

基于积分算子空间显式描述的框架核选择方法

刘勇, 廖士中*

天津大学计算机科学与技术学院, 天津 300072

* 通信作者. E-mail: szliao@tju.edu.cn

收稿日期: 2015-05-05; 接受日期: 2015-06-08; 网络出版日期: 2015-07-24

国家自然科学基金(批准号: 61170019)资助项目

摘要 积分算子特征值在核方法(kernel methods)学习理论分析中有重要作用. 但对于大多数常用核函数, 分析其积分算子特征值十分困难. 针对该问题, 提出了一种基于积分算子空间显式描述的框架核选择方法. 首先, 基于框架理论(frame theory)提出一种新的框架核构造方法, 该框架核能结合框架表示与核方法的优点. 然后, 显式描述框架核函数对应积分算子的特征值与特征函数, 并构造其对应的再生核希尔伯特空间(reproducing kernel Hilbert space, RKHS)和特征映射(feature mapping), 由此精确刻画核方法解的结构, 增强核方法模型可解释性. 在此基础上, 基于框架核所对应的积分算子特征值推导具有 $O(\frac{1}{m})$ 收敛率的泛化误差界, 而经典的Rademacher和半径间隔(radius-margin)泛化误差界的收敛率为 $O(\frac{1}{\sqrt{m}})$, m 为样本规模. 应用该泛化误差界, 提出基于积分算子特征值的框架核选择方法, 并将其形式化为易处理的凸优化问题, 可利用随机次梯度(stochastic sub-gradient)算法高效求解. 最后, 实验验证所提方法的精确性与有效性. 理论分析与实验结果表明, 所提出的基于积分算子特征值的框架核选择方法是合理且高效的.

关键词 框架理论 框架核 模型选择 核选择 积分算子 特征值

1 引言

由核函数诱导的积分算子特征值在核方法(kernel methods)理论分析中扮演重要的角色. Smale和Zhou^[1]最先考虑利用积分算子代替覆盖数(covering number)来研究核方法学习理论. Vito等^[2]应用积分算子度量正则化最小二乘方法的泛化误差上界, 并将正则化参数选择问题归约为求解偏差—方差问题(bias-variance). Vito等^[3]进一步考虑利用积分算子研究Tikhonov正则化的逆问题(inverse problem). Guo和Zhou^[4]推广了上述结果, 研究值域无界情况下的正则化最小二乘学习理论, 并应用积分算子得到了与有界样本情况下相同的学习率(learning rate). Steinwart等^[5]发展了上述理论, 利用积分算子特征值研究核方法学习理论, 得到了在积分算子特征值满足多项式衰减假设下正则化最小二乘方法的最优学习率. Eberts和Steinwart^[6]推广上述结果, 基于Gauss核函数对应积分算子特征值的指数衰减率, 得到了支持向量机最优学习率. 上述学习理论只考虑平方损失(2范数), Lin等^[7]将

引用格式: 刘勇, 廖士中. 基于积分算子空间显式描述的框架核选择方法. 中国科学: 信息科学, 2016, 46: 165–178, doi: 10.1360/N112015-00095

上述结果推广到 q 范数, $0 < q < \infty$. 作者前期建立了积分算子特征值稳定性理论^[8,9], 并给出了紧的基于积分算子特征值的泛化误差界^[10], 同时考虑应用稳定性理论来缓解交叉验证的过拟合现象^[11,12].

虽然积分算子特征值在核方法学习理论研究中有重要作用, 但对于大多数常见的核函数, 如 Gauss 核和多项式核, 描述其积分算子特征值十分困难, 阻碍了积分算子特征值学习理论的推广与应用. 框架 (frame) 是正交基的一种推广形式, 其冗余性使其具有更好的鲁棒性^[13,14]. 近年来, 基于框架理论 (frame theory) 来构造核函数受到关注. Gao 等^[15] 研究正则化算子 (regularization operator) 与支持向量核 (support vector kernel) 的关系, 提出了一种框架核构造方法. 文献^[16,17] 建立框架与再生核希尔伯特空间 (RKHS) 的联系, 继而给出了一种框架核构造方法. Opfer^[18] 推广上述结果, 提出了紧框架核函数构造方法.

上述框架核构造方法都需要计算对偶框架, 而对偶框架的计算十分困难^[16]. 针对这一问题, 提出一种新的框架核构造方法. 该方法无需计算对偶框架, 而且所构造的框架核还具有如下性质: (1) 解析表示核函数对应的积分算子特征值与特征函数, 为核方法理论分析与应用带来便利. (2) 显式描述核函数诱导 RKHS, 可获得假设空间 (hypothesis space) 解的清晰结构, 从而增强核方法模型的可解释性. (3) 显式构造框架核特征映射, 从而可利用线性支持向量机高效求解非线性分类问题^[19~21]; 进一步, 基于框架核对应的积分算子特征值可显式描述这一特性, 推导收敛率可达 $O(\frac{1}{m})$ 的泛化误差界, 其中 m 为样本规模. 而经典的 Rademacher 泛化误差界^[22] 和半径间隔 (radius-margin) 泛化误差界^[23] 的收敛率为 $O(\frac{1}{\sqrt{m}})$. 依据所建立的紧泛化误差界, 提出一种基于积分算子特征值的框架核选择方法. 该方法不但能保证泛化性能, 同时能形式化为凸优化问题, 可应用随机次梯度算法高效求解. 最后, 在多个标准数据集上验证了所提方法的精确性与有效性.

2 预备知识

样本 $S = \{z_i = (\mathbf{x}_i, y_i)\}_{i=1}^m$ 依定义在 $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ 上的分布 P 生成, $\mathcal{X} \subseteq \mathbb{R}^d$, $\mathcal{Y} \subseteq \mathbb{R}$. 假设 $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ 为核函数, 即对 \mathcal{X} 中的任意有限集 $\{\mathbf{x}_i\}_{i=1}^m$, 对应的核矩阵 $\mathbf{K} = [K(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^m$ 为对称半正定矩阵. 核函数 K 对应的 RKHS 为 $\{K(\mathbf{x}, \cdot) : \mathbf{x} \in \mathcal{X}\}$ 线性张成空间的完备集, 记为 $\mathcal{H}_K = \overline{\text{span}\{K(\mathbf{x}, \cdot) : \mathbf{x} \in \mathcal{X}\}}$, 其内积 $\langle \cdot, \cdot \rangle_K$ 满足

$$\left\langle \sum_{i=1}^m \alpha_i K(\mathbf{x}_i, \cdot), \sum_{j=1}^m \beta_j K(\mathbf{x}'_j, \cdot) \right\rangle_K = \sum_{i,j=1}^m \alpha_i \beta_j K(\mathbf{x}_i, \mathbf{x}'_j).$$

核函数 K 对应的积分算子 $T_K : L^2(\mathcal{X}) \rightarrow L^2(\mathcal{X})$ 定义为

$$(T_K f)(\cdot) = \int_{\mathcal{X}} K(\cdot, \mathbf{x}) f(\mathbf{x}) dP_{\mathcal{X}}(\mathbf{x}), \quad (1)$$

其中, $P_{\mathcal{X}}$ 为 \mathcal{X} 上的边缘分布, $L^2(\mathcal{X})$ 为平方可积空间. 积分算子 T_K 为自伴随紧算子, 特征值满足 $\lambda_1(T_K) \geq \lambda_2(T_K) \geq \dots \geq \lambda_i(T_K) \geq \dots \geq 0$. 当样本趋于无穷时, $\frac{1}{m} \mathbf{K}$ 的特征值和特征向量收敛于积分算子 T_K 的特征值和特征函数^[24]. 因此, 分析积分算子特征值对理论分析和实际应用核方法有重要作用^[5,25].

