ISSN 1000-0054 清华大学学报（自然科学版）2007 年 第47 卷 第10 期 34/40
CN 11-2223/N J Tsinghua Univ（Sci & Tech），2007，Vol.47，No.10 1689-1691

支持向量机方法中加权后验概率建模方法

张 翔[1,2]，肖小玲[3]，徐光[illegible][1]

（1. 清华大学 计算机科学与技术系，北京 100084；2. 长江大学 地球物理与石油资源学院，荆州 434023；
3. 武汉理工大学 计算机科学与技术学院，武汉 430063）

摘 要：为解决传统支持向量机方法不提供概率输出的问题，在支持向量机多类分类问题输出概率建模中，提出了加权后验概率建模方法。该方法在对多个两类支持向量机分类器的输出概率进行组合时，充分考虑了各个两类支持向量机分类器的差异。依据Bayesian 理论，采用后验概率作为各个两类支持向量机分类器的权系数。实验结果表明，与投票法及Pairwise Coupling 方法相比，加权后验概率方法具有较低的分类错误率，不仅具有较好的分类性能，而且得到的后验概率具有较好的概率分布形态。该方法有效地解决了实际多类分类问题中支持向量机的概率建模问题。

关键词：支持向量机；概率建模；多类分类器；后验概率

中图分类号：TP 391 **文献标识码**：A

文章编号：1000-0054(2007)10-1689-03

Weighted posterior probability output for support vector machines

ZHANG Xiang[1,2]，XIAO Xiaoling[3]，XU Guangyou[1]

（1. Department of Computer Science and Technology，Tsinghua University，Beijing 100084，China；
2. Department of Geophysics and Oil Resources，Yangtze University，Jingzhou 434023，China；
3. School of Computer Science and Technology，Wuhan University of Technology，Wuhan 430063，China）

Abstract：A weighted posterior probability method is presented to calculate the probability outputs of support vector machines（SVMs）for multi-class cases. The differences and weights for combination of the probabilty output among these two-class classifiers calculated from the posterior probability are given based on the Bayesian theory. Tests show that the weighted posterior probability method has less classification errors，better classification ability，and a better probability distribution of the posterior probability than the voting method or the Pairwise Coupling method. This method effectively provides probability outputs of SVMs in the multi-class case.

Key words：support vector machines；probability modeling；multi-class classifier；the posterior probability

在解决样本分类的不确定性时，一般对分类结果采用概率的方式输出。传统的支持向量机方法不提供后验概率的输出，最早考虑支持向量机（SVM）样本后验概率的学者是Platt [1]。两类分类问题的概率建模已经得到较好的解决，主要有两大类方法。一类是采用Bayes 框架理论，先求各类的类条件概率密度，假设其满足Gauss 分布，再依据Bayes 理论求出其后验概率[2,3]。另一类不计算类概率密度，直接拟合后验概率[1]。实际的分类问题主要是多类分类问题，在传统支持向量机方法的多类分类问题中，最成熟最有效的算法为“一对一”分类法[4]。对多类分类问题的概率建模，目前主要采用将多个两类分类的概率结果进行组合的方式进行概率建模[5-7]。最常用的组合方法是投票法[5]，该方法在对各两类分类器的分类结果进行组合时，将各分类器的重要性同等看待，即在合并每个两类分类器得到的后验概率时，没有考虑另一类样本出现的概率情况，因此，该方法得到的是一种近似的后验概率。文[7] 提出了一种Pairwise Coupling 方法，该方法由于要解一个优化问题，计算代价比较大，不利于支持向量机方法的实时处理。

本文针对上述存在的问题，在多类问题的输出概率建模中，根据Bayesion 理论，提出了加权后验概率方法。

1 加权后验概率方法

由于各两类分类器中样本以及样本的分布情况

收稿日期：2006-09-15

基金项目：国家自然科学基金资助项目（60273005）；
中国博士后科学基金项目（2005038351）；
湖北省自然科学基金项目（2004ABA043）；
湖北省教育厅科学技术研究重点项目（D200612002）

作者简介：张翔（1969—），男（汉），湖北，博士，副教授，博士后。
E-mail：zx-jr-xl@163.com

不一样,在统计每个两类支持向量机分类器中,样本属于某类的后验概率之和时,需要考虑各个两类支持向量机分类器之间的差异。而投票法在确定测试样本x属于第i类的最终后验概率为

$$P(i|x)=\frac{\sum_{j=1,j\neq i}^{N}P_{iaj}(i|j;x)}{\sum_{k=1}^{N}\sum_{j=1,j\neq k}^{N}P_{kaj}(k|j;x)},\quad i=1,2,\cdots,N. \tag{1}$$

