核密度估计在分类问题中带宽参数的优化研究

李泽中 白 勇
(重庆电力高等专科学校计算机系 重庆 400053)

摘 要 核密度估计可用于贝叶斯分类器类条件概率密度估计,其关键是带宽参数的确定。为此,提出了通过使受试者特征(ROC)曲线下的面积AUC最大而优化带宽参数的方法,建立了用于二进制特征的BKD分类方法和用于连续值特征的CKD分类方法。将这两种方法分别用于UCI数据集Promoter和Diabetics,得到的预测准确率与文献报道最佳结果接近,表明提出的带宽参数优化方法用于核密度分类具有较好的分类预测能力。

关键词 贝叶斯分类器,核密度估计,带宽参数,受试者特征曲线

Study on the Optimization of the Bandwidth of the Kernel Function for Kernel Discrimination

LI Ze-zhong BAI Yong
(Department of Computer, Chongqing Electric Power College, Chongqing 400053, China)

Abstract Kernel density estimation can be used to estimate the conditional class probability density functions and the key problem is to choose the bandwidth of the kernel. In this study, a method that the bandwidth is optimized through maximizing the area under the Receiver Operating Characteristic curve was proposed. Two classification systems called BKD and CKD, which are applicable to binary features and continuous features, respectively, were developed. The prediction accuracies for UCI Promoter dataset and UCI Diabetics dataset by the two classification systems are close to the best literature reported results, showing that application of our proposed method for the optimization of the bandwidth to kernel discrimination has a good prediction ability.

Keywords Bayes classifier, Kernel density estimation, Bandwidth parameter, ROC curve

1 引言

基于数据的机器学习是现代智能技术中的重要方面,研究从观测数据出发寻找统计规律,并利用这些规律对未来数据或无法观测的数据进行预测,是机器学习的主要任务。机器学习问题就是根据 n 个独立同分布观测样本$\{(X_i,y_i),i=1,2,3,\cdots,n\}$,对 y 与 X 之间的依赖关系进行估计[1]。机器学习的根本目的是对未知样本从其观测性质 X 预测响应性质 y,按响应性质是离散的类别变量还是连续变量可将机器学习方法划分为分类问题和回归问题。分类机器学习方法又称为模式识别。现今人们已提出了许多的分类机器学习方法,如支撑矢量机(SVM)、人工神经网络(ANN)、决策树(DT)、逻辑回归(LR)、贝叶斯分类器(Bayes classifier)、k-最近邻(k-NN)等。机器学习已广泛的用于如人脸识别、生物信息学中蛋白质功能和结构的预测、计算机辅助药物设计、入侵检测等各个领域,已成为一些交叉学科的重要支撑技术。之所以人们提出了各种各样的机器学习方法,是因为按"没有免费午餐原理",不存在一种机器学习方法在所有问题上均优于其它机器学习方法。

在众多分类方法中,贝叶斯分类器由于具有坚实的数学理论基础及综合先验信息和数据样本信息的能力,而且简单有效,所以得到了广泛的应用。在贝叶斯分类器的应用中,需要确定先验概率和类条件概率密度,类条件概率密度是贝叶斯方法的一个重要问题。对于离散化的 x 变量可以假定每一变量分量相互独立,根据多项分布估计每一变量分量的类条件概率密度,并采用朴素贝叶斯分类器。对连续的 x 变量,主要有3种处理方法。第一种方法是先对每一变量分量离散化,再用朴素贝叶斯分类器。由于连续变量离散化方法很多,其结果直接决定了分类结果的好坏[2]。而且,朴素贝叶斯分类器基于"独立性假设"前提,而现实世界中,这种独立性假设经常不满足,因此影响了朴素贝叶斯分类器的分类精确度。第二种方法是参数化估计法,即假定样本密度服从于某种分布,如Gauss密度,再根据样本训练集估计分布函数的参数。而在实际应用中,样本密度分布很难预先知道,因而采用假定分布往往得不到理想的结果。因此,现在更普遍采用的方法是第三种方法,即非参数密度估计,如核密度估计(Kernel Density Estimation)[3]。核密度估计方法中最关键的一个问题是对带宽参数的估计,AITCHISON 和 AITKEN[4]提出了留一法似然函数(Leave-one-out likelihood method)估计带宽参数。最近,Harper[5]及 Willett 等人[6-10]将核密度估计用于计算机辅助药物筛选,得到令人满意的结果。他们提出了在核密度估计用于筛选问题时,以正样本的排名顺序之和为最小,优化带宽参数。本文针对筛选问题,提出了以受试者特征曲线(ROC)下面积为目标函数优化带宽参数的方法。