本文考虑正则化最小二乘学习器 (核岭回归)

$$f_S = \arg \min_{f \in \mathcal{H}_K} \left\{ \frac{1}{m} \sum_{i=1}^m \ell(f(\mathbf{x}_i), y_i) + g \|f\|_K^2 \right\}, \quad (2)$$

其中, $\ell(f(\mathbf{x}), y) = (f(\mathbf{x}) - y)^2$ 为平方损失, g 为平衡系数 (trade off coefficient). 学习器性能优劣一般由泛化误差 $R(f_S)$ 度量,

$$R(f_S) = \int_{z \in \mathcal{Z}} \ell(f_S(\mathbf{x}), y) dP(z).$$

由于概率分布 P 未知, $R(f_S)$ 无法直接计算. 经验误差 $R_{\text{emp}}(f_S)$ 是其有效估计,

$$R_{\text{emp}}(f_S) = \frac{1}{m} \sum_{i=1}^m \ell(f_S(\mathbf{x}_i), y_i).$$

框架理论. 集合 $\{\phi_i\}_{i \in \Gamma}$ 称为希尔伯特空间 \mathcal{H} 的框架, 如果满足 $\forall f \in \mathcal{H}, A\|f\|_{\mathcal{H}}^2 \leq \sum_{i \in \Gamma} |\langle f, \phi_i \rangle_{\mathcal{H}}|^2 \leq B\|f\|_{\mathcal{H}}^2$. 其中, A, B 称为框架常数. 如果 $A = B$, 框架称为紧框架. 假设 $\{\phi_i\}_{i \in \Gamma}$ 为 \mathcal{H} 的框架, 令算子 $U: \mathcal{H} \rightarrow \mathcal{H}$ 为 $U(f) = \sum_{i \in \Gamma} \langle f, \phi_i \rangle_{\mathcal{H}} \phi_i$, 则任取 $f \in \mathcal{H}$,

$$f = \sum_{i \in \Gamma} \langle f, \bar{\phi}_i \rangle_{\mathcal{H}} \phi_i = \sum_{i \in \Gamma} \langle f, \phi_i \rangle_{\mathcal{H}} \bar{\phi}_i, \quad (3)$$

其中, $\{\bar{\phi}_i = U^{-1}\phi_i\}_{i \in \Gamma}$ 为 $\{\phi_i\}_{i \in \Gamma}$ 的对偶框架^[16]. 当 $A = B = 1, \|\phi_i\|_{\mathcal{H}} = 1$ 时, $\{\phi_i\}_{i \in \Gamma}$ 为 \mathcal{H} 的标准正交基. 框架是正交基的推广, 其冗余性使其具有更好的鲁棒性^[16, 18].

本文将利用框架理论提出一种全新的框架核构造方法, 并依据框架核积分算子特征值可显式表示这一特性, 度量 $R(f_S)$ 与 $R_{\text{emp}}(f_S)$ 的偏差上界, 最后基于该偏差上界给出框架核选择方法.

3 框架核构造及显式描述

本节将给出一种全新的框架核构造方法并显式描述该核对应积分算子的特征值与特征函数.

3.1 框架核构造

核方法利用核诱导的隐式特征映射 $\phi: \mathcal{X} \rightarrow \mathcal{F}$, 将输入空间 \mathcal{X} 的数据映射到高维特征空间 \mathcal{F} , 继而在 \mathcal{F} 中学习线性学习器. 但该隐式描述也带来如下问题: (1) 不能清晰地描述解的结构; (2) 很难确定核函数的优劣^[26]. 这节将通过框架构造一类新的核函数. 由于框架是一种广义基, 该核函数能结合框架表示 (frame representation) 与核方法的优点.

定理1 假设 $\{\phi_i\}_{i \in \Gamma}$ 为希尔伯特空间 $\mathcal{H} \subseteq L^2(\mathcal{X})$ 的框架, $\lambda_i \in \mathbb{R}_0^+$, $\sum_{i \in \Gamma} \lambda_i \leq \infty$. 令函数空间

$$\mathcal{H}_K := \left\{ f = \sum_{i \in \Gamma} \alpha_i \sqrt{\lambda_i} \phi_i : \alpha_i \in \mathbb{R} \right\}, \quad (4)$$

内积 $\langle \cdot, \cdot \rangle_{\mathcal{H}_K}$ 定义为 $\langle f, g \rangle_{\mathcal{H}_K} = \sum_{i \in \Gamma} \langle f, \tilde{\phi}_i \rangle_{L^2} \langle g, \tilde{\phi}_i \rangle_{L^2}, \forall f, g \in \mathcal{H}_K$, 其中, $L^2(\mathcal{X})$ 为平方可积空间, $\langle \cdot, \cdot \rangle_{L^2}$ 为 $L^2(\mathcal{X})$ 上的内积, $\{\tilde{\phi}_i\}_{i \in \Gamma}$ 为 $\{\sqrt{\lambda_i} \phi_i\}_{i \in \Gamma}$ 的对偶框架. 则 $(\mathcal{H}_K, \langle \cdot, \cdot \rangle_{\mathcal{H}_K})$ 为 RKHS, 其相应的再生核 (reproducing kernel) 为

$$K(\mathbf{x}, \mathbf{x}') = \sum_{i \in \Gamma} \lambda_i \phi_i(\mathbf{x}) \phi_i(\mathbf{x}'), \quad (5)$$

称为框架核.

证明 必须证明如下两点:

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

(2) 再生性, 即 $\langle f(\cdot), K(\mathbf{x}, \cdot) \rangle_{\mathcal{H}_K} = f(\mathbf{x})$.

因为 $K(\mathbf{x}, \cdot) = \sum_{i \in \Gamma} \lambda_i \phi_i(\cdot) \phi_i(\mathbf{x})$, 易证 $K(\mathbf{x}, \cdot) \in \mathcal{H}_K$.

