计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2014.20130825
51(10):2171-2177,2014

基于近似高斯核显式描述的大规模 SVM 求解

刘 勇 江沙里 廖士中
(天津大学计算机科学与技术学院 天津 300072)
(szliao@tju.edu.cn)

Approximate Gaussian Kernel for Large-Scale SVM

Liu Yong, Jiang Shali, and Liao Shizhong
(*School of Computer Science and Technology, Tianjin University, Tianjin* 300072)

Abstract Training support vector machine (SVM) with nonlinear kernel functions on large-scale data is usually very time consuming. In contrast, there exist faster solvers to train the linear SVM. To utilize the computational efficiency of linear SVM without sacrificing the accuracy of nonlinear ones, in this paper, we present a method for solving large-scale nonlinear SVM based on an explicit description of an approximate Gaussian kernel. We first give the definition of the approximate Gaussian kernel, and establish the connection between approximate Gaussian kernel and Gaussian kernel, and also derive the error bound between these two kernel functions. Then, we present an explicit description of the reproducing kernel Hilbert space (RKHS) induced by the approximate Gaussian kernel. Thus, we can exactly depict the structure of the solutions of SVM, which can enhance the interpretability of the model and make us more deeply understand this method. Finally, we explicitly construct the feature mapping induced by the approximate Gaussian kernel, and use the mapped data as input of linear SVM. In this way, we can utilize existing efficient linear SVM to solve non-linear SVM on large-scale data. Experimental results show that the proposed method is efficient, and can achieve comparable classification accuracy to a normal nonlinear SVM.

Key words support vector machine; linear support vector machine; kernel methods; approximate Gaussian kernel; reproducing kernel Hilbert space

摘 要 大规模数据集上非线性支持向量机(support vector machine, SVM)的求解代价过高,然而对于线性 SVM 却存在高效求解算法.为了应用线性 SVM 高效求解算法求解非线性 SVM,并保证非线性 SVM 的精确性,提出一种基于近似高斯核显式描述的大规模 SVM 求解方法.首先,定义近似高斯核并建立其与高斯核的关系,推导近似高斯核与高斯核的偏差上界.然后给出近似高斯核对应的再生核希尔伯特空间(reproducing kernel Hilbert space, RKHS)的显式描述,由此可精确刻画 SVM 解的结构,增强 SVM 方法的可解释性.最后显式地构造近似高斯核对应的特征映射,并将其作为线性 SVM 的输入,从而实现了用线性 SVM 算法高效求解大规模非线性 SVM.实验结果表明,所提出的方法能提高非线性 SVM 的求解效率,并得到与标准非线性 SVM 相近的精确性.

关键词 支持向量机;线性支持向量机;核方法;近似高斯核;再生核希尔伯特空间

中图法分类号 TP181; TP301

支持向量机(support vector machine, SVM)[1-3]是重要的机器学习方法.该方法利用核诱导的特征映射将输入空间中的数据映射到特征空间,继而在特征空间中训练线性学习器[2].

收稿日期:2013-06-18;**修回日期**:2013-08-23
基金项目:国家自然科学基金项目(61170019);天津市自然科学基金项目(11JCYBJC00700)
通信作者:廖士中(szliao@tju.edu.cn)

近年来,大规模线性 SVM 的高效求解方法得到长足发展. Zhang[4] 提出应用随机梯度下降算法来求解线性 SVM;Shalev-Shwartz 等人[5] 进一步应用次梯度(sub-gradient)下降算法给出更有效的求解算法;Joachims[6] 基于截平面方法(cutting plane)提出 SVM-Perf 算法;Smola 等人[7] 进一步提出丛方法(bundle method);Lin 等人[8] 应用置信域(trust region)来求解大规模逻辑斯蒂回归(logistic regression);Chang 等人[9] 采用坐标下降(coordinate descent)法来求解 L2-SVM;Hsieh 等人[10] 应用对偶坐标梯度算法(dual coordinate)求解线性 SVM;Yu 等人[11] 将上述方法推广到逻辑斯蒂回归上.

虽然已有线性 SVM 高效求解算法,但缺少处理非线性 SVM 的高效算法[12-14]. 如何应用线性 SVM 求解算法来设计非线性 SVM 求解算法得到关注[12,15-17]. 将特征映射显式地作为线性 SVM 的输入,能将非线性 SVM 转化为特征空间中的线性 SVM,从而可应用高效线性 SVM 算法求解非线性 SVM. Vedaldi 和 Zisserman[17] 研究可加核(additive kernel function)特征映射的近似构造;Chang 等人[12] 给出多项式核特征映射的显式构造;Rahimi 和 Recht[15-16] 基于随机特征(random features)来构造平移不变核的特征映射;Yang 等人[18] 应用低阶泰勒展开来近似高斯核,并加速共轭梯度优化算法;Cao 等人[19] 同样应用低阶泰勒展开来近似高斯核,加速 SVM 预测. 这些工作没有给出该近似核特征映射的显示构造,从而无法直接应用线性 SVM 算法来求解非线性 SVM.

本文提出一种基于近似高斯核显示描述的大规模 SVM 求解方法. 首先,给出近似高斯核的定义,推导近似高斯核与高斯核偏差的上界. 然后显式描述近似高斯核再生核希尔伯特空间的结构,精确描述 SVM 解的空间. 最后,显式构造近似高斯核的特征映射,并将特征数据(mapped data)作为线性 SVM 的输入,应用线性 SVM 高效求解算法来求解非线性 SVM.

1 预备知识

本节介绍再生核希尔伯特空间(reproducing kernel Hilbert space, RKHS)与再生核(reproducing kernel)的基本概念.

记 $\mathbb{R}_+=\{x:x\in\mathbb{R},x>0\}$, $\mathbb{N}_+=\{x:x\in\mathbb{N},x>0\}$. 对于 $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_n)^{\mathrm{T}}\in(\mathbb{N}_+\cup\{0\})^n$ 和 $\boldsymbol{x}=(x_1,\cdots,x_n)$, 记 $|\boldsymbol{\alpha}|=\sum_{i=1}^{n}\alpha_i$, $\boldsymbol{x}^{\boldsymbol{\alpha}}=\prod_{i=1}^{n}x_i^{\alpha_i}$ 及 $C_{\boldsymbol{\alpha}}^{k}=k!\Big/\prod_{i=1}^{n}\alpha_i!$.

