

具有间隔分布的抗噪声非平行支持向量机

刘历铭, 李 平, 储茂祥, 蔡宏斌

(辽宁科技大学电子与信息工程学院, 辽宁鞍山 114051)

摘要: 针对非平行支持向量机(NonParallel Support Vector Machine, NPSVM)对噪声敏感和忽略了数据分布结构的问题, 提出了一种具有间隔分布的抗噪声非平行支持向量机(Anti-Noise NPSVM with Margin Distribution, MD-ANPSVM)分类模型. 在MD-ANPSVM模型中, 每个优化问题同时最小化两类样本的基于L1范数的绝对损失和改进的铰链损失, 这可以保证模型的稳定性, 减小噪声和异常值的影响. 此外, 在MD-ANPSVM模型中, 采用一阶和二阶统计量来描述训练数据的间隔分布信息, 并试图同时最大化间隔均值和最小化间隔方差, 这进一步提高了模型的泛化性能. 最终, 我们在不同的数据集上进行了对比实验. 实验结果显示, MD-ANPSVM模型具有较强的泛化能力和强鲁棒性.

关键词: 模式识别; 非平行支持向量机; 噪声不敏感; 间隔分布; 结构信息

基金项目: 国家自然科学基金(No.61673199); 辽宁省自然科学基金(No.2022-MS-353); 辽宁省高等学校基本科研资助项目(No.LJKMZ20220640); 辽宁省科技厅博士启动基金(No.2021-BS-246); 辽宁省教育厅自然科学基金青年基金(No.2020LNQN22)

中图分类号: TP18

文献标识码: A

文章编号: 0372-2112(2023)07-1889-09

电子学报URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20220268

Anti-Noise Nonparallel Support Vector Machine with Margin Distribution

LIU Li-ming, LI Ping, CHU Mao-xiang, CAI Hong-bin

(School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, Liaoning 114051, China)

Abstract: Because nonparallel support vector machine (NPSVM) is sensitive to noise and ignores the distributing structure of data, an anti-noise NPSVM with margin distribution (MD-ANPSVM) model is proposed. In MD-ANPSVM, each optimization problem simultaneously minimizes the L1-norm loss and improved hinge loss, which can ensure the stability of the model and reduce the adverse impact of noise and outliers. In addition, in MD-ANPSVM, the margin distribution described by the first- and second-order statistics is introduced. Each optimization problem simultaneously maximizes the margin mean and minimizes the margin variance, which results in better generalization performance. The experimental results on the UCI datasets and steel surface defects dataset show that MD-ANPSVM can achieve better generalization ability and strong robustness.

Key words: pattern recognition; nonparallel support vector machine; noise insensitive; margin distribution; structural information

Foundation Items: National Natural Science Foundation of China (No.61673199); Natural Science Foundation of Liaoning Province (No.2022-MS-353); Basic Scientific Research Project of Education Department of Liaoning Province (No.LJKMZ20220640, No.2020LNQN22); Natural Science Foundation of Liaoning Province (2021-BS-246)

1 引言

支持向量机(Support Vector Machine, SVM)^[1,2]是由Vapnik及其合作伙伴提出的一种监督学习方法. 其基本思想是通过最大化两个平行边界超平面之间的间隔来寻找一个最优分类超平面. SVM遵循结构风险最

小化原则来控制其泛化误差的上界, 从而提高泛化能力. 因其在二分类问题上的良好性能而受到广泛关注. 迄今为止, SVM已被广泛应用于机器学习、统计、模式识别等领域. 但是, SVM的学习能力有限, 且计算复杂度. 为了克服上述弊端, 文献[3]提出了孪生支持向

量机(Twin Support Vector Machine, TSVM)算法. TSVM求解两个较小规模的凸二次规划问题(Quadratic Programming Problems, QPPs),而不是SVM中的一个较大尺寸的QPP.这意味着TSVM的学习效率大约是SVM的四倍^[4].然而,TSVM在准确性方面并不令人满意,仍有一些改进的空间.

为了改进TSVM,Tian等人^[5]通过引入 ε -不敏感损失函数代替TSVM中的二次损失函数,提出了一种新的非平行超平面分类器,称为非平行支持向量机(Non-Parallel SVM, NPSVM).NPSVM是基于非平行超平面的所有模型中的一个有效模型,具有很多优点.相比于TSVM,NPSVM在有效性和泛化性等方面都有了不同程度的提高.

