ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.18, No.11, November 2007, pp.2691−2699
DOI: 10.1360/jos182691

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

覆盖算法的概率模型*

张 铃[1,2+], 吴 涛[1,2], 周 瑛[1,2], 张燕平[1,2]

[1](安徽大学 计算智能与信号处理教育部重点实验室,安徽 合肥 230039)
[2](安徽大学 人工智能研究所,安徽 合肥 230039)

Probabilistic Model for Covering Algorithm

ZHANG Ling[1,2+], WU Tao[1,2], ZHOU Ying[1,2], ZHANG Yan-Ping[1,2]

[1](Key Laboratory of Intelligent Computing & Signal Processing of the Ministry of Education, Anhui University, Hefei 230039, China)
[2](Institute of Artificial Intelligence, Anhui University, Hefei 230039, China)
+ Corresponding author: Phn: +86-551-5106397, E-mail: zling@ahu.edu.cn, http://210.45.212.15/INDEX.ASP/

Zhang L, Wu T, Zhou Y, Zhang YP. Probabilistic model for covering algorithm. *Journal of Software*, 2007,18(11):2691−2699. http://www.jos.org.cn/1000-9825/18/2691.htm

Abstract: It is necessary to bring global optimization in covering algorithm to improve its precision of classification. So a probabilistic model of covering algorithm is put forward in this paper. Firstly, the covering algorithm is ameliorated to kernel covering model (Gaussian function is the kernel function), then a kind of finite mixture probabilistic model for kernel covering model is introduced according to the probabilistic meaning of Gaussian function. Finally, the global optimization calculation is inducted based on maximum likelihood theory and Expectation Maximization Algorithm. Therefore, the algorithm optimizes the covering network broadens the application domain of covering algorithm and improves its robustness. The experimental results show that the optimized probabilistic model of covering algorithm can improve the accuracy of classification.
Key words: machine learning; neural network; covering algorithm; finite mixture probabilistic model

摘 要: 要从本质上提高覆盖算法的精度,必须在算法中引入全局的优化计算.为此,先将覆盖算法扩展成核覆盖算法(以高斯函数为核函数),再利用高斯函数的概率意义(高斯分布),为核覆盖算法建立一个有限混合概率模型,在此基础上,利用“最大似然原理”引入全局优化计算,并利用 EM(expectation maximization)方法进行求解,完成对覆盖算法的全局优化计算,从而扩大覆盖方法的使用范围并提高算法的精度,且将它从确定的模型扩展成概率的模型,后

* Supported by the National Natural Science Foundation of China under Grant Nos.60475017, 60675031 (国家自然科学基金); the National Basic Research Program of China under Grant No.2004CB318108 (国家重点基础研究发展计划(973)); the National Research Foundation for the Doctoral Program of Higher Education of the Ministry of Education of China under Grant No.20040357002 (国家教育部高等学校博士学科点专项科研基金); the Natural Science Foundation of Anhui Province of China under Grant No.050420208 (安徽省自然科学基金); the Provincial Project of Natural Science Research for Colleges and Universities of Anhui Province of China under Grant Nos.2006KJ015A, 2006KJ244B (安徽高等学校省级自然科学研究项目); the Innovative Research Team of 211 Project in Anhui University of China (安徽大学学术创新团队); the Program of Superior Teacher Team in Anhui University of China (安徽大学人才队伍建设经费)

Received 2006-04-24; Accepted 2006-08-22

者更具抗噪声干扰的能力.最后给出模拟实验,实验比较结果表明,经优化后的概率模型确实提高了算法的精度.
关键词: 机器学习;神经网络;覆盖算法;有限混合概率模型
中图法分类号: TP181 **文献标识码**: A

在文献[1]中,我们利用 M-P 神经元的新的几何意义,提出了一种前向神经网络的学习算法——覆盖算法.该算法使得处理海量数据的分类、聚类问题成为可能.它随后被推广成核覆盖算法、多侧面覆盖算法、模糊覆盖算法、增量覆盖算法等[2−5],并被成功地应用于金融预测、模式识别、手写汉字、文本分类、网络上图像检索等问题[6−11].

这些推广和改进扩大了覆盖算法的应用范围,在一定程度上也提高了算法的精度.但要进一步提高算法的精度,还要解决如何在算法中引入全局的优化计算.只有合理地引入全局的优化计算,才能在本质上提高算法的精度.

