



基于能量的结构化最小二乘孪生支持向量机

史颂辉, 丁世飞

引用本文:

史颂辉, 丁世飞. 基于能量的结构化最小二乘孪生支持向量机[J]. 智能系统学报, 2020, 15(5): 1013–1019.

SHI Songhui, DING Shifei. Energy-based structural least square twin support vector machine[J]. *CAAI Transactions on Intelligent Systems*, 2020, 15(5): 1013–1019.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.201906030>

您可能感兴趣的其他文章

基于人工鱼群算法的孪生支持向量机

Twin support vector machine based on artificial fish swarm algorithm

智能系统学报. 2019, 14(6): 1121–1126 <https://dx.doi.org/10.11992/tis.201905025>

基于模糊核聚类粒化的粒度支持向量机

Granular support vector machine based on fuzzy kernel clustering granulation

智能系统学报. 2019, 14(6): 1271–1277 <https://dx.doi.org/10.11992/tis.201904048>

核对齐多核模糊支持向量机

Kernel-target alignment multi-kernel fuzzy support vector machine

智能系统学报. 2019, 14(6): 1163–1169 <https://dx.doi.org/10.11992/tis.201904050>

基于测度学习支持向量机的钢琴乐谱难度等级识别

Recognition of difficulty level of piano score based on metric learning support vector machine

智能系统学报. 2018, 13(2): 196–201 <https://dx.doi.org/10.11992/tis.201612012>

可拓支持向量分类机

Extension support vector classification machine

智能系统学报. 2018, 13(1): 147–151 <https://dx.doi.org/10.11992/tis.201610019>

运用核聚类和偏最小二乘回归的歌唱声音转换

Conversion of singing voice based on kernel clustering and partial least squares regression

智能系统学报. 2016, 11(1): 55–60 <https://dx.doi.org/10.11992/tis.201506022>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.201906030

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20200320.0956.006.html>

基于能量的结构化最小二乘孪生支持向量机

史颂辉^{1,2}, 丁世飞^{1,2}

(1. 中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116; 2. 矿山数字化教育部工程研究中心, 江苏 徐州 221116)

摘要: 针对最小二乘孪生支持向量机对噪声和离群值非常敏感的问题, 本文提出了一种基于能量的结构化最小二乘孪生支持向量机。首先对每个类进行聚类分析, 然后计算类中各个簇的协方差矩阵并将其引入到目标函数中。其次, 为了降低噪声和离群值对算法的影响, 本文为每个超平面引入能量因子, 在最小二乘的基础上将等式约束转换为基于能量的形式。最后采用“多对一”的策略将提出的算法用于处理多分类问题。研究结果表明: 本文提出的基于能量的结构化最小二乘孪生支持向量机具有良好的分类性能。

关键词: 孪生支持向量机; 最小二乘; 结构信息; 聚类; 协方差矩阵; 能量因子; “多对一”策略; 多分类问题

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2020)05-1013-07

中文引用格式: 史颂辉, 丁世飞. 基于能量的结构化最小二乘孪生支持向量机 [J]. 智能系统学报, 2020, 15(5): 1013-1019.

英文引用格式: SHI Songhui, DING Shifei. Energy-based structural least square twin support vector machine[J]. CAAI transactions on intelligent systems, 2020, 15(5): 1013-1019.

Energy-based structural least square twin support vector machine

SHI Songhui^{1,2}, DING Shifei^{1,2}

(1. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China; 2. Engineering Research Center (Ministry of Education) of Mine Digitization, Xuzhou 221116, China)

Abstract: The least square twin support vector machine (TWSVM) is very sensitive to noise and outlier. To solve this problem, we propose an energy-based structured least square TWSVM (ES-LSTWSVM). First, we perform a cluster analysis on each class, then compute the covariance matrix of each cluster in the classes, and introduce it into the objective function. To reduce the influence of noises and outliers on the algorithm, an energy factor is introduced to each hyperplane, and the equality constraint is converted into an energy-based form on the basis of least squares. Finally, we adopt the “all-versus-one” strategy to apply the proposed algorithm in solving a multi-class classification problem. The experimental results show that ES-LSTWSVM has good classification performance.

Keywords: twin support vector machine; least square; structural information; cluster; covariance matrix; energy factor; “all-versus-one” strategy; multi-class classification problem

支持向量机 (support vector machine, SVM)^[1] 是由 Vapnik 提出的基于结构风险最小化原理的二分类分类器。其基本思想是在 2 个平行支持超平面之间找到最大化边界。与此同时, 对于非线性分类问题, 内核技巧也可以应用到 SVM 中。然而, SVM 的计算复杂度却很高。

收稿日期: 2019-06-18. 网络出版日期: 2020-03-20.

基金项目: 国家自然科学基金项目 (61672522, 61976216, 61379101).

通信作者: 丁世飞. E-mail: dingsf@cumt.edu.cn.