尽管如此,仍然有两个关键问题限制了NPSVM的有效性.第一个问题是, ε -不敏感损失和铰链损失容易引起噪声敏感性和重采样不稳定性.在生成最优分类模型中,损失函数起着重要作用.鉴于此,许多学者^[6,7]开始研究具有噪声不敏感特性的损失函数.第二个问题是,NPSVM没有有效地利用先验数据分布信息,忽略了一个好的分类器应该对数据分布的结构敏感这一事实^[8].出于这种考虑,模式识别中出现了挖掘数据分布信息的方法以及将这些信息放入经典分类器的技术^[9,10].

为了克服NPSVM中的上述限制,本文提出了一种具有间隔分布的抗噪声非平行支持向量机(Anti-Noise NPSVM with Margin Distribution, MD-ANPSVM).本文的主要贡献总结如下:(1)建立了MD-ANPSVM框架,在MD-ANPSVM中,每个优化问题同时最小化两类样本的基于L1范数的绝对损失和改进的铰链损失,从而产生不同的对偶问题,进而降低了模型的计算复杂度;(2)基于L1范数的绝对损失比 ε -不敏感损失更不敏感,这可以保证模型的稳定性,减轻噪声和异常值的影响;(3)在MD-ANPSVM模型中,我们用一阶和二阶统计量来描述间隔分布,并试图同时最大化间隔均值和最小化间隔方差,从而充分利用数据分布信息,进一步提高了NPSVM的性能.

2 基于间隔分布的抗噪声非平行支持向量机

在本节中,提出了一种新的非平行超平面分类器,称为MD-ANPSVM.首先,为了更合理地将数据分布的结构应用到NPSVM中,重新定义了间隔分布,包括间隔均值和间隔方差.其次,将重新定义的间隔分布嵌入到NPSVM的目标函数中,建立了MD-ANPSVM的优化模型.最后,根据对偶理论和Karush-Kuhn-Tucker(KKT)优化理论,推导出MD-ANPSVM的最优解.

2.1 间隔分布

众所周知,大间隔分布机(Large margin Distribution

Machine, LDM)^[10]的分类原则是最大化间隔均值和最小化间隔方差.其目的是使分类超平面介于正负类样本之间,使分类器具有更大的间隔分布和更强的泛化能力.

通过分析发现,最小化间隔方差包括最小化正类样本的方差、最小化负类样本的方差和最小化正负类样本的方差.对于NPSVM而言,首先,我们希望将正类样本或负类样本尽可能地聚集在一起.这一点可以分别通过最小化正类样本的方差和负类样本的方差实现.其次,我们希望将负类或正类样本尽可能多地远离相对应的超平面.由于支持超平面 $(\mathbf{w}^+)^T \mathbf{x} + b^+ = -1$ 和 $(\mathbf{w}^-)^T \mathbf{x} + b^- = +1$ 不必位于边界,因此可以通过优化间隔分布来改变支持超平面的位置.

综上所述,对于NPSVM中的支持超平面 $(\mathbf{w}^+)^T \mathbf{x} + b^+ = -1$,间隔分布可以重新定义为

$$\bar{y}^+ = \frac{1}{l} \sum_{i=1}^l y_i f_s^+(\mathbf{x}_i) \quad (1)$$

$$\hat{y}^+ = \frac{1}{l^2} \sum_{i=1}^l \sum_{j=1}^l (f_s^+(\mathbf{x}_{i1}) - f_s^+(\mathbf{x}_{i2}))^2 + \frac{1}{l^2} \sum_{j=1}^l \sum_{i=1}^l (f_s^+(\mathbf{x}_{j1}) - f_s^+(\mathbf{x}_{j2}))^2 \quad (2)$$

其中, $f_s^+(\mathbf{x}_i) = (\mathbf{w}^+)^T \mathbf{x}_i + b^+ + 1$, $\mathbf{x}_i \in \mathbf{R}^d$ 为任意样本, $y_i \in \{+1, -1\}$ 为样本标签, l 是样本的总个数, d 是输入空间的维度.