其中,$P_{iaj}(i|j;x)$表示由第i类和第j类构成的两类支持向量机分类器,计算得到的x属于第i类的后验概率。

由于

$$\sum_{k=1}^{N}\sum_{j=1,j\neq k}^{N}P_{kaj}(k|j;x)=\frac{N(N-1)}{2}, \tag{2}$$

则投票法中后验概率公式(1)变为

$$P(i|x)=\frac{2}{N(N-1)}\sum_{j=1,j\neq i}^{N}P_{iaj}(i|j;x),\quad i=1,2,\cdots,N. \tag{3}$$

通过对式(1)的分析可以看出,可将$P_{iaj}(i|j;x)$看作为:在第i类与第j类构成的两类支持向量机分类器中,在第j类的条件下,样本x属于第i类的条件后验概率。在统计每个两类支持向量机分类器中,样本属于第i类的后验概率之和时,没有考虑两类支持向量机中第j类样本出现的概率情况。因此,本文提出对式(1)进行修正。在对多个两类支持向量机分类器的输出概率进行组合时,充分考虑各个两类支持向量机分类器的差异,并采用另一类(第j类)样本的后验概率作为各个两类支持向量机分类器的权系数。则测试样本属于第i类的最终后验概率为

$$P(i|x)=\sum_{j=1,j\neq i}^{N}P_{iaj}(i|j;x)P(j|x),\quad i=1,2,\cdots,N. \tag{4}$$

将式(4)与式(3)进行比较,可以看出,投票法可看作是本文提出的加权后验概率方法的特例,即投票法可以看作是权系数都等于$\frac{2}{N(N-1)}$的加权近似方法。相对于投票法,在对各两类分类器的分类结果进行组合时,加权后验概率方法考虑了各两类分类器重要性的差异,并用样本的另一类后验概率作为其权系数。

由Bayesian 公式可知,测试样本x属于第i类的最终后验概率为

$$P(i|x)=\sum_{j=1,j\neq i}^{N}P_{iaj}(i|j,x)P(j|x),\quad i=1,2,\cdots,N. \tag{5}$$

简记$P_i=P(i|x)$,$P_j=P(j|x)$,$P(i|j)=P_{iaj}(i|j,x)$,则式(5)可简化为

$$P_i=\sum_{j=1,j\neq i}^{N}P(i|j)P_j,\quad i=1,2,\cdots,N. \tag{6}$$

可以看出,式(6)即为式(4)的简化表示形式。

将式(6)进行整理变为

$$\begin{pmatrix}1 & -P(1|2) & \cdots & -P(1|N)\\ -P(2|1) & 1 & \cdots & -P(2|N)\\ \vdots & \vdots & \cdots & \vdots\\ -P(N|1) & -P(N|2) & \cdots & 1\end{pmatrix}\times\begin{pmatrix}P_1\\P_2\\\vdots\\P_N\end{pmatrix}=\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix}. \tag{7}$$

式(7)可看作是求解N个未知变量P_i,$i=1,2,\cdots,N$的N个方程组。其中:P_i,$i=1,2,\cdots,N$为样本x属于第i类的后验概率;$P(i|j)$与$P(j|i)$分别表示为在由第i类和第j类构成的两类分类器中,样本x属于第i类和第j类的条件后验概率,由文[8]提出的两类分类器中输出概率建模中的直接拟合方法得到。

$P(i|j)$与$P(j|i)$之间满足以下关系:

$$P(i|j)=1-P(j|i). \tag{8}$$

在式(7)中,P_i,$i=1,2,\cdots,N$需要满足约束条件:

$$\sum_{i=1}^{N}P_i=1. \tag{9}$$

在式(9)约束条件下,通过直接求解式(7),即可得到样本x在各类中的后验概率P_i,$i=1,2,\cdots,N$。

具体在对式(7)进行求解时,本文采用将式(7)与式(9)组合为一个超定方程组的求解问题,采用线性最小二乘问题求其最小二乘解,从而得到多类问题的输出概率建模中,样本x在各类中的后验概率P_i,$i=1,2,\cdots,N$。

由于式(7)中矩阵系数为由不同的两类分类器计算得到的,样本x属于某类的条件后验概率,因此,由式(7)与式(9)组合为一个超定方程组,其系数矩阵的列向量组是线性无关,则得到的最小二乘解是唯一的。

2 实验结果及分析

为了评价本文提出加权后验概率方法的分类性能,本文采用文[9]提供的具有不同分类数,不同特

征数的4种测试数据(Segment, Waveform, Usps及Mnist), 将它们与硬输出、概率建模输出中的投票法及文[7]提出的Pairwise Coupling方法等4种方法分别进行了对比实验。

表1为采用文[9]提供的4种测试数据, 多类分类中3种支持向量机概率输出与其硬输出的分类错误率对比表。在支持向量机中, 选择Gauss核函数, 其参数如表1所示, 惩罚因子C= 1000。