2007 年, Jayadeva 等^[2] 提出了孪生支持向量机 (twin support vector machine, TWSVM), 它为每一类构造一个超平面, 使每一类样本尽可能距离本类的超平面近, 而尽可能地远离另一类的超平面。TWSVM 的 2 个非平行的超平面是通过求解 2 个二次规划问题 (quadratic programming problems, QPPs) 得到的, 每个 QPP 的约束条件都只与一类样本有关。TWSVM 不但保持了 SVM 的优点, 而且训练速度比传统 SVM 要快 4 倍。为了进

一步提升 TWSVM 的性能, 学者们已经提出了不少改进算法, 例如: 孪生有界支持向量机^[3](twin bounded support vector machine, TBSVM)、最小二乘孪生支持向量机^[4](least square twin support vector machine, LSTWSVM)、非平行支持向量机^[5](nonparallel support vector machine, NPSVM)。具体来说, LSTWSVM 使用平方损失函数将 TWSVM 中的凸二次规划问题转化为解线性方程组, 从而实现极快的训练速度。但是, LSTWSVM 对噪声很敏感, 因为它要求超平面与其他类的点距离恰好为 1。近些年来出现了许多改进的算法。Nasiri 等^[6]提出了一种基于能量的 LSTWSVM (energy-based least square twin support vector machine, ELS-TWSVM) 模型, 该模型为每一类引入一个能量因子以减少噪声的影响。马跃峰等^[7]提出了一种基于全局代表点的快速最小二乘支持向量机稀疏化算法 (global-representation-based sparse least squares support vector machine, GRS-LSSVM), 利用数据全局代表性指标, 在全部数据中, 一次性地选择出其中最具有全局代表性的数据并构成稀疏化后的支持向量集, 然后在此基础上求解决策超平面。吴青等^[8]提出了一种最小二乘大间隔孪生支持向量机 (least squares large margin twin support vector machine, LSLMTSVM), 该方法在 TWSVM 的目标函数中引入间隔均值和间隔方差, 通过优化间隔分布来获得具有更强泛化能力的模型。

实际上, 在大多数分类问题中, 不同的类可能有不同的结构信息。然而, 传统的 SVM 和 TWSVM 都没有完全应用类中样本的分布信息。最流行的方法之一是基于假设的聚类算法^[9], 利用聚类算法提取类中的结构信息。例如, 结构化最大边缘机 (structured large margin machine, SLMM)^[10]、最小最大概率机 (minimax probability machine, MPM)^[11] 和最大最小边缘机 (maxi-min margin machine, M⁴)^[12]。但是, 这些方法的计算复杂度高于经典的 SVM。因此, Xue 等^[13]提出了一种结构正则化支持向量机 (structural regularized SVM, SRSVM)。此方法将集群粒度导入到获得数据结构的正则化项。遵循 SLMM 和 SRSVM 的策略, Qi 等^[14]提出了一种新的结构化的 TWSVM (structural twin support vector machine, S-TWSVM)。与 SRSVM 类似, S-TWSVM 使用聚类算法提取每个类的结构信息, 并将得到的结构信息引入到 S-TWSVM 模型中。

在本文中, 受文献^[6]的启发, 首先对结构化的最小二乘孪生支持向量机 (structural least square

twin support vector machine, S-LSTWSVM) 进行改进, 提出了一种基于能量的结构最小二乘孪生支持向量机 (energy-based structural least square twin support vector machine, ES-LSTWSVM)。为每个超平面引入能量因子, 将等式约束条件转换为基于能量的形式。最后采用“多对一”的策略^[15]解决多分类问题。

本文的主要贡献如下:

- 1) 提出了一种基于能量的结构化最小二乘孪生支持向量机;
- 2) 采用“多对一”的策略将提出的算法扩展到多分类问题中;
- 3) 选择几种有代表性的相关算法进行对比实验, 在 UCI 数据集上的实验结果表明提出的算法具有良好的分类性能。

1 相关工作

本节介绍 TWSVM 和 S-TWSVM 的数学模型。假设在 \mathbf{R}^n 空间中训练数据集定义为 $T = \{(x_i, y_i) | i = 1, 2, \dots, m\}$, 其中 x_i 是输入, $y_i \in \{+1, -1\}$ 是对应的输出。它们共有 m 个训练样本, 每个样本有 n 个属性, 其中有 m_1 个样本属于正类, m_2 个样本属于负类, 分别用矩阵 \mathbf{A} 、 \mathbf{B} 来表示。

1.1 TWSVM

TWSVM 通过训练数据集寻找 2 个不平行的分离超平面:

$$\mathbf{x}^T \mathbf{w}_1 + b_1 = 0, \mathbf{x}^T \mathbf{w}_2 + b_2 = 0$$

它们一般通过求解下面 2 个二次规划问题来得到:

$$\begin{aligned} \min_{w_1, b_1, \xi_2} \quad & \frac{1}{2} (\mathbf{A} \mathbf{w}_1 + \mathbf{e}_1 b_1)^T (\mathbf{A} \mathbf{w}_1 + \mathbf{e}_1 b_1) + c_1 \mathbf{e}_1^T \xi_2 - \\ \text{s.t.} \quad & (\mathbf{B} \mathbf{w}_1 + \mathbf{e}_2 b_1) \geq \mathbf{e}_2 - \xi_2, \xi_2 \geq 0 \\ \min_{w_2, b_2, \xi_1} \quad & \frac{1}{2} (\mathbf{B} \mathbf{w}_2 + \mathbf{e}_2 b_2)^T (\mathbf{B} \mathbf{w}_2 + \mathbf{e}_2 b_2) + c_2 \mathbf{e}_2^T \xi_1 - \\ \text{s.t.} \quad & (\mathbf{A} \mathbf{w}_2 + \mathbf{e}_1 b_2) \geq \mathbf{e}_1 - \xi_1, \xi_1 \geq 0 \end{aligned}$$

式中: $c_1, c_2 > 0$ 是 2 个惩罚参数; $\mathbf{e}_1, \mathbf{e}_2$ 是元素全为 1 的列向量; ξ_1, ξ_2 是松弛变量。

TWSVM 通过如下的决策函数判定新样本的类别标签:

$$\text{Label}(\mathbf{x}) = \arg \min_{l=1,2} |\mathbf{x}^T \mathbf{w}_l + b_l|$$