设 $\mathcal{X}\subseteq\mathbb{R}^n$ 为非空集合,若存在一个希尔伯特空间 $\mathcal{H}$ 及映射 $\boldsymbol{\Phi}:\mathcal{X}\to\mathcal{H}$ 使得对 $\forall\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}$,都有 $K(\boldsymbol{x},\boldsymbol{x}')=\langle\boldsymbol{\Phi}(\boldsymbol{x}'),\boldsymbol{\Phi}(\boldsymbol{x})\rangle$,则称 K 为 $\mathcal{X}$ 上的核函数,$\boldsymbol{\Phi}(\cdot)$ 为其特征映射.

定义 1. 设 $\mathcal{X}$ 为非空集合,$\mathcal{H}=\{f:\mathcal{X}\to\mathbb{R}\}$ 为函数 f 构成的希尔伯特空间.

1) 若对所有 $x\in\mathcal{X}$,Dirac 泛函 $\delta_x:\mathcal{H}\to\mathbb{R}$,

$$\delta_x(f):=f(\boldsymbol{x})$$

连续,则称空间 $\mathcal{H}$ 为 RKHS.

2) 若 $\forall\boldsymbol{x}\in\mathcal{X}$,有 $K(\cdot,\boldsymbol{x})\in\mathcal{H}$,且 $\forall f\in\mathcal{H}$, $\forall\boldsymbol{x}\in\mathcal{X}$ 均满足再生性:

$$f(\boldsymbol{x})=\langle f,K(\cdot,\boldsymbol{x})\rangle,$$

则称 K 为 $\mathcal{H}$ 的再生核.

具有再生核的希尔伯特空间一定为 RKHS. 反之亦然.

2 近似高斯核

本节首先定义近似高斯核,推导近似高斯核与高斯核偏差上界;然后显式描述近似高斯核的 RKHS、显示构造近似高斯核的特征映射.

2.1 近似高斯核定义

定义 2. 近似高斯核. 设 $\mathcal{X}\subseteq\mathbb{R}^n$ 为非空集,近似高斯核 $K:\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ 定义如下:

$$K(\boldsymbol{x},\boldsymbol{y})=\mathrm{e}^{-\frac{\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2}{2\sigma_0}}\sum_{k=0}^{m}\frac{1}{k!\rho_k}\sum_{|\boldsymbol{\alpha}|=k}C_{\boldsymbol{\alpha}}^{k}\boldsymbol{x}^{\boldsymbol{\alpha}}\boldsymbol{y}^{\boldsymbol{\alpha}},\tag{1}$$

其中,$\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_n)^{\mathrm{T}}\in(\mathbb{N}\cup\{0\})^n$, $m\in\mathbb{N}$, $\sigma_0,\rho_k\in\mathbb{R}_+$.

下面定理说明:近似高斯核收敛于高斯核,近似高斯核是高斯核的近似.

定理 1. 若 $m=\infty$ 且 $\rho_k=\sigma_0^k$, $k=0,1,\cdots$,则近似高斯核为

$$K(\boldsymbol{x},\boldsymbol{y})=\mathrm{e}^{-\frac{\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2}{2\sigma_0}}\sum_{k=0}^{\infty}\frac{1}{k!\sigma_0^k}\sum_{|\boldsymbol{\alpha}|=k}C_{\boldsymbol{\alpha}}^{k}\boldsymbol{x}^{\boldsymbol{\alpha}}\boldsymbol{y}^{\boldsymbol{\alpha}}$$

为高斯核,即 $K(\boldsymbol{x},\boldsymbol{y})=\mathrm{e}^{-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{2\sigma_0}}$.

证明. 令 $\delta:=\sum_{k=0}^{\infty}\frac{1}{k!\sigma_0^k}\sum_{|\boldsymbol{\alpha}|=k}C_{\boldsymbol{\alpha}}^{k}\boldsymbol{x}^{\boldsymbol{\alpha}}\boldsymbol{y}^{\boldsymbol{\alpha}}$,则近似高斯核可表示为 $\exp\left(-\frac{\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2}{2\sigma_0}\right)\delta$. 由泰勒展开可得:

$$\exp\left(\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\sigma_0}\right)=1+\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\sigma_0}+\cdots+\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle^m}{m!\sigma_0^m}+\cdots=\sum_{k=0}^{\infty}\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle^k}{k!\sigma_0^k}.$$

由于 $\langle \boldsymbol{x},\boldsymbol{y}\rangle^k=(x_1y_1+x_2y_2+\cdots+x_ny_n)^k=\sum_{|\boldsymbol{\alpha}|=k}C_{\boldsymbol{\alpha}}^k\boldsymbol{x}^{\boldsymbol{\alpha}}\boldsymbol{y}^{\boldsymbol{\alpha}}$,易得:

$$\exp\left(\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\sigma_0}\right)=\sum_{k=0}^{\infty}\frac{1}{k!\sigma_0^k}\sum_{|\boldsymbol{\alpha}|=k}C_{\boldsymbol{\alpha}}^k\boldsymbol{x}^{\boldsymbol{\alpha}}\boldsymbol{y}^{\boldsymbol{\alpha}}=\delta. \quad (2)$$

由式(2)可得:

$$K(\boldsymbol{x},\boldsymbol{y})=\mathrm{e}^{\left(-\frac{\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2}{2\sigma_0}\right)}\mathrm{e}^{\left(\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\sigma_0}\right)}=\mathrm{e}^{\left(-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{2\sigma_0}\right)}.$$

证毕.

下面给出偏差 $|K(\boldsymbol{x},\boldsymbol{y})-K_{\mathrm{gauss}}(\boldsymbol{x},\boldsymbol{y})|$ 的上界,其中,$K(\boldsymbol{x},\boldsymbol{y})$ 为近似高斯核,$K_{\mathrm{gauss}}(\boldsymbol{x},\boldsymbol{y})$ 为高斯核.

