第 34 卷 第 4 期 计 算 机 工 程 2008 年 2 月
Vol.34 No.4 Computer Engineering February 2008
·博士论文· 文章编号：1000—3428(2008)04—0048—03 文献标识码：A 中图分类号：TP391

支持向量的信息冗余和 SVM 改进方法

彭 兵，周建中，安学利，向秀桥，罗志猛
(华中科技大学水电与数字化工程学院，武汉 430074)

摘 要：在研究 RBF 核函数的几何特性和分析 SVM 数据依赖性改进方法的基础上，提出了支持向量携带数据冗余信息的论点。冗余信息掩盖了所研究对象的特征，影响 SVM 的性能。基于黎曼几何的 SVM 数据依赖性改进方法能够剔除支持向量携带的冗余信息，改进 SVM 的性能。理论分析和实验研究表明，该方法能够有效提高 SVM 的分类能力和分类速度。

关键词：核函数；冗余信息；支持向量机；黎曼几何

Redundant Information in Support Vectors and Improved Support Vector Machine

PENG Bing, ZHOU Jian-zhong, AN Xue-li, XIANG Xiu-qiao, LUO Zhi-meng
(College of Hydropower & Information Engineering, Huazhong University of Science and Technology, Wuhan 430074)

【Abstract】 This paper proposes support vectors including redundant information after analyzing geometrical structure of RBF kernel function and data dependent way for improved Support Vector Machine(SVM). Redundant information confuses the law of a learning problem. It can be excluded with data dependent way based on Riemannian geometry for improved SVM. Experimental results show remarkable improvement on classification ability and classification speed of SVM, supporting this idea.

【Key words】 kernel function; redundant information; support vector machine; Riemannian geometry

1 概述

支持向量机(Support Vector Machine, SVM)是Vapnik于20 世纪 90 年代提出的一种智能学习机器。它以统计学习理论为基础，用结构风险最小化评价标准取代传统的经验风险最小化评价标准，从而实现了容量控制，提高了泛化推广能力[1-2]。支持向量机是引入核函数来解决非线性问题的。由于核函数是以映射函数的内积形式出现，因此不需要求得显式的映射函数，这就相当于直接在输入空间解决了非线性问题。核函数是由分类器决策函数的平滑度假定所决定的[3]。如果数据输入空间的平滑度先验知识可知，那么就可以利用这些知识来选择一个性能优良的核函数。否则，只能通过数据依赖性方法来改进核函数[4]。文献[5-6]以黎曼几何为基础提出了SVM数据依赖性改进方法，其基本思想是扩大支持向量(Support Vector, SV)附近的体积微元以提高SVM的分类能力。试验结果表明该方法确实有效提高了SVM的分类能力。随着研究的深入，提出了两个问题：(1)该方法是否对分类速度有影响？(2)该方法对支持向量机性能的改善是否有其他原因？本文针对RBF核函数，对SVM数据依赖性改进方法进行了分析。通过分析，发现剔除冗余支持向量是该方法改进SVM性能的重要原因。运用该方法能够改善SVM的分类能力，同时也能够提高支持向量机的分类速度。

2 机器学习的基本理论

2.1 机器学习的实质

机器学习问题模型可以用 3 个要素来描述[1]：

(1)随机输入向量 $\boldsymbol{x}$。它是多维空间中的独立点，具有固定而未知的分布特性 $P(\boldsymbol{x})$。

(2)观测器。它会根据固定而未知的条件概率 $P(\boldsymbol{y}|\boldsymbol{x})$输出对应输入向量 $\boldsymbol{x}$ 的输出向量 $\boldsymbol{y}$。

(3)学习机器的容量。它包含了一系列的决策函数 $f(x,\alpha),\alpha\in\Lambda$，$\Lambda$ 是参数集。

学习问题的实质就是以 $P(\boldsymbol{x},\boldsymbol{y})=P(\boldsymbol{x})P(\boldsymbol{y}|\boldsymbol{x})$概率分布的随机独立样本 $(x_1,y_1),(x_2,y_2),\cdots(x_n,y_n)$ 训练学习机器，从决策函数集 $f(x,\alpha),\alpha\in\Lambda$ 中选择最优的决策函数，使观测器的输出结果最接近实际情况。