对于支持超平面 $(\mathbf{w}^-)^T \mathbf{x} + b^- = +1$,间隔分布可以重新定义为

$$\bar{y}^- = \frac{1}{l} \sum_{i=1}^l y_i f_s^-(\mathbf{x}_i) \quad (3)$$

$$\hat{y}^- = \frac{1}{l^2} \sum_{i=1}^l \sum_{j=1}^l (f_s^-(\mathbf{x}_{i1}) - f_s^-(\mathbf{x}_{i2}))^2 + \frac{1}{l^2} \sum_{j=1}^l \sum_{i=1}^l (f_s^-(\mathbf{x}_{j1}) - f_s^-(\mathbf{x}_{j2}))^2 \quad (4)$$

其中, $f_s^-(\mathbf{x}_i) = (\mathbf{w}^-)^T \mathbf{x}_i + b^- - 1$.

2.2 线性情况

为了实现经验风险,MD-ANPSVM考虑了以下两个损失函数:

$$R_1^{\text{emp}} = c_1 \sum_{i=1}^l |(\mathbf{w}^+)^T \mathbf{x}_i^+ + b^+ - 0| + c_2 \sum_{j=1}^l \max \left(0, \frac{(\mathbf{w}^+)^T \mathbf{x}_j^- + b^+ + 1}{t_1} \right) \quad (5)$$

$$R_2^{\text{emp}} = c_3 \sum_{j=1}^l |(\mathbf{w}^-)^T \mathbf{x}_j^- + b^- - 0| + c_4 \sum_{i=1}^l \max \left(0, \frac{1 - (\mathbf{w}^-)^T \mathbf{x}_i^+ - b^-}{t_2} \right) \quad (6)$$

其中, $c_k \geq 0$, $k=1,2,\dots,4$ 是惩罚参数, $t_1, t_2 > 0$ 是可调节参数.此外,为进一步实现结构风险,MD-ANPSVM中

引入了额外的正则化项.

因此,对于线性情况,MD-ANPSVM 求解如下两个凸规划问题:

$$\begin{aligned} \min_{w^+, b^+} & \frac{1}{2} \|w^+\|^2 + c_1 \|\xi^+\|_1 + c_2 (e^-)^T \eta^- - v_1 \bar{y}^+ + v_2 \hat{y}^+ \\ \text{s.t.} & \quad x^+ w^+ + e^+ b^+ = \xi^+, \\ & \quad x^- w^+ + e^- b^+ \leq -e^- + t_1 \eta^-, \quad \eta^- \geq 0, \end{aligned} \quad (7)$$

$$\begin{aligned} \min_{w^-, b^-} & \frac{1}{2} \|w^-\|^2 + c_3 \|\xi^-\|_1 + c_4 (e^+)^T \eta^+ - v_3 \bar{y}^- + v_4 \hat{y}^- \\ \text{s.t.} & \quad x^- w^- + e^- b^- = \xi^-, \\ & \quad x^+ w^- + e^+ b^- \geq e^+ - t_2 \eta^+, \quad \eta^+ \geq 0. \end{aligned} \quad (8)$$

其中, $v_k \geq 0, k=1, 2, \dots, 4$ 是惩罚参数, ξ^+ 和 η^+ 是具有适当维度的松弛向量, $\|\cdot\|_1$ 是 L1 范数, e^+ 和 e^- 是具有适当维度的全 1 列向量.

首先,该分类器利用基于 L1 范数的损失和改进的铰链损失来实现 MD-ANPSVM 的经验风险最小化. 也就是说,在每个优化问题中,它同时考虑对应类的 L1 范数的绝对损失和其他类的改进的铰链损失. 与 NPSVM 相比,MD-ANPSVM 不仅取得了较高的精度,而且对噪声和异常值的鲁棒性也更高. 其次,MD-ANPSVM 采用了一阶和二阶统计量来捕获数据的结构分布. 将最大化间隔均值和最小化间隔方差嵌入到 MD-ANPSVM 的目标函数,这导致数据结构被充分利用,进一步提高了模型的泛化性能.

定理 1^[11] 对于任意一个向量 θ , 如果令 $\theta = \theta^+ - \theta^-$, 且 θ^\pm 满足 $\theta^\pm \geq 0$, 那么 $\min \{\|\theta\|_1\} = \min \{\theta^+ - \theta^-\}$ 成立.