定理 2. 若令 $\rho_k=\sigma_0^k$,$k=0,1,\cdots,m$,则存在 $\xi\in[0,\langle \boldsymbol{x},\boldsymbol{y}\rangle/\sigma_0^2]$,使得:

$$|K(\boldsymbol{x},\boldsymbol{y})-K_{\mathrm{gauss}}(\boldsymbol{x},\boldsymbol{y})|\leqslant\frac{1}{(m+1)!}\left|\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\sigma_0}\right|^{m+1}e^{\xi}.$$

证明. 由于:

$$K_{\mathrm{gauss}}(\boldsymbol{x},\boldsymbol{y})-K(\boldsymbol{x},\boldsymbol{y})=\mathrm{e}^{-\frac{\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2}{2\sigma_0}}\left(\mathrm{e}^{\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\sigma_0}}-\sum_{k=0}^{m}\frac{1}{k!}\left(\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\sigma_0}\right)^k\right).$$

由泰勒展开余项定理可知,存在 $\xi\in[0,\langle \boldsymbol{x},\boldsymbol{y}\rangle/\sigma_0^2]$,使得:

$$\exp\left(\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\sigma_0}\right)=\sum_{k=0}^{m}\frac{1}{k!}\left(\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\sigma_0}\right)^k+R_m\left(\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\sigma_0}\right), \quad (3)$$

其中,$R_m(\langle \boldsymbol{x},\boldsymbol{y}\rangle/\sigma_0)=(\langle \boldsymbol{x},\boldsymbol{y}\rangle/\sigma_0)^{(m+1)}\mathrm{e}^{\xi}/(m+1)!$.

因此,$K_{\mathrm{gauss}}(\boldsymbol{x},\boldsymbol{y})-K(\boldsymbol{x},\boldsymbol{y})=\mathrm{e}^{-\frac{\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2}{2\sigma_0}}\times R_m\left(\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\sigma_0}\right)$.

由于 $-\frac{\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2}{2\sigma_0}\leqslant 0$,有 $\mathrm{e}^{-\frac{\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2}{2\sigma_0}}\leqslant 1$,可得:

$$|K_{\mathrm{gauss}}(\boldsymbol{x},\boldsymbol{y})-K(\boldsymbol{x},\boldsymbol{y})|\leqslant\left|R_m\left(\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\sigma_0}\right)\right|\leqslant\frac{1}{(m+1)!}\left|\frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\sigma_0}\right|^{m+1}\mathrm{e}^{\xi}.$$

证毕.

注 1. 若 $\mathcal{X}$ 有界,即 $\exists C\geqslant 0$,$\forall \boldsymbol{x}\in\mathcal{X}$,$\|\boldsymbol{x}\|\leqslant C$,则

$$|K(\boldsymbol{x},\boldsymbol{y})-K_{\mathrm{gauss}}(\boldsymbol{x},\boldsymbol{y})|\leqslant\frac{1}{(m+1)!}\left(\frac{\|\boldsymbol{x}\|\cdot\|\boldsymbol{y}\|}{\sigma_0}\right)^{m+1}\mathrm{e}^{\xi}\leqslant\frac{1}{(m+1)!}\left(\frac{C^2}{\sigma_0}\right)^{m+1}\mathrm{e}^{\xi}.$$

因此,当阶数 $m\to\infty$,$|K(\boldsymbol{x},\boldsymbol{y})-K_{\mathrm{gauss}}(\boldsymbol{x},\boldsymbol{y})|\to 0$. 这表明近似高斯核的构造是一致的.

2.2　近似高斯核的 RKHS

下面显示描述近似高斯核的 RKHS.

定理 3. 设 $\mathcal{X}\subseteq\mathbb{R}^n$,空间 $\mathcal{H}_K$ 为

$$\mathcal{H}_K=\left\{f(x)=\mathrm{e}^{-\frac{\|x\|^2}{2\sigma_0}}\sum_{k=0}^{m}\frac{1}{k!}\sum_{|\boldsymbol{\alpha}|=k}f_{\boldsymbol{\alpha}}\boldsymbol{x}^{\boldsymbol{\alpha}},\boldsymbol{x}\in\mathcal{X}\right\}, \quad (4)$$

$\mathcal{H}_K$ 上的内积 $\langle\cdot,\cdot\rangle_K$ 定义为

$$\langle f,g\rangle_K=\sum_{k=0}^{m}\frac{\rho_k}{k!}\sum_{|\boldsymbol{\alpha}|=k}\frac{f_{\boldsymbol{\alpha}}g_{\boldsymbol{\alpha}}}{C_{\boldsymbol{\alpha}}^k}, \quad (5)$$

其中,$f,g\in\mathcal{H}_K$,有:

$$f=\mathrm{e}^{-\frac{\|x\|^2}{2\sigma_0}}\sum_{k=0}^{m}\frac{1}{k!}\sum_{|\boldsymbol{\alpha}|=k}f_{\boldsymbol{\alpha}}x^{\boldsymbol{\alpha}},$$

$$g=\mathrm{e}^{-\frac{\|x\|^2}{2\sigma_0}}\sum_{k=0}^{m}\frac{1}{k!}\sum_{|\boldsymbol{\alpha}|=k}g_{\boldsymbol{\alpha}}x^{\boldsymbol{\alpha}},$$

$f_{\boldsymbol{\alpha}}$ 和 $g_{\boldsymbol{\alpha}}$ 为刻画 f 和 g 的系数. 则近似高斯核:

$$K(\boldsymbol{x},\boldsymbol{y})=\mathrm{e}^{-\frac{\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2}{2\sigma_0}}\sum_{k=0}^{m}\frac{1}{k!\rho_k}\sum_{|\boldsymbol{\alpha}|=k}C_{\boldsymbol{\alpha}}^k\boldsymbol{x}^{\boldsymbol{\alpha}}\boldsymbol{y}^{\boldsymbol{\alpha}}$$

为 $\mathcal{H}_K$ 的再生核.

证明.

