

DOI: 10.11992/tis.201904050

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.tp.20190722.1039.002.html>

核对齐多核模糊支持向量机

何强, 张娇阳

(北京建筑大学理学院, 北京 100044)

摘要: 支持向量机 (SVMs) 是当前被广泛使用的机器学习技术, 其通过最优分割超平面来提高分类器的泛化能力, 在实际应用中表现优异。然而 SVM 也存在易受噪声影响, 以及核函数选择等难题。针对以上问题, 本文将基于核对齐的多核学习方法引入到模糊支持向量机 (fuzzy support vector machine, FSVM) 中, 提出了模糊多核支持向量机模型 (multiple kernel fuzzy support vector machine, MFSVM)。MFSVM 通过模糊粗糙集方法计算每一样例隶属度; 其次, 利用核对齐的多核方法计算每一单核权重, 并将组合核引入到模糊支持向量机中。该方法不仅提高了支持向量机的抗噪声能力, 也有效避免了核选择难题。在 UCI 数据库上进行实验, 结果表明本文所提方法具有较高的分类精度, 验证了该方法的可行性与有效性。

关键词: 核函数; 支持向量机; 粗糙集理论; 监督学习; 模糊分类; 模糊隶属函数; 鲁棒性; 噪声

中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2019)06-1163-07

中文引用格式: 何强, 张娇阳. 核对齐多核模糊支持向量机 [J]. 智能系统学报, 2019, 14(6): 1163-1169.

英文引用格式: HE Qiang, ZHANG Jiaoyang. Kernel-target alignment multi-kernel fuzzy support vector machine[J]. CAAI transactions on intelligent systems, 2019, 14(6): 1163-1169.

Kernel-target alignment multi-kernel fuzzy support vector machine

HE Qiang, ZHANG Jiaoyang

(School of Science, Beijing University of Civil Engineering and Architecture, Beijing 100044, China)

Abstract: Support vector machines (SVMs) are widely used machine learning techniques. They are used to construct an optimal hyper-plane and have an extraordinary generalization capability and good performance. However, SVMs are sensitive to noise, and it is difficult to select an appropriate kernel for SVMs. In this paper, we introduce kernel-target alignment-based multi-kernel learning method into fuzzy support vector machine (FSVM) and propose the kernel-target alignment-based multi-kernel fuzzy support vector machine (MFSVM). First, we assign the corresponding membership degree to each sample point by the fuzzy rough set method, and then calculate the kernel weight by the multi-kernel learning based on the kernel alignment. Then, the combined kernel is introduced into the fuzzy SVM. The proposed method not only improves the anti-noise ability of the SVM but also effectively avoids the problem of kernel selection. Experiments on nine datasets of the UCI database show that the proposed method has a high classification accuracy, which verifies its feasibility and effectiveness.

Keywords: kernels; support vector machines; rough set theory; supervised learning; fuzzy classification; fuzzy set membership functions; robustness; noise

支持向量机 (SVM) 是 Vapnik 等人于 1995 年提出, 以统计学习理论为基础建立起来的一类新的机器学习算法^[1]。它以最小化结构风险^[2]为原则来提高学习机泛化能力, 实现期望风险的最小

化, 从而达到在统计样本量较少的情况下, 亦能获得良好统计规律的目的。SVM 能够很好地解决小样本、非线性、高维度、局部极小等问题, 现已被应用到很多领域, 如文本分类、语音识别、情感分析、回归分析等。

随着 SVM 的不断发展和应用, 其部分局限性也逐渐显露出来, 在很多方面的研究还有待探索

收稿日期: 2019-04-20. 网络出版日期: 2019-07-23.

基金项目: 国家自然科学基金项目 (61473111); 北京建筑大学科学研究基金项目 (KYJJ2017017).

通信作者: 何强. E-mail: heqiang@bucea.edu.cn.

和改进。例如, SVM 易受到噪声和孤立点影响, 为解决此问题, Lin 等^[3] 提出了模糊支持向量机 (fuzzy support vector machine, FSVM) 的概念, 即将模糊隶属度引入到支持向量机中。FSVM 降低了噪声和孤立点对最终决策函数的影响, 提高了分类精度, 现在也应用到了风险预测、故障诊断、手写字符串识别等领域。

此外, SVM 中核函数和核参数的选择对最终的学习结果有着重要影响, 然而目前还没有关于核函数以及核参数选取的有效手段。作为核方法的重要成果, 多核学习 (multiple kernel learning, MKL)^[4-5] 近年来已成为机器学习领域广大学者的研究热点。单一核函数往往无法充分刻画数据间的相似性, 尤其是复杂数据间的相似性, 不同的核函数对应于不同的相似性表示, 多核相结合可以较准确的表达数据间的相似性, 同时也克服了核函数选择这一难题。受模糊支持向量机和多核学习的启发, 为了同时解决这两大问题, 本文提出模糊多核支持向量机模型。