到稿日期:2008-12-03 返修日期:2009-03-18

李泽中(1955-),男,副教授,主要研究方向为计算机应用等,E-mail:lizz6@163.com;白 勇(1973-),男,副教授,主要研究方向为软件理论与技术。

2 基本原理

2.1 贝叶斯分类器与核密度估计

设有训练集$\{(X_i,Y_i),i=1,2,3,\cdots,n\}$共$n$个样本，其中$X_i\in R^m$为$m$维观察矢量，$Y_i\in(1,2,3,\cdots,K)$为$X_i$的类别标识。根据贝叶斯原理，样本$X$属于第$K$类的概率为：

$$P(Y=k|X)=\frac{P(X|Y=k)\cdot P(Y=k)}{P(X)}=\frac{P(X|Y=k)\cdot P(Y=k)}{\sum_{k=1}^{K}P(X,Y=k)}=\frac{P(X|Y=k)\cdot P(Y=k)}{\sum_{k=1}^{K}P(X|Y=k)P(Y=k)} \tag{1}$$

其中$P(Y=k)$和$P(X|Y=k)$分别称为先验概率和类条件概率密度。

设 $$\pi_k=p(Y=k),f_k(X)=P(X|Y=k) \tag{2}$$

则式(1)化为：

$$P(Y=k|X)=\frac{f_k(X)\cdot\pi_k}{\sum_{k=1}^{K}f_k(X)\cdot\pi_k} \tag{3}$$

其中各量可从样本训练集估计，对先验概率π_k可从样本训练集中各类样本的数目估计：

$$\pi_k=\frac{n_k}{n} \tag{4}$$

其中，n_k为第k类样本的数目。而对类条件概率密度$f_k(X)$则可由核密度估计：

$$f_k(X)=\frac{1}{n_k}\sum_{i=1}^{n}K_{h_k}(X,X_i) \tag{5}$$

$K_{h_k}(X,X_i)$为核密度函数，h_k称为核密度函数的带宽(Bandwidth)或平滑参数。则

$$P(Y=k|X)=\frac{\frac{\pi_k}{n_k}\sum_{i=1}^{n_k}K_{hk}(X,X_i)}{\sum_{k=1}^{K}\frac{\pi_k}{n_k}\sum_{i=1}^{n_k}K_{hk}(X,X_i)} \tag{6}$$

对于X为非连续的二进制0-1变量，核密度函数通常选为Aitchison和Aitkin提出的如下形式的核密度函数[4]：

$$K_h(X,X_i)=h^{m-d(X,X_i)}(1-h)^{d(X,X_i)} \tag{7}$$

其中带宽参数h取值范围：$0.5<h<1$，$d(X_i,X_j)$为欧氏距离的平方：

$$d(X,X_i)=(X-X_i)^{\mathrm{T}}(X-X_i) \tag{8}$$

对连续的X变量，核密度函数通常选为Gauss核密度函数[10]：

$$K_h(X,X_i)=\frac{1}{h\sqrt{2\pi}}e^{-\frac{d(X_i,X)}{2h^2}} \tag{9}$$

其带宽参数h取值范围为：$0<h<\infty$，$d(X,X_i)$仍为欧氏距离的平方。

对二类分类问题，设$y=\{+1,-1\}$，且有n_+个正样本，n_-个负样本，则有：

$$\pi_+=\frac{n_+}{n},\pi_-=\frac{n_-}{n} \tag{10}$$

正、负样本的分布的核密度估计分别为：

$$f_+(X)=\frac{1}{n_+}\sum_{i=1}^{n_+}K_{h+}(X,X_i) \tag{11}$$

$$f_-(X)=\frac{1}{n_-}\sum_{i=1}^{n_-}K_{h-}(X,X_i) \tag{12}$$

则样本X属于正样本类和负样本类的概率分别为：

$$P(Y=+1|X)=\frac{\frac{\pi_+}{n_+}\sum_{i=1}^{n_+}K_{\lambda_+}(X,X_i)}{\frac{\pi_+}{n_+}\sum_{i=1}^{n_+}K_{\lambda_+}(X,X_i)+\frac{\pi_-}{n_-}\sum_{i=1}^{n_-}K_{\lambda_-}(X,X_i)} \tag{13}$$

$$P(Y=-1|X)=\frac{\frac{\pi_-}{n_-}\sum_{i=1}^{n_-}K_{\lambda_-}(X,X_i)}{\frac{\pi_+}{n_+}\sum_{i=1}^{n_+}K_{\lambda_+}(X,X_i)+\frac{\pi_-}{n_-}\sum_{i=1}^{n_-}K_{\lambda_-}(X,X_i)} \tag{14}$$

