一种改进的显性多核支持向量机

张凯军 [1] 梁 循 [1]

摘 要 在支持向量机 (Support vector machine, SVM) 中, 对核函数的定义非常重要, 不同的核会产生不同的分类结果. 如何充分利用多个不同核函数的特点, 来共同提高 SVM 学习的效果, 已成为一个研究热点. 于是, 多核学习 (Multiple kernel learning, MKL) 方法应运而生. 最近, 有的学者提出了一种简单有效的稀疏 MKL 算法, 即 GMKL (Generalized MKL) 算法, 它结合了 L_1 范式和 L_2 范式的优点, 形成了一个对核权重的弹性限定. 然而, GMKL 算法也并没有考虑到如何在充分利用已经选用的核函数中的共有信息. 另一方面, MultiK-MHKS 算法则考虑了利用典型关联分析 (Canonical correlation analysis, CCA) 来获取核函数之间的共有信息, 但是却没有考虑到核函数的筛选问题. 本文模型则基于这两种算法进行了一定程度的改进, 我们称我们的算法为改进的显性多核支持向量机 (Improved domain multiple kernel support vector machine, IDMK-SVM). 我们证明了本文的模型保持了 GMKL 的特性, 并且证明了算法的收敛性. 最后通过模拟实验, 本文证明了本文的多核学习方法相比于传统的多核学习方法有一定的精确性优势.

关键词 多核学习, 分类精度, 典型关联分析, 支持向量机

引用格式 张凯军, 梁循. 一种改进的显性多核支持向量机. 自动化学报, 2014, **40**(10): 2288−2294

DOI 10.3724/SP.J.1004.2014.02288

An Improved Domain Multiple Kernel Support Vector Machine

ZHANG Kai-Jun[1] LIANG Xun[1]

Abstract In support vector machine (SVM), it is critical to define the kernel function and a different kernel would cause different classification accuracy. People have started pursuing how to make the most use of multiple kernels harmoniously to improve the SVM performance, hence, the multiple kernel learning (MKL). Recently, an efficient generalized multiple kernel learning (GMKL) method was presented, which combines the advantages of L_1-norm and L_2-norm. However, the GMKL algorithm does not make the most use of the common information among the selected kernels. On the other hand, the MultiK-MHKS algorithm uses the canonical correlation analysis (CCA) to get the common information among the kernels while ignoring the selecting of kernels. So this paper tries to combine them and an improved domain multiple kernel support vector machine (IDMK-SVM) is presented. Simulation experiments demonstrate that the IDMK-SVM gets a higher classification precision than the existing typical MKL algorithms.

Key words Multiple kernel learning (MKL), classification precision, canonical correlation analysis (CCA), support vector machine (SVM)

Citation Zhang Kai-Jun, Liang Xun. An improved domain multiple kernel support vector machine. *Acta Automatica Sinica*, 2014, **40**(10): 2288−2294

多核学习 (Multiple kernel learning, MKL) 模型[1] 是近年来机器学习领域提出的一种新型的支持向量机 (Support vector machine, SVM)[2−4] 学习模型, 许多学者已经证明多核学习模型能够拥有较高的分类精度, 会对不同的样本采用不同的核组合, 其较高的灵活性能够在一定程度上提高模型的分类精确度. 实际上, 多核学习就是寻求一个基于 M 个基本核的线性组合作为核函数, 从而获取更优的分类结果. MKL 最简单也最常用的一种模型是 L_1 范式多核学习模型. L_1 范式多核学习模型使用的是多个基本核函数的凸组合, 其核权重域可表示为 $K = \sum_{j=1}^{M} \theta_j K_j$, $\theta_j \geq 0$, $\sum_{j=1}^{M} \theta_j = 1$, 其中 K_j 是第 j 个基本核函数, θ_j 为核权重, M 为核函数的个数. 通过将异构数据的不同特征分量分别通过对应的核函数进行映射, 使数据在新的特征空间得到更好的表达, 多核学习模型显著提高了预测精度[5−9]. 多核学习的过程, 很重要的一个部分是设定核函数中核权重. 如何合理有效地解决求解核权重的过程, 也成为了多核学习研究的重点之一. Duan 等提出的

收稿日期 2013-09-08 录用日期 2014-03-06
Manuscript received September 8, 2013; accepted March 6, 2014
国家自然科学基金 (71271211), 北京市自然科学基金 (4132067), 中国人民大学品牌计划 (10XNI029) 资助
Supported by National Natural Science Foundation of China (71271211), National Natural Science Foundation of Beijing (4132067), and Brand Plan of Renmin University of China (10XNI029)
本文责任编委 刘成林
Recommended by Associate Editor LIU Cheng-Lin
1. 中国人民大学 北京 100872
1. Renmin University of China, Beijing 100872

DTMKL 对多核学习的交叉验证的方法做了新一步的改进[10]; Wang 等提出的 MultiK-MHKS 的多核学习算法, 该算法从对每个不同核的样本的映射入手, 对支持向量机中的超平面向量 $\boldsymbol{w}$ 和偏移 b 的更新提出了新的一种思路[11].

在各种 MKL 模型中, 常用的方法有 L_1 范式的 MKL[12]、L_2 范式的 MKL[13] 等.

