分类性能, 本文通过 3 组实验共 5 个数据集进行测试, 并与 AIRS 算法、AP 算法和一些经典分类算法进行比较。实验中将待处理数据集分为两部分: 训练抗原和测试抗原, 采用十重交叉验证法来检验算法的分类性能。

实验 1 人工数据集实验。为直观地显示实验结果, 使用平面坐标上的点模拟具有 5 个簇的人工二维数据集, 它的结构特征为轻微重叠, 结构松散, 其分布如图 1 所示, 样本横坐标范围 $[0, 14]$, 纵坐标范围 $[0, 12]$, 样本总数为 1 100, 噪声样本数为 60。训练时, 随机选择 800 个样本作为训练集, 其余样本作为测试集。免疫网络训练时的几个常系数取值为 $k_d = 1, k_a = 1, k_s = 1, k'_s = 1$, 算法中偏向参数 p 的变化区间为 $[-100 \quad -0.1]$, 步长为 0.1。样本选择参数 $\xi = 50\%$, 网络采用迭代次数为 50 作为训练终止条件, 某一次训练结束后形成的记忆抗体集合样本数为 336, 如图 2 所示。因此, 聚类类别数的范围为 $2 \leq n \leq 18$ 。对于正确类

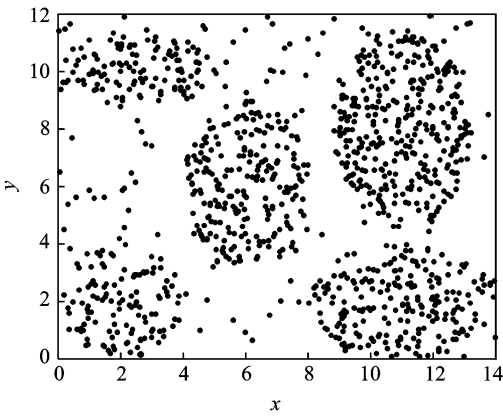


图 1 人工数据集

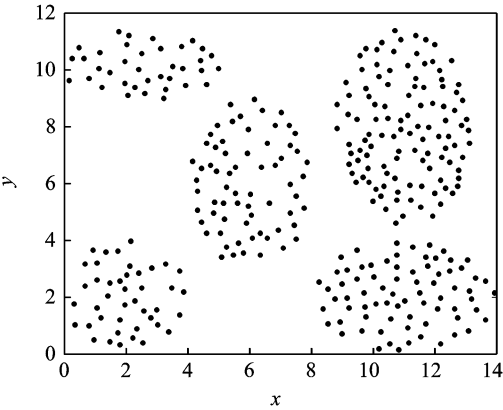


图 2 人工数据集的记忆抗体集合

数为 5 类的人工数据集, 此时运用 BWP 指标确定最佳聚类数的实验结果如图 3 所示。从中可以看出, 最高 BWP 指标得到的最佳聚类 5 是正确的。然后取参数 $K = 15$, 按记忆抗体集合构造 K 近邻分类器对测试样本进行验证, 采用十重交叉验证法, 测试结果与直接采用 AIRS 分类器和 AP 分类器的比较如表 1 所示。从表 1 中结果可看出, 本文算法(AIS-AP)对这类特征的人工数据集体现出较好的分类能力, 同时具备一定的去噪能力。

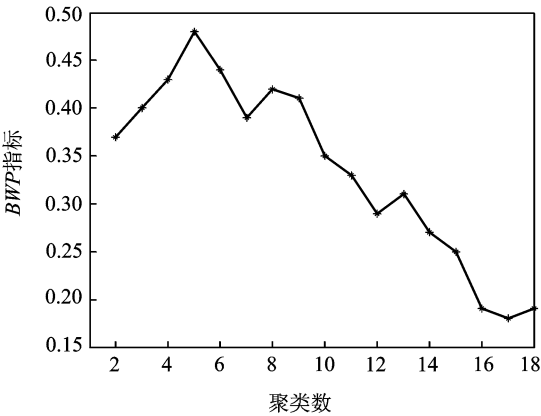


图 3 人工数据集的聚类数- BWP 指标关系图

表 1 不同方法对人工数据集的分类能力及性能指标

实验方法	样本数	训练样本数	测试样本数	平均记忆细胞数	偏向参数 p	K 近邻分类 K 值	平均分类正确率/%
AIS-AP	1 100	800	300	321	-11.6	15	99.37
AIRS	1 100	800	300	347		14	95.06
AP	1 100	800	300		-17.2	15	96.41