核覆盖算法在某种意义下可以看成是多分类的(核)支持向量机(support vector machine,简称 SVM)方法(SVM 方法本质上是针对二分类).其次,在应用(核)SVM 方法时,最大的困难在于如何选取适当的核函数,而覆盖算法利用“覆盖”很好地解决了核函数选取的问题.但 SVM 方法有一个全局优化过程,而覆盖算法只是进行局部优化运算,故一般地说,SVM 方法具有更高的精度.

本文的目的是,通过为(核)覆盖算法建立一个概率(有限混合概率)模型,在此基础上利用“最大似然原理”引入全局优化计算,完成对核覆盖算法的全局优化.希望通过全局优化,提高算法的精度.

本文第 1 节简单介绍覆盖算法.第 2 节简单介绍有限混合概率模型.第 3 节给出覆盖算法的概率模型.第 4 节给出模拟比较的结果.最后是结论.

1 覆盖算法

1.1 覆盖算法(分类的学习问题)

问题的提出:给定一个训练样本集 $K=\{(x^1,y^1),\ldots,(x^p,y^p),x^i\in R^n,y^i\in\{0,1\}^m\}$,要求构造一个前向神经网络 N,满足 $N(x^i)=y^i,i=1,\ldots,p$(即将 n 维空间中的 p 个样本分成指定的 m 类).

(1) 覆盖算法

利用变换:$f:R^n\rightarrow S^{n+1},\ f(x)=\left(x,\sqrt{R^2-|x|^2}\right)$.

求一组覆盖:$\boldsymbol{C}=\{C_1^1,C_2^1,\ldots,C_{n_1}^1,C_1^2,C_2^2,\ldots,C_{n_2}^2,\ldots,C_1^m,C_2^m,\ldots,C_{n_m}^m\}$.

其中每一覆盖子组 $\{C_1^i,C_2^i,\ldots,C_{n_i}^i\}$ 只覆盖第 i 类的样本点,则 $\boldsymbol{C}$ 就是覆盖算法得到的解.其判别(决策)函数可以表示为

$$F_i(x)=\sum_j \sigma_j^i(x),$$

其中,$\sigma_j^i(x)$ 是 C_j^i 的特征函数,$F_i(x)$是第 i 类的判别函数,$i=1,\ldots,m$.

样本测试:给定一个测试样本 x,先给定一个阈值 a,当 x 落在距各覆盖的距离均大于 a 时,则认为是拒识状态,否则按“就近原则”进行判别.

(2) 模糊覆盖算法

在上面的覆盖算法中,其判别函数是 C_j^i 的特征函数,我们也可以将特征函数改为其他函数,构成模糊覆盖.其判别函数可以表示为

$$F_i(x)=\sum_j f_j^i(x),$$

其中,$f_j^i(x)$ 是 C_j^i 的模糊覆盖函数,$F_i(x)$是第 i 类的判别函数,$i=1,\ldots,m$.

(3) 核覆盖算法

给定核函数 $K(x,y)$,设 C_j^i 覆盖的中心为 a_j^i,令对应的判别函数为

$$f_j^i(x)=K(a_j^i,x),F_i(x)=\sum_j K(a_j^i,x),i=1,...,m.$$

若将各个覆盖中所覆盖的样本点的个数考虑进去,上式也可以写成

$$f_j^i(x)=K(a_j^i,x),F_i(x)=\sum_j \beta_j^i K(a_j^i,x) \tag{1}$$

β_j^i 由 C_j^i 覆盖的样本点的密度决定.

若取 $K(x,y)$为高斯核函数,则式(1)为

$$F_i(x)=\sum_j \beta_j^i \exp(-|x-a_j^i|^2/{\sigma_j^i}^2) \tag{1′}$$

1.2 SVM核函数法

对二类分类问题,核支持向量机方法可以简述如下:

问题:给定样本集(二类分类样本) $K=\{(x^1,y^1),…,(x^p,y^p),x^i\in R^n,y^i\in\{0,1\}^m\}$.