二者满足关系：

$$P(Y=+1|X)+P(Y=-1|X)=1 \tag{15}$$

核密度估计分类方法是一种懒散的算法。它不需要训练，将所有训练样本保存在内存中。对未知类别的样本数据X，由式(13)和式(14)求出它属于正负样本的概率。通常，当$P(Y=+1|X)>P(Y=-1|X)$，即$P(Y=+1|X)>0.5$时，预测X为正样本，否则X为负样本。这里，0.5即是通常的分类截断值。核密度估计应用于分类问题最关键的一个问题是对带宽参数h的估计。对此，人们已提出了各种各样的计算方法。如Aitchison和Aitkin[4]用留一法极大似然函数估计量：

$$W(h|D)=\prod_{i=1}^{n}P(X_i|D_i,h) \tag{16}$$

估计带宽参数h。其中，D_i为训练集D排除样本X_i后的样本集。Willett等人将核密度估计分类方法用于计算机辅助药物筛选的非连续的二进制X变量和连续的X变量情形，分别称为BKD(Binary Kernel Discrimination)和CKD(Continuous Kernel Discrimination)方法。在许多分类问题，如计算机辅助药物筛选，人们不仅需要预测未知化合物的类别，还需要知道这种预测的可信度。对贝叶斯分类器，即为式(12)和式(13)预测的概率输出。Willett等人提出了带宽参数h的估计方法：对训练集中每一样本，根据留一法预测的概率输出结果的大小，进行样本由高到低的排序，优化h使正样本的排序之和最小。很显然，这里以正样本的排序之和为h优化的目标数并非最好选择。本文中我们采用受试者特征(ROC)曲线下的面积AUC为目标函数，优化带宽参数。

2.2 受试者特征曲线及带宽参数的优化

对于分类器性能的评估，人们提出了许多评价指标，如灵敏度(sensitivity，Se)：又称真阳性率，为正样本中被正确预测的样本所占比例；特异度(specificity，Sp)：又称真阴性率，为负样本中被正确预测的样本所占比例；总体精度(overall accuracy，Q)：指所有样本中被正确预测的样本所占比例。用这些单个数值指标不能很好地完全描述分类器性能，为此人们提出了采用ROC曲线描述的方法。ROC曲线早先主要用医学领域的临床试验，最近才被用于分类器性能的评估。将所有样本的预测结果按某一打分函数(Scoring function)由大到小排序，使预测为正样本的可能性越大的越排在前面。如由式(11)计算的样本预测为正样本的概率即可作为打分函数。在对样本作分类预测时，需要一截断阈值，当打分函数大于该截断阈值时预测为正样本，反之为负样本。根据打分函数，采用不同的截断阈值，计算预测的敏感度和特异度，并以敏感度为纵坐标，以1-特异度(即假阳性率)为横坐标给出的各点联成的曲线，即为ROC曲线。由ROC曲线可以判断出最佳截断阈值，又称最佳临界值。ROC曲线上距左上角最近的一

点，即为最佳临界值。用该点数值区分正常与异常，其敏感度及特异度都比较高，而误诊及漏诊例数之和最小。同时，由ROC曲线下的面积(Area under the Curve，AUC)可很好地评价分类器性能，越接近1.0则预测越准确，低于0.5则预测结果不如随机预测好。

在本文中，我们以留一法预测的样本属于正样本的概率作为打分函数，构造ROC曲线，并以ROC曲线下的面积为目标函数，优化核密度函数的带宽参数 h。ROC曲线下的面积是按Wilcoxon-Mann-Whitney统计公式[11]计算的：

$$AUC=\frac{\sum_{i=1}^{n_+}\sum_{j=1}^{n_-}I(X_i,X_j)}{n_+n_-} \tag{17}$$

其中，$f(X)$为打分函数。本研究中，对核密度估计分类 $f(X)$ 即为 $P(Y=+1|X)$。式中求和函数为：

$$I(X_i,X_j)=\begin{cases}1 & f(X_i)>f(X_j)\\ 0 & f(X_i)<f(X_j)\end{cases} \tag{18}$$