实验 2 UCI 基准数据集实验。在 UCI 机器学习库中选择 4 个广泛用于测试分类算法性能的数据集,分别是 Iris, Breast-cancer-wisconsin (Bcw), Wine 和 Glass,各数据集的特征描述如表 2 所示。在训练前,所有数据均被标度到区间 $[0,1]$ 。免疫网络训练时的几个常系数设置为 $k_d=1,k_d=1,k_s=1,k'_s=1$,算法中偏向参数 p 的变化区间为 $[-200\ -0.1]$,步长为 0.1。采用十重交叉验证法对 4 个数据集分别利用 AIS-AP 分类器、AIRS 分类器和 AP 分类器的分类性能比较如表 3 所示。从表 3 中结果可看出,本文所提算法不仅能够获得

表 2 UCI 基准数据集的特征描述

数据集	聚类结构特征	类数	样本数	维数
Iris	靠近,松散	3	150	4
Bcw	完全分离,松散	2	699	9
Wine	轻微重叠,松散	3	178	13
Glass	部分重叠,松散	6	214	9

相对较高的分类精度,平均记忆细胞数相对要少,同时网络的迭代次数也是最少的。对于正确类数为 3 类的 Iris 数据集,此时运用 BWP 指标确定最佳聚类数的实验结果如图 4 所示。从中可以看出,最高 BWP 指标得到的最佳聚类 3 是正确的。

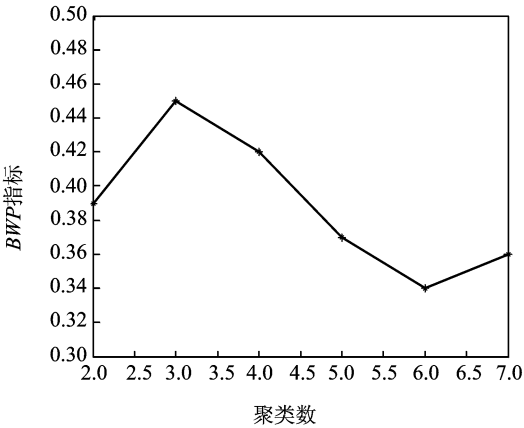


图 4 Iris 数据集的聚类数-BWP 指标关系图

表 3 对 UCI 基准数据集的分类结果

数据集	方法	训练样本	测试样本	样本选择参数 ξ /%	网络迭代次数	平均记忆细胞数	偏向参数 p	K 近邻分类 K 值	平均分类精度/%
Iris	AIS-AP	120	30	50	80	51	-18.1	12	98.67
	AIRS	120	30	60	100	88		12	95.33
	AP	120	30		500		-9.4	16	92.19
Bcw	AIS-AP	559	140	60	60	211	-21.8	10	97.38
	AIRS	559	140	55	150	239		14	94.82
	AP	559	140		400		-26.4	15	93.55
Wine	AIS-AP	142	36	50	40	58	-14.2	14	94.43
	AIRS	142	36	50	100	101		15	90.75
	AP	142	36		250		-16.9	12	92.22
Glass	AIS-AP	171	43	50	100	71	-8.7	19	86.96
	AIRS	171	43	60	300	126		21	80.54
	AP	171	43		250		-10.3	17	81.77