L_1 范式的 MKL 模型在确定核权重方面体现出独到的优越性, 其对核权重的限制较为简单, 故其模型的结果是稀疏的. 核函数的稀疏性是衡量模型优劣的重要指标, 因为如果一个模型的核权重解空间是非稀疏的, 那么意味着该模型将会利用大量的核函数, 从而导致算法的复杂性增加. 相比之下, L_2 范式的 MKL 及其扩展的 L_p 范式的 MKL 的产生, 正好弥补了 L_1 范式 MKL 的一些缺点, 通过在大规模数据集下与 L_1 范式 MKL 的对比实验, 结果显示 L_2 范式 MKL 在抗噪声和特征集冗余方面具有较强的鲁棒性. 然而, L_2 范式 MKL 不能合理的区分核函数的有效性, 会导致核权重的非稀疏解. 有鉴于此, 文献 [14] 通过结合 L_1 范式 MKL 和 L_2 范式 MKL, 给出 $\{\theta|\theta_j \geq 0, v\sum_{j=1}^{M}\theta_j+(1-v)\sum_{j=1}^{M}\theta_j^2 \leq 1\}$ 的限定, 提出了一种新的稀疏多核学习算法, 即 GMKL (Generalized MKL) 模型, 实验证明很好地保持了 L_1 范式 MKL 和 L_2 范式 MKL 各自的优点. 该模型能有效地选取核函数, 并剔除掉对样本泛化性能不高的核函数.

另一方面, 文献 [11] 提出的 MultiK-MHKS 基于 MHKS 模型, 通过多核学习的方法, 提出了一种多核学习的显性支持向量机模型, 通过典型关联分析 (Canonical correlation analysis, CCA)[15] 获取核映射空间的共有信息, 从而提升模型的分类精度. 为获取核映射, 该模型提出了一种粗放式的从核矩阵到核映射的转化. 然而, MultiK-MHKS 模型的问题在于没有考虑剔除掉对样本泛化性能不高的核函数, 从而进一步提升模型的泛化能力.

本文在 GMKL 及 MultiK-MHKS 的基础上, 对 MultiK-MHKS 引用了 GMKL 的核权重限定, 提出了一种改进的显性多核支持向量机模型. 本文模型既有 GMKL 模型的特点, 即能剔除掉泛化性能不高的核函数, 同时在已选择的核函数中, 能获取这些核映射的共有信息, 从而提升算法的泛化能力.

本文模型通过结合 GMKL 模型, 能保持核函数选取的稀疏性, 筛选出对分类效果较好的核函数, 在此基础上, 结合 Multi-MHKS 提出的多核学习模型中利用 CCA 的思想, 尽可能获取核函数间的共性, 从而既保持了 GMKL 模型筛选核函数的稀疏性, 又提升了核函数间的共性. 本文结构如下: 第 1 节对本文模型进行阐述; 第 2 节对算法的求解进行分析; 第 3 节对核权重的唯一性和核权重的稀疏性及算法的收敛性进行证明; 第 4 节对改进的显性多核支持向量机 (Improved domain multiple kernel support vector machine, IDMK-SVM) 进行模拟实验, 对其分类效率进行分析; 第 5 节对以上工作进行总结.

1 模型

对于给定的数据集 $\{\boldsymbol{x}_i, y_i\}_{i=1}^{l}$, $\boldsymbol{x}_i \in \mathbf{R}^d$, $y_i \in \{-1,+1\}$. 根据文献 [8], MKL 模型可表示为

$$\begin{aligned}&\min_{\theta}\min_{\boldsymbol{w},b}\left(\frac{1}{2}\sum_{j=1}^{M}\frac{\boldsymbol{w}_j^{\mathrm{T}}\boldsymbol{w}_j}{\theta_j}+C\sum_{i=1}^{l}\xi_i^2\right)\\ &\text{s.t.}\quad y_i\left(\sum_{j=1}^{M}\boldsymbol{w}_j^{\mathrm{T}}\phi_j(\boldsymbol{x}_i)+b\right)\geq 1-\xi_i,\ \ \xi_i\geq 0,\\ &\qquad\qquad i=1,\cdots,l\end{aligned}\tag{1}$$

在 GMKL 中, 核权重 θ 的限定为 $\{\theta|\theta_j \geq 0, v\sum_{j=1}^{M}\theta_j+(1-v)\sum_{j=1}^{M}\theta_j^2 \leq 1\}$. 其中 $0<v<1$ 表示 GMKL 中 L_1 范式和 L_2 范式弹性的限定, 当 $v=1$ 时, GMKL 退化为 L_1 范式 MKL, $v=0$ 时, GMKL 退化为 L_1 范式 MKL.