3 实验

3.1 数据集及数据处理

UCI数据集是一个常用于测试机器学习方法的标准数据集，收集了各个学科领域的分类或回归问题的数据，可免费下载[12]。本文选用UCI中的“Promoter Gene Sequences”和“Pima Indians Diabetes”两个数据集作为实验数据集。这两个数据集是生物和医学中的二类分类问题，这里简称为Promoter和Diabetes数据集。Promoter数据集包含106个DNA序列片段样本，含53个正样本和53个负样本，分别对应于大肠肝菌DNA序列的启动子和非启动子。启动子作为DNA聚合酶结合的靶序列，对转录起始有调节和控制作用，决定着基因表达过程是否开始以及在什么条件下开始。启动子的识别是研究转录调节网络的前提。该数据集含有每个序列的类别(“+”为启动子，“−”为非启动子)，序列名及长度为57的DNA序列片段。DNA序列片段由AGCT四种碱基组成。我们将DNA序列片段进行如下二进制编码构成特征矢量：$A=\{1000\}$，$G=\{0100\}$，$C=\{0010\}$，$T=\{0001\}$。这样，长度为57的DNA序列片段就由维数为228的二进制变量的矢量表示。本文将Promoter数据集用于检验Aitchison-Aitkin核密度函数带宽参数优化。Diabetes数据集为检测印地安妇女是否患有糖尿病的测试数据，含有768个样本，含268个正样本和500个负样本。每个样本由8个连续特征及一个类别变量描述。由于这8个特征取值范围不一样，所以我们先对变量进行自标度化处理，即每一特征变量减去其平均值再除以标准偏差，使处理后变量均值为0、标准偏差为1。

3.2 模型验证与数据集的划分

在本文的实验中，模型验证分别用留一法(LOO)和10重交叉验证(10-fold CV)方法。LOO方法依此将数据集中每一个样本作为测试，余下样本作为训练集，这样对每一个样本都作了预测，分类器的预测性能为对所有样本预测结果的评价；10-fold CV方法将数据集随机地分为样本数相同或接近的10个组，然后依此将每一组样本作为测试集，余下各组样本一起作为训练集，这样对每一个样本或每一组样本都作了预测，分类器预测性能为对所有样本预测结果的评价。表1中列出了Promoter数据集和Diabetes数据集用10重交叉验证方法时数据集划分情况。

表1 数据集的划分

Validation Method	Set	Number of samples					
		Total samples		Positive samples		Negative samples	
		Promoter	Diabetes	Promoter	Diabetes	Promoter	Diabetes
10-fold CV	Set 1	12	77	6	27	6	50
	Set2	12	77	6	27	6	50
	Set 3	12	77	6	27	6	50
	Set 4	10	77	5	27	5	50
	Set 5	10	77	5	27	5	50
	Set 6	10	77	5	27	5	50
	Set 7	10	77	5	27	5	50
	Set 8	10	77	5	27	5	50
	Set 9	10	76	5	26	5	50
	Set 10	10	76	5	26	5	50

本文的计算均以自行编制的Fortran程序进行。为方便描述起见，我们采用Willett等人[6-10]的符号约定，分别将二进制特征矢量和连续值特征矢量的核密度分类方法称为BKD方法和CKD方法。

3.3 实验结果

表2和表3分别列出了用不同验证方法对Promoter和Diabetes数据集进行分类预测的结果及文献报道结果。结果包括正样本预测准确率(Se)、负样本预测准确率(Sp)、总体精度(Q)、ROC曲线下面积(AUC)、优化的带宽(Bandwidth)和分类切断值(cutoff)。优化的分类截断值是在采用最优带宽参数时，选取分类截断值，尽可能使总体精度达最大的同时，使正、负样本预测准确率达最大。