由此可知，数据样本包含了反映所研究对象的特征信息 $P(\boldsymbol{x},\boldsymbol{y})$。但是，数据样本还包含了冗余信息。由这样的数据样本训练建立的学习机器反映的规律与真实的规律总是存在偏差，这种偏差可以用风险函数 $R(\alpha)$来表示，如式(1)所示。机器学习的目标就是在 $P(\boldsymbol{x},\boldsymbol{y})$未知、所有信息都包含在训练样本中的情况下，剔除冗余信息，使风险函数的值达到最小。

$$R(\alpha)=\int L(y,f(x,\alpha))\mathrm{d}P(\boldsymbol{x},\boldsymbol{y}) \tag{1}$$

2.2 核的几何特性

已经知道，SVM 通过映射函数 $\Phi(x)$可以将数据由输入空间映射到特征空间，则非线性 SVM 的决策函数在输入空间可以描述为

$$f(x)=\sum_{i\in SV}\alpha_i y_i K(x,x_i)+b \tag{2}$$

基金项目：高等学校博士学科点专项科研基金资助项目(20050487062)；国家自然科学基金资助项目(50579022)；国家自然科学基金资助重点项目(50539140)

作者简介：彭 兵(1978－)，男，博士研究生，主研方向：机器学习，故障诊断；周建中，教授、博士生导师；安学利、向秀桥、罗志猛，博士研究生

收稿日期：2007-02-28 **E-mail：**prof.zhou.hust@263.net

$$K(x,x_i)=\Phi(x)\bullet\Phi(x_i) \tag{3}$$

其中，α_i是一个正数，代表第i个支持向量SV的概率分布；b是阈值。

下面首先分析核函数作用下输入空间的几何结构。映射函数$\Phi(x)$通过定义曲面流形把输入空间S嵌入到特征空间F。当F是Euclidean空间或者 Hilbert空间时，输入空间S中就产生了黎曼度规$g_{ij}(x)$，它使得输入空间S中的微元dx在特征空间中得到放大[5]。

令 z 为 x 在特征空间中的映射值，则微元在特征空间中的映射值可以表示为

$$\mathrm{d}z=\nabla\Phi(x)\bullet\mathrm{d}x=\sum_i\left(\frac{\partial}{\partial x_i}\Phi(x)\right)\mathrm{d}x_i \tag{4}$$

$|\mathrm{d}z|^2$的二次项形式如下：

$$|\mathrm{d}z|^2=\sum_\alpha(\mathrm{d}z_\alpha)^2=\sum_{i,j}g_{ij}(x)\mathrm{d}x_i\mathrm{d}x_j \tag{5}$$

$$g_{ij}(x)=\left(\frac{\partial}{\partial x_i}\Phi(x)\right)\bullet\left(\frac{\partial}{\partial x_j}\Phi(x)\right) \tag{6}$$

称为黎曼度规。

由式(3)可知

$$\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j'}K(x,x')=\nabla\Phi(x)\bullet\nabla\Phi(x') \tag{7}$$

所以，黎曼度规和核函数有如下关系：

$$g_{ij}(x)=\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j'}K(x,x') \tag{8}$$

在黎曼空间中，体积微元可表示为

$$\mathrm{d}V=\sqrt{g(x)}\mathrm{d}x_1,\mathrm{d}x_2,\cdots,\mathrm{d}x_n \tag{9}$$

其中，$g(x)$=det$|g_{ij}|$；因子 $\sqrt{g(x)}$ 代表了输入空间S中的局部区域通过映射函数向特征空间F映射时的缩放过程，因此，它可以被称为缩放因子。