1.2 S-TWSVM

首先, S-TWSVM 通过“沃氏链接”(ward's linkage) 聚类算法^[16]提取类内的结构信息。然后它使用 2 个非平行超平面来判定新数据的类别, 其中每个超平面尽可能离本类数据点近, 并且在目标函数中只考虑该类的结构信息, 而约束条件规定该超平面要离其他类数据点至少距离为 1 个单位。假设通过聚类之后正类中有 C_p 个簇, 负类

中有 C_N 个簇。那么 S-TWSVM 的优化问题可以表述为

$$\begin{aligned} \min_{\mathbf{w}_+, b_+, \xi} \quad & \frac{1}{2} \|\mathbf{A}\mathbf{w}_+ + \mathbf{e}_+ b_+\|^2 + c_1 \mathbf{e}_+^T \xi + \frac{1}{2} c_2 (\|\mathbf{w}_+\|^2 + b_+^2) + \\ & \frac{1}{2} c_3 \mathbf{w}_+^T \sum_+ \mathbf{w}_+ \\ \text{s.t.} \quad & -(\mathbf{B}\mathbf{w}_+ + \mathbf{e}_- b_+) + \xi \geq \mathbf{e}_-, \xi \geq 0 \\ \min_{\mathbf{w}_-, b_-, \eta} \quad & \frac{1}{2} \|\mathbf{B}\mathbf{w}_- + \mathbf{e}_- b_-\|^2 + c_4 \mathbf{e}_-^T \eta + \frac{1}{2} c_5 (\|\mathbf{w}_-\|^2 + b_-^2) + \\ & \frac{1}{2} c_6 \mathbf{w}_-^T \sum_- \mathbf{w}_- \\ \text{s.t.} \quad & (\mathbf{A}\mathbf{w}_- + \mathbf{e}_+ b_-) + \eta \geq \mathbf{e}_+, \eta \geq 0 \end{aligned}$$

式中: $c_i (i=1, 2, \dots, 6)$ 是惩罚参数; ξ, η 是松弛变量; $\mathbf{e}_+, \mathbf{e}_-$ 是元素全为 1 的列向量; $\sum_+ = \sum_{P_1} + \dots + \sum_{P_{C_P}}$; $\sum_- = \sum_{N_1} + \dots + \sum_{N_{C_N}}$; \sum_{P_i} 和 \sum_{N_j} 是 2 类中的每一个簇的相应协方差矩阵, 其中, $i=1, 2, \dots, C_P, j=1, 2, \dots, C_N$ 。

新的数据点被分配给正类还是负类取决于得到的 2 个超平面 $\mathbf{w}_+^T \mathbf{x} + b_+ = 0$ 和 $\mathbf{w}_-^T \mathbf{x} + b_- = 0$ 中它距离哪一个超平面更近, 即

$$\text{Label}(\mathbf{x}) = \arg \min_{+,-} \{|\mathbf{w}_+^T \mathbf{x} + b_+|, |\mathbf{w}_-^T \mathbf{x} + b_-|\}$$

2 ES-LSTWSVM

在本节中, 提出了一种基于能量的结构化孪生支持向量机 (ES-LSTWSVM)。该算法主要分为 2 步: 1) 用某种聚类方法提取类内的结构信息; 2) 最小化每个类中各个簇之间的紧密度并将其嵌入到目标函数中。此外, 利用核技巧使模型能够解决非线性可分问题。在下面的小节中, 将具体讨论这些步骤。

2.1 线性 ES-LSTWSVM

有很多聚类方法可被用于提取类内的结构信息, 例如: k 均值算法^[17]、模糊聚类和密度聚类^[18-21]。使用拐点图^[22]来确定最佳聚类数目。聚类之后, 通过计算类中各个簇的协方差矩阵将结构信息引入到优化问题中。因此, 聚类之后的簇应该是紧凑的并且是球形的。鉴于此, 采用 Ward's linkage 聚类算法, 这是一种层次聚类方法, 为协方差矩阵的计算提供了基础。

假设从正类和负类中分别得到了 C_P 和 C_N 个簇, 即 $A = P_1 \cup \dots \cup P_i \cup \dots \cup P_{C_P}$, $B = N_1 \cup \dots \cup N_i \cup \dots \cup N_{C_N}$ 。对于线性可分的二分类问题, ES-LSTWSVM 模型可表示为

$$\begin{aligned} \min_{\mathbf{w}_+, b_+, \xi} \quad & \frac{1}{2} \|\mathbf{A}\mathbf{w}_+ + \mathbf{e}_+ b_+\|^2 + \frac{c_1}{2} \xi^T \xi + \\ & \frac{c_2}{2} (\|\mathbf{w}_+\|^2 + b_+^2) + \frac{c_3}{2} \mathbf{w}_+^T \sum_+ \mathbf{w}_+ \\ \text{s.t.} \quad & -(\mathbf{B}\mathbf{w}_+ + \mathbf{e}_- b_+) + \xi = \mathbf{E}_- \end{aligned} \quad (1)$$

$$\begin{aligned} \min_{\mathbf{w}_-, b_-, \eta} \quad & \frac{1}{2} \|\mathbf{B}\mathbf{w}_- + \mathbf{e}_- b_-\|^2 + \frac{c_4}{2} \eta^T \eta + \\ & \frac{c_5}{2} (\|\mathbf{w}_-\|^2 + b_-^2) + \frac{c_6}{2} \mathbf{w}_-^T \sum_- \mathbf{w}_- \\ \text{s.t.} \quad & \mathbf{A}\mathbf{w}_- + \mathbf{e}_+ b_- + \eta = \mathbf{E}_+ \end{aligned} \quad (2)$$