通过令 $\boldsymbol{w}_j \rightarrow (\boldsymbol{w}_j; b)$, $\phi_j(\boldsymbol{x}_i) \rightarrow (\phi_j(\boldsymbol{x}_i); 1)$, 可消去偏置量 b, 此时限定变为 $y_i(\sum_{j=1}^{M}\boldsymbol{w}_j^{\mathrm{T}}\phi_j(\boldsymbol{x}_i)) \geq 1-\xi_i$. 接下来, 采用文献 [8] 中的核映射矩阵的定义及 CCA 抽取核函数两两间共有信息. 核映射矩阵的定义为 $U_j=[y_1\phi_j(\boldsymbol{x}_1),\cdots,y_l\phi_j(\boldsymbol{x}_l)]$, $j=1,\cdots,M$. 根据文献 [7], 对任意两个核矩阵采用 CCA, 找到相应的 $\boldsymbol{w}_s, \boldsymbol{w}_t$, 使得 $\|U_s\boldsymbol{w}_s-U_t\boldsymbol{w}_t\|$ 最小, 能够有效提高算法的泛化能力. 同时定义 $\mathbf{1}=(1,\cdots,1)^{\mathrm{T}}$, $\boldsymbol{c}=(\boldsymbol{c}_1,\cdots,\boldsymbol{c}_{l+1})^{\mathrm{T}}$, $\boldsymbol{c}_i\geq 0$, $i=1,\cdots,l+1$, 由文献 [12] 可知, 式 (1) 可转化为

$$\begin{aligned}\min_{\theta}\min_{\boldsymbol{w},c}&\left(\frac{1}{2}\sum_{j=1}^{M}\frac{\boldsymbol{w}_j^{\mathrm{T}}\boldsymbol{w}_j}{\theta_j}+\right.\\ &C\left(\sum_{j=1}^{M}U_s\boldsymbol{w}_s-\mathbf{1}-\boldsymbol{c}\right)^{\mathrm{T}}\times\\ &\left(\sum_{j=1}^{M}U_s\boldsymbol{w}_s-\mathbf{1}-\boldsymbol{c}\right)+\\ &D\sum_{s,t=1}^{M}\Big(U_s\boldsymbol{w}_s-U_t\boldsymbol{w}_t\Big)^{\mathrm{T}}\times\end{aligned}$$

$$\left(U_s\boldsymbol{w}_s-U_t\boldsymbol{w}_t\right)\Bigg) \tag{2}$$

式 (2) 与 MultiK-MHKS 的区别在于, MultiK-MHKS 模型未考虑核函数的权重, MultiK-MHKS 模型中所有的核函数均按同一的权重选取, 本文的 IDMK-SVM 模型是要在 MultiK-MHKS 使用 GMKL 的核权重提升算法的泛化能力. 通过使用 $\sum_{j=1}^M a_j\sum_{j=1}^M b_j \geq \sum_{j=1}^M a_jb_j$, $\boldsymbol{c}\rightarrow \frac{1}{M}\boldsymbol{c}$, 可得 IDMK-SVM 的标准形式:

$$\begin{aligned}\min_{\theta}\min_{\boldsymbol{w},c}\Bigg(&\frac{1}{2}\sum_{j=1}^M\frac{\boldsymbol{w}_j^{\mathrm{T}}\boldsymbol{w}_j}{\theta_j}+\\&C\left(\sum_{j=1}^M U_s\boldsymbol{w}_s-\boldsymbol{1}-\boldsymbol{c}\right)^{\mathrm{T}}\times\\&\left(\sum_{j=1}^M U_s\boldsymbol{w}_s-\boldsymbol{1}-\boldsymbol{c}\right)+\\&D\sum_{s,t=1}^M\left(U_s\boldsymbol{w}_s-U_t\boldsymbol{w}_t\right)^{\mathrm{T}}\times\\&\left(U_s\boldsymbol{w}_s-U_t\boldsymbol{w}_t\right)\Bigg)\\ \text{s.t.}\quad &\left\{\theta|\theta_j\geq 0, v\sum_{j=1}^M\theta_j+(1-v)\sum_{j=1}^M\theta_j^2\leq 1\right\}\end{aligned} \tag{3}$$

式 (3) 即本文提出的 IDMK-SVM 模型. 其与 GMKL 模型最大的区别是不再利用核矩阵, 而是回归到利用核映射; 本文模型与 MultiK-MHKS 模型的区别是充分利用了核权重这个概念, 能够有效合理地选取核函数.

由于 IDMK-SVM 是利用了核映射, 那么如何利用核矩阵 K_j 求取 U_j 是本文必须解决的问题. 由于核函数对应的映射往往维度很大, 显然要得到一个核函数的精确核映射非常困难, 文献 [7] 提出了一种经验估计核映射的方法, 本文采用这种转变方法. 设核矩阵为 K_j, $j=1,\cdots,M$, 样本的映射 $\phi_j:\boldsymbol{x}\rightarrow\phi_j(\boldsymbol{x})$, $j=1,\cdots,M$. 对于任意一个 $l\times l$ 矩阵, 都可映射为 $K_j=P_{l\times r}S_{r\times r}P_{r\times l}^{\mathrm{T}}$, 其中 r 是 K_j 的秩, $S_{r\times r}$ 为对角阵, $P_{l\times r}$ 是 K_j 特征向量所构成矩阵.

文献 [7] 证明了 $\phi_j(\boldsymbol{x}_i)$ 可粗略表示为

$$\Phi_j(\boldsymbol{x}_i)=S^{-\frac{1}{2}}P^{\mathrm{T}}(K_j(\boldsymbol{x}_i,\boldsymbol{x}_1),\cdots,K_j(\boldsymbol{x}_i,\boldsymbol{x}_l)) \tag{4}$$

通过式 (4) 即可根据一些常用的核函数求得其对应的映射, 从而求得核映射矩阵 U_j.