本文的主要工作有: 1) 将基于核对齐^[6] 的多核学习方法引入到模糊支持向量机中, 提出了基于核对齐的模糊多核支持向量机, 同时解决了核函数选择难题和对噪声数据敏感问题; 2) 利用基于 Gaussian 核的模糊粗糙集下近似算子来确定每个样本点的隶属度^[7], 更有效地减小噪声点和孤立点对分类超平面的影响; 3) 根据最大化组合核与理想核之间的相似性, 计算核权重, 使组合核更精确地刻画数据间的相似性; 4) 实验结果验证了本文所提方法比经典支持向量机 (SVM)、模糊支持向量机 (FSVM) 和多核支持向量机 (MSVM) 表现更为优异。

1 相关工作

1.1 经典支持向量机

对于给定数据 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$, 其中 $x_i \in \mathbf{R}^n$, $y_i \in \{-1, 1\}$, $i = 1, 2, \dots, l$ 。SVM 的目标是得到最优超平面 $f(x) = \omega^T x + b$, 其不仅能将这组数据正确分为两类, 同时能够保证两类样本到分类超平面的距离之和最大。根据结构风险最小化原则, 可将寻找最优超平面的过程归结为如下的优化问题^[1-2]:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s.t.} \quad & y_i (\omega^T x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, 2, \dots, l \end{aligned} \quad (1)$$

式中: ξ_i 为误差项; C 为惩罚系数。

利用 Lagrange 乘子法, 上述优化问题可转化为如下对偶问题:

$$\begin{aligned} \max \quad & \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \\ \text{s.t.} \quad & \sum_{i=1}^l \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i, j = 1, 2, \dots, l \end{aligned} \quad (2)$$

式中: α_i 是样本点 x_i 的 Lagrange 乘子; $\langle \cdot, \cdot \rangle$ 表示内积。最终得到的分类决策函数有如下形式:

$$f(x) = \text{sign} \left(\sum_{i=1}^l \alpha_i y_i \langle x, x_i \rangle + b \right) \quad (3)$$

对于线性不可分数据, 虽然通过引入误差项 ξ_i 可以得到分割超平面, 然而其并无法实现对所有样例的正确划分。引入非线性映射 $\varphi: \mathbf{R}^n \rightarrow H$, 将数据从原空间映射到高维特征空间中, 并在该空间中构造最优分割超平面。由 SVM 模型可知, 只需要计算出 $\varphi(x)$ 和 $\varphi(x')$ 的内积。通过核函数 $k(x, x')$ ^[2] 可以获得两点在特征空间中的内积:

$$k(x, x') = \langle \varphi(x), \varphi(x') \rangle \quad (4)$$

则 (2) 式可转化为

$$\begin{aligned} \max \quad & \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j k(x_i, x_j) \\ \text{s.t.} \quad & \sum_{i=1}^l \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i, j = 1, 2, \dots, l \end{aligned} \quad (5)$$

所以在实际应用中并不需要知道 φ , 只通过核函数 $k(x, x')$ 就可以得到相应的决策函数:

$$f(x) = \text{sgn} \left(\sum_{i=1}^l y_i \alpha_i k(x, x_i) + b \right) \quad (6)$$

常用的核函数主要有线性核、多项式核和 Gaussian 核^[1-2]。

1.2 模糊支持向量机

在实际问题中, 数据易受到噪声等各种因素的干扰, SVM 往往难以获得令人满意的学习效果。为了有效地排除不确定环境中野点 (标签错误的训练样本) 的干扰, 2002 年 Lin 将样本隶属度引入 SVM, 使样本由于隶属度的不同而对超平面有不同的影响, 得到模糊支持向量机 (FSVM)^[3,8]。野点和噪声数据往往被赋予较小的隶属度, 从而削弱其对于分类超平面的影响。FSVM 对应如下优化问题:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l s_i \xi_i \\ \text{s.t.} \quad & y_i (\omega^T x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, 2, \dots, l \end{aligned} \quad (7)$$

这里 s_i 是样例 x_i 属于类别 y_i 的隶属度。显然 s_i 对目标函数中的 ξ_i 起加权作用, 使得噪声点对最终得到的超平面有较小的影响。该算法增强了 SVM 的鲁棒性。