给定核函数 $K(x,y)$,则最后得到的最优决策函数为

$$F(x)=\mathrm{sgn}\left(\sum_{\text{支持向量}} y^i\alpha_i K(x^i,x)-b\right) \tag{2}$$

取高斯核函数,则式(2)为

$$F(x)=\mathrm{sgn}\left(\sum_{\text{支持向量}} y^i\alpha_i \exp(-|x^i-x|^2/\sigma_i^2)-b\right) \tag{2′}$$

2 概率有限混合模型

在概率论中有如下命题成立:任何连续分布均可由等方差 Gauss 密度的有限混合任意逼近.

换句话说,在近似的意义下,只要研究有限 Gauss 混合密度就足够了.本文正是本着这种思想,利用有限高斯混合密度为我们的覆盖算法建立概率模型.下面简单介绍有限混合模型的性质.

2.1 有限混合模型的定义[12,13]

设 $X_1,…,X_p$ 是样本量为 p 的独立同分布随机样本,其中 X_i 是 n-维随机向量,其概率密度函数为 $f(x)$ (probability density function,简称 p.d.f).设 $f(x_i)$可以表示为

$$\begin{cases} f(x_i)=\sum_{j=1}^{g}\alpha_j f_j(x_i) \\ 0\le\alpha_j\le 1 (j=1,...,g) \\ \sum_{j=1}^{g}\alpha_j=1 \end{cases} \tag{3}$$

其中 $f_j(x_i)$表示观察值 x_i 来自第 j 分量时的条件 p.d.f.

称由式(3)定义的密度函数为 g-分量有限混合密度(g-component finite mixture density).

若取 $f_j(x)$为 1-维高斯分布,则式(3)可以表示为

$$f(x_i)=\sum_j \beta_j \exp(-|x_i-x|^2/\sigma_j^2) \tag{3′}$$

从形式上看,式(1′)、式(2′)和式(3′)非常相似,特别是式(1′)和式(3′)形式上完全相同.这就启发我们是否可以利用有限混合概率模型来研究覆盖算法.

本文的目的就是利用有限混合概率模型为覆盖算法建立一个概率模型,利用有限混合模型的极大似然拟合方法为覆盖算法提供全局优化方法.

2.2 混合模型的极大似然拟合

设给定样本$\{x_1,\dots,x_p\}$,令其概率密度函数为

$$\begin{cases} f(x_i)=\sum_{j=1}^{g}\alpha_j f_j(x_i) \\ 0\le \alpha_j \le 1 (j=1,\dots,g) \\ \sum_{j=1}^{g}\alpha_j=1 \end{cases}.$$

$f_j(x)$属于某个参数分布族$\{f_j(x,\theta)\}$,记为

$$f(x_i,\Theta)=\sum_{j=1}\alpha_j f_j(x_i,\theta_j),\ \theta\in\Omega.$$

令

$$L(\Theta)=\prod_i f(x_i,\Theta) \tag{4}$$

称其为(样本$\{x_1,\dots,x_p\}$)的似然函数(likelihood function).$L(\Theta)$在Ω中某一点$\hat{\Theta}$达到最大值,称$\hat{\Theta}$是Θ的最大似然估计(maximum likelihood estimator,简称 MLE).

对式(4)取对数,再对Θ求偏导数,可得

$$\hat{\alpha}_j=\frac{1}{p}\sum_{j=1}\beta_j(x_i,\hat{\Theta}),$$

$$\sum_{i=1}^{p}\beta_j(x_i,\hat{\Theta})(\partial\ln f_j(x_i,\hat{\theta}_j)/\partial\theta_j)|_{\theta_j=\hat{\theta}_j}=0,$$

对$j=1,\dots,g$成立,其中,

$$\beta(x_i,\hat{\Theta})=\frac{\hat{\alpha}_j f_j(x_i,\hat{\theta}_j)}{\sum\hat{\alpha}_t f_t(x_i,\hat{\theta}_t)}.$$

2.3 最大似然求解方法

Dempsteret 在文献[12]中给出了求解最大似然的迭代 EM(expectation maximization)算法.算法步骤如下:

E 步:

第 1 步:给定初始值$\Theta^{(0)}$,求均值:

$$Q(\Theta,\Theta^{(0)})=E_{\Theta^{(0)}}[\ln L(\Theta\,|\,x)],$$

其中,E是取均值.