2 算法

类似于 MutiK-SVM, 我们采用迭代法对 IDMK-SVM 求解. 首先固定 θ 和 β, 其中 β 是式 (3) 用拉格朗日变换产生的变量, 对 $\boldsymbol{w}$ 和 $\boldsymbol{c}$ 进行更新; 再根据更新的 $\boldsymbol{w}$ 和 $\boldsymbol{c}$, 对 θ 和 β 进行更新; 算法不断循环, 直至满足停止准则. 对式 (3) 采用拉格朗日变换, 可得

$$\begin{aligned}O(\boldsymbol{w},\boldsymbol{c},\boldsymbol{\theta},\beta)=&\frac{1}{2}\sum_{j=1}^M\frac{\boldsymbol{w}_j^{\mathrm{T}}\boldsymbol{w}_j}{\theta_j}+\\&C\sum_{j=1}^M\left(U_s\boldsymbol{w}_s-\frac{1}{M}\boldsymbol{1}-\boldsymbol{c}\right)^{\mathrm{T}}\times\\&\left(U_s\boldsymbol{w}_s-\frac{1}{M}\boldsymbol{1}-\boldsymbol{c}\right)+\\&D\left(\sum_{s,t=1}^M\left(U_s\boldsymbol{w}_s-U_t\boldsymbol{w}_t\right)^{\mathrm{T}}\right)\left(U_s\boldsymbol{w}_s-U_t\boldsymbol{w}_t\right)+\\&\beta\left(v\sum_{j=1}^M\theta_j+\left(1-v\right)\sum_{j=1}^M\theta_j^2-1\right)\end{aligned} \tag{5}$$

在式 (5) 中分别对 $\boldsymbol{w}$, $\boldsymbol{c}$ 和 θ 求偏导可得

$$\begin{aligned}\frac{\partial O}{\partial\boldsymbol{w}}=&\frac{\boldsymbol{w}_j}{\theta_j}+2C\left(U_s\boldsymbol{w}_s-\frac{1}{M}\boldsymbol{1}-\boldsymbol{c}\right)+\\&2D((M-1)U_s\boldsymbol{w}_s-\sum_{t=1,t\neq j}^M U_t\boldsymbol{w}_t)\end{aligned} \tag{6}$$

$$\frac{\partial O}{\partial\boldsymbol{c}}=-C\sum_{j=1}^M\left(U_s\boldsymbol{w}_s-\frac{1}{M}\boldsymbol{1}-\boldsymbol{c}\right) \tag{7}$$

$$\frac{\partial O}{\partial\theta_j}=-\frac{1}{2}\frac{\boldsymbol{w}_j^{\mathrm{T}}\boldsymbol{w}_j}{\theta_j}+\beta\Big(v+2\left(1-v\right)\theta_j\Big) \tag{8}$$

根据式 (7) 和式 (8) 可知, $\frac{\partial^2 O}{\partial\boldsymbol{c}^2}=MC$, $\frac{\partial^2 O}{\partial\theta_j^2}=\frac{\boldsymbol{w}_j^{\mathrm{T}}\boldsymbol{w}_j}{\theta_j^3}+2(1-v)\beta$. 其迭代步长设为以上得到公式的负倒数.

通过求解式 (6), 可得

$$\begin{aligned}\boldsymbol{w}_j=&\left(2\theta_j\Big(C+D\Big(M-1\Big)U_j^TU_j+I\Big)\right)^{-1}U_j^{\mathrm{T}}\times\\&\left(\frac{2\theta_jC}{M}\boldsymbol{1}+2\theta_jC\boldsymbol{c}+2\theta_jD\sum_{t=1,t\neq j}^M U_t\boldsymbol{w}_t\right)\end{aligned} \tag{9}$$

采用如下方法对 β 进行更新迭代

$$\beta^{(k+1)}=\frac{1}{M}\sum_{j=1}^{M}\frac{\left(\boldsymbol{w}_j^{\mathrm{T}}\right)^{(k)}\left(\boldsymbol{w}_j\right)^{(k)}}{2\left(\theta_j^2\right)^{(k)}\left(v+2\left(1-v\right)\left(\theta_j\right)^{(k)}\right)} \tag{10}$$

算法 1. IDMK-SVM 算法

输入. 设 $\{\boldsymbol{x}_i,y_i\}_{i=1}^{l}$, $\boldsymbol{x}_i\in\mathbf{R}^d$, $y_i\in\{-1,+1\}$. M 个核函数, 惩罚权重 C, D, 核权重比 $0<v<1$, 阀值 γ.

初始化.

由 $\phi_j(\boldsymbol{x}_i)=S^{-\frac{1}{2}}P^{\mathrm{T}}(K_j(\boldsymbol{x}_i,\boldsymbol{x}_1),\cdots,K_j(\boldsymbol{x}_i,\boldsymbol{x}_l))$, 计算 U_j, $j=1,\cdots,M$. 求解 $vs+(1-v)s^2=1/M$, 令

$$\beta^{(1)}=\frac{1}{M}\sum_{j=1}^{M}\frac{(\boldsymbol{w}_j^{\mathrm{T}})^{(1)}(\boldsymbol{w}_j)^{(1)}}{2(\theta_j^2)^{(1)}(v+2(1-v)(\theta_j)^{(1)})},\ \ k=1$$

Do

$k'=1$

Do

$$\boldsymbol{c}^{(k'+1)}=\frac{1}{M}\sum_{j=1}^{M}(U_j\boldsymbol{w}_j^{(k')}-\frac{1}{M}\boldsymbol{1}-\boldsymbol{c}^{(k')})+\boldsymbol{c}^{(k')}$$