式中: $c_i (i=1, 2, \dots, 6)$ 是惩罚参数; ξ, η 是松弛变量; \mathbf{E}_+ 和 \mathbf{E}_- 是超平面的能量因子; $\sum_+ = \sum_{P_1} + \dots + \sum_{P_{C_P}}$; $\sum_- = \sum_{N_1} + \dots + \sum_{N_{C_N}}$; \sum_{P_i} 和 \sum_{N_j} 分别是对应于 2 个类中的第 i 个簇和第 j 个簇的协方差矩阵, $i=1, 2, \dots, C_P, j=1, 2, \dots, C_N$ 。

将等式约束条件代入目标函数式 (1) 中得到:

$$\begin{aligned} \min_{\mathbf{w}_+, b_+} \quad & \frac{1}{2} \|\mathbf{A}\mathbf{w}_+ + \mathbf{e}_+ b_+\|^2 + \frac{c_1}{2} \|\mathbf{B}\mathbf{w}_+ + \mathbf{e}_- b_+ + \mathbf{E}_-\|^2 + \\ & \frac{c_2}{2} (\|\mathbf{w}_+\|^2 + b_+^2) + \frac{c_3}{2} \mathbf{w}_+^T \sum_+ \mathbf{w}_+ \end{aligned} \quad (3)$$

分别对式 (3) 中 \mathbf{w}_+ 和 b_+ 求导, 并且由 KKT 条件得:

$$\mathbf{A}^T(\mathbf{A}\mathbf{w}_+ + \mathbf{e}_+ b_+) + c_1 \mathbf{B}^T(\mathbf{B}\mathbf{w}_+ + \mathbf{e}_- b_+ + \mathbf{E}_-) + c_2 \mathbf{w}_+ + c_3 \sum_+ \mathbf{w}_+ = 0 \quad (4)$$

$$\mathbf{e}_+^T(\mathbf{A}\mathbf{w}_+ + \mathbf{e}_+ b_+) + c_1 \mathbf{e}_-^T(\mathbf{B}\mathbf{w}_+ + \mathbf{e}_- b_+ + \mathbf{E}_-) + c_2 b_+ = 0 \quad (5)$$

将式 (4)、(5) 整理成矩阵的形式为

$$\begin{aligned} & \begin{bmatrix} \mathbf{A}^T \mathbf{A} & \mathbf{A}^T \mathbf{e}_+ \\ \mathbf{e}_+^T \mathbf{A} & \mathbf{e}_+^T \mathbf{e}_+ \end{bmatrix} \begin{bmatrix} \mathbf{w}_+ \\ b_+ \end{bmatrix} + c_1 \begin{bmatrix} \mathbf{B}^T \mathbf{B} & \mathbf{B}^T \mathbf{e}_- \\ \mathbf{e}_-^T \mathbf{B} & \mathbf{e}_-^T \mathbf{e}_- \end{bmatrix} \begin{bmatrix} \mathbf{w}_+ \\ b_+ \end{bmatrix} + \\ & c_1 \begin{bmatrix} \mathbf{B}^T \\ \mathbf{e}_-^T \end{bmatrix} \mathbf{E}_- + c_2 \begin{bmatrix} \mathbf{w}_+ \\ b_+ \end{bmatrix} + c_3 \begin{bmatrix} \sum_+ \mathbf{w}_+ \\ 0 \end{bmatrix} = 0 \\ & \begin{bmatrix} \mathbf{A}^T \\ \mathbf{e}_+^T \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{e}_+ \end{bmatrix} \begin{bmatrix} \mathbf{w}_+ \\ b_+ \end{bmatrix} + c_1 \begin{bmatrix} \mathbf{B}^T \\ \mathbf{e}_-^T \end{bmatrix} \begin{bmatrix} \mathbf{B} & \mathbf{e}_- \end{bmatrix} \begin{bmatrix} \mathbf{w}_+ \\ b_+ \end{bmatrix} + \\ & c_2 \begin{bmatrix} \mathbf{w}_+ \\ b_+ \end{bmatrix} + c_3 \begin{bmatrix} \sum_+ \mathbf{w}_+ \\ 0 \end{bmatrix} = -c_1 \begin{bmatrix} \mathbf{B}^T \\ \mathbf{e}_-^T \end{bmatrix} \mathbf{E}_- \end{aligned}$$

使 $\mathbf{P} = \begin{bmatrix} \mathbf{A} & \mathbf{e}_+ \end{bmatrix}$, $\mathbf{Q} = \begin{bmatrix} \mathbf{B} & \mathbf{e}_- \end{bmatrix}$, 并且

$\sum_1 = \begin{bmatrix} \sum_+ & 0 \\ 0 & 0 \end{bmatrix}$ 。则解可表示为

$$\mathbf{z}_1 = \begin{bmatrix} \mathbf{w}_+ \\ b_+ \end{bmatrix} = -c_1 (\mathbf{P}^T \mathbf{P} + c_1 \mathbf{Q}^T \mathbf{Q} + c_2 \mathbf{I} + c_3 \sum_1)^{-1} \mathbf{Q}^T \mathbf{E}_-$$

同理, 式 (2) 可表示为

$$\mathbf{z}_2 = \begin{bmatrix} \mathbf{w}_- \\ b_- \end{bmatrix} = c_4 (\mathbf{Q}^T \mathbf{Q} + c_4 \mathbf{P}^T \mathbf{P} + c_5 \mathbf{I} + c_6 \sum_2)^{-1} \mathbf{P}^T \mathbf{E}_+$$