给定$\Theta^{(k)}$,求得

$$\beta_j(x_i,\Theta^{(k)})=\frac{\alpha_j^{(k)}f_j(x_i,\theta_j^{(k)})}{\sum_{t=1}\alpha_t^{(k)}f_t(x_i,\theta_t^{(k)})}.$$

求期望:

$$Q(\Theta,\Theta^{(k)})=\sum_{t=1}^{p}\sum_{j=1}^{g}\beta_j(x_i,\Theta^{(k)})(\ln\alpha_j\ln f_j(x_i,\theta_j)).$$

M 步:

第$(k+1)$步要求:

$$\Theta^{(k+1)}=\arg\max_{\Theta\in\Omega}Q(\Theta,\Theta^{(k)}).$$

用 Lagrange 乘子法,可得

$$Q(\Theta,\Theta^{(k)})-\lambda\left(\sum_j\alpha_j-1\right)=0.$$

易解得$\lambda=n$.

$$\alpha_j^{(k+1)}=\frac{1}{p}\sum_{i=1}^{p}\beta_j(x_i,\Theta^{(k)}),\ j=1,\dots,g.$$

新估计 $\theta_j^{(k+1)}$,取它为以下方程的一个适当解:

$$\sum_{i=1}^{p}\beta_j(x_i,\theta^{(k)})(\partial\ln f_j(x_i,\theta_j)/\partial\theta_j)=0\ .$$

注:M 步的第 $k+1$ 迭代只要求满足下式,也可证明迭代的收敛性.

$$Q(\Theta^{(k+1)},\Theta^{(k)})\geq Q(\Theta^{(k)},\Theta^{(k)}).$$

3 覆盖算法的概率模型

3.1 覆盖算法的概率模型

给定分类问题,$K=\{(x^1,y^1),\dots,(x^p,y^p),x^i\in R^n,y^i\in\{0,1\}^m\}$,即共有 p 个 n 维样本,分成 m 类.

我们用覆盖算法求得覆盖组为

$$\{C_1^1,C_2^1,\dots,C_{g_1}^1,C_1^2,C_2^2,\dots,C_{g_2}^2,\dots,C_1^m,C_2^m,\dots,C_{g_m}^m\}\ .$$

设覆盖第 i 类点的覆盖组 $C^i=\{C_1^i,C_2^i,\dots,C_{g_i}^i\}$.

简单的覆盖算法通过对每个覆盖取特征函数来表示.这种表示未能反映样本在覆盖中的分布情况.为了克服这个缺点,应用某种分布函数来代替特征函数,最容易想到的就是高斯分布函数.若对每个覆盖 C_j^i 引入高斯核函数(以覆盖的中心为高斯核函数的均值,取半径为方差 σ_j^i,记为 $N(a_j^i,\sigma_j^i),j=1,\dots,g_i$.

对应于第 i 类的决策函数为

$$F_i(x)=\sum_j\beta_j^i\exp(-|x-a_j^i|^2/{\sigma_j^i}^2)\ ,\ i=1,\dots,p \tag{5}$$

其中,β_j^i 是与各覆盖中样本点的密度有关的参数.

式(5)又可以理解为取核函数为高斯函数的核覆盖算法得到的决策函数.而核函数法的一个未能很好地解决的问题,就是如何合理地给定核函数的参数.这里利用由覆盖算法求到的覆盖作为所取的核函数,这样就解决了如何选择核函数参数的难题.

全局优化问题的最大难度在于如何引入“优化原则”,如果将核覆盖算法看成是多类分类的 SVM 方法,就很难将“最优分离超平面”的概念改造成多类分类的最优原则.若将最优原则定为“覆盖个数最少”,则可能将问题引入死胡同,因为最少覆盖问题本身就是一个 NPC 问题.而且覆盖数最少是不是就能保证算法的泛化能力最强?未必.故必须对覆盖算法建立适当的模型后,才能合理地引入“优化原则”.

经过上面的一系列变换,我们就可以从概率的角度来考察它,因为式(5)可以看成是一个有限混合概率模型,于是可以利用第 2 节给出的求模型的最大似然的 EM 算法,求得式(5)的最优参数.这样就从概率角度为核函数的确定参数问题,给出一个合理的解决方法.

也就是说,当我们将核函数法中的核函数理解成概率分布函数时,核函数覆盖算法为我们提供了一个新的视角——概率模型.于是可将概率统计中现成的方法引入分类学习,也就是把核覆盖算法与统计模型结合起来,为覆盖算法找到全局优化的途径.