3 数据依赖性改进方法

3.1 一种 SVM 数据依赖性改进方法

基于黎曼几何的数据依赖性改进方法的基本思想是：通过黎曼度规构成的缩放因子，在特征空间中扩大分类边界处的体积，使得分类间隔扩大，从而提高 SVM 的分类能力。实际上，获取分类边界的精确描述并不容易，支持向量 SV 的位置决定了分类边界的形状。所以，扩大 SV 附近的空间体积，而缩小其他点处的空间体积，就能够扩大分类间隔，实现对 SVM 的改进。

在实现算法中，运用了保角变换 $D(x)$来改进核函数。运用保角变换的好处是整个空间的角度不会发生变化，各个数据点之间的空间关系不会受到影响。经过保角变换，核函数变为

$$\widetilde{K}(x,x')=D(x)D(x')K(x,x') \tag{10}$$

RBF核函数经过保角变换，其黎曼度规$g_{ij}(x)$变为

$$\widetilde{g}_{ij}(x)=D^2(x)g_{ij}(x)+D_i(x)D_j(x) \tag{11}$$

其中，$D(x)$是一个正的标量函数：

$$D(x)=\sum_{i\in SV}\exp\left(\frac{-\|x-\boldsymbol{x}_i\|^2}{\tau_i^2}\right) \tag{12}$$

$$D_i(x)=\frac{\partial D(x)}{\partial\boldsymbol{x}_i} \tag{13}$$

其中，τ_i是控制参数，表示支持向量$\boldsymbol{x}_i$与其邻近的M个同类SV之间距离的均值：

$$\tau_i^2=\frac{1}{M}\sum_\alpha\|\boldsymbol{x}_\alpha-\boldsymbol{x}_i\|^2 \tag{14}$$

由式(12)和式(13)可知，输入向量$\boldsymbol{x}$越靠近支持向量$\boldsymbol{x}_i$，则对应的$D(x)$和$D_i(x)$的值就会越大，黎曼度规 $\widetilde{g_{ij}}(x)$ 也就越大。所以，x距离分类边界越近，x附近的空间体积就扩张得越大。该数据依赖性改进方法具体算法如下：

(1)用核函数 $K(x,x')$ 训练 SVM，获取 SV 的相关信息；

(2)运用式(10)、式(12)、式(14)计算获取改进核函数 $\tilde{K}(x,x')$ ；

(3)重复执行步骤(1)和步骤(2)，直到获得最好的结果。

3.2 关于数据依赖性改进方法的新观点

支持向量 SV 在支持向量机 SVM 的分类边界上的分布不是均匀的。在 SV 分布密集的地方就存在冗余 SV。冗余 SV 指对分类边界描述作用不大的 SV。在第 3 节的实验中，可以清楚地看到冗余 SV 的存在。

如式(2)所示，SVM 的分类计算涉及 SV 的乘积和求和运算。如表 1 和表 2 所示，SV 的个数越多，分类计算的复杂度越高，分类速度就越低，如果能够剔除冗余 SV，就能够提高 SVM 的分类速度，见表 2，表中：

$$a_m=\frac{1}{6}\sum_{i=1}^{6}a_i,\quad \Delta a=a_0-a_m,\quad r=\frac{\Delta a}{a_0},\quad \Delta a_{\max}=a_0-a_{\min},\quad r_{\max}=\frac{\Delta a_{\max}}{a_0}。$$

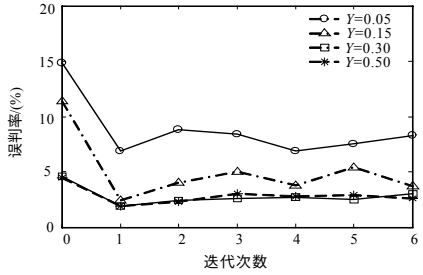
表 1　分类能力改进情况(误判率)