1) $K(\boldsymbol{x},\cdot)\in\mathcal{H}_K$;

2) 再生性,即对所有 $f\in\mathcal{H}_K$ 和 $\boldsymbol{x}\in\mathcal{X}$,$\langle f,K(\cdot,\boldsymbol{x})\rangle=f(\boldsymbol{x})$.

给定 $\boldsymbol{x}$,$K(\boldsymbol{x},\boldsymbol{y})$ 为 $\boldsymbol{y}$ 的函数,$K(\boldsymbol{x},\boldsymbol{y})$ 可以表示为

$$K(\boldsymbol{x},\cdot)(\boldsymbol{y})=\mathrm{e}^{-\frac{\|\boldsymbol{y}\|^2}{2\sigma_0}}\sum_{k=0}^{m}\frac{1}{k!}\sum_{|\boldsymbol{\alpha}|=k}\frac{\mathrm{e}^{-\frac{\|\boldsymbol{x}\|^2}{2\sigma_0}}C_{\boldsymbol{\alpha}}^k\boldsymbol{x}^{\boldsymbol{\alpha}}}{\rho_k}y^{\boldsymbol{\alpha}}.$$

令 $f_{\boldsymbol{\alpha}}=\mathrm{e}^{-\frac{\|\boldsymbol{x}\|^2}{2\sigma_0}}C_{\boldsymbol{\alpha}}^k\boldsymbol{x}^{\boldsymbol{\alpha}}/\rho_k$,因此可知 $K(\boldsymbol{x},\cdot)\in\mathcal{H}_K$. 对于 $\forall f\in\mathcal{H}_K$,由 $\mathcal{H}_K$ 中的内积定义可得:

$$\langle K(\boldsymbol{x},\cdot),f\rangle_K=\sum_{k=0}^{m}\frac{\rho_k}{k!}\sum_{|\boldsymbol{\alpha}|=k}\frac{(\mathrm{e}^{-\|\boldsymbol{x}\|^2/2\sigma_0}C_{\boldsymbol{\alpha}}^k\boldsymbol{x}^{\boldsymbol{\alpha}}/\rho_k)f_{\boldsymbol{\alpha}}}{C_{\boldsymbol{\alpha}}^k}=\mathrm{e}^{-\frac{\|\boldsymbol{x}\|^2}{2\sigma_0}}\sum_{k=0}^{m}\frac{1}{k!}\sum_{|\boldsymbol{\alpha}|=k}f_{\boldsymbol{\alpha}}\boldsymbol{x}^{\boldsymbol{\alpha}}=f(\boldsymbol{x}).$$

证毕.

注 2. 上述定理给出近似高斯 RKHS 的显示描述,精确刻画假设空间(hypothesis space)形式,增强核方法的可解释性,奠定近似高斯核的理论基础.

下面考虑阶数 $m=\infty$ 的情况.

推论 1. 设 $\mathcal{X}\subseteq\mathbb{R}^n$,空间 $\mathcal{H}_K$ 为

$$\mathcal{H}_K=\left\{f(\boldsymbol{x})=e^{-\frac{\|\boldsymbol{x}\|^2}{2\sigma_0}}\sum_{k=0}^{\infty}\frac{1}{k!}\sum_{|\boldsymbol{\alpha}|=k}f_{\boldsymbol{\alpha}}\boldsymbol{x}^{\boldsymbol{\alpha}}:\|f\|_K^2=\sum_{k=0}^{m}\frac{\rho_k}{k!}\sum_{|\boldsymbol{\alpha}|=k}\frac{f_{\boldsymbol{\alpha}}^2}{C_{\boldsymbol{\alpha}}^k}<\infty,\boldsymbol{x}\in\mathcal{X}\right\},$$

$\mathcal{H}_K$ 上的内积 $\langle\cdot,\cdot\rangle_K$ 定义为

$$\langle f,h\rangle_K=\sum_{k=0}^{\infty}\frac{\rho_k}{k!}\sum_{|\boldsymbol{\alpha}|=k}\frac{f_{\boldsymbol{\alpha}}g_{\boldsymbol{\alpha}}}{C_{\boldsymbol{\alpha}}^k},$$

其中,$f,g\in\mathcal{H}_K$,有:

$$f=\mathrm{e}^{-\frac{\|x\|^2}{2\sigma_0}}\sum_{k=0}^{\infty}\frac{1}{k!}\sum_{|\boldsymbol{\alpha}|=k}f_{\boldsymbol{\alpha}}\boldsymbol{x}^{\boldsymbol{\alpha}},$$

$$g=\mathrm{e}^{-\frac{\|x\|^2}{2\sigma_0}}\sum_{k=0}^{\infty}\frac{1}{k!}\sum_{|\boldsymbol{\alpha}|=k}g_{\boldsymbol{\alpha}}\boldsymbol{x}^{\boldsymbol{\alpha}},$$

$f_{\boldsymbol{\alpha}}$ 和 $g_{\boldsymbol{\alpha}}$ 为刻画 f 和 g 的系数. 那么:

$$K(\boldsymbol{x},\boldsymbol{y})=\mathrm{e}^{-\frac{\|x\|^2+\|y\|^2}{2\sigma_0}}\sum_{k=0}^{\infty}\frac{1}{k!\rho_k}\sum_{|\boldsymbol{\alpha}|=k}C_{\boldsymbol{\alpha}}^{k}\boldsymbol{x}^{\boldsymbol{\alpha}}\boldsymbol{y}^{\boldsymbol{\alpha}}$$

为 $\mathcal{H}_K$ 的再生核.

证明. 令 $m=\infty$,基于定理 3 可证明该结论.

注 3. 文献[20]已经给出高斯核 RKHS 的显式描述. 由定理 1 可知,当 $m=\infty,\rho_k=\sigma_0^k$ 时,近似高斯核为高斯核. 因此本文扩展文献[20]相关结论,而且采用 Weyl 内积使证明更为简洁.

2.3 近似高斯核显式特征映射

下面显式构造近似高斯核的特征映射.