下面验证再生性. 由 \mathcal{H}_K 的内积定义可得

$$\begin{aligned} \langle f(\cdot), K(\mathbf{x}, \cdot) \rangle_{\mathcal{H}_K} &= \sum_{i \in \Gamma} \left\langle f(\cdot), \tilde{\phi}_i(\cdot) \right\rangle_{L^2} \left\langle \sum_{j \in \Gamma} \lambda_j \phi_j(\cdot) \phi_j(\mathbf{x}), \tilde{\phi}_i(\cdot) \right\rangle_{L^2} \\ &= \sum_{i \in \Gamma} \left\langle f(\cdot), \tilde{\phi}_i(\cdot) \right\rangle_{L^2} \sum_{j \in \Gamma} \langle \tilde{\phi}_i(\cdot), \sqrt{\lambda_j} \phi_j(\cdot) \rangle_{L^2} \sqrt{\lambda_j} \phi_j(\mathbf{x}). \end{aligned}$$

依据文献 [15] 定理 3 可知 $\sum_{j \in \Gamma} \langle \tilde{\phi}_i(\cdot), \sqrt{\lambda_j} \phi_j(\cdot) \rangle_{L^2} \sqrt{\lambda_j} \phi_j(\mathbf{x}) = \sqrt{\lambda_i} \phi_i(\mathbf{x})$. 所以

$$\langle f(\cdot), K(\mathbf{x}, \cdot) \rangle_{\mathcal{H}_K} = \sum_{i \in \Gamma} \left\langle f(\cdot), \tilde{\phi}_i(\cdot) \right\rangle_{L^2} \sqrt{\lambda_i} \phi_i(\mathbf{x}).$$

根据 (3) 式有 $\sum_{i \in \Gamma} \left\langle f(\cdot), \tilde{\phi}_i(\cdot) \right\rangle_{L^2} \sqrt{\lambda_i} \phi_i(\mathbf{x}) = f(\mathbf{x})$. 则 $\langle f(\cdot), K(\mathbf{x}, \cdot) \rangle_{\mathcal{H}_K} = f(\mathbf{x})$. 证毕.

若 Γ 为无穷可数集, 可采用有限集 $\{\phi_i\}_{i=1}^n \subseteq \{\phi_i\}_{i \in \Gamma}$ 来构造框架核. 基于定理 1, 易得如下推论.

推论 1 假设 $\{\phi_i\}_{i \in \Gamma}$ 为希尔伯特空间 $\mathcal{H} \subseteq L^2(\mathcal{X})$ 的框架, $\lambda_i \in \mathbb{R}_0^+$, 令

$$\mathcal{H}_K = \left\{ \sum_{i=1}^n \alpha_i \sqrt{\lambda_i} \phi_i : \alpha_i \in \mathbb{R} \right\}.$$

内积定义为 $\langle f, g \rangle_{\mathcal{H}_K} = \sum_{i=1}^n \langle f, \tilde{\phi}_i \rangle_{L^2} \langle g, \tilde{\phi}_i \rangle_{L^2}$. 则 $(\mathcal{H}_K, \langle \cdot, \cdot \rangle_{\mathcal{H}_K})$ 为 RKHS, 其对应再生核为

$$K(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^n \lambda_i \phi_i(\mathbf{x}) \phi_i(\mathbf{z}). \quad (6)$$

考虑在正则化最小二乘学习器中使用本文所构框架核, 即

$$f_S = \arg \min_{f \in \mathcal{H}_K} \frac{1}{m} \sum_{i=1}^m \ell(f(\mathbf{x}_i), y_i) + g \|f\|_{\mathcal{H}_K}^2,$$

其中, \mathcal{H}_K 为框架核对应的 RKHS. 则 f_S 可表示为 $f_S = \sum_{i=1}^n \langle f, \tilde{\phi}_i \rangle \phi_i$. 解由框架元素表示, 增强了正则化最小二乘方法的可解释性.

3.2 具体实例

应用推论 1, 将给出框架核的一些具体实例.

3.2.1 Fourier 框架核 (Fourier frame kernels)

例 1 假设 $\{\cos(iw_i x), \sin(iw_i x)\}_{i \in \Gamma}$ 为 $L^2(\mathcal{X})$ 的框架, 基于推论 1 易证如下 $K_{\text{FFK}}(x, z)$ 为框架核

$$K_{\text{FFK}}(x, z) = \sum_{i=1}^n \begin{bmatrix} \lambda_i \cos(iw_i x) \\ \lambda'_i \sin(iw_i x) \end{bmatrix}^T \begin{bmatrix} \cos(iw_i z) \\ \sin(iw_i z) \end{bmatrix}.$$

3.2.2 小波框架核 (Wavelet frame kernels)

例 2 假设 $\{\phi_{j,k}(x) = \frac{1}{\sqrt{a^j}} \Upsilon(\frac{x-k\tau a^j}{a^j}), j, k \in \mathbb{Z}\}$ 为 $L^2(\mathcal{X})$ 的框架, 其中, $\Upsilon(\cdot)$ 为一母小波, a 和 τ 为常数, 则框架核 $K_{\text{WFK}}(x, z)$ 如下:

$$K_{\text{WFK}}(x, z) = \sum_{j=j_{\min}}^{j_{\max}} \sum_{k=k_{\min}}^{k_{\max}} \lambda_{j,k} \phi_{j,k}(x) \phi_{j,k}(z).$$

例 1 和 2 只考虑输入空间为标量的情况, 即 $\mathcal{X} \subseteq \mathbb{R}$. 可以通过简单的加法 (additive) 或者乘法 (multiplicative) 组合将框架核推广到多维情况 ($\mathcal{X} \subseteq \mathbb{R}^d$)

$$K(\mathbf{x}, \mathbf{x}') = \sum_{l=1}^d K(x_l, x'_l) \text{ 或 } K(\mathbf{x}, \mathbf{x}') = \prod_{l=1}^d K(x_l, x'_l).$$

实验中发现这两种组合得到相似的精度. 为计算方便, 实验中将采用加法组合方法.

3.3 积分算子特征值和特征函数

本小节将显式描述框架核对应积分算子的特征值与特征函数.

定理 2 假设 $\{\phi_i\}_{i \in \Gamma}$ 为 \mathcal{H} 的紧框架, 框架常数为 A . 则框架核 $K(\mathbf{x}, \mathbf{z}) = \sum_{i \in \Gamma} \lambda_i \phi_i(\mathbf{x}) \phi_i(\mathbf{z})$ 对应积分算子的特征值为 $\lambda_i(T_K) = A\lambda_i$, 相应特征函数为 $\phi_i, i \in \Gamma$.

证明 由积分算子的定义可知

$$(T_K \phi_i)(\mathbf{x}') = \int_{\mathcal{X}} \sum_{j \in \Gamma} \lambda_j \phi_j(\mathbf{x}') \phi_j(\mathbf{x}) \phi_i(\mathbf{x}) dP_{\mathcal{X}}(\mathbf{x}) = \sum_{j \in \Gamma} \lambda_j \phi_j(\mathbf{x}') \int_{\mathcal{X}} \phi_j(\mathbf{x}) \phi_i(\mathbf{x}) dP_{\mathcal{X}}(\mathbf{x}).$$

由 $\{\phi_i\}_{i \in \Gamma}$ 为 \mathcal{H} 的紧框架, 框架常数为 A , 可知 $\int_{\mathcal{X}} \phi_j(\mathbf{x}) \phi_i(\mathbf{x}) P_{\mathcal{X}}(\mathbf{x}) = A$, 当 $j = i$, 其他为 0. 因此, $(T_K \phi_i)(\mathbf{x}') = A\lambda_i \phi_i(\mathbf{x}')$. 证毕.