//更新 $\boldsymbol{c}$

$$\boldsymbol{w}_j^{(k'+1)}=(2\theta_j^{(k)}(C+D(M-1)\times U_j^{\mathrm{T}}U_j+I))^{-1}U_j^{\mathrm{T}}\times(\frac{2\theta_j^{(k)}C}{M}\boldsymbol{1}+2\theta_j^{(k)}C\boldsymbol{c}^{(k')}+2\theta_j^{(k)}D\sum_{t=1,t\neq j}^{M}U_t\boldsymbol{w}_t^{(k')}),\quad j=1,\cdots,M$$

//更新 $\boldsymbol{w}$

While $(|\boldsymbol{c}^{(k'+1)}-\boldsymbol{c}^{(k')}|\leq\gamma)$

$k=k'$

$$\theta_j^{(k+1)}=\theta_j^{(k)}-\frac{1}{\frac{(\boldsymbol{w}_j^{\mathrm{T}})^{(k)}(\boldsymbol{w}_j)^{(k)}}{\theta_j^3}+2(1-v)\beta}\times(-\frac{1}{2}\frac{(\boldsymbol{w}_j^{\mathrm{T}})^{(k)}(\boldsymbol{w}_j)^{(k)}}{\theta_j^2}+\beta(v+2(1-v)\theta_j)),\quad j=1,\cdots,M$$

If $(\theta_j^{(k+1)}\leq 0)$, $\theta_j^{(k+1)}=0$, 并且接下来的循环不再使用这个核　　//不使用泛化能力弱的核

$$\beta^{(k+1)}=\frac{1}{M}\times\sum_{j=1}^{M}\frac{(\boldsymbol{w}_j^{\mathrm{T}})^{(k)}(\boldsymbol{w}_j)^{(k)}}{2(\theta_j^2)^{(k)}(v+2(1-v)(\theta_j)^{(k)})},\quad j=1,\cdots,M$$

$k=k+1$

While $(|O_{k+1}-O_k|\leq\gamma)$

End.

3　定理

定理 1. 当 $0<v<1$ 时, 核权重 θ_j 是唯一的, 并且核权重的选取是稀疏的, 即泛化性能较弱的核将不会被选取.

证明. 设 $\boldsymbol{w}_j^*$ 和 θ_j^* 为第 j 个核函数取得最优值时 $\boldsymbol{w}_j$ 和 θ_j 的取值. 通过式 (8) 可知

$$-\frac{1}{2}\frac{\boldsymbol{w}_j^{\mathrm{T}^*}\boldsymbol{w}_j^*}{\theta_j^{2^*}}+\beta(v+2(1-v)\theta_j^*)=0 \tag{11}$$

将式 (11) 变形, 结合 $\beta\neq 0$, $1-v\neq 0$, 可得

$$\theta_j^{3^*}+\frac{v}{2(1-v)}\theta_j^{2^*}-\frac{\boldsymbol{w}_j^{\mathrm{T}^*}\boldsymbol{w}_j^*}{4\beta(1-v)}=0 \tag{12}$$

视 θ_j^* 为未知数, 求解该方程, 可得式 (12) 的解:

$$\theta_j^*=\sqrt[3]{\frac{\left(\frac{v}{72(1-v)^3}+\frac{\boldsymbol{w}_j^{\mathrm{T}^*}\boldsymbol{w}_j^*}{4\beta(1-v)}\right)+\sqrt{\Delta}}{2}}+\sqrt[3]{\frac{\left(\frac{v}{72(1-v)^3}+\frac{\boldsymbol{w}_j^{\mathrm{T}^*}\boldsymbol{w}_j^*}{4\beta(1-v)}\right)-\sqrt{\Delta}}{2}}-\frac{v}{6(1-v)} \tag{13}$$

其中, Δ 是一个正常数. 从式 (13) 中可以看出 θ_j 具有不确定性, 当 $\theta_j<0$ 时, 在迭代的过程中, $\theta_j=0$, 说明会有一部分泛化性能较弱的核函数不会被选中. 因此, IDMK-SVM 算法保留了 GMKL 算法的稀疏性. □

推论 1. 当 $v=1$ 时, 核权重的解是稀疏并且不是唯一的; 当 $v=0$ 时, 核权重的解是非稀疏的并且是唯一的.

证明. 当 $v=1$ 时, 核权重的限定退化为 $\sum_{j=1}^{M}\theta_j\leq 1$, 即多核学习算法的 L_1 范式模型. 此时式 (11) 退化为

$$-\frac{1}{2}\frac{\boldsymbol{w}_j^{\mathrm{T}^*}\boldsymbol{w}_j^*}{\theta_j^{2^*}}+\beta=0 \tag{14}$$

从式 (14) 中不难看出, 在采用迭代法对 θ_j 进行更新的过程中, θ_j 既可能向大于 0 的方向进行变化, 也可能向小于 0 的方向进行变化, 从而造成了解的不

唯一. 其中在迭代过程中小于 0 的部分将会被设置为 0, 从而保证了泛化能力弱的核函数不会被选取.