定理 4. 设 $\mathcal{X}\subseteq\mathbb{R}^n$,近似高斯核 $K(\boldsymbol{x},\boldsymbol{y})$ 的特征映射:

$$\boldsymbol{\Phi}_K(\boldsymbol{x})=[\boldsymbol{\Phi}_0(\boldsymbol{x}),\boldsymbol{\Phi}_1(\boldsymbol{x}),\cdots,\boldsymbol{\Phi}_m(\boldsymbol{x})],\quad(6)$$

即 $K(\boldsymbol{x},\boldsymbol{y})=\langle\boldsymbol{\Phi}_K(\boldsymbol{x}),\boldsymbol{\Phi}_K(\boldsymbol{y})\rangle$,其中,

$$\boldsymbol{\Phi}_i(\boldsymbol{x})=\left[\mathrm{e}^{-\frac{\|x\|^2}{2\sigma_0}}\sqrt{C_{\boldsymbol{\alpha}}^{i}/(i!\rho_i)}\,x^{\boldsymbol{\alpha}}\,|\,|\boldsymbol{\alpha}|=i\right]\quad(7)$$

为向量,$i=0,1,\cdots,m$.

证明. 由式(7)容易验证:

$$\mathrm{e}^{\left(-\frac{\|x\|^2+\|y\|^2}{2\sigma_0}\right)}\frac{1}{k!\rho_k}\sum_{|\boldsymbol{\alpha}|=k}C_{\boldsymbol{\alpha}}^{k}\boldsymbol{x}^{\boldsymbol{\alpha}}\boldsymbol{y}^{\boldsymbol{\alpha}}=\langle\boldsymbol{\Phi}_k(\boldsymbol{x}),\boldsymbol{\Phi}_k(\boldsymbol{y})\rangle.$$

因而有 $K(\boldsymbol{x},\boldsymbol{y})=\sum_{k=0}^{m}\langle\boldsymbol{\Phi}_k(\boldsymbol{x}),\boldsymbol{\Phi}_k(\boldsymbol{y})\rangle=\langle\boldsymbol{\Phi}_K(\boldsymbol{x}),\boldsymbol{\Phi}_K(\boldsymbol{y})\rangle$. 证毕.

现举例说明上述近似高斯核的显式特征映射. 设输入空间维数 $n=2$,则 2 阶近似高斯核($m=2$)为

$$K(\boldsymbol{x},\boldsymbol{y})=\mathrm{e}^{-\frac{\|x\|^2+\|y\|^2}{2\sigma_0}}\sum_{k=0}^{2}\frac{1}{k!\rho_k}\sum_{|\boldsymbol{\alpha}|=k}C_{\boldsymbol{\alpha}}^{k}\boldsymbol{x}^{\boldsymbol{\alpha}}\boldsymbol{y}^{\boldsymbol{\alpha}}=$$

$$\mathrm{e}^{-\frac{\|x\|^2+\|y\|^2}{2\sigma_0}}\left(\frac{1}{\rho_0}+\frac{1}{\rho_1}x_1y_1+\frac{1}{\rho_1}x_2y_2+\right.$$

$$\left.\frac{1}{\rho_2}x_1^2y_1^2+\frac{2}{\rho_2}x_1x_2y_1y_2+\frac{1}{\rho_2}x_2^2y_2^2\right),$$

其中,$\boldsymbol{x}=[x_1,x_2],\boldsymbol{y}=[y_1,y_2]\in\mathbb{R}^2$. 由式(6)(7)可知:

$$\boldsymbol{\Phi}_K(\boldsymbol{x})=\mathrm{e}^{-\frac{\|x\|^2}{2\sigma_0}}\left[\sqrt{\frac{1}{\rho_0}},\sqrt{\frac{1}{\rho_1}}x_1,\sqrt{\frac{1}{\rho_1}}x_2,\right.$$

$$\left.\sqrt{\frac{1}{\rho_2}}x_1^2,\sqrt{\frac{2}{\rho_2}}x_1x_2,\sqrt{\frac{1}{\rho_2}}x_2^2\right].$$

容易验证 $K(\boldsymbol{x},\boldsymbol{y})=\langle\boldsymbol{\Phi}_K(\boldsymbol{x}),\boldsymbol{\Phi}_K(\boldsymbol{y})\rangle$.

注 4. 设训练数据 $S=\{\boldsymbol{x}_i,y_i\}_{i=1}^{\ell}$,通过近似高斯核的显式特征映射,将 $\{\boldsymbol{\Phi}_K(\boldsymbol{x}_i),y_i\}_{i=1}^{\ell}$ 作为线性 SVM 的输入,从而可应用线性 SVM 来求解非线性分类问题.

3 基于近似高斯核的高效非线性 SVM

本节首先简要介绍线性 SVM 高效求解算法. 然后,基于近似高斯核的显式特征映射,应用线性 SVM 高效算法来求解非线性 SVM.

3.1 高效线性 SVM

设训练数据为 $\{(\boldsymbol{x}_i,y_i)\}_{i=1}^{\ell},\boldsymbol{x}_i\in\mathbb{R}^n,y_i\in\{-1,+1\}$,线性 SVM 求解如下优化问题:

$$\min\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}+C\sum_{i=1}^{\ell}\max(1-y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b),0).$$

不失一般性,可通过在输入数据中增加一维使偏置项 b 包含在 $\boldsymbol{w}$ 中,从而将 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b$ 替换成 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i$,由此得式(8)优化问题:

$$\min\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}+C\sum_{i=1}^{\ell}\max(1-y_i\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i,0).\quad(8)$$

优化问题式(8)称为 SVM 的原始形式,其对偶形式为

$$\begin{aligned}\min_{\boldsymbol{\alpha}}\quad&\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\alpha}-\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\alpha},\\ \text{s.t.}\quad&0\leqslant\alpha_i\leqslant C,i=1,\cdots,\ell,\end{aligned}\quad(9)$$

其中,$\boldsymbol{Q}_{ij}=\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_jy_iy_j,\boldsymbol{e}=[1,\cdots,1]^{\mathrm{T}}$.