依据定理 2, 易将上述结果推广到框架基有限情况.

推论 2 假设 $\{\phi_i\}_{i \in \Gamma}$ 为 \mathcal{H} 的紧框架, 框架常数为 A , $\{\phi_i\}_{i=1}^n \subseteq \{\phi_i\}_{i \in \Gamma}$. 则框架核 $K(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^n \lambda_i \phi_i(\mathbf{x}) \phi_i(\mathbf{x}')$ 对应积分算子的特征值为 $\lambda_i(T_K) = A\lambda_i$, 相应特征函数为 $\phi_i, i = 1, \dots, n$.

上述结果表明框架核对应的积分算子特征值与特征函数能显式描述, 为核方法理论分析带来便利. 在第 4 节中, 将利用这一特性建立紧的泛化误差界.

3.4 特征映射

基于框架核的定义, 见 (6) 式, 容易验证

$$\Psi(\mathbf{x}) = \left(\sqrt{\lambda_1} \phi_1(\mathbf{x}), \dots, \sqrt{\lambda_n} \phi_n(\mathbf{x}) \right)^T, \quad (7)$$

为框架核的特征映射, 即满足 $K(\mathbf{x}, \mathbf{x}') = \langle \Psi(\mathbf{x}), \Psi(\mathbf{x}') \rangle$.

文献 [19~21] 提出了解决大规模线性 SVM 的方法. 因此, 可以用 $\Psi(\mathbf{x})$ 作为线性 SVM 的输入来求解大规模非线性问题.

4 框架核选择

本节将利用积分算子的特征值来推导泛化误差上界, 并提出一种全新的框架核选择方法, 即选择框架核参数 $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)^T$. 下文中, 假设 $|y| \leq M, \forall f \in \mathcal{H}_K, |f(\mathbf{x})| \leq D$.

4.1 基于积分算子特征值的泛化误差界

基于作者已有工作^[10], 可以得到如下基于积分算子特征值的泛化误差界.

定理3 假设核函数 K 对应的积分算子特征值满足

$$\lambda_i(T_K) \leq ci^{-\gamma}, \quad (8)$$

其中, $c > 0$ 和 $\gamma > 1$ 为常数. 则至少以 $1 - \delta$ 的概率, $\forall k \geq 2, t \geq 1, k, t \in \mathbb{N}$, 满足

$$R(fs) \leq \frac{k}{k-1} R_{\text{emp}}(fs) + \frac{C_1}{m},$$

其中, $C_1 = 12ktBd^2 + \frac{12kct^{1-\gamma}Bd^2}{(\gamma-1)} + \log \frac{1}{\delta}(22M + 5Bk)$, $B = (D + M)^2$.

证明 为了证明上述定理, 先引入 Rademacher 复杂度和局部 Rademacher 复杂度概念. 对于由 P 生成的样本 $\{\mathbf{x}_i\}_{i=1}^m$ 和域为 \mathcal{X} 的实值函数类 \mathcal{H} . \mathcal{H} 的 Rademacher 复杂度定义为

$$R_m(\mathcal{H}) := \mathbb{E}_{\{\mathbf{x}_i\}_{i=1}^m \in \mathcal{X}^m} \mathbb{E}_{\boldsymbol{\sigma}} \left[\sup_{f \in \mathcal{H}} \frac{1}{m} \sum_{i=1}^m \sigma_i f(\mathbf{x}_i) \right].$$

对于任意 $r > 0$, \mathcal{H} 的局部 Rademacher 复杂度为

$$R_m(\mathcal{H}, r) := R_m(\{f \in \mathcal{H} : \mathbb{E}[f^2] \leq r\}),$$

其中, $\sigma_1, \dots, \sigma_m$ 是独立的随机变量, $P(\sigma = 1) = P(\sigma = -1) = \frac{1}{2}$. 令与 \mathcal{H}_K 相关的损失函数类为

$$\ell_{\mathcal{H}_K} = \{\ell_f : f \in \mathcal{H}_K\} = \{z \mapsto \ell(f(\mathbf{x}), y) : f \in \mathcal{H}_K\}.$$

由 $\ell(f(\mathbf{x}), y)^2 = (f(\mathbf{x}) - y)^4 \leq (D + M)^2 \ell(f(\mathbf{x}), y)$, 可得 $\mathbb{E}_z \ell^2(f(\mathbf{x}), y) \leq (D + M)^2 \mathbb{E}_z \ell(f(\mathbf{x}), y)$. 根据文献 [27] 中定理 3.3 可知至少以 $1 - \delta$ 的概率,

$$\forall k > 1, R(fs) \leq \frac{k}{k-1} R_{\text{emp}}(fs) + \frac{6kr^*}{B} + \frac{\log \frac{1}{\delta}(22M + 5Bk)}{m}, \quad (9)$$

其中, r^* 为 $BR_n(\ell_{\mathcal{H}_K}, r)$ 的固定点 (fixed point)^[27]. 由于 $\forall f, g \in \mathcal{H}_K$,

$$|\ell(f(\mathbf{x}), y) - \ell(g(\mathbf{x}), y)| \leq |f(\mathbf{x}) - g(\mathbf{x})| \cdot |f(\mathbf{x}) + g(\mathbf{x}) - 2y| \leq 2(M + D)|f(\mathbf{x}) - g(\mathbf{x})|.$$

则 $\ell(f(\mathbf{x}), y) = (f(\mathbf{x}) - y)^2$ 为 $d = 2(M + D)$ -Lipschitz 连续损失函数. 根据文献 [28], 可得 $R_m(\ell_{\mathcal{H}_K}, r) \leq dR_m(\mathcal{H}_K, r)$. 由文献 [25] 可知

$$R_m(\mathcal{H}_K, r) \leq \sqrt{\frac{2}{m} \min_{\theta \geq 0} \left(\theta r + \sum_{j>\theta} \lambda_j(T_K) \right)}.$$

从而可得

$$BR_m(\ell_{\mathcal{H}_K}, r) \leq dB \sqrt{\frac{2}{m} \left(t \cdot r + \sum_{j>t} \lambda_j(T_K) \right)}.$$

令 $a = \frac{2tB^2d^2}{n}$, $b = \frac{2B^2d^2}{n} \sum_{j>t} \lambda_j(T_K)$. $BR_n(\ell_{\mathcal{H}_K}, r)$ 的固定点为 $r^2 - ar - b = 0$ 的正数根 (positive root). 容易验证

$$r^* = \frac{a + \sqrt{a^2 + 4b^2}}{2} \leq \frac{a + (a + 2b)}{2} \leq a + b.$$

由于 $\lambda_i(T_K) \leq ci^{-\gamma}$, 则

$$\sum_{i=t+1}^{\infty} \lambda_i(T_K) = \sum_{i=t+1}^{\infty} ci^{-\gamma} \leq \int_t^{\infty} cx^{-\gamma} dx = \frac{cx^{1-\gamma}}{1-\gamma} \Big|_t^{\infty} = \frac{ct^{1-\gamma}}{\gamma-1}.$$

故可得

$$r^* \leq a + b = \frac{2tB^2d^2}{m} + \frac{2ct^{1-\gamma}B^2d^2}{(\gamma-1)m}. \tag{10}$$

将 (10) 式代入 (9) 式, 可得

$$R(f_S) \leq \frac{k}{k-1} R_{\text{emp}}(f_S) + \frac{12ktBd^2}{m} + \frac{12kct^{1-\gamma}Bd^2}{(\gamma-1)m} + \frac{\log \frac{1}{\delta}(22M + 5Bk)}{m}.$$

证毕.