因此,通过将 ξ^+ 和 ξ^- 分别设置为 $\xi^+ = \xi_1^+ - \xi_2^+$ 和 $\xi^- = \xi_1^- - \xi_2^-$, 且满足 $\xi_1^+ \geq 0, \xi_2^+ \geq 0, \xi_1^- \geq 0$ 和 $\xi_2^- \geq 0$, 则式(7)和(8)可改写为如下优化问题:

$$\begin{aligned} \min_{w^+, b^+} & \frac{1}{2} \|w^+\|^2 + c_1 (e^+)^T (\xi_1^+ + \xi_2^+) + c_2 (e^-)^T \eta^- \\ & \quad - v_1 \bar{y}^+ + v_2 \hat{y}^+ \\ \text{s.t.} & \quad x^+ w^+ + e^+ b^+ = \xi_1^+ - \xi_2^+, \quad \xi_1^+ \geq 0, \quad \xi_2^+ \geq 0, \\ & \quad x^- w^+ + e^- b^+ \leq -e^- + t_1 \eta^-, \quad \eta^- \geq 0, \end{aligned} \quad (9)$$

$$\begin{aligned} \min_{w^-, b^-} & \frac{1}{2} \|w^-\|^2 + c_3 (e^-)^T (\xi_1^- + \xi_2^-) + c_4 (e^+)^T \eta^+ \\ & \quad - v_3 \bar{y}^- + v_4 \hat{y}^- \\ \text{s.t.} & \quad x^- w^- + e^- b^- = \xi_1^- - \xi_2^-, \quad \xi_1^- \geq 0, \quad \xi_2^- \geq 0, \\ & \quad x^+ w^- + e^+ b^- \geq e^+ - t_2 \eta^+, \quad \eta^+ \geq 0. \end{aligned} \quad (10)$$

接下来,根据优化理论^[12],利用拉格朗日乘子法将上述原始优化问题转化为对偶问题来求解. 以式(9)为例,对应的拉格朗日函数为

$$\begin{aligned} L = & \frac{1}{2} \|w^+\|^2 + c_1 (e^+)^T (\xi_1^+ + \xi_2^+) + c_2 (e^-)^T \eta^- \\ & - \frac{v_1}{l} y^T (X w^+ + e b^+ + e) \\ & + \frac{2v_2}{l^2} (l^+ (x^+ w^+)^T (x^+ w^+) - (x^+ w^+)^T (x^+ w^+)) \\ & + \frac{2v_2}{l^2} (l^- (x^- w^+)^T (x^- w^+) - (x^- w^+)^T (x^- w^+)) \\ & - (\alpha^+)^T (x^+ w^+ + e^+ b^+ - \xi_1^+ + \xi_2^+) \\ & + (\beta^-)^T (x^- w^+ + e^- b^+ + e^- - t_1 \eta^-) \\ & - (\rho_1^+)^T \xi_1^+ - (\rho_2^+)^T \xi_2^+ - (\lambda^-)^T \eta^- \end{aligned} \quad (11)$$

其中, $\alpha^+, \beta^-, \rho_1^+, \rho_2^+$ 和 λ^- 是拉格朗日乘子.

将拉格朗日函数 L 与 $w^+, b^+, \xi_1^+, \xi_2^+$ 和 η^- 进行微分, 得到以下 KKT 条件:

$$\begin{aligned} \frac{\partial L}{\partial w^+} = & w^+ - \frac{v_1}{l} X^T y - (x^+)^T \alpha^+ + (x^-)^T \beta^- \\ & + \frac{4v_2}{l^2} (l^+ (x^+)^T (x^+ w^+) - (x^+)^T (x^+ w^+)) \\ & + \frac{4v_2}{l^2} (l^- (x^-)^T (x^- w^+) - (x^-)^T (x^- w^+)) = 0 \end{aligned} \quad (12)$$

$$\frac{\partial L}{\partial b^+} = -\frac{v_1}{l} e^T y - (e^+)^T \alpha^+ + (e^-)^T \beta^- = 0 \quad (13)$$

$$\frac{\partial L}{\partial \xi_1^+} = c_1 e^+ + \alpha^+ - \rho_1^+ = 0 \quad (14)$$

$$\frac{\partial L}{\partial \xi_2^+} = c_1 e^+ - \alpha^+ - \rho_2^+ = 0 \quad (15)$$

$$\frac{\partial L}{\partial \eta^-} = c_2 e^- - t_1 \beta^- - \lambda^- = 0 \quad (16)$$

根据 KKT 条件,式(9)的对偶问题可以表述为

$$\begin{aligned} \min_{\pi^+} & \frac{1}{2} (\Theta^+ \pi^+)^T ((Q^+)^{-1})^T \Theta^+ \pi^+ + ((S^+)^T ((Q^+)^{-1})^T \Theta^+ - f^+) \pi^+ \\ \text{s.t.} & \quad L^+ \leq \pi^+ \leq U^+ \\ & \quad \Omega^+ \pi^+ = \frac{v_1}{l} e^T y \end{aligned} \quad (17)$$