式中: $\sum_2 = \begin{bmatrix} \sum_- & 0 \\ 0 & 0 \end{bmatrix}$; \mathbf{I} 是适当维数的单位矩阵。

ES-LSTWSVM 通过如下的决策函数判定新样本的类别标签:

$$\text{Label}(\mathbf{x}) = \left\{ \frac{|\mathbf{x}^T \mathbf{w}_+ + b_+|}{\|\mathbf{w}_+\|}, \frac{|\mathbf{x}^T \mathbf{w}_- + b_-|}{\|\mathbf{w}_-\|} \right\}$$

2.2 非线性 ES-LSTWSVM

对于非线性分类问题, 利用核技巧进行处理。首先将原始特征空间中的数据映射到希尔伯特空间中: $\Phi: \mathbf{R}^m \rightarrow H$ 。类似于线性情形, 决策函

数可表示为 $f_+(x) = (\mathbf{w}_+ \Phi(x)) + b_+$ 和 $f_-(x) = (\mathbf{w}_- \Phi(x)) + b_-$ 。

由希尔伯特空间理论^[23]知:

$$\mathbf{w}_+ = \sum_{i=1}^{m_1+m_2} (\lambda_+)_i \Phi(x_i) = \Phi(M) \lambda_+$$

$$\mathbf{w}_- = \sum_{i=1}^{m_1+m_2} (\lambda_-)_i \Phi(x_i) = \Phi(M) \lambda_-$$

ES-LSTWSVM 构造如下 2 个基于核函数的超平面:

$$K(\mathbf{x}^T, \mathbf{M}^T) \lambda_+ + b_+ = 0, \quad K(\mathbf{x}^T, \mathbf{M}^T) \lambda_- + b_- = 0$$

式中: $\mathbf{M} = [\mathbf{A}^T \quad \mathbf{B}^T]$; $K(x_i, x_j) = (\Phi(x_i) \Phi(x_j))$ 。

在非线情形下, ES-LSTWSVM 可表示为

$$\min_{\mathbf{w}_+, b_+, \xi} \quad \frac{1}{2} \|K(\mathbf{A}, \mathbf{M}^T) \lambda_+ + \mathbf{e}_+ b_+\|^2 + \frac{c_1}{2} \xi^T \xi + \frac{c_2}{2} (\|\lambda_+\|_2^2 + b_+^2) + \frac{c_3}{2} \lambda_+^T \Phi(\mathbf{M})^T \sum_+^\Phi \Phi(\mathbf{M}) \lambda_+ \quad (6)$$

$$\text{s.t.} \quad -(K(\mathbf{B}, \mathbf{M}^T) \lambda_+ + \mathbf{e}_+ b_+) + \xi = \mathbf{E}_-$$

$$\min_{\mathbf{w}_-, b_-, \eta} \quad \frac{1}{2} \|K(\mathbf{B}, \mathbf{M}^T) \lambda_- + \mathbf{e}_- b_-\|^2 + \frac{c_4}{2} \eta^T \eta + \frac{c_5}{2} (\|\lambda_-\|_2^2 + b_-^2) + \frac{c_6}{2} \lambda_-^T \Phi(\mathbf{M})^T \sum_-^\Phi \Phi(\mathbf{M}) \lambda_- \quad (7)$$

$$\text{s.t.} \quad (K(\mathbf{A}, \mathbf{M}^T) \lambda_- + \mathbf{e}_- b_-) + \eta = \mathbf{E}_+$$

式中: $\sum_+^\Phi = \sum_{p_1}^\Phi + \dots + \sum_{p_{C_p}}^\Phi$; $\sum_-^\Phi = \sum_{N_1}^\Phi + \dots + \sum_{N_{C_n}}^\Phi$ 。
 $\sum_{p_i}^\Phi$ 和 $\sum_{N_j}^\Phi$ 分别是由核 Ward's linkage 聚类算法得到的 2 个类中第 i 个簇和第 j 个簇对应的协方差矩阵^[13-14], $i = 1, 2, \dots, C_p, j = 1, 2, \dots, C_n$ 。

类似于线性情形, 式 (6)、(7) 的解可表示为

$$\mathbf{z}_1 = \begin{bmatrix} \mathbf{w}_+ \\ b_+ \end{bmatrix} = -c_1 \left(\mathbf{P}_\phi^T \mathbf{P}_\phi + c_1 \mathbf{Q}_\phi^T \mathbf{Q}_\phi + c_2 \mathbf{I} + c_3 \sum_+^\Phi \right)^{-1} \mathbf{Q}_\phi^T \mathbf{E}_- \quad (8)$$

$$\mathbf{z}_2 = \begin{bmatrix} \mathbf{w}_- \\ b_- \end{bmatrix} = c_4 \left(\mathbf{Q}_\phi^T \mathbf{Q}_\phi + c_4 \mathbf{P}_\phi^T \mathbf{P}_\phi + c_5 \mathbf{I} + c_6 \sum_-^\Phi \right)^{-1} \mathbf{P}_\phi^T \mathbf{E}_+$$

式中: $\mathbf{P}_\phi = [K(\mathbf{A}, \mathbf{M}^T) \quad \mathbf{e}_+]$; $\mathbf{Q}_\phi = [K(\mathbf{B}, \mathbf{M}^T) \quad \mathbf{e}_-]$,

$$\sum_+^\Phi = \begin{bmatrix} \Phi(\mathbf{M})^T \sum_+^\Phi \Phi(\mathbf{M}) & 0 \\ 0 & 0 \end{bmatrix}$$

$$\sum_-^\Phi = \begin{bmatrix} \Phi(\mathbf{M})^T \sum_-^\Phi \Phi(\mathbf{M}) & 0 \\ 0 & 0 \end{bmatrix}$$