定理 3 给出了收敛率为 $O(\frac{1}{m})$ 的泛化误差界. 而对于经典的 Rademacher 泛化误差界 [22],

$$R(f_S) \leq R_{\text{emp}}(f_S) + \frac{R_n(\mathcal{H}_K)}{2} + \sqrt{\frac{\ln(1/\delta)}{2m}},$$

radius-margin 泛化误差界 [23],

$$R(f_S) \leq R_{\text{emp}}(f_S) + \sqrt{\frac{c \left(\frac{R^2 \log^2 m}{\rho^2} - \log \delta \right)}{m}},$$

以及最新的基于谱扰动的泛化误差界 [8], 收敛率为 $O\left(\frac{1}{\sqrt{m}}\right)$. 上述结果表明当积分算子特征值满足 $\lambda_i(T_K) \leq ci^{-\gamma}$ 时, 能保证紧的误差界. 假如没有上述限制, 在经典的 Rademacher 泛化误差理论下, 只能得到收敛率为 $O\left(\frac{1}{\sqrt{m}}\right)$ 的泛化误差界 [25].

4.2 框架核参数选择

由定理 3 可知, 当积分算子特征值满足 (8) 式, 即 $\lambda_i(T_K) \leq ci^{-\gamma}$, 能保证学习器的快速收敛率. 但对于大多数核函数, 如 Gauss 核和多项式核, 分析其对应积分算子特征值十分困难, 很难确定 (8) 式是否成立, 可能导致所选核函数对应的假设空间过大, 从而不能保证学习器的泛化性能. 但是, 对于本文所构框架核 $K(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^n \lambda_i \phi_i(\mathbf{x}) \phi_i(\mathbf{x}')$, 积分算子特征值能显式描述. 可选择满足 (8) 式的框架核, 从而保证学习器泛化性能

$$\min_{\lambda, f \in \mathcal{H}_K} \frac{1}{m} \sum_{i=1}^m (f(\mathbf{x}_i) - y_i)^2 + g \|f\|_{\mathcal{H}_K}^2, \text{ s.t. } \lambda_i^\gamma \leq h_i, i = 1, \dots, n, \tag{11}$$

其中, $\gamma, h_i, i = 1, \dots, n$ 用来限制特征值的取值使其满足 (8) 式. 优化问题 (11) 的解可表示为 $f(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\Psi}(\mathbf{x})$, 其中, $\boldsymbol{\Psi}(\mathbf{x}) = (\sqrt{\lambda_1} \phi_1(\mathbf{x}), \dots, \sqrt{\lambda_n} \phi_n(\mathbf{x}))^T$. 由文献 [29] 可知 $\|f\|_{\mathcal{H}_K}^2 = \|\mathbf{w}\|_2^2$, 容易验证优化问题 (11) 等价于

$$\min_{\mathbf{w}, \boldsymbol{\lambda}} \frac{1}{m} \sum_{i=1}^m (\mathbf{w}^T \boldsymbol{\Psi}(\mathbf{x}_i) - y_i)^2 + g \|\mathbf{w}\|_2^2 + \sum_{j=1}^n b_j \lambda_j^\gamma, \quad (12)$$

其中, b_j 为 Lagrange 系数.

下面定理表明可将优化问题 (12) 转化为更容易处理的优化形式.

定理4 假设 $\mathbf{w}^*, \boldsymbol{\lambda}^*$ 为问题 (12) 的最优解, $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_n^*)^T$ 为问题 (13) 的最优解

$$\min_{\boldsymbol{\mu}} \frac{1}{m} \sum_{i=1}^m (\boldsymbol{\mu}^T \boldsymbol{\Phi}(\mathbf{x}_i) - y_i)^2 + \sum_{j=1}^n c_j |\mu_j|^{\frac{2\gamma}{\gamma+1}}, \quad (13)$$

其中, $\boldsymbol{\Phi}(\mathbf{x}) = (\phi_1(\mathbf{x}), \dots, \phi_n(\mathbf{x}))^T, c_j = 2 \left(g b_j^{1/\gamma} \right)^{\frac{\gamma}{\gamma+1}}$. 则

$$\lambda_j^* = \left(g \mu_j^{*2} / b_j \right)^{\frac{1}{\gamma+1}}, w_j^* = \mu_j^* / \sqrt{\lambda_j^*}, j = 1, \dots, n.$$

证明 令 $\mu_j = \sqrt{\lambda_j} w_j, j = 1, \dots, n$, 优化问题 (12) 可表示为

$$\min_{\boldsymbol{\mu}, \boldsymbol{\lambda}} \frac{1}{m} \sum_{i=1}^m (\boldsymbol{\mu}^T \boldsymbol{\Phi}(\mathbf{x}_i) - y_i)^2 + \sum_{j=1}^n \left\{ \frac{g \mu_j^2}{\lambda_j} + b_j \lambda_j^\gamma \right\}.$$

由于 $\frac{1}{m} \sum_{i=1}^m (\boldsymbol{\mu}^T \boldsymbol{\Phi}(\mathbf{x}_i) - y_i)^2$ 不依赖于 $\boldsymbol{\lambda}$, 上述优化问题等价于

$$\min_{\boldsymbol{\mu}} \frac{1}{m} \sum_{i=1}^m (\boldsymbol{\mu}^T \boldsymbol{\Phi}(\mathbf{x}_i) - y_i)^2 + \min_{\boldsymbol{\lambda}} \sum_{j=1}^n \left\{ \frac{g \mu_j^2}{\lambda_j} + b_j \lambda_j^\gamma \right\}. \quad (14)$$

当 $\lambda_j = (g \mu_j^2 / b_j)^{\frac{1}{\gamma+1}}$ 时,

$$\frac{g \mu_j^2}{\lambda_j} + b_j \lambda_j^\gamma = 2 (g \mu_j^2 b_j^{\frac{1}{\gamma}})^{\frac{\gamma}{\gamma+1}} = c_j |\mu_j|^{\frac{2\gamma}{\gamma+1}}.$$

故优化问题 (14) 可转化为

$$\min_{\boldsymbol{\mu}} \sum_{i=1}^m \frac{1}{m} (\boldsymbol{\mu}^T \boldsymbol{\Phi}(\mathbf{x}_i) - y_i)^2 + \sum_{j=1}^n c_j |\mu_j|^{\frac{2\gamma}{\gamma+1}},$$

其中, $\lambda_j = (g \mu_j^2 / b_j)^{\frac{1}{\gamma+1}}$. 证毕.

上述定理表明, 可以通过求解优化问题 (13) 来求解优化问题 (12). 下节中将给出优化问题 (13) 的高效求解算法.