文献[10]给出高效的对偶坐标下降算法求解优化问题. 该算法每次只对一个坐标方向进行优化. 若 $\boldsymbol{\alpha}$ 是当前迭代点,则对它的第 i 个坐标方向进行优化相当于求解如下单变量的优化问题:

$$\begin{aligned}\min\quad&\frac{1}{2}\boldsymbol{Q}_{ii}d^2+(\boldsymbol{Q}\boldsymbol{\alpha}-\boldsymbol{e})_id+constant,\\ \text{s.t.}\quad&0\leqslant\alpha_i+d\leqslant C.\end{aligned}\quad(10)$$

其最优解为

$$\alpha_{i,\mathrm{new}}=\min\left(\max\left(\alpha_i-\frac{y_i\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i-1}{\boldsymbol{Q}_{ii}},0\right),C\right).\quad(11)$$

若 α_i 为当前值,而 $\alpha_{i,\mathrm{new}}$ 是更新后的值,则:

$$\boldsymbol{w}\leftarrow\boldsymbol{w}+(\alpha_{i,\mathrm{new}}-\alpha_i)y_i\boldsymbol{x}_i.\quad(12)$$

3.2 基于显式特征映射的高效非线性 SVM

将 $\{\boldsymbol{\Phi}_K(\boldsymbol{x}_i),y_i\}_{i=1}^{\ell}$ 作为线性 SVM 的输入,从而可将线性 SVM 的高效求解算法应用到非线性 SVM 上. 这种情况下,决策函数 f 为 $f(x)=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Phi}_K(\boldsymbol{x})$. 相应的迭代为

$$\alpha_{i,\mathrm{new}}=\min\left(\max\left(\alpha_i-\frac{y_i\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Phi}_K(x_i)-1}{\boldsymbol{Q}_{ii}},0\right),C\right),$$

$$\boldsymbol{w}\leftarrow\boldsymbol{w}+(\alpha_{i,\mathrm{new}}-\alpha_i)y_i\boldsymbol{\Phi}_K(x_i),$$

其中,$\boldsymbol{Q}_{ii}=\langle\boldsymbol{\Phi}_K(\boldsymbol{x}_i),\boldsymbol{\Phi}_K(\boldsymbol{x}_i)\rangle y_i y_i$.

注 5. 假设 $\boldsymbol{\Phi}_K(\boldsymbol{x})$的维数为 p,那么上述方法每次迭代需要 $O(p)$次操作.而标准的非线性 SVM 的复杂度为 $O(\ell^3)$,ℓ 为样本的数目.因此对于大规模分类问题,当特征映射 $\boldsymbol{\Phi}_K(\boldsymbol{x})$的维数适中时,相比于标准非线性 SVM 算法,本文所提出的算法求解效率将显著提高.

3.3　显式特征映射的稀疏保持性

由 $\boldsymbol{\Phi}_K(\boldsymbol{x})$的定义可知其维数为 C_m^{n+m},故当原始数据的维数 n 很大或近似高斯核的阶数 m 很大时,特征映射后的数据的维数较大.如果原始数据本身具有一定的稀疏性,那么映射后的数据仍会保持一定的稀疏性.假定阶数 $m=2$,设维数为 n 维的实例 $\boldsymbol{x}$ 中有 $\bar{n}$ 维的属性值为 0.由式(7)可知,$\boldsymbol{\Phi}_1(\boldsymbol{x})$中为 0 的特征的个数为 $\bar{n}$.在 $\boldsymbol{\Phi}_2(\boldsymbol{x})$中则有 $n\bar{n}-C_2^{\bar{n}}$ 为 0.因此 $\boldsymbol{\Phi}_K(\boldsymbol{x})$有 $\bar{n}+n\bar{n}-C_2^{\bar{n}}$ 维的值为 0.如假设 $n=20$,当数据稠密时,用 2 阶近似高斯核映射后维数为 231.若该实例中有 5 个属性为 0,则映射后的特征向量中将有 95 个值为 0.由此可见,显式特征映射具有稀疏保持性.可使特征映射的计算和 SVM 的求解简便有效.

3.4　预　测

如果采用近似高斯核的显式特征映射,其决策函数为 $f(\boldsymbol{x})=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Phi}_K(\boldsymbol{x})$.对测试样本 $\boldsymbol{x}$ 进行预测需要 $O(\hat{n})$次操作,$\hat{n}$ 为 $\boldsymbol{\Phi}_K(\boldsymbol{x})$中非 0 特征的个数.如果采用高斯,决策函数为 $f(\boldsymbol{x})=\sum_{i=1}^{\#\mathrm{SV}}K_{\mathrm{gauss}}(\boldsymbol{x}_i,\boldsymbol{x})\alpha_i$,预测复杂度为 $O(n\times\#\mathrm{SV})$,$\#\mathrm{SV}$ 表示支持向量的数目.可见,采用高斯核的非线性 SVM 的预测复杂度随着支持向量的数目线性增长.对于大规模训练数据,支持向量的规模可能较大.而通过选择适当的近似阶数 m,基于近似高斯核显式映射的非线性 SVM 的预测复杂度比普通高斯核的非线性 SVM 低很多.

4　实验结果与分析

本节将通过实验验证所提出的方法的精确性与有效性.实验均是在 2.2 GB AMD Opteron Processor 6174 CPU,30 GB DDR3 内存的机器上运行的.

本文采用与文献[12]相同的实验设定,并应用相同的数据集.数据集的描述如表 1 所示.其中,n 是属性维数,$\bar{n}$ 是非 0 属性数目的平均值.ℓ 是训练数据的数目.t 为测试数据的数目.本文实验考虑阶数 $m=2$ 的情况,则近似高斯核的形式为

$$K_{\mathrm{Gauss\text{-}2}}(\boldsymbol{x},\boldsymbol{y})=\mathrm{e}^{-\gamma(\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2)}\sum_{k=0}^{2}2^k\gamma^k/k!\sum_{|\boldsymbol{\alpha}|=k}C_{\boldsymbol{\alpha}}^k\boldsymbol{x}^{\boldsymbol{\alpha}}\boldsymbol{y}^{\boldsymbol{\alpha}},\quad(13)$$

其中,$\gamma=1/(2\sigma)$.以下实验将对比 LINEAR[21],POLY-2[12],KERNEL[22]和本文所提出 GAUSS-2,详情如下:

1) LINEAR、线性 SVM.代码为 LIBLINEAR[21].