当 $v=0$, 核权重的限定退化为 $\sum_{j=1}^{M}\theta_j^2\le 1$, 即多核学习算法的 L_2 范式模型. 此时式 (11) 退化为

$$-\frac{1}{2}\frac{\boldsymbol{w}_j^{\mathrm{T}^*}\boldsymbol{w}_j^*}{\theta_j^{2^*}}+2\beta\theta_j^*=0 \tag{15}$$

对式 (15), 其解 $\theta_j^*>0$ 总是成立. 说明 L_2 范式模型没有稀疏性, 即所有的核函数都会被选取, 同时核权重是唯一的. 一方面虽然该模型保证了不会错误的丢掉有用的核函数, 但另外一方面也可能导致由于无用核过多而降低了算法的分类效果. □

从定理 1 可以看出, 本文的 IDMK-SVM 模型及 GMKL 模型结合了 L_1 范式和 L_2 范式多核学习模型的优点.

定理 2. IDMK-SVM 模型是收敛的.

证明. 我们只需证明内层循环收敛即可. 在迭代过程中, $\boldsymbol{w}, c$ 交替进行更新, 我们先视其他变量为常量, 分析变量 $\boldsymbol{w}_j, c$ 的变化. 由 IDMK-SVM 更新模型可知

$$\begin{aligned}\boldsymbol{c}^{(k+1)}=\ &\boldsymbol{c}^{(k)}+\\ &\frac{1}{M}\sum_{j=1}^{M}\left(U_j\boldsymbol{w}_j^{(k)}-\frac{1}{M}\boldsymbol{1}-\boldsymbol{c}^{(k)}\right)=\\ &\frac{1}{M}\sum_{j=1}^{M}\left(U_j\boldsymbol{w}_j^{(k)}-\frac{1}{M}\boldsymbol{1}\right)=\\ &\frac{1}{M}U_j\boldsymbol{w}_j^{(k)}+S_1\end{aligned} \tag{16}$$

$$\begin{aligned}\boldsymbol{w}_j^{(k+1)}=\ &\Big(2\theta_j^{(k)}(C+D(M-1)U_j^{\mathrm{T}}U_j+\\ &I)\Big)^{-1}U_j^{\mathrm{T}}\frac{2\theta_j^{(k)}C}{M}\boldsymbol{1}+2\theta_j^{(k)}C\boldsymbol{c}^{(k)}+\\ &2\theta_j^{(k)}D\sum_{t=1,t\neq j}^{M}U_t\boldsymbol{w}_t^{(k)}=\\ &S_2\left(S_3+2\theta_j^{(k)}C\boldsymbol{c}^{(k)}\right)\end{aligned} \tag{17}$$

其中, $S_1=\frac{1}{M}\sum_{t=1,t\neq j}^{M}U_t\boldsymbol{w}_t^{(k)}-\frac{1}{M}\boldsymbol{1}$ 可视为常量; $S_2=(2\theta_j^{(k)}(C+D(M-1)U_j^{\mathrm{T}}U_j+I))^{-1}U_j^{\mathrm{T}}$, $S_3=\frac{2\theta_j^{(k)}C}{M}\boldsymbol{1}+2\theta_j^{(k)}D\sum_{t=1,t\neq j}^{M}U_t\boldsymbol{w}_t^{(k)}$ 也可视为常量. 从式 (16) 可以看出 $\boldsymbol{c}$ 的更新取决于 $\boldsymbol{w}_j$ 的变化; 从式 (17) 可以看出 $\boldsymbol{w}_j$ 的更新取决于 $\boldsymbol{c}$ 的变化. 利用式 (16) 和 (17), 消掉 $\boldsymbol{w}_j$, 可得

$$\begin{aligned}\boldsymbol{c}^{(k+2)}=\ &S_1+\frac{1}{M}U_j\boldsymbol{w}_j^{(k+1)}=\\ &S_1+\frac{1}{M}U_j(S_2(S_3+2\theta_j^{(k)}C\boldsymbol{c}^{(k)}))=\\ &\frac{2\theta_j^{(k)}}{M}U_jS_2\boldsymbol{c}^{(k)}+\frac{1}{M}U_jS_2S_3+S_1\end{aligned} \tag{18}$$

从式 (18) 不难看出, $\boldsymbol{c}$ 是否收敛完全取决于 $\frac{2\theta_j^{(k)}C}{M}U_jS_2$. 通过对 U_jS_2 进行变形, 可得

$$\begin{aligned}&U_jS_2=\\ &\quad U_j(2\theta_j^{(k)}(C+D(M-1)U_j^{\mathrm{T}}U_j+I))^{-1}U_j^{\mathrm{T}}\end{aligned} \tag{19}$$

设

$$U_j'=\sqrt{2\theta_j^{(k)}(C+D(M-1))}U_j$$

式 (19) 变成

$$\begin{aligned}U_jS_2=\ &\frac{1}{2\theta_j^{(k)}(C+D(M-1))}U_j'\times\\ &(U_j'^{\mathrm{T}}U_j'+I)^{-1}U_j'^{\mathrm{T}}\end{aligned} \tag{20}$$

由 Woodbury 公式 $(C+VV^{\mathrm{T}})^{-1}=C^{-1}-C^{-1}V(I+V^{\mathrm{T}}C^{-1}V)^{-1}V^{\mathrm{T}}C^{-1}$ 可知, 式 (20) 可变为

$$\begin{aligned}&U_jS_2=\\ &\frac{1}{2\theta_j^{(k)}(C+D(M-1))}(I-(I+U_j'U_j^{\mathrm{T}})^{-1})\ll\\ &\frac{1}{2\theta_j^{(k)}(C+D(M-1))}I\end{aligned} \tag{21}$$