其中,

$$\pi^+ = [(\alpha^+)^T, (\beta^-)^T]^T, \quad U^+ = [(c_1 e^+)^T, (c_2 e^- / t_1)^T]$$

$$L^+ = [-(c_1 e^+)^T, \theta^{-T}], \quad \Omega^+ = [-(e^+)^T, (e^-)^T]$$

$$\Theta^+ = [(x^+)^T, -(x^-)^T], \quad S^+ = \frac{v_1}{l} X^T y, \quad f^+ = [\theta^{+T}, (\beta^-)^T],$$

$$Q^+ = I + \frac{4v_2}{l^2} (l^+ - 1)(x^+)^T x^+ + \frac{4v_2}{l^2} (l^- - 1)(x^-)^T x^-.$$

同样地,可以得到式(10)的对偶问题为

$$\begin{aligned} \min_{\pi^-} & \frac{1}{2} (\Theta^- \pi^-)^T ((Q^-)^{-1})^T \Theta^- \pi^- \\ & + ((S^-)^T ((Q^-)^{-1})^T \Theta^- \pi^- - f^-) \pi^- \\ \text{s.t.} & \quad L^- \leq \pi^- \leq U^- \\ & \quad \Omega^- \pi^- = \frac{v_3}{l} e^T y \end{aligned} \quad (18)$$

其中,

$$\begin{aligned}\pi^- &= [(\alpha^-)^T, (\beta^+)^T]^T, U^- = [(c_3 e^-)^T, (c_4 e^+ / t_2)^T]^T, \\ L^- &= [-(c_3 e^-)^T, \theta^+]^T, \Omega^- = [-(e^-)^T, -(e^+)^T]^T, \\ \Theta^- &= [(x^-)^T, (x^+)^T]^T, S^- = \frac{v_3}{l} X^T y, f^- = [\theta^{-T}, (\beta^+)^T]^T, \\ Q^- &= I + \frac{4v_4}{l^2} (l^+ - 1)(x^+)^T x^+ + \frac{4v_4}{l^2} (l^- - 1)(x^-)^T x^-\end{aligned}$$

对于式(17),根据KKT条件,可以得到与文献[13, 14]中类似的结论.

定理 2 假设 $\pi^* = [(\alpha^*)^T, (\beta^*)^T]^T$ 是对偶问题(26)的最优解. 如果 π^* 中有一个分量的值在区间 (L^+, U^+) 内,则可以计算原始问题(9)的解 $[w^+, b^+]$, 其中,

$$\begin{aligned}w^+ &= (Q^+)^{-1} \left(\frac{v_1}{l} X^T y - (x^-)^T \beta^* + (x^+)^T \alpha^* \right) \quad (19) \\ b^+ &= \begin{cases} -\frac{1}{|\hat{T}^+|} \sum_{i \in \hat{T}^+} (w^+)^T x_i^+, & \alpha_i^+ \in (-c_1, c_1) \\ -\frac{1}{|\hat{T}^-|} \sum_{j \in \hat{T}^-} 1 + (w^+)^T x_j^-, & \beta_j^- \in (0, c_2 / t_1) \end{cases} \quad (20)\end{aligned}$$

其中, $\hat{T}^+ = \{i | -c_1 < \alpha_i^+ < c_1\}$, $|\hat{T}^+|$ 是 \hat{T}^+ 中的元素个数; $\hat{T}^- = \{j | 0 < \beta_j^- < c_2 / t_1\}$, $|\hat{T}^-|$ 是 \hat{T}^- 中的元素个数.

类似地,可以推导出优化问题(10)的最优解 $[w^-, b^-]$, 如下:

$$\begin{aligned}w^- &= (Q^-)^{-1} \left(\frac{v_3}{l} X^T y + (x^+)^T \beta^* + (x^-)^T \alpha^* \right) \quad (21) \\ b^- &= \begin{cases} -\frac{1}{|\bar{T}^-|} \sum_{j \in \bar{T}^-} (w^-)^T x_j^-, & \alpha_j^- \in (-c_3, c_3) \\ \frac{1}{|\bar{T}^+|} \sum_{i \in \bar{T}^+} 1 - (w^-)^T x_i^+, & \beta_i^+ \in (0, c_4 / t_2) \end{cases} \quad (22)\end{aligned}$$

其中, $\bar{T}^- = \{j | -c_3 < \alpha_j^- < c_3\}$, $|\bar{T}^-|$ 是集合 \bar{T}^- 中的元素个数; $\bar{T}^+ = \{i | 0 < \beta_i^+ < c_4 / t_2\}$, $|\bar{T}^+|$ 是集合 \bar{T}^+ 中的元素个数.