将覆盖问题化成有限混合概率模型后,就可将“最大似然原理”引入覆盖算法.“最大似然原理”从直观上看,就是对给定的数据,在所有可能提供的参数中求出一个对所给定的数据全体具有最大概率的参数.因为是对“所给定的数据全体”求最大,故这个优化具有“全局性”.

3.2 混合模型的最大似然拟合

算法 1. 概率混合覆盖算法.

给定 m 类分类的训练样本集 $K=\{K_1,\ldots,K_m\}$.

S1:利用覆盖算法,求出各类的覆盖组$\{C^1,\ldots,C^m\}$.

对每个覆盖组 $C^i=\{C_1^i,C_2^i,\ldots,C_{g_i}^i\}$,$i=1,\ldots,m$.建立有限混合概率模型,下面为书写方便,将上标略去,记为 $C=\{C_1,\ldots,C_g\}$,并设第 i 类样本共有 p 个.

其有限混合模型为

$$F(x)=\sum_j \alpha_j \varphi(x,\mu_j,\Sigma_j)\ ,\ j=1,\ldots,g.$$

S2:利用最大似然拟合法进行拟合.

(2.1) 取 $\alpha_j^{(0)}=d_j$,d_j是覆盖 C_j 含有第 i 类点占第 i 类点总数的比例.

$\sigma_j^{(0)}=r_j$,r_j 为 C_j 的半径.

$\mu_j^{(0)}=a_j$,a_j 为 C_j 的中心.

$\Sigma_j^{(0)}=(\sigma_j^{(0)})^2 I_n$,$I_n$ 是 n 维单位矩阵.

(2.2) 给定初始值后按 EM 方法进行迭代.

公式如下:

E 步:

第$(k+1)$次迭代计算观测样本 x^i 来自第 j 个分量的后验概率

$$\beta_j(x^i,\Theta^{(k)})=\frac{\alpha_j^{(k)}\varphi(x^i,\mu_j^{(k)},\Sigma_j^{(k)})}{\sum_{j=1}^{g}\alpha_j^{(k)}\varphi(x^i,\mu_j^{(k)},\Sigma_j^{(k)})},\ j=1,\ldots,g,\ i=1,\ldots,p \tag{E-1}$$

$$\alpha_j^{(k+1)}=\frac{1}{p}\sum_{i=1}^{p}\beta_{ij}^{(k)},\ j=1,\ldots,g \tag{E-2}$$

M 步:

迭代求均值和方差矩阵:

$$\mu_j^{(k+1)}=\frac{\sum_{i=1}^{p}\beta_{ij}^{(k)}x^i}{\sum_{i=1}^{p}\beta_{ij}^{(k)}}\ ,\text{其中},\ \beta_{ij}^{(k)}=\beta_j(x^i,\Theta^{(k)}),\ j=1,\ldots,g,\ i=1,\ldots,p \tag{M-1}$$

$$\Sigma_j^{(k+1)}=\frac{\sum_{i=1}^{p}\beta_{ij}^{(k)}(x^i-\mu_j^{(k+1)})(x^i-\mu_j^{(k+1)})^T}{\sum_{i=1}^{p}\beta_{ij}^{(k)}},\ j=1,\ldots,p \tag{M-2}$$

当$\varphi(x,\mu_j,\sigma_j)$取为 1-维高斯分布时,有

$$(\sigma_j^2)^{(k+1)}=\frac{\sum_{i=1}^{p}\beta_{ij}^{(k)}\,|(x^i-\mu_j^{(k+1)})|^2}{\sum_{i=1}^{p}\beta_{ij}^{(k)}} \tag{M-2$'$}$$

上面的 EM 方法,就简化为均值、方差(矩阵)和线性组合系数的迭代计算.其结果就是对覆盖迭代改进(不断变换其中心、半径和组合系数).

利用 EM 方法求解最大似然问题,其难点在于如何合理地选取混合模型的分量个数问题(这就像 k-均值法中如何取 k 的问题).对覆盖算法,这个问题可以利用覆盖算法求得的覆盖,作为 EM 算法的迭代起始值,得到很好地解决.因为利用覆盖算法求得的覆盖组基本上已是次优的覆盖,在此基础上,再利用 EM 算法进行迭代,就能很

快求到最优解.这也是对覆盖算法的概率模型引入“最大似然原则”成功的原因.