γ	初始值 a_0/(%)	平均值 a_m/(%)	差值 Δa/(%)	改进率 r/(%)	最小值 $a_{\min}$/(%)	最大差值 $\Delta a_{\max}$/(%)	最大改进率 $r_{\max}$/(%)
0.05	14.8	7.8	7.0	47.3	6.9	7.9	53.4
0.15	11.4	4.1	7.3	64.0	2.4	9.0	78.9
0.30	4.6	2.5	2.1	45.7	1.9	2.7	58.7
0.50	4.5	2.6	1.9	42.2	1.9	2.6	57.8

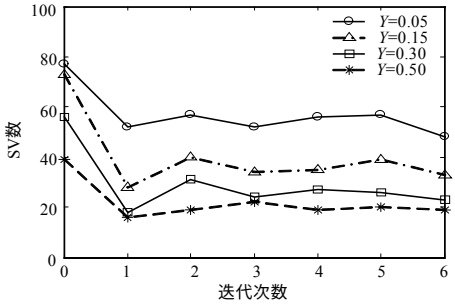
表 2　分类速度改进情况(SV 的个数)

γ	初始值 a_0(%)	平均值 a_m(%)	差值 Δa(%)	改进率 r/(%)(%)	最小值 $a_{\min}$(%)	最大差值 $\Delta a_{\max}$(%)	最大改进率 $r_{\max}$/(%)
0.05	77	53.7	23.3	30.3	48	29	37.7
0.15	73	34.8	38.2	52.3	28	45	61.6
0.30	56	24.8	31.2	55.7	18	38	67.9
0.50	39	19.2	19.8	50.8	16	23	60.0

为改进分类功能，笔者采用了图 1 所示的改进过程。



(a)识别率



(b)SV 数

图 1　SVM 改进过程

冗余 SV 对正确描述类型边界不起作用，它们只会使类型边界的描述模糊不清，使得 SVM 的分类能力低下。通过本文的改进方法，采用保角变换，增大了 SV 附近的空间体积，剔除了冗余 SV，起到了类型边界的特征提取作用，使得分类能力得到提高。

4 实验与讨论

为了证实第 2 节中的论述，研究工作进行了分类实验，获取支持向量存在信息冗余的证据。在[-0.5, 0.5]×[-0.5, 0.5]的区域内，随机抽取孤立点作为样本。这些样本由正弦曲线 y=0.5 sin(2 πx)分为正类和负类。其中 200 组样本作为训练样本，1 000 组样本作为测试样本。保角变换的同类邻近 SV 个数 M 取 3。本文采用 SVM 的误判率来表示 SVM 的分类能力，误判率即误判的样本数与总的测试样本数之比。SV 的个数代表 SVM 的分类速度。

实验采用了 Libsvm-2.82 作为 SVM 的训练和测试工具，运用 VC++6.0 编程实现核函数的迭代改进。SVM 采用的核函数是 RBF 核函数，如式(15)所示。

$$K(x,x_i)=\exp(-\gamma\bullet\|x-x_i\|^2) \tag{15}$$

实验讨论 1 实验结果表明，SV 个数的减少与 SVM 分类能力的提高具有相关性，这说明冗余 SV 的存在。根据表 1 、表 2 和图 1 可知，在不同的核参数值 γ 的作用下，SVM 通过改进，SV 个数都大幅减少。以 γ=0.05 为例，SV 个数减少 30.3%，同时，SVM 的分类能力明显提高，误判率降低 47.3%。SV 个数变化和 SVM 分类能力变化的相异性可以说明冗余 SV 的存在。

实验讨论 2 本文的方法在 SVM 分类能力和分类速度上都有良好的改进效果。改进后的 SVM 的误判率和 SV 的个数都低于初始值。

改进后的 SVM 在分类能力和分类速度上都有较大改进。见表 1，改进后 SVM 分类能力最大提高了 78.9%(γ=0.15)。见表 2 ，SV 个数大幅减少，SVM 的分类速度大幅提高，分类速度最大提高 67.9%(γ=0.3)。