一旦得到解 $[w^+, b^+]$ 和 $[w^-, b^-]$, 即可通过下列决策函数对一个未知标签数据 x_i 进行分类.

$$f_d(x_i) = \arg \min_{+, -} \{d^+, d^-\} \quad (23)$$

其中, $d^+ = |(w^+)^T x_i + b^+|$ 和 $d^- = |(w^-)^T x_i + b^-|$.

2.3 非线性情况

当样本在输入空间中呈线性不可分时,线性分类器的性能会显著下降. 对于这种情况,可通过引入非线性映射函数 $\varphi(\cdot)$ 将线性不可分离样本从原始输入空间映射到一个维数足够高的特征空间,以使非线性分类问题转化为线性分类问题. 在本小节中,使用核技巧将线性 MD-ANPSVM 扩展到非线性情况.

具体来说,非线性 MD-ANPSVM 优化了以下两个优化问题:

$$\begin{aligned}\min_{w^+, b^+} & \frac{1}{2} \|w^+\|^2 + c_1 \|\xi^+\|_1 + c_2 (e^-)^T \eta^- - v_1 \bar{y}^+ + v_2 \hat{y}^+ \\ \text{s.t.} & \varphi(x^+) w^+ + e^+ b^+ = \xi^+, \\ & \varphi(x^-) w^+ + e^- b^+ \leq -e^- + t_1 \eta^-, \eta^- \geq 0,\end{aligned} \quad (24)$$

$$\begin{aligned}\min_{w^-, b^-} & \frac{1}{2} \|w^-\|^2 + c_3 \|\xi^-\|_1 + c_4 (e^+)^T \eta^+ - v_3 \bar{y}^- + v_4 \hat{y}^- \\ \text{s.t.} & \varphi(x^-) w^- + e^- b^- = \xi^-, \\ & \varphi(x^+) w^- + e^+ b^- \geq e^+ - t_2 \eta^+, \eta^+ \geq 0.\end{aligned} \quad (25)$$

与线性情况类似,通过引入拉格朗日函数,可以得到式(24)和式(25)的对偶问题,如下所示:

$$\begin{aligned}\min_{\pi^+} & \frac{1}{2} (\Theta^{\varphi^+} \pi^+)^T ((Q^{\varphi^+})^{-1})^T \Theta^{\varphi^+} \pi^+ \\ & + ((S^{\varphi^+})^T ((Q^{\varphi^+})^{-1})^T \Theta^{\varphi^+} - f^+) \pi^+ \\ \text{s.t.} & L^+ \leq \pi^+ \leq U^+, \\ & \Omega^+ \pi^+ = \frac{v_1}{l} e^T y,\end{aligned} \quad (26)$$

$$\begin{aligned}\min_{\pi^-} & \frac{1}{2} (\Theta^{\varphi^-} \pi^-)^T ((Q^{\varphi^-})^{-1})^T \Theta^{\varphi^-} \pi^- \\ & + ((S^{\varphi^-})^T ((Q^{\varphi^-})^{-1})^T \Theta^{\varphi^-} - f^-) \pi^- \\ \text{s.t.} & L^- \leq \pi^- \leq U^-, \\ & \Omega^- \pi^- = \frac{v_3}{l} e^T y.\end{aligned} \quad (27)$$