实验 3 与经典分类算法对比实验。为进一步验证本文所提算法的有效性,仍然采用表 2 所示的基准数据集,选择人工免疫网络分类算法(Artificial immune network classifier, AINC)^[14]、克隆选择算法(Clonal selection algorithm, CSA)^[15]以及多值免疫网络算法(Multiple-valued immune

network, MVIN)^[16]等经典分类算法进行比较实验。各算法中的参数设置如下,本文算法(AIS-AP):仍选择实验 2 中的参数。AINC:变异率 $p=0.5$,克隆规模 $p_{\text{clone}}=10$,网络的增殖规模 $p_g=10$,最大迭代次数为 400。CSA:采用实数进行编码,变异率 $p=0.6$,克隆规模 $p_{\text{clone}}=12$,初始抗体数等

于样本数,最大迭代次数设为 400。MVIN:采用实数进行编码,初始记忆细胞从每个类训练样本中随机选取,每个类仅有一个记忆细胞,耐受值取 0.05。采用十重交叉验证法对 4 个数据集分别利用 4 种不同算法的分类性能比较如表 4 所示。

从表 4 中可看出,对这 4 个基准数据集,在分类精度上,本文算法除了在 Wine 数据集上比 AINC 和 CSA 稍低点外,相比其他 3 个数据集,本文算法都有明显的提高。对于 Glass 数据集,AINC 和 CSA 的分类精度很不理想,而本文算法可以达到 86.96%。在标准差方面,本文算法的稳定性比 CSA 稍差,但比 AINC 和 MVIN 明显要好。从图 5 可看出,AIS-AP 算法在 Iris 数据集 10

次实验中得到的分类精度曲线平稳,波动较小,说明算法的分类稳定性较其他算法好。

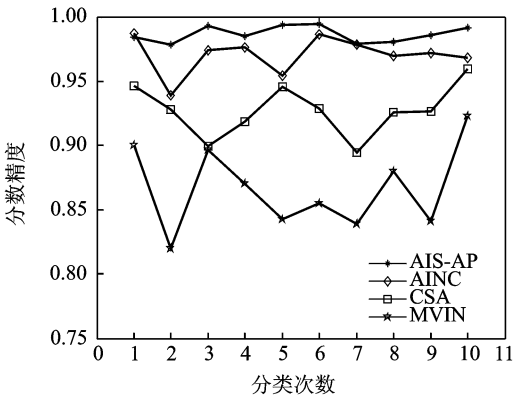


图 5 Iris 数据集十重实验的分类精度

表 4 对比实验结果

数据集	方法	平均分类精度/%	标准差	数据集	方法	平均分类精度/%	标准差
Iris	AIS-AP	98.67	1.42	Wine	AIS-AP	95.43	2.77
	AINC	97.06	1.54		AINC	96.04	2.55
	CSA	92.74	0.76		CSA	96.62	0.81
	MVIN	86.67	3.39		MVIN	92.44	4.85
Bcw	AIS-AP	97.38	1.27	Glass	AIS-AP	86.96	8.74
	AINC	96.66	3.35		AINC	70.51	22.81
	CSA	94.83	0.61		CSA	71.28	7.53
	MVIN	95.17	12.16		MVIN	80.49	26.31

5 结束语

常规人工免疫系统在处理分类问题时,能够快速提取数据集的分布特征,但不能有效确定数据集的类别数。而基于 BWP 指标的 AP 聚类算法在不需初始化聚类中心的前提下,能够在较短时间里得到很好的聚类结果。为此,本文提出了一种基于人工免疫系统与 AP 算法相结合的分类算法。实验结果表明,相对直接采用 Iris 算法和 AP 算法进行分类,本文算法在分类正确率和识别性能方面具备一定的优势。但是本文算法在处理一些结构复杂的大规模数据集时,算法的计算复杂度改进并不明显,这将是下一步的工作重点。

参考文献:

[1] Carter J H. The immune system as a model for pattern recongnition and classifiication[J]. Journal of the American Medical Informatics Association,2000, 7(3):28-41.

[2] Castro P D, Coelho G P, Caetano M F, et al. Designing ensembles of fuzzy classification systems: an

immune-inspired approach[C]//International Conference on Artificial Immune Systems, Canada: Springer,2005:462-482.