3 实验结果与分析

使用“多对一”的策略将 ES-LSTWSVM 模型扩展到多类分类中。为了证明该算法的有效性, 给出了该算法在多个 UCI 数据集上的实验结果。在实验中, 选择了“一对一”最小二乘孪生支持向量机 (1-v-1 LSTWSVM)、“一对多”最小二乘

孪生支持向量机 (1-v-r LSTWSVM)、改进的最小二乘孪生多分类支持向量^[24] (improvements on least squares twin multi-class classification support vector machine, ILST-KSVC), 多生最小二乘支持向量机^[25] (multiple birth least squares support vector machine, MBLSSVM) 与本文算法进行比较。实验结果为 10 折交叉验证的平均值。在非线情形下, 使用高斯核函数 $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2)$ 。

3.1 实验结果

在本节中, 使用 UCI 数据集集中的 Iris、Wine、Vowel、Balance、Vehicle、Zoo 和 Segment。这些数据集的特征如表 1 所示。

表 1 数据集的具体特征
Table 1 The detail features of the datasets

数据集	样本个数	特征维数	类别数
Iris	150	4	3
Wine	178	13	3
Vowel	528	10	11
Balance	625	4	3
Vehicle	846	18	4
Zoo	101	16	7
Segment	2310	18	7

表 2 列出了上述算法在使用线性核的情况下的分类结果。表 3 给出了上述算法在使用高斯核的情况下的分类结果。从表 2 可以看出, 提出的算法在 Wine、Balance、Vehicle、Zoo 这 4 个数据集上有最高的分类精度, 并且在其他数据集上也有不错的表现。从表 3 可以看出, 所提出的算法在 Iris、Vowel、Balance、Vehicle、Zoo、Segment 这 6 个数据集上的分类准确率最高, 并且在 Wine 数据集上也有较高的准确率。图 1 是对实验结果更直观表示。

表 2 线性核函数下精确度的比较

数据集	Table 2 Comparison of accuracies using linear kernel				本文 算法
	1-v-1 LSTWSVM	1-v-r LSTWSVM	ILST- KSVC	MBLSSVM	
Iris	95.35	91.63	76.67	85.73	90.45
Wine	97.26	97.67	97.90	98.80	98.99
Vowel	80.66	44.12	50.00	47.38	49.60
Balance	86.39	85.87	88.47	86.63	88.65
Vehicle	80.04	77.18	45.80	78.70	81.14
Zoo	94.87	92.02	92.85	96.51	98.23
Segment	94.32	90.51	60.60	89.72	90.45

表3 径向基核函数下精确度的比较

Table 3 Comparison of accuracies using RBF kernel

数据集	1-v-1 LSTWSVM	1-v-r LSTWSVM	ILST- KSVC	MBLSSVM	本文 算法
Iris	95.03	96.12	80.00	94.87	98.00
Wine	74.34	76.69	91.89	71.77	75.17
Vowel	97.86	98.64	98.28	96.09	98.93
Balance	88.91	90.12	91.06	87.81	91.40
Vehicle	80.79	80.25	58.77	75.71	81.20
Zoo	96.93	97.11	96.72	95.61	97.62
Segment	84.54	85.32	75.76	93.31	93.79

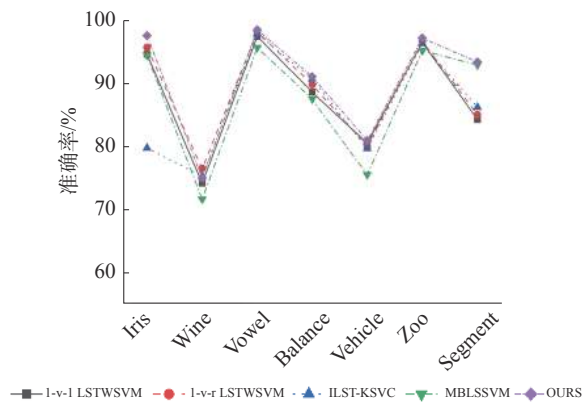
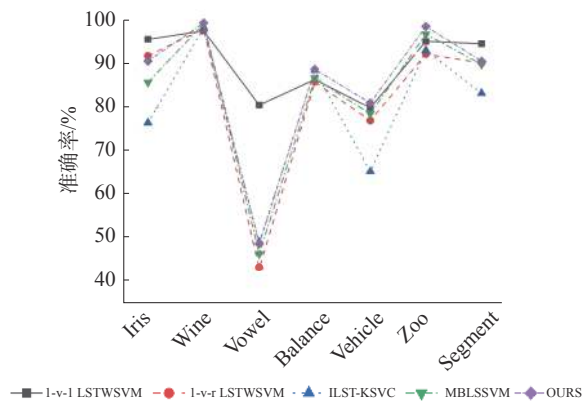


图1 算法精确度的比较

Fig. 1 Comparison of algorithm accuracies

3.2 能量参数 E_+ 和 E_- 对算法性能的影响

为了分析能量参数 E_+ 和 E_- 对 ES-LSTWSVM 的分类性能的影响,在 Heart、Liver、Breast 和 Ionosphere 数据集上进行了实验。通过引入可变能量参数,所提出的分类器将做出相应的调整以减少噪声干扰。并使用网格搜索对参数进行选择。为了只考虑能量参数对实验结果的影响,将 ES-LSTWSVM 中的其他参数设置为固定值。图2显示了分类性能在 Australian、Heart 和 Diabetes 数据集上随能量参数 E_+ 和 E_- 的变化而变化。图2(a)显示当 $E_+ = 0.7$, $E_- = 1.0$ 时,算法在 Heart 数据集上