表1 4种支持向量机概率输出与其硬输出的分类错误率

数据类型	分类错误率/%				σ^2
	硬输出	投票法	Pairwise Coupling 法	加权后验概率法	
Segment	7.0	2.5	2.5	2.3	1.0
Waveform	15.0	7.3	6.3	6.2	1.0
Usps	9.7	4.0	4.0	3.6	50.0
Mnist	14.4	5.3	5.3	4.7	50.0

由表1数据对比可以看出, 相对于支持向量机硬输出,3种概率输出方法具有较低的分类错误率, 都取得了比硬输出更好的分类结果。在3种输出概率建模方法中, 与投票法与Pairwise Coupling方法的分类效果相比, 本文提出的加权后验概率方法具有最好的分类精度。

下面将采用本文提出的加权后验概率方法与投票法及Pairwise Coupling方法等3种输出概率建模方法, 得到每类的概率分布进行比较。图1为采用3种概率输出方法, Mnist数据中某个测试样本在10个类别中的概率分布。

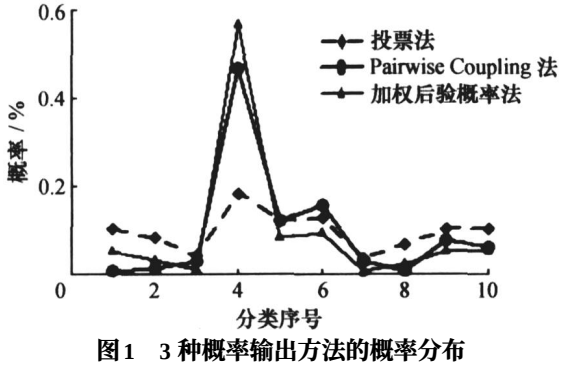

图1 3种概率输出方法的概率分布

由图1可以看出, 即使由3种输出概率建模方法, 确定样本的最终分类结果相同, 但样本属于各类的概率分布不相同。相对于投票法与Pairwise Coupling方法, 本文提出的加权后验概率方法具有较好的概率分布, 主要表现在样本确定的类中具有较高的概率, 而在其他类中的概率相对较低。如图1中, 采用加权后验概率方法进行概率建模, 样本属于第4类的概率为57%, 属于其他类的概率都低于10%。这种概率分布有利于解决不易确定样本类别的问题。

3 结 论

本文在对支持向量机多类分类问题进行概率建模时, 在对投票法进行分析的基础上, 研究了加权后验概率建模方法。相对于投票法, 加权后验概率方法不仅具有更好的分类精度, 还具有更好的概率分布形式, 主要表现在样本确定的类中具有较高的概率, 而在其他类中的概率相对较低, 这种概率分布有利于解决当样本属于各类的概率出现相同时不易确定样本类别的问题。该方法不仅使支持向量机的分类精度得到了提高, 还提供了样本属于所在类中的可信程度。

参考文献 (References)

[1] Platt J.Probabilities for support vector machines [C]// Smola A, Bartlett P, Scholkopf B. Advances in Large Margin Classiers. Cambridge MA: MIT Press, 2000: 61-74.

[2] Sollich P. Bayesian methods for support vector machines: Evidence and predictive class probabilities [J]. *Machine Learning*, 2002, **46**: 21-52.

[3] Kwok J T Y. Moderating the outputs of support vector machine classifiers [J]. *IEEE Trans on Neural Networks*, 1999, **10**(5): 1018-1031.

[4] Hsu C W, Lin C J. A comparison of methods for multi-class support vector machines [J]. *IEEE Transactions on Neural Networks*, 2002, **13**(2): 415-425.

[5] Moreira M, Mayoraz E. Improved pairwise coupling classification with correcting classifiers [C]// Proc Of 10th European Conference on Machine Learning. Chemnitz: Springer-Verlag, 1998: 160-171.

[6] Xin D, Wu Z H. Speaker recognition using continuous density support vector machines [J]. *Electronics Letters*, 2001, **37**(17): 1099-1101.

[7] Wu T F, Lin C J, Weng R C. Probability estimates for multi-class classification by pairwise coupling [J]. *Journal of Machine Learning Research*, 2004, **5**: 975-1005.

[8] 张翔, 肖小玲, 徐光祐. 基于最大熵估计的支持向量机概率建模 [J]. 控制与决策, 2006, **21**(7): 767-770.
ZHANG Xiang, XIAO Xiaoling, XU Guangyou. Probabilistic outputs for support vector machines based on the maximum entropy estimation [J]. *Control and Decision*, 2006, **21**(7): 767-770. (in Chinese)

[9] Hsu C W. Department of Computer Science [EB/OL]. 2006. National Taiwan University, Taipei 106, China. http://www.csien.tue.dut.w/~cjlin/libsvmtools/.