由 Lagrange 乘子法可将式 (7) 转化为对偶形式:

$$\max \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \quad (8)$$

$$\text{s.t. } \sum_{i=1}^l \alpha_i y_i = 0, 0 \leq \alpha_i \leq C s_i, i, j = 1, 2, \dots, l$$

分类决策函数为

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^l \alpha_i y_i \langle \mathbf{x}_i, \mathbf{x} \rangle + b \right) \quad (9)$$

经过不断发展,基于上述 FSVM 算法,其他学者也提出了各种模糊 SVM 来处理不同的具体问题。这些方法都是针对实际问题中的某种不确定性而提出的,是对传统 SVM 的改进与完善。

2 多核学习

在支持向量机处理分类问题时,核函数的选择对分类效果有非常重要的影响。在核函数选择的过程中,普遍采用试算方法,核函数与核参数选择的计算代价较大。多核学习将多个核函数组合在一起,取代了经典支持向量机的单个核函数。多核组合往往能够更加准确地刻画样本间的相似性,从而使得 SVM 在相对较小的计算开销前提下,获得了较好的分类性能。

多核支持向量机 (MSVM) 主要研究问题为核函数的组合方式和相关权系数的计算。目前来说,核函数的组合方式可以分为以下 3 种:

1) 线性组合方式^[9],也是现在使用最广泛的方式,如简单单核相加和加权线性组合:

$$k_\eta(\mathbf{x}_i, \mathbf{x}_j) = \sum_{m=1}^p \eta_m k_m(\mathbf{x}_i^m, \mathbf{x}_j^m) \quad (10)$$

式中: η_m 为核权重, η_m 都为 1 时即为简单单核相加; p 为核函数个数。

2) 非线性组合方式^[10],例如乘积、点积和幂:

$$k_\eta(\mathbf{x}_i, \mathbf{x}_j) = \prod_{m=1}^p k_m(\mathbf{x}_i^m, \mathbf{x}_j^m) \quad (11)$$

3) 数据相关组合方式^[11],这种方式会对每个数据样本分配一个特定的权重,这样可以得到数据的局部分布状况,进而根据分布情况对每个区域学习到合适的组合规则。

多核权系数的计算在多核学习研究中是关键问题,目前,相关权系数的计算方式具体来说有以下 5 类:固定规则学习方式^[12]、启发式学习方式^[6,13]、最优化求解方式^[14]、贝叶斯学习方式^[11]和 Boosting 学习方式^[15]。

本文中采用启发式学习方式来获得核权重,通过核对齐^[6,16-18]的方式计算多核权系数,即计算核函数相应核矩阵之间的相似性来确定权系数。基于训练集 \mathbf{T} ,两个核函数 k_1 和 k_2 对应的核矩阵 \mathbf{K}_1 与 \mathbf{K}_2 之间的核对齐定义为

$$A(\mathbf{T}, \mathbf{K}_1, \mathbf{K}_2) = \frac{\langle \mathbf{K}_1, \mathbf{K}_2 \rangle_F}{\sqrt{\langle \mathbf{K}_1, \mathbf{K}_1 \rangle_F \langle \mathbf{K}_2, \mathbf{K}_2 \rangle_F}} \quad (12)$$

$$\text{其中 } \langle \mathbf{K}_1, \mathbf{K}_2 \rangle_F = \sum_{i=1}^l \sum_{j=1}^l \mathbf{K}_1(\mathbf{x}_i, \mathbf{x}_j) \mathbf{K}_2(\mathbf{x}_i, \mathbf{x}_j).$$

多核 SVM 的优化问题由 (1) 转化为

$$\begin{aligned} \min & \frac{1}{2} \|\omega_\eta\|_2^2 + C \sum_{i=1}^l \xi_i \\ \text{s.t.} & y_i (\omega_\eta^T \cdot \Phi_\eta(\mathbf{x}_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, 2, \dots, l \end{aligned} \quad (13)$$

式中: ω_η 为高维特征空间中分类超平面的法向量,映射 Φ_η 将数据由原空间映射到高维特征空间。分类决策函数变为

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^l \alpha_i y_i k_\eta(\mathbf{x}_i, \mathbf{x}) + b \right) \quad (14)$$

3 模糊多核支持向量机

针对核函数选择难题和对噪声数据敏感问题,本文将基于核对齐加权求和多核学习方法引入到模糊支持向量机模型中,并利用模糊粗糙集方法得到样例隶属度,提出了基于核对齐的模糊多核支持向量机 (multiple kernel fuzzy support vector machine, MFSVM)。