2) POLY-2、构造 2 次多项式核 $K_{\mathrm{Poly\text{-}2}}=(\gamma\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}+1)^2$ 的显式特征映射,将数据映射到显式表示的特征空间,再采用高效线性 SVM 求解算法.文献[12]提供该方法的代码.该代码是 LIBLINEAR 的扩展,计算 $\boldsymbol{\Phi}(\boldsymbol{x})$是在训练时进行的.

3) KERNEL、普通高斯核的非线性 SVM,即 $K_{\mathrm{Gauss}}=\exp(-\gamma\|\boldsymbol{x}-\boldsymbol{y}\|^2)$,代码为 LIBSVM[22].

4) GAUSS-2、本文提出的算法.调用 LIBLINEAR.但与 POLY-2 不同的是,训练前先计算所有 $\boldsymbol{\Phi}(\boldsymbol{x}_1),\boldsymbol{\Phi}(\boldsymbol{x}_2),\cdots,\boldsymbol{\Phi}(\boldsymbol{x}_\ell)$,不需要对 LIBLINEAR 代码进行修改.这样实现仅仅是为了简单,尽管由于缓存-内存-CPU 速度问题,可能使效率受一定影响(参见文献[12]中 3.4 节的分析).

4 种算法中参数 $C\in\{2^0,\cdots,2^{10}\}$和 $\gamma\in\{2^{-10},\cdots,2^5\}$都通过 5 折交叉验证进行选择(LINEAR 没有参数 γ).

Table 1　Benchmark Datasets

表 1　实验所用标准数据集

Datasets	n	$\bar{n}$	ℓ	t
a9a	123	13.9	32 561	16 281
ijcnn	22	13.0	49 990	91 701
covtype	54	11.9	464 810	116 202
mnist38	752	168.2	11 982	1 984
real-sim	20 958	51.5	57 848	14 461
webspam	254	85.1	280 000	70 000

4.1　精度与效率

4 种方法的测试精度和训练时间分别如表 2 和表 3 所示.首先考虑 LINEAR 和 KERNEL 的结果.从表 2 可知,在所有 6 个数据集上,KERNEL 的精度都高于 LINEAR;而从表 3 可知,KERNEL 的训练时间比 LINEAR 长很多.

Table 2 Testing Accuracies of the LINEAR, GAUSS-2, POLY-2 and KERNEL

表 2 4 种方法的测试精度

Datasets	LINEAR		GAUSS-2			POLY-2			KERNEL		
	C	Acc/%	C	σ	Acc/%	C	γ	Acc/%	C	σ	Acc/%
a9a	32	84.98	32	0.125	84.78	8	0.031 25	85.05	8	0.031 25	85.03
ijcnn	32	92.17	128	0.5	**97.88**	0.125	32	97.83	32	2	98.67
covtype	0.062 5	76.34	512	2	**80.08**	8	32	79.99	32	32	95.99
mnist38	0.031 25	96.82	32	0.007 812 5	**99.44**	2	0.312 5	99.29	2	0.031 25	99.69
real-sim	1	97.41	8	2	**98.07**	0.031 25	8	98.06	8	0.5	97.82
webspam	32	93.21	512	0.5	98.06	8	8	98.44	8	32	99.20

Table 3 Training time (in seconds) for LINEAR, GAUSS-2, POLY-2 and KERNEL

表 3 4 种方法的训练时间 s

Datasets	LINEAR	GAUSS-2	POLY-2	KERNEL
a9a	10.00	4.81	2.31	165.30
ijcnn	2.93	**11.09**	14.38	35.99
covtype	1.68	**374.96**	5 985.36	46 303.98
mnist38	0.18	22.86	13.88	43.05
real-sim	0.28	**44.06**	66.76	1 053.31
webspam	24.71	**3 917.06**	4 437.30	17 283.35

然后再看 GAUSS-2 的性能. 从表 2 可以看出, GAUSS-2 与 KERNEL 的分类性能相近, 且从表 3 可看出, 相比于 KERNEL, GAUSS-2 的计算效率显著提高. 具体地, 在 a9a, ijcnn, mnist38, real-sim 和 webspam 数据集上, 本文方法的精度与 KERNEL 相差无几. 在 covtype 数据集上, 精度不如 KERNEL. 原因是阶数 $m=2$ 对这个数据集是不够的. 实际上, 当考虑 $m=3$ 时, 此时 GAUSS-2 在该数据集上的精度为 82.48%, 提高 2.4%. 在个别数据集上精度不如普通高斯核 SVM, 但训练时间上的优势是显著的, 因而本文所提出的方法提供了一个效率和精度之间的很好折中.

最后比较算法 GAUSS-2 和 POLY-2. 从表 2 可以看出, 在所有 6 个数据集上有 4 个 GAUSS-2 在精度上比 POLY-2 高. 尽管前面所述本文计算特征映射的实现方式不如 POLY-2 高效, 但 GAUSS-2 的训练时间在 4/6 个数据集上比 POLY-2 的短. 原因是 GAUSS-2 的收敛速度往往 POLY-2 快, 这可以从图 1 得到验证. 从图 1 可知 GAUSS-2 的迭代次数在 5/6 个数据集上比 POLY-2 少. 上述实验表明, 本文方法能提高非线性 SVM 的求解效率, 并保证算法的精确性.

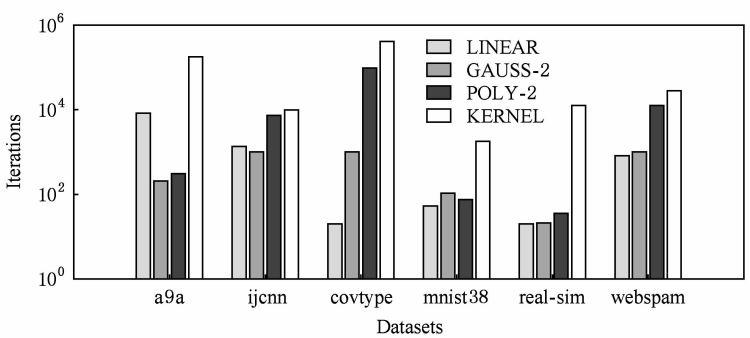


Fig. 1 The number of iterations of the LINEAR, GAUSS-2, POLY-2 and KERNEL.