另外,为了防止迭代过程发散,在迭代过程中,对出现下面情况进行调整:

(1) 给定ε_1,当$|\alpha_j^{(k+1)}|<\varepsilon_1$,将$\alpha_j^{(k+1)}$删去;

(2) 给定δ,当$|\mu_j^{(k+1)}-\mu_i^{(k+1)}|<\delta$,合并为$\mu_j^{(k+1)}=\dfrac{\mu_j^{(k+1)}+\mu_i^{(k+1)}}{2}$;

(3) 给定ε_2,对$\sigma_j^2<\varepsilon_2$的覆盖C_j,将覆盖C_j保留,在下一轮迭代时对该覆盖不进行迭代.

给出一个迭代停止的条件(如达到一定迭代精度或达到一定的迭代次数).最后迭代停止时,所得到有限概率模型即所求.

注:在(3)中保留小方差的覆盖,主要是为了防止 EM 迭代的发散,方差很小的覆盖,所覆盖的点多是“离群点”.

4 模拟结果

利用覆盖算法的概率模型,得到改进的覆盖算法(probabilistic model for covering algorithm,简称 PCA),并与原覆盖算法、SVM 算法[14]进行模拟比较.

4.1 实验结果

本文使用 http://www.ics.uci.edu/mlearn/mlrepository.html 的 5 个数据库(见表 1),分别选用覆盖算法及其改进算法 PCA 进行实验,并调用 http://www.ece.osu.edu/~maj/osu_svm/上的 OSU SVM Classifier Matlab Toolbox,将实验结果与其他分类器的结果进行比较,见表 2.本文的所有实验都是在 CPU 为 Intel Pentium 4 2.6GHz,1.00GB 的内存,编程环境为 MATLAB 6.5.1 下完成的.

Table 1 Databases for experiments

表 1 实验数据库

Name of databases	Number of training samples	Number of testing samples	Dimension of input	Class identification number
Waveform	4 500	500	21	3
Wine	178	18	13	3
Pima indians diabetes	691	77	8	2
Glass	214	21	10	6
Ionosphere	316	35	34	2

对 5 个数据库取同样的训练集和测试集,用 10-交叉进行实验得到的测试精度(%)如表 2 及图 1 所示.

Table2 Comparison table of testing accuracy on five databases

表 2 5 个数据库的测试精度比较表

Algorithm \ Name of databases			Wine	Waveform	Pima	Glass	Ionosphere
Covering algorithm without optimization			96.111	81.417	71.948	57.92	89.225
Covering algorithm without optimization	Iterations	1	100	82.22	73.1	61.9	91.06
		5	100	83.1	77.32	61.9	93.83
		10	100	84.59	78.22	71.43	94.11
		20	100	85.1	79.2	71.43	94.43
		30	100	85.12	79.21	71.43	94.45
Linear SVM			97.778	87.046	76.753	57.727	87.778
RBF SVM			98.333	87.026	77.532	66.364	94.444

在迭代几次以后,算法 3.2 的M步中的$(\sigma_j^2)^{(k)}$就可能为 0,在下一次迭代中,我们用两种方法进行处理:一种方法是将$(\sigma_j^2)^{(k)}$人为地赋以一个很小的值δ,以便能进行下一次循环;另一种方法是将该覆盖保留,在下一轮迭代时对该覆盖不进行迭代.比较两种结果,第 2 种方法要优于第 1 种,所以在实验中,我们的结果都是选用第 2 种方法进行处理的.

从实验的结果得出经优化后的结果提高了算法的精度,在 5 个模拟比较中有 4 个的结果比 SVM 方法好.

而未优化的原覆盖算法的 5 个数据的模拟结果的精度都不如 SVM 方法.