4.3 算法

本文将采用次梯度下降算法 (sub-gradient algorithm) 高效求解优化问题 (13) (详情参见文献 [21]). 首先, 初始化参数 $\boldsymbol{\mu}^0$, 计算目标函数次梯度, 并使用一维线性搜索确定步长 η_t , 然后更新 $\boldsymbol{\mu}^t$,

$$\boldsymbol{\mu}^{t+1} \leftarrow \boldsymbol{\mu}^t - \eta_t \Delta_t,$$

算法 1 框架核选择算法 (frame kernel selection, FKS)**Require:** $\{(\mathbf{x}_i, y_i)\}_{i=1}^m$, $c_1 = \dots = c_n = c$, γ ;1: 初始化 $\boldsymbol{\mu}^1 = 0$, $T = 0$;2: **for** $t = 1, 2, \dots$ **do**

3: 计算次梯度

$$\Delta_t = \frac{2}{m} \sum_{i=1}^m (\boldsymbol{\mu}^T \boldsymbol{\Phi}(\mathbf{x}_i) - y_i) \boldsymbol{\Phi}(\mathbf{x}_i) + \boldsymbol{\nu};$$

4: 计算步长 η_t ;5: $\boldsymbol{\mu}^{t+1} \leftarrow \boldsymbol{\mu}^t - \eta_t \Delta_t$, $T = T + 1$;6: **if** $\|\boldsymbol{\mu}^{t+1} - \boldsymbol{\mu}^t\|/\boldsymbol{\mu}^t \leq 0.01$ **then**

7: 算法终止;

8: **end if**9: **end for**10: **return** $\boldsymbol{\mu}^T$, $\boldsymbol{\lambda}^*$, $\lambda_j^* = (g\mu_j^2/b_j)^{\frac{1}{\gamma+1}}$ 其中, η_t 为步长, Δ_t 为次梯度

$$\Delta_t = \frac{2}{m} \sum_{i=1}^m (\boldsymbol{\mu}^T \boldsymbol{\Phi}(\mathbf{x}_i) - y_i) \boldsymbol{\Phi}(\mathbf{x}_i) + \boldsymbol{\nu},$$

$\boldsymbol{\nu} = (\nu_1, \dots, \nu_n)^T$, $\nu_j = \frac{2\gamma c_j}{\gamma+1} \text{sgn}(\mu_j) |\mu_j|^{\frac{\gamma-1}{\gamma+1}}$, 其中, $\text{sgn}(c) = 1$, 如果 $c \geq 0$, 其他 $\text{sgn}(c) = 0$. 详细描述见算法 1.

由上述算法可知, 每次迭代的时间复杂度为 $O(m\hat{n})$, m 为样本规模, \hat{n} 为 $\boldsymbol{\Phi}(\mathbf{x})$ 的平均非零个数. 由文献 [21] 可知上述框架核选择算法是收敛的, 收敛率为 $O\left(\frac{1}{\sqrt{T}}\right)$. 利用框架核特征映射可显式表示这一特性, 本文采用高效的次梯度下降算法. 实际上, 由于框架核是一类全新的核函数, 因而可将框架核嵌入到别的学习器中, 如 SMO-MKL [30], Simple MKL [31] 等.

5 实验

本节将在标准数据集上验证框架核选择方法 (FKS) 的精确性与有效性. 对比的算法包括支持向量回归算法 ϵ -SVR, ν -SVR [29] 和多核学习方法 SMO-MKL [30].

实验数据取自标准数据集¹⁾, 详情见表 1. 为避免数值问题, 分别将输入数据 $\{\mathbf{x}_i\}_{i=1}^m$ 和标签数据 $\{y_i\}_{i=1}^m$ 放缩到 $[-1, 1]$ 和 $[0, 1]$. SVR 和 SMO-MKL 的 C++ 代码分别来自 LibSVM²⁾ 和文献 [30]. 本文 FKS 方法用 C++ 实现.

在 FKS 中, 参数 $\gamma = 1$. 利用 5 折交叉验证确定 FKS 中参数 c 和平衡系数 g , 其中, $c, g \in \{2^{-7}, 2^{-5}, \dots, 2^9\}$. 实验中将采用 Fourier 框架核和两种常用核函数:

$$(1) \text{ Fourier 框架核 } K_{\text{FFK}}(\mathbf{x}, \mathbf{x}') = \sum_{l=1}^d \left(\frac{1}{2} + \sum_{i=1}^n \lambda_i \cos(ix_l) \cos(ix'_l) + \lambda'_i \sin(ix_l) \sin(ix'_l) \right);$$

$$(2) \text{ Gauss 核 } K_{\text{Gauss}}(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|_2^2);$$

$$(3) \text{ 多项式核 } K_{\text{Poly}}(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + 1)^s.$$

在 SVR 中使用 Gauss 核和多项式核. 其中, 核参数 γ 和 s 由 5 折交叉验证确定, $\gamma \in \{2^{-9}, 2^{-8}, \dots, 2^3\}$, $s \in \{1, 2, \dots, 10\}$. 对于 SMO-MKL, 用 13 个 Gauss 核 $\gamma \in \{2^{-9}, 2^{-8}, \dots, 2^3\}$ 和 10 个多项式核

1) UCI, LIBSVM, StatLib 和 Delve 数据集.

2) Libsvm. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html>.

表 1 数据集

Table 1 Summary of the datasets

Datasets	Sample size	Training size	Test size	Feature
Mg	1385	692	692	6
Balloon	2001	1000	1001	2
Space-ga	3107	1558	1559	6
Abalone	4177	2088	2089	8
Kinematics	8192	4096	4096	8
Bank ¹	8192	4096	4096	32
Puma32H	8192	4096	4096	32
ICB ²	9822	4911	4911	86
Ailerons	13750	6875	6875	40
Elevators	16599	8299	8230	18
Cal-housing ³	20460	10230	10230	8