图 1 LINEAR, GAUSS-2, POLY-2 和 KERNEL 的迭代次数

5 结 语

为了应用高效线性 SVM 算法求解大规模非线性 SVM 分类问题, 本文引入近似高斯核. 首先, 度量近似高斯核与高斯核的偏差上界, 从理论上说明近似高斯核的合理性. 然后, 对近似高斯核的再生核希尔伯特空间进行精确刻画, 由此能对 SVM 解的结构进行精确描述, 增强模型的可解释性, 为刻画 SVM 的学习性能带来帮助. 最后, 显式构造近似高斯核的特征映射, 从而能应用线性 SVM 高效算法来求解非线性 SVM. 实验进一步验证所提出的方法合理性及高效性.

进一步工作将研究近似高斯核在其他核方法中的应用及其泛化理论.

参　考　文　献

[1] Vapnik V. The Nature of Statistical Learning Theory [M]. Berlin: Springer, 2000

[2] Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods [M]. Cambridge, UK: Cambridge University Press, 2000

[3] Zhang Xuegong. Introduction to statistical learning theory and support vector machine [J]. Acta Automatica Sinica, 2000, 26(1): 32-42 (in Chinese)
(张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32-42)

[4] Zhang Tong. Solving large scale linear prediction problems using stochastic gradient descent algorithms [C] //Proc of the 21st Int Conf on Machine Learning. San Francisco, CA: Morgan Kaufmann, 2004: 16-123

[5] Shalev-Shwartz S, Singe Y, Srebro N, et al. Pegasos: primal estimated sub-gradient solver for SVM [J]. Mathematical Programming, 2011, 127(1): 3-30

[6] Joachims T. Training linear SVMs in linear time [C] //Proc of the 12th ACM Conf on Knowledge Discovery and Data Mining. New York: ACM: 2006: 217-226

[7] Smola A, Vishwanathan S, Le Q. Bundle methods for machine learning [C] //Advances in Neural Information Processing Systems 20. Cambridge, MA: MIT Press, 2008: 1377-1384

[8] Lin C J, Weng R C, Keerthi S. Trust region newton method for large-scale logistic regression [J]. Journal of Machine Learning Research, 2008, 9: 627-650

[9] Chang K W, Hsieh C J, Lin C J. Coordinate descent method for large-scale L2-loss linear Support Vector Machines [J]. Journal of Machine Learning Research, 2008, 9: 1369-1398

[10] Hsieh C J, Chang K W, Lin C J. A dual coordinate descent method for Large-scale linearSVM [C] //Proc of the 25th Int Conf on Machine Learning. San Francisco, CA: Morgan Kaufmann, 2008: 408-415

[11] Yu H F, Huang F L, Lin C J. Dual coordinate descent methods for logistic regression and maximum entropy models [J]. Machine Learning, 2011, 85(1/2): 41-75

[12] Chang Y W, Hsieh C J, Chang K W, et al. Training and testing low-degree polynomial data mappings via linearSVM [J]. Journal of Machine Learning Research, 2010, 11: 1471-1490

[13] Ding Lizhong, Liao Shizhong. Approximate model selection on regularization path for support vector machines [J]. Journal of Computer Research and Development, 2012, 49(6): 1248-1255 (in Chinese)
(丁立中, 廖士中. 基于正则化路径的支持向量机近似模型选择[J]. 计算机研究与发展, 2012, 49(6): 1248-1255)

[14] Ding Lizhong, Liao Shizhong. KMA-α: A kernel matrix approximation algorithm for support vector machines. Journal of Computer Research and Development, 2012, 49(4): 746-753 (in Chinese)
(丁立中, 廖士中. KMA-α: 一个支持向量机核矩阵的近似计算算法[J]. 计算机研究与发展, 2012, 49(4): 746-753)

[15] Rahimi A, Recht B. Random features for large-scale kernel machines [C] //Advances in Neural Information Processing Systems 19. Cambridge, MA: MIT Press, 2007: 1177-1184

[16] Rahimi A, Recht B. Uniform approximation of functions with random bases [C] //Proc of the 46th Annual Allerton Conf on Communication, Control, and Computing. Los Alamitos, CA: IEEE Computer Society, 2008: 555-561

[17] Vedaldi A, Zisserman A. Efficient additive kernels via explicit feature maps [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2012, 34(3): 480-492

[18] Yang Changjiang, Duraiswami R, Davis L. Efficient kernel machines using the improved fast Gauss transform [C] //Advances in Neural Information Processing Systems 16. Cambridge, MA: MIT Press, 2004: 1561-1568

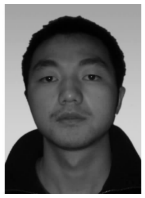
[19] Cao Hui, Naito T, Ninomiya Y. Approximate RBF kernel SVM and Its applications in pedestrian classification [C] //Proc of the 1st Int Workshop on Machine Learning for Vision-based Motion Analysis. Berlin: Springer, 2008: 120-129

[20] Steinwart I, Hush D, Scovel C. An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels [J]. IEEE Trans on Information Theory, 2006, 52(10): 4635-4643

[21] Fan R, Chang K W, Hsieh C J. LIBLINEAR: A library for large linear classification [J]. Journal of Machine Learning Research, 2008, 9: 1871-1874

[22] Chang C C, Lin C J. LIBSVM: A library for support vector machines [EB/OL]. 2001 [2007-08-06]. http://www.csie.ntu.edu.tw/~cjlin/libsvm

Liu Yong, born in 1986. PhD candidate. Student member of China Computer Federation. His main research interests include machine learning and model selection (yongliu@tju.edu.cn).



Jiang Shali, born in 1989. Master candidate. His main research interests include machine learning and sparse coding (sljiang@tju.edu.cn).

Liao Shizhong, born in 1964. PhD, professor, and PhD supervisor. Member of China Computer Federation. His main research interests include artificial intelligence and theoretical computer science.