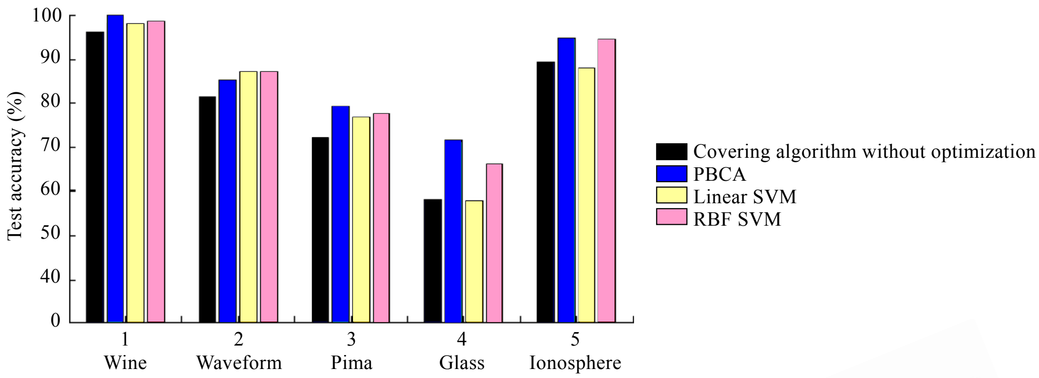

Fig.1 Histogram of test accuracy using four algorithms on five databases

图 1 5 个数据库 4 种不同算法的测试精度比较图

5 结 论

本文所给出的覆盖算法的概率模型解决了覆盖算法的全局优化问题,即将算法推广到概率模型而且引入了全局的优化过程(求最大似然拟合).

若将核覆盖算法看成是多分类的(核)SVM 方法,则在一定意义下解决了如何选取适当核函数参数的难题.

实际上,我们是利用已知的训练样本的分布情况,先利用覆盖算法近似地求出各类的的分界“面”(就是所得到的覆盖组),各类覆盖构成的区域大体上就是各类对应的分布的主要定义域.然后利用有限混合模型的最大似然拟合方法求出各类的最后分布.这样分两步走,可以大幅度减少计算量.当然,这种方法也存在一些不足之处,如本方法利用覆盖算法求得的覆盖后,有限混合模型的分量个数基本上就确定下来,在后面的求解过程中只有少数的删除和合并过程,这个分量个数可能不是最合适的.故进一步可以在 EM 步骤后加上一个调整分量个数的计算,其思想是,定义两分量的相关系数,并给出相应的阈值,当两分量的相关系数小于阈值时,就将它们合并成一个分量.另外,也可定义两分量的偏差并给出对应的阈值,当偏差大于阈值时,就将一个分量分解为两个分量.这样不断地调整分量的个数,直到不必再分解和合并为止.

覆盖算法与 SVM 方法相比的优点是简单、直观、计算量低,另外,它能同时处理 k 类分类问题.而 SVM 方法基本上是二类分类的方法,要解决 k 类分类问题,就要将 k 类分类问题化成 $k-1$ 个二类分类问题.从覆盖算法的概率模型可以看出,对各类决策函数,在求最大似然拟合时基本上是分开进行的,这就大幅度降低了计算的复杂性.现将覆盖算法扩展成概率模型(即将确定型的模型扩展成概率模型),以提高算法抗噪声干扰的能力.而 SVM 算法虽然是从统计学习理论中推导出来的,但其算法本身(化成求二次规划问题)却是确定型的,故缺乏抗噪声干扰的能力.SVM 方法的优点是精度较高(因为它引入全局优化计算).

我们希望通过引入覆盖算法的概率模型并对它进行全局优化来提高覆盖算法的精度.从初步的模拟结果已看出概率覆盖算法的优越性,当然,仍须在今后的应用实践中继续进行验证和改进.另外,对于利用 k-维高斯(k>1)分布构造有限混合模型,在迭代过程中出现当方差矩阵Σ的逆矩阵不存在情况的处理,我们将另文加以讨论.

References:

[1] Zhang L, Zhang B. A geometrical representation of McCulloch-Pitts neural model and its applications. IEEE Trans. on Neural Networks, 1999,10(4):925−929.

[2] Wu T, Zhang L, Zhang YP. Kernel covering algorithm for machine learning. Chinese Journal of Computers, 2005,28(8): 1295−1301 (in Chinese with English abstract).

[3] Zhang YP, Zhang L,Wu T. A multi-side increase by degrees algorithm at machine learning. Acta Electronica Sinica, 2005,33(2):

327−331 (in Chinese with English abstract).