表 2 平均测试误差与方差

Table 2 The testing mean square errors (TMSE) with standard deviations

Methods	FKS (Ours)	ϵ -SVR + Ploy	ϵ -SVR + Gauss	ν -SVR + Ploy	ν -SVR + Gauss	SMO-MKL
Mg	0.0068 \pm 0.0024	0.0086 \pm 0.0009	0.0082 \pm 0.0008	0.0082 \pm 0.0016	0.0087 \pm 0.0022	0.0057 \pm 0.0007
Balloon	0.0016 \pm 0.0001	0.0031 \pm 0.0000	0.0046 \pm 0.0002	0.0020 \pm 0.0001	0.0033 \pm 0.0000	0.0018 \pm 0.0011
Space-ga	0.0305 \pm 0.0016	0.0961 \pm 0.0151	0.0971 \pm 0.0152	0.0385 \pm 0.0023	0.0401 \pm 0.0239	0.0289 \pm 0.0073
Abalone	0.0030 \pm 0.0001	0.0045 \pm 0.0013	0.0035 \pm 0.0005	0.0035 \pm 1.29e-5	0.0039 \pm 0.0001	0.0036 \pm 0.0007
Kinematics	0.3041 \pm 0.1050	0.2959 \pm 0.0834	0.3253 \pm 0.1841	0.3056 \pm 0.0634	0.3124 \pm 0.1521	0.2699 \pm 0.0007
Bank	0.0184 \pm 0.0003	0.0220 \pm 0.0002	0.0237 \pm 0.0013	0.0213 \pm 0.0023	0.0194 \pm 0.0032	0.0218 \pm 0.0002
Puma32H	0.0299 \pm 0.0008	0.0322 \pm 0.0012	0.0303 \pm 0.0047	0.0292 \pm 0.0004	0.0343 \pm 0.0043	0.0272 \pm 0.0005
ICB	0.0533 \pm 0.0023	0.0696 \pm 0.0044	0.0937 \pm 0.0026	0.1283 \pm 0.0074	0.0523 \pm 0.0012	0.0059 \pm 0.0035
Ailerons	0.0046 \pm 0.0001	0.0052 \pm 0.0021	0.0054 \pm 0.0002	0.0042 \pm 0.0001	0.0043 \pm 0.0002	0.0043 \pm 0.0001
Elevators	0.0042 \pm 0.0003	0.0039 \pm 0.0001	0.0040 \pm 0.0000	0.0056 \pm 3.34e-6	0.0043 \pm 0.0002	0.0043 \pm 0.0001
Cal-housing	0.0204 \pm 0.0008	0.0239 \pm 0.0003	0.0252 \pm 0.0013	0.0244 \pm 0.0026	0.0254 \pm 0.0002	0.0212 \pm 0.0002

$s \in \{1, 2, \dots, 10\}$ 作为其基核 (basis kernel). 数据集随机等分成训练数据和测试数据. 所有方法重复 30 次. 所有实验在 2.2G AMD Opteron Processor machine 上完成.

5.1 精确性

本实验中 $n = 30$ (在 5.3 小节中将研究 n 对性能的影响). 平均测试误差与方差由表 2 给出. 表 2 表明 FKS 和 SMO-MKL 在绝大多数数据集上错误率要低于其他算法. 具体地, 在 11 个数据集中, FKS 在 7 个数据上优于 SVR 算法, 余下 4 个数据上错误率也与 SVR 接近. FKS 与 SMO-MKL 具有相似的精度. 该实验表明基于积分算子特征值的框架核选择方法能保证学习器泛化性能.

5.2 有效性

训练时间由表 3 给出. 表 3 表明在训练时间上 FKS 要显著优于其他方法. 具体地, 在 10/11 个数

表 3 训练时间
Table 3 Training time

Methods	FKS	ϵ -SVR		ν -SVR		SMO-MKL
Kernels	FFK	Gauss	Poly	Gauss	Poly	Gauss+Ploy
Mg	<u>0.0178</u>	0.0227	0.0250	0.0246	0.0326	2.0823
Balloon	<u>0.0046</u>	0.0268	0.0218	0.3157	0.0221	1.0110
Space-ga	<u>0.0410</u>	0.3806	0.7991	0.4498	0.3266	3.6633
Abalone	<u>0.0657</u>	0.4282	0.4333	1.0791	1.1061	5.9031
Kinematics	<u>0.1424</u>	6.8483	4.8830	4.0044	7.5481	25.8653
Bank	<u>0.5774</u>	2.2274	3.6189	8.6997	6.2177	30.6685
Puma32H	<u>0.5754</u>	2.6501	2.3772	8.1454	4.2144	27.9336
ICB	<u>1.8176</u>	12.1914	14.3642	14.8264	18.9364	35.4679
Ailerons	<u>1.1621</u>	18.7596	19.6241	18.1001	20.1009	42.7537
Elevators	<u>0.6492</u>	25.5926	24.0981	27.2582	22.7802	54.0931
Cal-housing	<u>0.3781</u>	36.1576	39.6419	40.8560	38.4607	67.5139

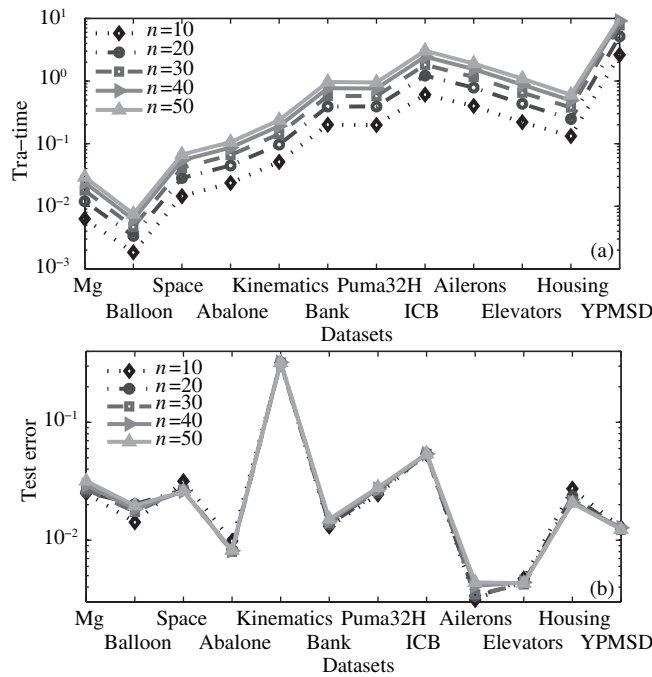


图 1 不同 n 下 FKS 训练时间与平均测试错误

Figure 1 Training time and test mean square errors of FKS with respect to the parameter n . (a) Training time with different n ; (b) TMSE with different n

据上, FKS 比 SVR 方法快 5 倍以上. 在全部数据集上 FKS 比 SMO-MKL 快 30 倍以上.

5.3 参数 n 的影响

在上述实验中, $n = 30$. 下面实验研究 n 的取值对算法性能的影响. 不同 n 下的训练时间与测试

错误由图 1 给出. 图 1 表明当 $n \geq 10$ 时, 测试错误率没有明显变化, 而训练时间随 n 的增大而线性增大. 实际中可选择较小的 n 以降低训练时间, 同时保证较高的精确度.

6 结语

本文提出一种新的框架核构造方法, 相对于常用核函数, 该框架核能很好地分析其对应的积分算子的特征值与特征函数, 并能显式描述其对应再生核希尔伯特空间与特征映射, 从而能精确地刻画假设空间的结构, 增强模型的可解释性. 进一步提出基于积分算子特征值的框架核选择方法, 该方法既能保证更紧的泛化误差界, 又存在高效的优化算法. 理论分析与实验验证阐明了所提出方法的精确性和有效性.

进一步工作拟将该方法推广到 SVM 或其他核方法, 并基于积分算子特征值来分析 SVM 与其他核方法的学习性能.