[3] Sahan S, Polat K, Kodaz H, et al. A new hybrid method based on fuzzy-artificial immune system and K-NN algorithm for breast cancer diagnosis [J]. Computers in Biology and Medicine,2006,37(3):415-423.

[4] 莫宏伟,吕淑萍,管凤旭,等. 基于人工免疫网络记忆的新型分类器研究[J]. 计算机工程与应用,2004,40(36):28-32.

Mo Hongwei , Lv Shuping , Guan Fengxu, et al. The research of new classifier based on artificial immune network memory [J]. Computer Engineering and Application, 2004,40(36):28-32.

[5] Nunes de Castro L, Von Zuben J F. An evolutionary immune network for data clustering[C]//Proceedings of the IEEE SBRN. [S.l.]:IEEE,2000:84-89.

[6] 黎昱,黄席樾,周欣. 基于免疫聚类与 HMM 的时序信息系统决策[J]. 信息与控制,2003,32(5):385-390.

Li Yu, Huang Xiyue, Zhou Xin. Decision-making in time-series information system based on immune

- clustering and HMM[J]. Information and Control, 2003,32(5):385-390.
- [7] Frey B J, Dueck D. Clustering by passing messages between data points[J]. Science, 2007, 315(5814): 972-976.
- [8] 肖宇, 于剑. 基于近邻传播算法的半监督聚类[J]. 软件学报, 2008, 19(11): 2803-2813.
- Xiao Yu, Yu Jian. Semi-supervised clustering based on affinity propagation algorithm[J]. Journal of Software, 2008, 19(11): 2803-2813.
- [9] 储岳中, 徐波. 基于流形分析与 AP 算法 RBF 神经网络分类器[J]. 华中科技大学学报: 自然科学版, 2012, 40(8): 93-97.
- Chu Yuezhong, Xu Bo. RBF neural network classifier based on manifold analysis and AP algorithm[J]. Journal of Huazhong University of Science and Technology: Natural Science Edition, 2012, 40(8): 93-97.
- [10] Dudoit S, Fridlyand J. A prediction-based resampling method for estimating the number of clusters in a dataset[J]. Genome Biology, 2002, 3(7): 1-21.
- [11] Chen G, Jaradat S A, Banerjee N, et al. Evaluation and comparison of clustering algorithms in analyzing ES cell gene expression data[J]. Statistica Sinica, 2002, 12(1): 241-262.
- [12] Kapp A V, Tibshirani R. Are clusters found in one dataset present in another dataset? [J]. Biostatistics, 2007, 8(1): 9-31.
- [13] 周世兵, 徐振源, 唐旭清. 一种基于近邻传播算法的最佳聚类数确定方法[J]. 控制与决策, 2011, 26(8): 1147-1152.
- Zhou Shibing, Xu Zhenyuan, Tang Xuqing. Method for determining optimal number of clusters based on affinity propagation clustering[J]. Control and Decision, 2011, 26(8): 1147-1152.
- [14] 刘若辰, 钮满春, 焦李成. 一种新的人工免疫网络算法及其在复杂数据分类中的应用[J]. 电子与信息学报, 2010, 32(3): 515-521.
- Liu Ruochen, Niu Manchun, Jiao Licheng. A new artificial immune network algorithm for classifying complex data[J]. Journal of Electronics & Information Technology, 2010, 32(3): 515-521.
- [15] Du Haifeng, Jiao Licheng, Wang Sunan. Clonal operator and antibody clone algorithms[C]//Proceedings of the First International Conference on Machine Learning and Cybernetics. Beijing: [s. n.], 2002: 506-510.
- [16] 钟燕飞, 张良培, 李平湘. 基于多值免疫网络的多光谱遥感影像分类[J]. 计算机学报, 2007, 30(12): 2181-2188.
- Zhong Yanfei, Zhang Liangpei, Li Pingxiang. Classification of multi-spectral remote sensing image based on multiple-valued immune network [J]. Chinese Journal of Computers, 2007, 30(12): 2181-2188.