粗糙集 (RS) 是 20 世纪 80 年代波兰学者 Pawlak 提出的一种不确定性数学理论^[19],可以对数据进行分析 and 推理,从中发现隐含的知识。经典的 RS 理论建立在等价关系基础上,其只能处理符号数据。作为 RS 的推广,模糊粗糙集 (FRS) 技术可以处理实值数据。FRS 理论最早是由 Dubois 和 Prade 提出^[20-21],之后得到了快速发展。现有的 FRS 方法已经被成功应用于很多实际问题。本文中,使用基于 Gaussian 核的 FRS 的下近似算子来作为事例的隶属度^[17,22]。取 $k(\mathbf{x}, \mathbf{x}')$ 为 Gaussian 核函数。对样例 $\mathbf{x}_i \in \mathbf{A}$,令 s_i 为 \mathbf{x}_i 属于 \mathbf{A} 的隶属度,则 $s_i = \inf_{\mathbf{x}' \notin \mathbf{A}} \sqrt{1 - (k(\mathbf{x}, \mathbf{x}'))^2}$ 。

本文采用基于核对齐加权求和的多核组合方式,多核组合公式为

$$\mathbf{K}_\eta(\mathbf{x}_i, \mathbf{x}_j) = \sum_{m=1}^p \eta_m \mathbf{K}_m(\mathbf{x}_i, \mathbf{x}_j) \quad (15)$$

式中: η_m 是核权重, $m = 1, 2, \dots, p$; \mathbf{K}_m 是一组基础核矩阵, $m = 1, 2, \dots, p$,优化目标为最大化组合核与理想核的相似性,即可得到如下公式^[6,23]:

$$\begin{aligned} \max & A(\mathbf{T}, \mathbf{K}_\eta, \mathbf{y}\mathbf{y}^T) = \\ & \frac{\left\langle \sum_{m=1}^p \eta_m \mathbf{K}_m, \mathbf{y}\mathbf{y}^T \right\rangle_F}{\sqrt{\left\langle \sum_{m=1}^p \eta_m \mathbf{K}_m, \sum_{m=1}^p \eta_m \mathbf{K}_m \right\rangle_F \langle \mathbf{y}\mathbf{y}^T, \mathbf{y}\mathbf{y}^T \rangle_F}} = \\ & \frac{\sum_{m=1}^p \eta_m \langle \mathbf{K}_m, \mathbf{y}\mathbf{y}^T \rangle_F}{l \sqrt{\sum_{m=1}^p \sum_{j=1}^p \eta_m \eta_j \langle \mathbf{K}_m, \mathbf{K}_j \rangle_F}} \end{aligned} \quad (16)$$

s.t. $\eta \in \mathbb{R}_+^p$

其中 $\mathbf{y} = [y_1, y_2, \dots, y_l]^T$, $\mathbf{y}\mathbf{y}^T$ 是基于训练集 \mathbf{T} 的二分类任务的理想核。为了解决此优化问题, 可以令 $A(\mathbf{T}, \mathbf{K}_\eta, \mathbf{y}\mathbf{y}^T)$ 中的分母为常数, 即式 (16) 等价于:

$$\begin{aligned} & \max_{\eta} \sum_{m=1}^p \eta_m \langle \mathbf{K}_m, \mathbf{y}\mathbf{y}^T \rangle_F \\ & \text{s.t.} \sum_{m=1}^p \sum_{j=1}^p \eta_m \eta_j \langle \mathbf{K}_m, \mathbf{K}_j \rangle_F = C, \\ & \eta_m \geq 0, m = 1, 2, \dots, p \end{aligned} \quad (17)$$

通过 Lagrange 乘子法, 转化为其对偶问题:

$$\begin{aligned} & \max_{\eta} \sum_{m=1}^p \eta_m \langle \mathbf{K}_m, \mathbf{y}\mathbf{y}^T \rangle_F - \\ & \lambda \left(\sum_{m=1}^p \sum_{j=1}^p \eta_m \eta_j \langle \mathbf{K}_m, \mathbf{K}_j \rangle_F - C \right) \\ & \text{s.t.} \eta_m \geq 0, m = 1, 2, \dots, p \end{aligned} \quad (18)$$

观察式 (18), λ 随着 C 的变化而变化, 根据核对齐的定义, 当核权重 η_m 做线性变化时, 核对齐的值保持不变, 所以可考虑 $\lambda = 1$, 那么式 (18) 优化问题变为

$$\begin{aligned} & \max_{\eta} \sum_{m=1}^p \eta_m \langle \mathbf{K}_m, \mathbf{y}\mathbf{y}^T \rangle_F - \sum_{m=1}^p \sum_{j=1}^p \eta_m \eta_j \langle \mathbf{K}_m, \mathbf{K}_j \rangle_F \\ & \text{s.t.} \eta_m \geq 0, m = 1, 2, \dots, p \end{aligned} \quad (19)$$