参考文献

- Smale S, Zhou D X. Learning theory estimates via integral operators and their approximations. *Constr Approx*, 2007, 26: 153–172
- Vito E D, Caponnetto A, Rosasco L. Model selection for regularized least-squares algorithm in learning theory. *Found Comput Math*, 2005, 5: 59–85
- Vito E D, Rosasco L, Caponnetto A, et al. Learning from examples as an inverse problem. *J Mach Learn Res*, 2005, 6: 883–904
- Guo Z C, Zhou D X. Concentration estimates for learning with unbounded sampling. *Adv Comput Math*, 2013, 38: 207–223
- Steinwart I, Hush D, Scovel C. Optimal rates for regularized least squares regression. In: *Proceedings of the 22nd Conference on Learning Theory, Montreal, 2009*. 79–93
- Eberts M, Steinwart I. Optimal learning rates for least squares SVMs using Gaussian kernels. In: *Advances in Neural Information Processing Systems 25, Granada, 2011*. 1539–1547
- Lin S B, Zeng J S, Fang J, et al. Learning rates of ℓ_q coefficient regularization learning with Gaussian kernel. *Neural Comput*, 2014, 26: 2350–2378
- Liu Y, Jiang S L, Liao S Z. Eigenvalues perturbation of integral operator for kernel selection. In: *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, San Francisco, 2013*. 2814–2820
- Liu Y, Liao S Z. Kernel selection with spectral perturbation stability of kernel matrix. *Sci China Inf Sci*, 2014, 57: 112103(10)
- Liu Y, Liao S Z. Eigenvalues ratio for kernel selection of kernel methods. In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin Texas, 2015*. 2814–2820
- Liu Y, Liao S Z. Preventing over-fitting of cross-validation with kernel stability. In: *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Nancy, 2014*. 290–305
- Liu Y, Jiang S L, Liao S Z. Efficient approximation of cross-validation for kernel methods using Bouligand influence function. In: *Proceedings of the 31st International Conference on Machine Learning, Beijing, 2014*. 324–332
- Huang Y D, Li Q F. Characterizations of minimum-energy wavelet frames on intervals with arbitrary integer dilation factors. *Sci Sin Inform*, 2013, 43: 469–487 [黄永东, 李秋富. α 进制最小能量区间小波框架的构造. *中国科学: 信息科学*, 2013, 43: 469–487]
- Li C R, Fu B, Lin J. Texture retrieval using copula multidimensional model on wavelet domain. *Sci Sin Inform*, 2014, 44: 1527–1541 [李朝荣, 付波, 林劫. 小波域 Copula 多维模型纹理检索. *中国科学: 信息科学*, 2014, 44: 1527–1541]
- Gao J, Harris C, Gunn S. On a class of support vector kernels based on frames in function Hilbert spaces. *Neural Comput*, 2001, 13: 1975–1994

- 16 Rakotomamonjy A, Canu S. Frames, reproducing kernels, regularization and learning. *J Mach Learn Res*, 2005, 6: 1485–1515
- 17 Zhang W F, Dai D Q, Yan H. Framelet kernels with applications to support vector regression and regularization networks. *IEEE Trans Syst Man Cybern B*, 2010, 40: 1128–1144
- 18 Opfer R. Tight frame expansions of multiscale reproducing kernels in Sobolev spaces. *Appl Comput Harmon A*, 2006, 20: 357–374
- 19 Hsieh C J, Chang K W, Lin C J. A dual coordinate descent method for large-scale linear SVM. In: *Proceeding of the 25th International Conference on Machine Learning, Helsinki, 2008*. 408–415
- 20 Chang K W, Hsieh C J, Lin C J. Coordinate descent method for large-scale ℓ_2 -loss linear support vector machines. *J Mach Learn Res*, 2008, 9: 1369–1398
- 21 Shalev-Shwartz S, Singer Y, Srebro N, et al. Pegasos: primal estimated sub-gradient solver for SVM. *Math Program*, 2011, 127: 3–30
- 22 Bartlett P L, Mendelson S. Rademacher and Gaussian complexities: risk bounds and structural results. *J Mach Learn Res*, 2002, 3: 463–482
- 23 Vapnik V. *The Nature of Statistical Learning Theory*. New York: Springer Verlag, 2000
- 24 Rosasco L, Belkin M, Vito E D. On learning with integral operators. *J Mach Learn Res*, 2010, 11: 905–934
- 25 Cortes C, Kloft M, Mohri M. Learning kernels using local Rademacher complexity. In: *Advances in Neural Information Processing Systems, Granada, 2013*. 2760–2768
- 26 Chapelle O, Vapnik V, Bousquet O, et al. Choosing multiple parameters for support vector machines. *Mach Learn*, 2002, 46: 131–159
- 27 Bartlett P L, Bousquet O, Mendelson S. Local Rademacher complexities. *Ann Stat*, 2005, 33: 1497–1537
- 28 Ledoux M, Talagrand M. *Probability in Banach Spaces: Isoperimetry and Processes*. Berlin: Springer-Verlag, 1991
- 29 Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. London: Cambridge University Press, 2000
- 30 Vishwanathan S V N, Sun Z, Theera-Ampornpant N. Multiple kernel learning and the SMO algorithm. In: *Advances in Neural Information Processing Systems, Granada, 2010*. 2361–2369
- 31 Rakotomamonjy A, Bach F, Canu S, et al. SimpleMKL. *J Mach Learn Res*, 2008, 9: 2491–2521

Frame kernel selection via explicit description of integral operator space

Yong LIU & Shizhong LIAO*

School of Computer Science and Technology, Tianjin University, Tianjin 300072, China

*E-mail: szliao@tju.edu.cn

Abstract The eigenvalues of an integral operator induced by a kernel function play an important role in the theoretical analysis of kernel methods. However, for most kernels, publicly depicting the eigenvalues of their corresponding integral operators is difficult. To address this problem, in this paper, we present an effective frame kernel selection method based on an explicit description of the integral operator space. We first present a new class of kernel functions (called frame kernels), derived from frame theory, that combine the power of frame representation with the merit of kernel methods. Then, we provide the analytical forms of their eigenvalues and eigenfunctions, and present an explicit description of their corresponding reproducing kernel Hilbert space and feature mapping. Thus, we can exactly depict the structure of the solutions of the kernel methods and enhance the interpretability of the model. Furthermore, with the eigenvalues of integral operators, we establish a generalization error bound of convergence rate $O\left(\frac{1}{m}\right)$ compared with $O\left(\frac{1}{\sqrt{m}}\right)$ for the traditional complexities of the Rademacher complexity and VC dimensionality. Finally, we propose a frame kernel selection method (FKS) based on the above-derived error bound. This method is reformulated as a tractable convex optimization that can be solved globally and

efficiently by the stochastic sub-gradient method. Both the efficacy and efficiency of our proposed method are evaluated on several benchmark datasets.

Keywords frame theory, frame kernel, model selection, kernel selection, integral operator, eigenvalues



Yong LIU was born in 1986. He received his B.S. degree in Information and Computing Science from Hebei University of Technology, Tianjin, in 2009. Currently, he is a Ph.D. candidate at the School of Computer Science and Technology, Tianjin University. His main research interests include kernel methods and model selection.



Shizhong LIAO was born in 1964. He received his Ph.D. degree in Computer Science from Tsinghua University, Beijing, China, in 1997. Currently, he is a professor and Head of the Computer Science Department with the School of Computer Science and Technology, Tianjin University, Tianjin, China. His research interests include artificial intelligence and theoretical computer science.