的分类准确率达 86.96%。图2(b)显示当 $E_+ : E_- = 4 : 3$ 时,算法在 Liver 数据集上的分类准确率达 71.15%。图2(c)显示当 $E_+ = 0.8, E_- = 0.7$ 时,算法在 Breast 数据集上的分类准确率达 77.38%。图2(d)显示当 $E_+ : E_- = 1 : 1$ 时,算法在 Ionosphere 数据集上的分类准确率达 88.68%。由此可知,可以通过寻找最优能量参数来确定最优超平面。

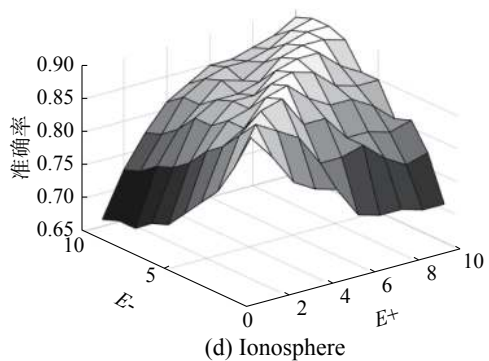
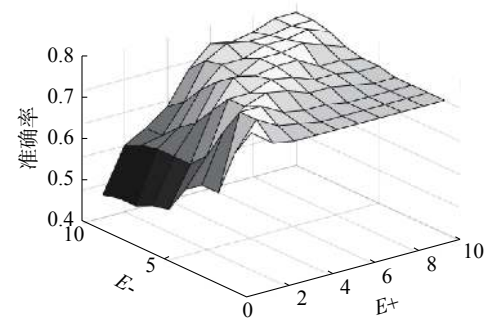
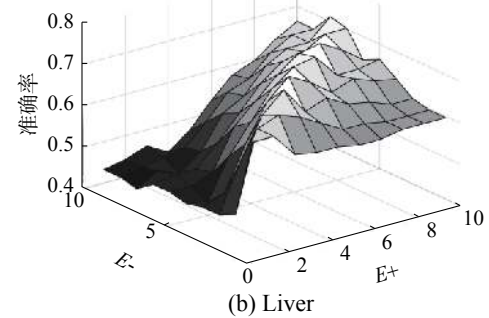
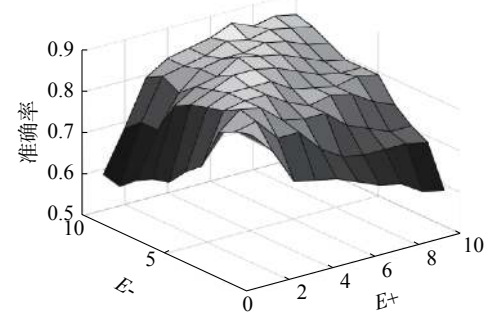


图2 能量参数 E_+ 和 E_- 对 ES-LSTWSVM 性能的影响
Fig. 2 Effect of energy parameters (E_+ and E_-) on the performance of the ES-LSTWSVM

4 结束语

在本文中,提出了一种基于能量的结构化最小二乘孪生支持向量机,称为ES-LTWSVM。首先使用聚类算法来获取类中的结构信息,然后为每个超平面引入能量因子,将传统TWSVM的不等式约束转换为基于能量模型的等式约束。这使得提出的ES-LTWSVM不仅有较低的计算复杂度,而且提高了模型的泛化能力。最后,使用“多对一”策略将ES-LTWSVM扩展到多类分类问题中。实验结果表明,所提出的算法具有良好的分类性能。

参考文献:

- [1] CORTES C, VAPNIK V N. Support-vector networks[J]. *Machine learning*, 1995, 20(3): 273–297.
- [2] JAYADEVA, KHEMCHANDANI R, CHANDRA S. Twin support vector machines for pattern classification[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2007, 29(5): 905–910.
- [3] SHAO Yuanhai, ZHANG Chunhua, WANG Xiaobo, et al. Improvements on twin support vector machines[J]. *IEEE transactions on neural networks*, 2011, 22(6): 962–968.
- [4] KUMAR M A, GOPAL M. Least squares twin support vector machines for pattern classification[J]. *Expert systems with applications*, 2009, 36(4): 7535–7543.
- [5] TIAN Yingjie, QI Zhiqian, JU Xuchan, et al. Nonparallel support vector machines for pattern classification[J]. *IEEE transactions on cybernetics*, 2014, 44(7): 1067–1079.
- [6] NASIRI J A, CHARKARI N M, MOZAFARI K. Energy-based model of least squares twin support vector machines for human action recognition[J]. *Signal processing*, 2014, 104: 248–257.
- [7] 马跃峰, 梁循, 周小平. 一种基于全局代表点的快速最小二乘支持向量机稀疏化算法[J]. *自动化学报*, 2017, 43(1): 132–141.
MA Yuefeng, LIANG Xun, ZHOU Xiaoping. A fast sparse algorithm for least squares support vector machine based on global representative points[J]. *Acta automatica sinica*, 2017, 43(1): 132–141.
- [8] 吴青, 齐韶维, 孙凯悦, 等. 最小二乘大间隔孪生支持向量机[J]. *北京邮电大学学报*, 2018, 41(6): 34–38.
WU Qing, QI Shaowei, SUN Kaiyue, et al. Least squares large margin twin support vector machine[J]. *Journal of Beijing University of Posts and Telecommunications*, 2018, 41(6): 34–38.
- [9] RIGOLLET P. Generalization error bounds in semi-supervised classification under the cluster assumption[J]. *The journal of machine learning research*, 2007, 8: 1369–1392.
- [10] YEUNG D S, WANG Defeng, NG W W Y, et al. Structured large margin machines: sensitive to data distributions[J]. *Machine learning*, 2007, 68(2): 171–200.
- [11] LANCKRIET G R G, EL GHAOU L, BHATTACHARYYA C, et al. A robust minimax approach to classification[J]. *The journal of machine learning research*, 2002, 3: 555–582.
- [12] HUANG Kaizhu, YANG Haiqin, KING I, et al. Maxi-min margin machine: learning large margin classifiers locally and globally[J]. *IEEE transactions on neural networks*, 2008, 19(2): 260–272.
- [13] XUE Hui, CHEN Songcan, YANG Qiang. Structural regularized support vector machine: a framework for structural large margin classifier[J]. *IEEE transactions on neural networks*, 2011, 22(4): 573–587.
- [14] QI Zhiqian, TIAN Yingjie, SHI Yong. Structural twin support vector machine for classification[J]. *Knowledge-based systems*, 2013, 43: 74–81.
- [15] 丁世飞, 张健, 张谢锴, 等. 多分类孪生支持向量机研究进展[J]. *软件学报*, 2018, 29(1): 89–108.
DING Shifei, ZHANG Jian, ZHANG Xiekai, et al. Survey on multi class twin support vector machines[J]. *Journal of software*, 2018, 29(1): 89–108.
- [16] WARD JR J H. Hierarchical grouping to optimize an objective function[J]. *Journal of the American statistical association*, 1963, 58(301): 236–244.
- [17] 陶莹, 杨锋, 刘洋, 等. K 均值聚类算法的研究与优化[J]. *计算机技术与发展*, 2018, 28(6): 90–92.
TAO Ying, YANG Feng, LIU Yang, et al. Research and optimization of K -means clustering algorithm[J]. *Computer technology and development*, 2018, 28(6): 90–92.
- [18] XU Xiao, DING Shifei, DU Mingjing, et al. DPCG: an efficient density peaks clustering algorithm based on grid[J]. *International journal of machine learning and cybernetics*, 2018, 9(5): 743–754.
- [19] DU Mingjing, DING Shifei, XUE Yu. A robust density peaks clustering algorithm using fuzzy neighborhood[J]. *International journal of machine learning and cybernetics*, 2018, 9(7): 1131–1140.
- [20] XU Xiao, DING Shifei, SHI Zhongzhi. An improved density peaks clustering algorithm with fast finding cluster centers[J]. *Knowledge-based systems*, 2018, 158: 65–74.
- [21] DING Shifei, DU Mingjing, SUN Tongfeng, et al. An entropy-based density peaks clustering algorithm for mixed type data employing fuzzy neighborhood[J]. *Knowledge-based systems*, 2017, 133: 294–313.
- [22] SALVADOR S, CHAN P. Determining the number of clusters/segments in hierarchical clustering/segmentation

algorithms[C]//Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence. Boca Raton, USA, 2004.

- [23] WILLIAMS C K I. Learning with kernels: support vector machines, regularization, optimization, and beyond[J]. Journal of the American statistical association, 2003, 98(462): 489.
- [24] DE LIMA M D, COSTA N L, BARBOSA R. Improvements on least squares twin multi-class classification support vector machine[J]. *Neurocomputing*, 2018, 313: 196–205.
- [25] CHEN Sugan, WU Xiaojun. Multiple birth least squares support vector machine for multi-class classification[J]. *International journal of machine learning and cybernetics*, 2017, 8(6): 1731–1742.

作者简介:



史颂辉, 硕士研究生, 主要研究方向为支持向量机、机器学习。



丁世飞, 教授, 博士生导师, 主要研究方向为人工智能、机器学习、模式识别、数据挖掘。主持国家重点基础研究计划项目 1 项、国家自然科学基金面上项目 3 项。发表学术论文 200 余篇, 出版专著 5 部。

2020 中国人工智能大会 Chinese Congress on Artificial Intelligence 2020

中国人工智能大会由中国人工智能学会创办于 2015 年, 每年举办一届。该会是我国最早发起举办的人工智能大会, 目前已经成为我国人工智能领域规格最高、规模最大、影响力最强的会议之一。

2020 年中国人工智能大会 (CCAI 2020) 由中国科学技术协会、中国科学院、南京市人民政府为指导单位, 中国人工智能学会、南京市建邺区人民政府、江苏省科学技术协会主办, 江苏南京生态科技岛经济开发区管理委员会、中新南京生态科技岛开发有限公司、马上科普教育科技 (北京) 有限公司承办, 江苏省人工智能学会、华为技术有限公司、北京中科闻歌科技股份有限公司协办。

CCAI 2020 以“智周万物”为主题, 探讨人工智能作为引领未来的战略性技术, 如何以周知万物的学术境界赋能未来, 带动时代发展, 实现互联世界的远大理想。

本次大会将延续过去五届的强大阵容, 汇聚全球人工智能领域学术界、产业界顶级科学家和业界精英, 围绕全球前沿热点、核心技术以及国家和社会热点话题进行重点探讨, 并着重关注我国当前人工智能产业发展现状和未来发展态势。

(相关会议信息参见 <http://ccai.caii.cn/>)