为了避免核函数在训练集中过拟合, 在优化问题中添加对 η_m 的约束, 即变为

$$\begin{aligned} & \max_{\eta} \sum_{m=1}^p \eta_m \langle \mathbf{K}_m, \mathbf{y}\mathbf{y}^T \rangle_F - \\ & \sum_{m=1}^p \sum_{j=1}^p \eta_m \eta_j \langle \mathbf{K}_m, \mathbf{K}_j \rangle_F - \delta \sum_{m=1}^p \eta_m^2 \\ & \text{s.t.} \eta_m \geq 0, m = 1, 2, \dots, p \end{aligned} \quad (20)$$

其中 $\delta > 0$, 定义一个矩阵 $\mathbf{O}_{mj} = \begin{cases} 1 & m = j \\ 0 & m \neq j \end{cases}$, 式 (20) 即可变为^[23]

$$\begin{aligned} & \max_{\eta} \sum_{m=1}^p \eta_m \langle \mathbf{K}_m, \mathbf{y}\mathbf{y}^T \rangle_F - \\ & \sum_{m=1}^p \sum_{j=1}^p \eta_m \eta_j (\langle \mathbf{K}_m, \mathbf{K}_j \rangle_F + \delta \mathbf{O}_{mj}) \\ & \text{s.t.} \eta_m \geq 0, m = 1, 2, \dots, p \end{aligned} \quad (21)$$

由于式 (21) 是凸二次规划问题, 所以可求得 η 的唯一解, 代入式 (15) 后得到合成核, 再将其引入到上述基于模糊粗糙集隶属度的模糊支持向量机中, 优化问题即变为

$$\begin{aligned} & \min \frac{1}{2} \|\omega_\eta\|^2 + C \sum_{i=1}^l s_i \xi_i \\ & \text{s.t.} y_i (\omega_\eta^T \cdot \Phi_\eta(x_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, 2, \dots, l \end{aligned} \quad (22)$$

引入 Lagrange 乘子, 其对偶问题由式 (8) 可转化为

$$\begin{aligned} & \max \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j k_\eta(x_i, x_j) \\ & \text{s.t.} \sum_{i=1}^l \alpha_i y_i = 0, 0 \leq \alpha_i \leq C s_i, i, j = 1, 2, \dots, l \end{aligned} \quad (23)$$

决策函数为

$$\begin{aligned} f(\mathbf{x}) &= \text{sgn} \left(\sum_{i=1}^l \alpha_i y_i k_\eta(x_i, \mathbf{x}) + b \right) = \\ & \text{sgn} \left(\sum_{i=1}^l \sum_{m=1}^p \alpha_i y_i \eta_m k_m(x_i, \mathbf{x}) + b \right) \end{aligned} \quad (24)$$

基于核对齐的模糊多核支持向量机的算法实现流程图如图 1 所示。

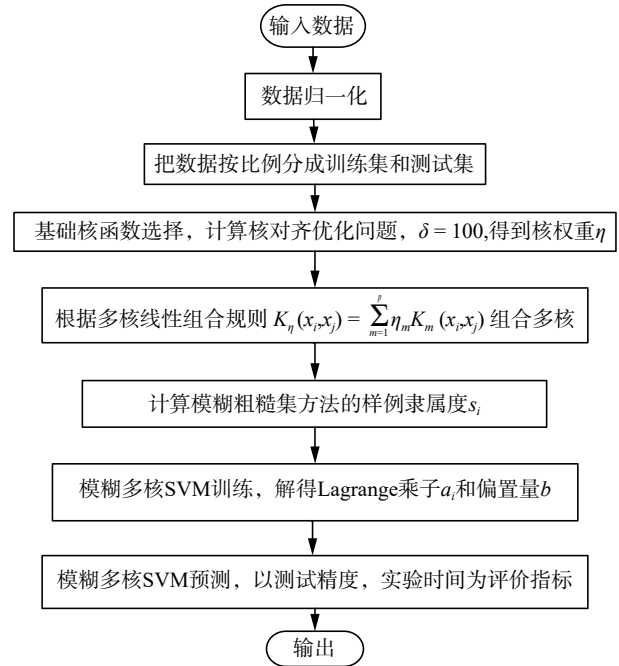


图 1 算法流程图

Fig. 1 Flow chart of algorithm

4 实验仿真

UCI 数据库是加州大学欧文分校提出的用于机器学习的数据库, 是常用的标准测试数据库集, 为了验证本文所提出方法的可行性与有效性, 本文在 UCI 中选取了 9 个数据集, 有关信息如表 1 所示。由于本文只考虑两类分类问题, 所以对于 wine 数据集, 将类别 2 与 3 视为一类处理。