于是易得 $\frac{2\theta_j^{(k)}\boldsymbol{c}}{M}U_jS_2\ll\frac{1}{M}I$. 这样有 $\boldsymbol{c}^{(k+2)}\ll\frac{1}{M}\boldsymbol{c}^{(k)}+\frac{1}{M}U_jS_2S_3+S_1$, 显然可得 $\boldsymbol{c}$ 是收敛的, 因此 $\boldsymbol{w}_j$ 也是收敛的. □

4 实验

我们采用 Matlab 软件进行模拟实验. 所有的多核函数由以下几种基本核构成: RBF 核 $K(\boldsymbol{x}_i,\boldsymbol{x}_j)=\mathrm{e}^{-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|/\sigma^2}$, 其中 σ 取值为 $\{2^{-3},\cdots,2^6\}$, 多项式核 $K(\boldsymbol{x}_i,\boldsymbol{x}_j)=(\boldsymbol{x}_i{}^{\mathrm{T}}\boldsymbol{x}_j+1)^s$, 其中 s 取值为 1, 2, 3. 这样基本核共有 13 个. 核矩阵的构造方法和 GMKL 模型相同: 对样本 $\boldsymbol{x}_i$, $i=1,\cdots,l$, 选取其中一个维度 $\boldsymbol{x}_i{}'$, $i=1,\cdots,l$ 构造核矩阵, 这样的核函数的核矩阵有 $13d$ 个, 其中 d 表示样本的维度. 然后再对样本的所有维度构造一个核矩阵. 因此核函数的数量为 $13(d+1)$.

为证明对比实验的可靠性, 本文的多核模型采用的终止条件 λ 设为 0.01, v 设为 0.5, 惩罚因子 C 设为 10, D 设为 10. 在对比实验中, 我们采用 L_1 范式的 MKL (SimpleMKL), L_1 范式的 MKL 及 GMKL 作为对比模型, 其模型参数均与 IDMK-SVM 相同. 训练集仿照 L_2 范式的 MKL、SimpleMKL 及 GMKL 随机抽取一半样本做为训练集, 剩余的一半样本做为测试集, 每个模型均重复 10 次, 最后取均值作为分类结果和训练时间. 实验采用的数据集为 UCI 数据集, 这些数据集均为二值的分类问题 (见表 1). 计算机的配置为 CPU: Intel Core2 T6400, 2.00 GHz, 2 G 内存, 256 G 硬盘.

表 1　数据集
Table 1　Database

名称	数据格式	维度
Autralian	690	14
Diabetes	768	9
Heart	270	13
Ionosphere	351	33
Liverdisorder	345	6
Sonar	208	60

表 2 和表 3 反映了分类模型的平均精确度、运行时间和选取的核函数个数. 实验表明, IDMK-SVM 的精确度高于 SimpleMKL 和 L_2-MKL, 略高于 GMKL 模型以及 MulitK-MHKS 模型, 平均高于其他多核学习模型 1% 左右. 在 5 种多核学习模型中, IDMK-SVM 表现出了较好的分类性能, 其精确度为: 87.14, 76.37, 84.99, 92.45, 69.10, 81.73. 可见, IDMK-SVM 对于分类模型的确有一定的精确度改进.

接下来本文讨论时间复杂度. 对于 IDMK-SVM 模型来说, 计算 U_j 所需的时间复杂度 $\mathrm{O}(Ml^2)$, 及迭代过程中耗费的时间, 其时间复杂度相对较高. 从表 3 看出, IDMK-SVM 模型与 SimpleMKL 的运行时间大致相同, 高于 GMKL, MulitK-MHKS 和 L_2-MKL. 因此, IDMK-SVM 仍然不适合样本量很大的分析. 在运行时间方面, 本文提出的模型并不具有运算优势, 这是在今后的研究中值得改进的地方.

5　结论

GMKL 是近年来提出的一种新型的多核学习模型, 具有较好的泛化性能. 与大多数多核学习模型一样, GMKL 没有考虑在已选用核函数中, 如何尽可能获取核函数间的共性. 另一方面, MultiK-MHKS 模型未能很好地考虑如何有效合理筛选泛化能力较高的核函数. 本文提出的 IDMK-SVM 模型, 是基于 MultiK-MHKS 和 GMKL 提出的一种新型的多核学习模型, 能够先筛选出对分类精确度贡献较大的核函数, 再利用 CCA 提升算法的泛化能力.