从表2可以看出，对Promoter数据集，用本文的BKD方法，当分类截断值取最佳截断值时，留一法(LOO)和10重交叉验证法的预测结果非常接近，正样本预测准确率、负样本预测准确率和总体精度均在85%～89%之间。此时，留一法(LOO)和10重交叉验证法的最佳切断值分别为0.64和0.65，即样本预测为正样本的概率大于0.64或0.65时划分为正样本，否则为负样本。当截断值选为0.5时，总体精度下降为84%～85%，虽然正样本的预测准确率提高到94%～96%，但负样本的预测准确率则下降到约74%。因此，通过ROC分析优化切断值，可以使总体精度提高，同时还可使正样本的预测准确率和负样本的预测准确率尽可能相同。而很多其他分类方法(如SVM)，没有直接的概率输出，往往很难同时兼顾正样本的预测准确率和负样本的预测准确率。因而大部分的文献报道都没有给出正样本的预测准确率和负样本的预测准确率。本研究中，最佳截断值偏离于0.5的一个重要原因，是正样本的先验概率 π_+ 和负样本的先验概率 π_- 是根据训练集中正样本和负样本的数量由式(10)计算出的，而真实的正负样本样本分布也许偏离于训练集中正负样本样本分布，而且很难预先估计，因此，在贝叶斯类分类器中，优化截断值会很大程度地改善预测能力。对比表2中文献报道结果还可以看出，我们的预测结果接近于文献报道最好结果。在本文中，我们没有进行特征选择(Feature Selection)，所有特征均用于分类模型的建立。而Polat[15]等人研究表明，经恰当的特征选择后，总体预测精度可大幅提高，由50%(表2中Fuzzy-AIRS结果)提高到90%(表2中Fuzzy-AIRS/FS结果)，表明这些特征中有些是多余的变量或躁声变量。因此，本文的预测结果在没有特征选择时就可给出较好的预测精度，表明我们提出的核密度估计用于分类时通过使AUC最大而优化带宽参数的方法是一可行的方法，用于Promoter数据集可以得到满意的结果。

从表 3 对 Diabetics 数据集的分类预测结果可以看出，当选取最佳切断值时，本文 LOO 结果和 10-fold CV 结果的总体精度均约为 73%，AUC 均约 0.82。同 Promoter 数据集一样，采用最佳切断值时，总体精度明显比采用切断值选为 0.5 时的高，而且，正样本的预测准确率和负样本的预测准确率接近，而切断值选为 0.5 时，负样本的预测准确率远远低于正样本的预测准确率。对比表 3 中文献报道结果，可以看出，本文的总体精度与文献报道用其它方法得到的最佳结果接近，表明本文提出的优化带宽参数的方法对于连续特征矢量的分类问题也可得到较好结果。

表 2　Promoter 数据集分类预测结果

Methods	Validation Method	Se (%)	Sp (%)	Q (%)	AUC	Bandwidth[a]	Cutoff[b]
BKD/This work[c]	LOO	88.67	86.79	87.73	0.9494	0.6	0.64
		94.33	73.58	83.96	0.9494	0.6	0.50
BKD/This work[c]	10-fold CV	88.67	84.90	86.79	0.9533	0.6	0.65
		96.22	73.58	84.90	0.9533	0.6	0.50
Naïve Bayes[13]	10-fold CV			90.56			
TAN[13]	10-fold CV			80.07			
BAN[13]	10-fold CV			88.11			
GBN[13]	10-fold CV			87.12			
C4.5[14]	LOO			84.0			
Fuzzy-AIRS[15]	10-fold CV			50%			
Fuzzy-AIRS/FS[15]	10-fold CV			90%			

[a]计算表明，正负样本的带宽取值一样时模型最优；[b] Cutoff：为分类切断值，即若某样本的打分函数大于该值划分为正样本，否则为负样本；[c]第一行数据对应采用最优带宽参数时分类截断值取最佳临界值时的预测结果，而第二行数据对应采用最优带宽参数时分类切断值取通常值 0.5 的预测结果。

表 3　Diabetes 数据集分类预测结果

Methods	Validation Method	Se (%)	Sp (%)	Q (%)	AUC	Bandwidth	Cutoff[a]
BKD/This work[b]	LOO	72.23	70.60	72.91	0.8197	1.4	0.294
		25.00	97.00	71.87	0.8197	1.4	0.500
BKD/This work[b]	10-fold CV	76.86	71.20	73.17	0.8198	1.3	0.291
		30.97	96.20	73.43	0.8198	1.3	0.500
SVM[16]	Hold-out			70.47			
kNN (k=5)[17]	10-fold CV			70.96			
kNN (k=3)[17]	10-fold CV			69.09			
kNN (k=1)[17]	10-fold CV			59.52			
Adboost-NN[18]	5-fold CV			73.5			
Adboost-SVM[18]	5-fold CV			75.2			
SVM[18]	5-fold CV			76.5			
AIRS1[19]	10-fold CV			74.1			
C4.5[20]				62.0			
SLIQ[20]				76.0			
EDTA[20]				64.0			
PV1[20]				72.0			
PV2[20]				74.0			
PV3[20]				70.0			
PV4[20]				66.0			
PV5/Ref17				64.0			