表 1 实验数据信息
Table 1 Data information

编号	数据集名称	样本个数	属性个数	类别个数
1	horse	368	23	2
2	hepatitis	155	20	2
3	heart	270	14	2
4	CT	221	37	2
5	breastcancer	683	10	2
6	ionosphere	351	34	2
7	lymphagraphy	148	19	2
8	wine	178	14	3
9	wdbc	198	34	2

实验在一台 PC 机(CPU: 2.6 GHz, 内存 4 GB)上进行, 操作系统为 Windows 8.1, 实验工具为 Matlab R2014b。实验中将所有数据都做归一化处理, 核函数采用的是不同核参数的 Gaussian 核。给定惩罚系数 $C = 100$, 解核对齐优化问题时的参数 $\delta = 10^2$ 。选择 7 个参数 σ 不同的 Gaussian 核作为基础核, 每个数据集对应的核参数如表 2 所示。所有精度都采用 10 折交叉验证方式获得, 其中单核支持向量机结果为不同参数 Gaussian 核精度最大值。实验对本文提出的基于粗糙集隶属度的模糊多核支持向量机 (MFSVM) 和经典支持向量机 (SVM)、模糊支持向量机 (FSVM)、多核支持向量机 (MSVM) 的分类性能进行对比。

无噪声情况下实验结果如表 3 所示。从实验结果可以看出, 在无噪声的情况下, 本文提出的 MFSVM 方法在 9 个数据集中分类精度最高, 其中在 lymphography 数据集和 wine 数据集中, MF-SVM 和 MSVM 分类精度一样, 并且比 SVM 和

FSVM 的分类精度高, 而且在计算效率方面, MS-VM 与 MFSVM 比 SVM 和 FSVM 显著提高。

表 2 高斯核参数设置
Table 2 Parameters setting of base kernel

编号	数据集名称	基核参数
1	horse	$2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6$
2	hepatitis	$2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0$
3	heart	$2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0$
4	CT	$2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0$
5	breastcancer	$2^1, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7$
6	ionosphere	$2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0$
7	lymphagraphy	$2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3$
8	wine	$2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6$
9	wdbc	$2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0$

表 3 无噪声情况下分类精度与训练时间比较

Table 3 Comparison of classification accuracy and training time without noise

编号	数据集名称	测试精度/%、训练时间/s			
		SVM	FSVM	MKVM	MFSVM
1	horse	78.00±3.62、	81.24±3.50、	81.78±3.87、	83.15±3.99、
		3.01	2.7	0.53	0.49
2	hepatitis	70.06±11.01、	69.77±8.62、	70.40±4.01、	71.65±5.21、
		1.13	1.2	0.17	0.21
3	heart	78.25±9.04、	80.37±6.31、	82.96±8.04、	83.70±8.04、
		4.3	4.35	0.26	0.24
4	CT	90.51±5.10、	90.95±5.60、	90.95±7.73、	91.88±3.49、
		2.81	2.86	0.13	0.15
5	breastcancer	96.92±2.14、	96.92±2.24、	96.93±1.44、	97.36±2.28、
		4.41	4.36	0.4	0.45
6	ionosphere	94.02±4.94、	94.87±4.43、	95.43±3.61、	96.00±4.30、
		8.99	8.97	0.31	0.35
7	lymphography	68.15±12.39、	69.59±10.02、	70.82±14.69、	70.82±14.69、
		0.95	0.98	0.15	0.17
8	wine	96.15±4.05、	97.69±3.71、	98.46±3.24、	98.46±3.24、
		0.48	0.55	0.05	0.05
9	wdbc	77.22±7.70、	80.17±7.83、	81.28±8.90、	82.89±7.42、
		2.13	2.19	0.21	0.24

为了进一步验证 MFSVM 在噪声环境中的表现, 接下来在所有 9 个数据集中都随机选择一定比例的分类超平面附近训练样本改变其类标签。通过这样的方式构造了带有 10%、20% 的噪声水平的数据。表 4 和表 5 分别是噪声比例为 10% 和 20% 时, 4 种算法的测试精度比较结果。