如图 1 所示，改进的迭代过程是一个收敛的过程，改进结果能够达到较好的水平。以最差情况 γ=0.05 为例，改进过程的平均误判率为 7.8%，比改进前降低了 47.3%。改进过程的平均 SV 个数为 53.7 个，比改进前减少了 30.3%。

图 2、图 3 示出了改进前后的 SV 分布。

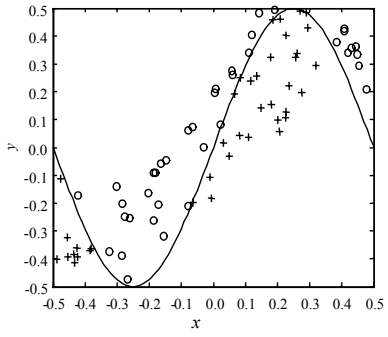


图 2 γ=0.05 时改进前 SV 分布

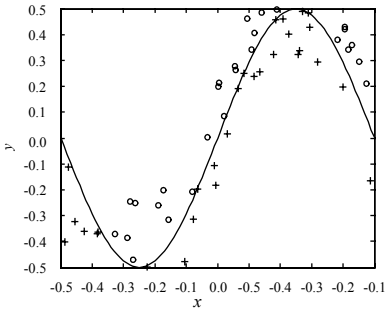


图 3 γ=0.05 时一次迭代改进后 SV 分布

实验讨论 3 通过改进，SV 沿分类边界分布更加均匀。如图 2 所示，在(-0.5, -0.4)×(-0.5, -0.35)区域内，以及区域(-0.3, -0.1)×(-0.2, 0)和区域(0.15, 0.3)×(0, 0.2)内，SV 分布密集，对类型边界的描述作用不大，属于冗余 SV。如图 3 所示，通过一次迭代改进，剔除了这些区域的冗余 SV，SV 沿分类边界的分布更均匀，使得分类边界的特征更突出，SVM 的误判率降低了 53.4%。

5 结束语

本文采用的改进方法以黎曼几何为理论基础，运用保角变换增大 SV 附近的空间体积，使得 SVM 分类间隔增大，从而提高了 SVM 的分类能力。

此外，该改进方法又起到剔除冗余 SV、提取类型边界特征的作用，因此，能够提高 SVM 的分类能力。同时，SV 参与了分类计算，SV 的个数决定了 SVM 的分类速度，因此，改进方法也提高了 SVM 的分类速度。

实验发现，SV 个数减少，SVM 的分类能力会提高，两者的变化具有相关性，这说明冗余 SV 的存在。实验表明，改进后的 SVM 在分类能力和分类速度上都有较大改进，并能够维持在较好的水平。

本文的改进方法引入了距同类邻近SV的平均距离τ_i作为控制参数，在SVM分类模型含有较多冗余SV的区域内，保角变换会使扩大因子增大，使SV沿分类边界分布更均匀，分类边界特征更明显，从而使SVM的分类能力提高。

参考文献

[1] Vapnik V N. An Overview of Statistical Learning Theory[J]. IEEE Trans. on Neural Networks, 1999, 10(5): 988-999.

[2] 周伟达, 张 莉, 焦李成. 一种改进的推广能力衡量准则[J]. 计算机学报, 2003, 26(5): 598-604.

[3] Seeger M. Bayesian Model Selection for Support Vector Machines, Gaussian Processes and Other Kernel Classifiers[M]. [S. l.]: MIT Press, 2000: 603-609.

[4] Scholkopf B, Simard P, Smola A, et al. Prior Knowledge in Support Vector Kernels[M]. [S. l.]: MIT Press, 1998.

[5] Amari S, Wu S. Improving Support Vector Machine Classifiers by Modifying Kernel Function[J]. Journal of Neural Networks, 1999, (12)6: 783-789.

[6] Wu Si. Conformal Transformation of Kernel Functions: A Data-dependent Way to Improve Support Vector Machine Classifiers[J]. Neural Processing Letters, 2002, 15(2): 59-67.