[a]同表 2；[b]Cutoff：同表 2；[c] 同表 2。

结束语　本文提出了将核密度估计用于贝叶斯分类器类条件概率密度的估计，以 ROC 曲线下的面积 AUC 为目标函数优化核函数的带宽参数，建立了二进制特征和连续值特征情形的 BKD 分类方法和 CKD 分类方法。由于以 AUC 为机器学习方法的性能评估，并通过 ROC 分析得到最佳切断值，可以在得到较好总体精度的同时，得到较好的正样本的预测准确率和负样本的预测准确率。

参 考 文 献

[1] 谭东宁，谭东汉. 小样本机器学习理论[J]. 统计学习理论. 南京理工大学学报，2001，25(1)：109-112

[2] Abraham R, Simha J B, Iyengar S S. A comparative analysis of discretization methods for Medical Datamining with Naïve Bayesian classifier[C]// Proceedings of the 9th International Conference on Information Technology. 2006：235-236

[3] Perez A, Larrañaga P, Inza I. Bayesian classifiers based on kernel density estimation: Flexible classifiers[J]. International Journal of Approximate Reasoning, Article in Press, 2008

[4] Aitchison J, Aitken G G. Multivariate binary discrimination by the kernel method[J]. Biometrika, 1976, 63(3)：413-420

[5] Harper G, Bradshaw J, Gittins J C, et al. Prediction of Biological Activity for High-Throughput Screening Using Binary Kernel [J]. J. Chem. Inf. Comput. Sci. , 2001, 41：1295-1300

[6] Hert J, Willett P, Wilton D J. New Methods for Ligand-Based Virtual Screening: Use of Data Fusion and Machine Learning to Enhance the Effectiveness of Similarity Searching[J]. J. Chem. Inf. Model, 2006, 46：462-470

[7] Willett P, Wilton D. Virtual Screening Using Binary Kernel Discrimination: Analysis of Pesticide Data[J]. J. Chem. Inf. Model, 2006, 46：471-477

[8] Willett P, Wilton D. Virtual Screening Using Binary Kernel Discrimination: Effect of Noisy Training Data and the Optimization of Performance[J]. J. Chem. Inf. Model, 2006, 46：478-486

[9] Willett P, Wilton D. Prediction of Ion Channel Activity Using Binary Kernel Discrimination[J]. J. Chem. Inf. Model, 2007, 47：1961-1966

[10] Chen B, Harrison R F, Papadatos G, et al. Evaluation of machine-learning methods for ligand-based virtual screening[J]. Comput Aided Mol Des, 2007, 21：53-62

[11] Mann H B, Whitney D R. On a test whether one of two random variables is stochastically larger than the other[J]. Ann. Math. Statist. , 1974, 18：50-60

[12] Asuncion A, Newman D J. UCI Machine Learning Repository [OL]. http://www. ics. uci. edu/~mlearn/MLRepository. html. Irvine, CA: University of California, School of Information and Computer Science, 2007

[13] 邓甦，付长贺. 四种贝叶斯分类器及其比较[J]. 沈阳师范大学学报：自然科学版，2008，26(1)：31-33

[14] Geamsakul W, Matsuda T, Yoshida T, et al. Constructing a Decision Tree for Graph-Structured Data and its Applications[J]. Fundamenta Informaticae, 2004, 66：131-160

[15] Polat K, Gunes S. A new method to forecast of Escherichia coli-promoter gene sequences: Integrating feature selection and Fuzzy-AIRS classifier system[J]. Expert Systems with Applications, 2009, 36：57-64

[16] 张鸿雁. 基于进化计算方法的支持向量机特征选择[J]. 煤矿机械，2008，29(5)：47-49

[17] 任江涛，卓晓岚，许盛灿，等. 基于 PSO 面向 K 近邻分类的特征权重学习算法[J]. 计算机科学，2007，34(5)：187-189

[18] 胡金海，谢寿生，杨帆，等. 基于支持向量机的组合分类方法及应用[J]. 推进技术，2007，28(6)：669-673

[19] Watkins A , Boggess L . Artificial Immune Recognition System (AIRS): An Immune-Inspired Supervised Learning Algorithm [J]. Genetic Programming and Evolvable Machines, 2004, 5：291-317

[20] Chandra B, Paul V P. A Robust Algorithm for Classification Using Decision Trees[C]// Cybernetics and Intelligent Systems, 2006 IEEE Conference. 2006：1-5

[21] 黄贤英，张丽芳. 基于粒子群优化的模糊聚类算法[J]. 重庆工学院学报，2008，22(11)：120-123