[4] Ye SZ, Zhang B, Wu MR, Zhang WB. A fuzzy classifier based on the constructive covering approach in neural networks. Journal of Software, 2003,14(3):429−434 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/14/429.htm

[5] Tao P, Zhang B, Ye Z. An incremental BiCovering learning algorithm for constructive neural network. Journal of Software, 2003,14(2):194−201 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/14/194.htm

[6] Wu MR, Zhang B. A neural network algorithm for large scale pattern recognition problems. Journal of Software, 2001,12(6): 851−855 (in Chinese with English abstract).

[7] Zhang M, Wu T, Wang LW, Cheng JX. The application of granularity of the quotient space theory in database and data warehouse. Journal of the Computer Engineering and Application, 2003,39(17):47−49,60 (in Chinese with English abstract).

[8] Wang LW, Wu T, Zhang M, Zhang L. Improvement on neighborhood covering algorithm and it’s application. Pattern Recognition and Artificial Intelligence, 2003,16(1):81−85 (in Chinese with English abstract).

[9] Zhang YP, Zhang L, Wu T, Xu F, Zhang M, Wang LW. A structural learning algorithm based on covering algorithm and its application in stock forecasting. Journal of Computer Research and Development, 2004,41(6):979−984 (in Chinese with English abstract).

[10] Zhang YP, Zhang L, Duan Z. A constructive kernel covering algorithm and applying it to image recognition. Journal of Image and Graphics, 2004,9(11):1304−1308 (in Chinese with English abstract).

[11] Zhou Y. A study of algorithm of neural networks as classifiers and their application in text classification [Ph.D. Thesis]. Hefei: Anhui University, 2006 (in Chinese with English abstract).

[12] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data using the EM algorithm (with discussion). Journal of the Royal Statistical Society (Series B), 1977,39(1):1−38.

[13] McLachlan GJ, Peel D. Finite Mixture Models. New York: John Wiley & Sons, 2000.

[14] Zhang L. Research on support vector machines and kernel methods [Ph.D. Thesis]. Xi’an: Xi’an University of Electronic Sciences, 2002 (in Chinese with English abstract).

附中文参考文献:

[2] 吴涛,张铃,张燕平.机器学习中的核覆盖算法.计算机学报,2005,28(8):1295−1231.

[3] 张燕平,张铃,吴涛.机器学习中的多侧面递进算法 MIDA.电子学报,2005,33(2):327−331.

[4] 叶少珍,张钹,吴鸣锐,郑文波.一种基于神经网络覆盖构造法的模糊分类器.软件学报,2003,14(3):429−434. http://www.jos.org.cn/1000-9825/14/429.htm

[5] 陶品,张钹,叶榛.构造型神经网络双交叉覆盖增量学习算法.软件学报,2003,14(2):194−201. http://www.jos.org.cn/1000-9825/14/194.htm

[6] 吴鸣锐,张钹.一种用于大规模模式识别问题的神经网络算法.软件学报,2001,12(6):851−855.

[7] 张旻,吴涛,王伦文,程家兴.商空间粒度计算理论在数据库和数据仓库中应用.计算机工程与应用,2003,39(17):47−49,60.

[8] 王伦文,吴涛,张旻,张铃.一种改进的领域覆盖算法与应用.模式识别与人工智能,2003,16(1):81−85.

[9] 张燕平,张铃,吴涛,徐锋,张旻,王伦文.基于覆盖的构造性学习算法(SLA)及在股票预测中的应用.计算机研究与发展,2004,41(6): 979−984.

[10] 张燕平,张铃,段震.构造性核覆盖算法在图像识别中的应用.中国图像图形学报,2004,9(11):1304−1308.

[11] 周瑛.神经网络作为分类器的算法研究及在信息检索中的应用[博士学位论文].合肥:安徽大学,2006.

[14] 张莉.支撑矢量机与核方法研究[博士学位论文].西安:西安电子科技大学,2002.

张铃(1937－),男,福建福清人,教授,博士生导师,主要研究领域为人工神经网络,智能计算,粒度计算的理论与应用.

周瑛(1968－),女,博士,教授,主要研究领域为模糊理论及应用,神经网络,信息检索.

吴涛(1970－),男,博士,副教授,主要研究领域为计算智能的理论与应用.

张燕平(1962－),女,博士,教授,主要研究领域为智能计算的理论及应用.