表 2　分类准确性 (%)
Table 2　Classification accuracy (%)

	Australian	Diabetes	Heart	Ionosphere	Liverdisorder	Sonar
SimpleMKL	84.30	73.44	83.70	91.48	43.35	76.70
GMKL	84.88	76.33	80.37	92.27	66.47	78.27
L2-MKL	83.63	75.32	82.44	90.05	67.98	77.44
MultiK-MHKS	85.80	71.46	82.96	90.05	68.20	75.96
IDMK-SVM	87.14	76.37	84.99	92.45	69.10	81.73

表 3　运行时间 (s)
Table 3　Runtime (s)

	Australian	Diabetes	Heart	Ionosphere	Liverdisorder	Sonar
SimpleMKL	55.25	21.83	14.36	28.95	10.95	153.41
GMKL	14.63	21.83	13.09	5.77	5.85	28.12
L2-MKL	24.94	17.46	10.06	7.14	7.41	30.76
MultiK-MHKS	37.24	21.13	13.43	10.76	13.23	128.11
IDMK-SVM	54.08	24.37	15.67	24.11	14.14	130.91

通过实验可以看出, 相比于传统的 L_1 范式和 L_2 范式的多核学习模型, IDMK-SVM 保持了 GMKL 的优秀分类性能, 且在所有数据集上表现都比 GMKL 分类效果更加好. 然而, 本文提出的 IDMK-SVM 模型运行时间较长, 一定程度上降低了实用性. 如何进一步提高模型运行速度是继续研究的重点.

References

1 Wang Hong-Qiao, Sun Fu-Chun, Cai Yan-Ning, Chen Ning, Ding Lin-Ge. On multiple kernel learning methods. *Acta Automatica Sinica*, 2010, **36**(8): 1037−1050
(汪洪桥, 孙富春, 蔡艳宁, 陈宁, 丁林阁. 多核学习方法. 自动化学报, 2010, **36**(8): 1037−1050)

2 Liu Jian-Wei, Li Shuang-Cheng, Luo Xiong-Lin. Classification algorithm of support vector machine via p-norm regularization. *Acta Automatica Sinica*, 2012, **38**(1): 76−87
(刘建伟, 李双成, 罗雄麟. p 范数正则化支持向量机分类算法. 自动化学报, 2012, **38**(1): 76−87)

3 Ying Wen-Hao, Wang Shi-Tong, Deng Zhao-Hong, Wang Jun. Support vector machine for domain adaptation based on class distributioned. *Acta Automatica Sinica*, 2012, **38**(12): 1−16
(应文豪, 王士同, 邓赵红, 王骏. 基于类分布的领域自适应支持向量机. 自动化学报, 2012, **38**(12): 1−16)

4 Liu Qiao, Qin Zhi-Guang, Chen Wei, Zhang Feng-Li. Zero-norm penalized feature selection support vector machine. *Acta Automatica Sinica*, 2011, **37**(2): 252−256
(刘峤, 秦志光, 陈伟, 张凤荔. 基于零范数特征选择的支持向量机模型. 自动化学报, 2011, **37**(2): 252−256)

5 Xu Z L, Jin R, Zhu S H, Lyu M R, King I. Smooth optimization for effective multiple kernel learning. In: Proceedings of the 24th the Association for the Advancement of Artificial Intelligence. California, America: AAAI, 2010. 543−549

6 Han Yan-Jun, Wang Jue. A bi-sparse relational learning algorithm based on multiple kernel learning. *Journal of Machine Learning Research Development*, 2010, **47**(3): 1400−1406
(韩彦军, 王珏. 基于多核学习的双稀疏关系学习算法. 计算机研究与发展, 2010, **47**(8): 1400−1406)

7 Sonnenburg S, Rätsch G, Schäfer C, Schölkopf B. Large scale multiple kernel learning. *Journal of Machine Learning Research*, 2006, **7**(1): 1531−1565

8 Bach F R, Lanckriet G R G, Jordan M I. Multiple kernel learning, conic duality, and the SMO algorithm. In: Proceedings of the 21st International Conference Machine Learning. New York, USA: ACM, 2004. 6−13

9 Xu Z L, Jin R, Yang H Q, King I, Lyu M R. Simple and efficient multiple kernel learning by group lasso. In: Proceedings of the 2010 International Conference Machine Learning. Haifa, Israel: ICML, 2010. 1−8

10 Duan L X, Tsang I W, Xu D. Domain transfer multiple kernel learning. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 2012, **34**(3): 123−131

11 Wang Z, Chen S C, Sun T K. MultiK-MHKS: a novel multiple kernel learning algorithm. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 2008, **30**(2): 12−18

12 Rakotomamonjy A, Bach F, Canu S, Grandvalet Y. SimpleMKL. *Journal of Machine Learning Research*, 2008, **9**(1): 2491−2521

13 Cortes C, Mohri M, Rostamizadeh A. L_2 regularization for learning kernels. In: Proceedings of the 25th Conference on Uncertainty Artificial Intelligence. Arlington, Virginia, United States: AUAI Press, 2009. 1−8

14 Yang H Q, Xu Z L, Ye J P, King I, Lyu M R. Efficient sparse generalized multiple kernel learning. *IEEE Transactions on Neural Networks*, 2011, **22**(3): 433−446

15 Hardoon D R, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: an overview with application to learning methods. *Neural Computering*, 2004, **16**(12): 2639−2664

张凯军 中国人民大学硕士研究生. 主要研究方向为数据挖掘. 本文通信作者. E-mail: zhangkaijun801210@163.com
(**ZHANG Kai-Jun** Master student at Renmin University of China. His main interest is data mining. Corresponding author of this paper.)

梁 循 中国人民大学信息学院教授. 主要研究方向为互联网信息分析, 数据挖掘, 商务智能, 社会计算. E-mail: xliang@ruc.edu.cn
(**LIANG Xun** Professor at the School of Information, Renmin University of China. His research interest covers internet information analysis, data mining, business intelligence, and social computing.)