从实验结果可以看出, 当加入 10% 的噪声时, MFSVM 方法在所有 9 个数据集中的分类精度均高于其他 3 种方法, 具有最高的分类精度。

在噪声比例为 20% 的情况下, 虽然 4 种方法的分类精度较于表 4 中都相对变低, 但在所有数据库上, MFSVM 都具有最高的分类精度。该结果也进一步验证了, 本文所提出的 MFSVM 方法在抗噪声能力与计算开销两方面集中了 MSVM 与 FSVM 的优势, 不仅避免了核选择难题, 而且具有较强的抗噪声能力。从而 MFSVM 不仅可行, 且有更广的应用范围。

表4 10%噪声下分类精度与训练时间比较

Table 4 Comparison of classification accuracy and training time with 10% noise

编号	数据集名称	测试精度/%、训练时间/s			
		SVM	FSVM	MKVM	MFSVM
1	horse	80.69± 2.91	81.77± 3.42	81.23± 2.94	82.31± 3.99
		2.24	1.8	0.37	0.34
2	hepatitis	67.56±10.21	69.43±10.05	71.65±10.26	72.22±13.09
		1.01	0.97	0.13	0.16
3	heart	75.93± 8.05	79.26± 5.84	79.63± 6.36	81.11± 4.76
		4.37	4.35	0.25	0.26
4	CT	87.77± 5.30	89.60± 3.69	90.95± 5.67	91.86± 5.98
		2.87	2.89	0.15	0.18
5	breastcancer	91.96± 3.76	95.33± 2.51	93.69± 3.04	96.78± 1.52
		5.36	6.71	0.79	0.71
6	ionosphere	89.17± 4.01	94.02± 3.92	91.44± 4.27	94.58± 3.68
		9.22	9.36	0.38	0.42
7	lymphography	63.79±16.69	64.46±16.35	68.36±15.51	70.05±11.36
		0.93	0.94	0.17	0.15
8	wine	89.23± 8.27	93.85± 4.86	96.15± 4.05	98.46± 3.24
		0.81	0.89	0.15	0.15
9	wpbc	76.39± 9.85	78.39± 8.01	81.94±12.00	82.28±8.01
		2	2.01	0.21	0.24

表5 20%噪声下分类精度与训练时间比较

Table 5 Comparison of classification accuracy and training time with 20% noise

编号	数据集名称	测试精度/%、训练时间/s			
		SVM	FSVM	MKVM	MFSVM
1	horse	73.69± 5.76	75.05± 6.01	76.37± 6.34	78.83± 5.01
		2.43	2.16	0.53	0.53
2	hepatitis	64.83±10.46	65.74± 9.85	70.11± 8.64	70.40± 7.71
		1.04	0.97	0.15	0.12
3	heart	78.15± 6.86	79.63± 8.42	82.22± 7.96	84.07± 9.08
		4.57	4.59	0.25	0.29
4	CT	83.81± 9.38	88.74± 8.22	89.64± 6.94	91.46± 6.71
		1.25	1.26	0.18	0.18
5	breastcancer	85.82± 3.82	88.73± 3.9	93.71± 2.15	95.18±3.00
		8.04	9.09	1.27	1.21
6	ionosphere	84.60± 6.08	89.46± 5.22	88.88± 4.57	91.17± 4.94
		8.56	8.61	0.42	0.48
7	lymphography	64.26±12.74	67.13± 9.22	68.46±12.67	70.82± 8.81
		1.06	1.08	0.17	0.2
8	wine	87.69± 9.73	90.00±11.49	93.85± 4.86	94.62± 7.30
		0.97	0.97	0.1	0.12
9	wpbc	73.28± 6.90	79.83± 3.09	82.11±10.86	83.89± 9.30
		2.07	2.1	0.24	0.22

5 结论

本文是在模糊支持向量机和多核学习的基础之上,通过将基于核对齐的多核方法引入到模糊支持向量机中,提出了基于核对齐的模糊多核支持向量机模型。实验仿真表明此方法综合了模糊支持向量机和多核学习的优点,在分类性能上有

一定程度的提升,特别是在有噪声的样本中表现更为突出,证实了本文方法在解决核函数选择难和对噪声数据敏感问题上的可行性和有效性。由于本文方法中涉及到了多核组合和隶属度的计算,因此下一步将进一步研究更优的核组合方式以及如何改进样本隶属度的计算。

参考文献:

- [1] VAPNIK V N. The nature of statistical learning theory[M]. 2nd ed. New York: Springer-Verlag, 1999: 69-110.
- [2] 邓乃扬, 田英杰. 支持向量机: 理论、算法与拓展 [M]. 北京: 科学出版社, 2009: 115-132.
- [3] LIN Chunfu, WANG Shengde. Fuzzy support vector machines[J]. *IEEE transactions on neural networks*, 2002, 13(2): 464-471.
- [4] SIAHROUDI S K, MOODI P Z, BEIGY H. Detection of evolving concepts in non-stationary data streams: a multiple kernel learning approach[J]. *Expert systems with applications*, 2018, 91: 187-197.
- [5] GONEN M, ALPAYDIN E. Multiple kernel learning algorithms[J]. *Journal of machine learning research*, 2011, 12: 2211-2268.
- [6] CRISTIANINI N, SHAW-TAYLOR J, ELISSEEFF A, et al. On kernel-target alignment[C]//Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic. Vancouver, Canada, 2001: 367-373.
- [7] HE Qiang, WU Congxin. Membership evaluation and feature selection for fuzzy support vector machine based on fuzzy rough sets[J]. *Soft computing*, 2011, 15(6): 1105-1114.
- [8] LIN Chunfu, WANG Shengde. Training algorithms for fuzzy support vector machines with noisy data[J]. *Pattern recognition letters*, 2004, 25(14): 1647-1656.
- [9] CRISTIANINI N, SHAW-TAYLOR J. An introduction to support vector machines and other kernel-based learning methods[M]. Cambridge: Cambridge University Press, 2000: 1-28.
- [10] DE DIEGO I M, MUÑOZ A, MOGUERZA J M. Methods for the combination of kernel matrices within a support vector framework[J]. *Machine learning*, 2010, 78(1/2): 137-174.
- [11] CHRISTOUDIAS C M, URTASUN R, DARRELL T. Bayesian localized multiple kernel learning. Technical Report No. UCB/EECS-2009-96[R]. Berkeley: University of California, 2009: 1531-1565.
- [12] BEN-HUR A, NOBLE W S. Kernel methods for predicting protein-protein interactions[J]. *Bioinformatics*, 2005, 21(S1).
- [13] QIU Shibin, LANE T. A Framework for multiple kernel support vector regression and its applications to siRNA efficacy prediction[J]. *IEEE/ACM transactions on computational biology and bioinformatics*, 2009, 6(2): 190-199.
- [14] VARMA M, RAY D. Learning the discriminative power-invariance trade-off[C]//Proceedings of 2007 IEEE 11th International Conference on Computer Vision. Rio de Janeiro, Brazil, 2007: 1-8.
- [15] 杜海洋. 简化多核支持向量机的研究 [D]. 北京: 北京交通大学, 2015: 13-25.
- DU Haiyang. Research of reduced multiple kernel support vector machine[D]. Beijing: Beijing Jiaotong University, 2015: 13-25.
- [16] WANG Tinghua, ZHAO Dongyan, TIAN Shengfeng. An overview of kernel alignment and its applications[J]. *Artificial intelligence review*, 2015, 43(2): 179-192.
- [17] ZHONG Shangping, CHEN Daya, XU Qianfen, et al. Optimizing the Gaussian kernel function with the formulated kernel target alignment criterion for two-class pattern classification[J]. *Pattern recognition*, 2013, 46(7): 2045-2054.
- [18] 刘向东, 骆斌, 陈兆乾. 支持向量机最优模型选择的研究 [J]. *计算机研究与发展*, 2005, 42(4): 576-581.
- LIU Xiangdong, LUO Bin, CHEN Zhaoqian. Optimal model selection for support vector machines[J]. *Journal of computer research and development*, 2005, 42(4): 576-581.
- [19] 周炜, 周创明, 史朝辉, 等. 粗糙集理论及应用 [M]. 北京: 清华大学出版社, 2015: 57-64.
- [20] DUBOIS D, PRADE H. Rough fuzzy sets and fuzzy rough sets[J]. *International journal of general systems*, 1990, 17(2/3): 191-209.
- [21] DUBOIS D, PRADE H. Putting rough sets and fuzzy sets together[M]//SŁOWIŃSKI R. Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory. Dordrecht, Netherlands: Springer, 1992: 203-232.
- [22] CHEN Degang, YANG Yongping, WANG Hui. Granular computing based on fuzzy similarity relations[J]. *Soft computing*, 2011, 15(6): 1161-1172.
- [23] CHEN Linlin, CHEN Degang, WANG Hui. Alignment based kernel selection for multi-label learning[J]. *Neural processing letters*, 2019, 49(3): 1157-1177.

作者简介:



何强, 男, 1977 年生, 副教授, 主要研究方向为多核学习、监督学习、确定性信息处理。主持国家自然科学基金、河北省自然科学基金等多项。发表学术论文 30 余篇。



张娇阳, 女, 1993 年生, 硕士, 主要研究方向为多核学习、支持向量机。参与国家基金项目 1 项, 发表学术论文 3 篇。