其中,

$$\begin{aligned}\Theta^{\varphi^+} &= [\varphi(x^+)^T, -\varphi(x^-)^T]^T, S^{\varphi^+} = \frac{v_1}{l} \varphi(X)^T y, \\ Q^{\varphi^+} &= I + \frac{4v_2}{l^2} (l^+ \varphi(x^+)^T \varphi(x^+) - \varphi(x^+)^T \varphi(x^+)) \\ & + \frac{4v_2}{l^2} (l^- \varphi(x^-)^T \varphi(x^-) - \varphi(x^-)^T \varphi(x^-)), \\ \Theta^{\varphi^-} &= [\varphi(x^-)^T, \varphi(x^+)^T]^T, S^{\varphi^-} = \frac{v_3}{l} \varphi(X)^T y, \\ Q^{\varphi^-} &= I + \frac{4v_4}{l^2} (l^+ \varphi(x^+)^T \varphi(x^+) - \varphi(x^+)^T \varphi(x^+)) \\ & + \frac{4v_4}{l^2} (l^- \varphi(x^-)^T \varphi(x^-) - \varphi(x^-)^T \varphi(x^-))\end{aligned}$$

因为特征空间的维度更高甚至无限大,所以 $\varphi(x)$ 不能明确表示. 针对该问题,可以采用内积运算来表示所有计算. 换句话说,核函数 $\psi(x_i, x_j) = \varphi(x_i) \varphi(x_j)^T$ 可以弥补 $\varphi(x)$ 的复杂性.

3 MD-ANPSVM 算法分析

3.1 与 LDM 和 FSVM 的关系

与 LDM 一样, MD-ANPSVM 计算所有训练样本的间隔分布,并将间隔分布引入到 NPSVM 模型中. LDM 和 MD-NANPSVM 都更关注于训练样本的间隔分布. 不同的是, LDM 生成分类超平面 $w^T x = 0$, 并使其位于正负类样本之间. 而 MD-ANPSVM 则是生成一对非平行的支持超平面 $(w^+)^T x + b^+ = -1$ 和 $(w^-)^T x + b^- = +1$, 使其可以不必位于边界,其位置可通过优化间隔分布来改变.

MD-ANPSVM 和模糊支持向量机 (Fuzzy SVM, FSVM)^[15]都是用来处理异常值和噪声对分类结果的影响. FSVM 可以看作是一种特殊的代价敏感学习算法,因为它为不同的样本分配不同的惩罚权重. 在 FSVM 中,惩罚权重首先由模糊隶属函数计算,然后存储到模糊向量中. 至于本文提出的 MD-ABPSVM,主要是通过损失函数进行研究,通过构造新的损失函数,来建立一个鲁棒的分类模型,使其能抑制噪声的不利影响,具有强鲁棒性特点.

除此之外,相比于 LDM 和 FSVM, MD-ANPSVM 具有非平行结构,所以,它能很好地解决交叉数据集的分类问题,具有更好的泛化性能.

3.2 与 TSVM 和 NPSVM 的关系

虽然 MD-ANPSVM、NPSVM 和 TSVM 都寻求两个非平行分类面,但它们的对偶问题是不同的. 首先,对于非线性情况,TSVM 需要考虑核生成曲面而不是超平面,并需要两个不同的原始问题,且没有合理的几何解释. 相反,本文提出的 MD-ANPSVM 和 NPSVM 在线性和非线性情况下都具有一致性. 也就是说,它们的对偶问题中只出现内积,因此核函数可以像 SVM 一样直接应用. 此外,与 TSVM 和 NPSVM 相比,MD-ANPSVM 可以通过优化间隔分布充分利用训练样本的结构信息,进而实现更好的泛化性能. 除此之外,相比于 NPSVM,MD-ANPSVM 可以保证模型的稳定性,减轻噪声和异常值的影响. 但是,我们的 MD-ANPSVM 需要计算矩阵的逆,且缺乏稀疏性,这需要进一步地改进.

3.3 MD-ANPSVM 的计算复杂度

为了公平起见,所有算法均采用拉格朗日乘数法来求解优化问题. 假设训练样本总数为 l , 正类样本数量等于负类样本数量,即 $l^+ = l^- = l/2$.

对于 SVM 而言,其计算复杂度是 $O(l^3)$. 类似地,FSVM 的计算复杂度大约为 $O(l^3)$. 对于 TSVM,它求解两个较小规模的优化问题,因此 TSVM 的计算复杂度大约是 $O(2 \times (l/2)^3)$. 显然,TSVM 的速度几乎是 SVM 的四倍. 至于 NPSVM,它同样求解两个优化问题,但每个问题的大小约为 $(3l/2)$, 所以其时间损失大约是 $O(2 \times (3l/2)^3)$. 可以看出,TSVM 比 NPSVM 快 27 倍. 也就是说,NPSVM 比 SVM 和 TSVM 都慢. 至于本文提出的 MD-ANPSVM,它求解两个优化问题,每个问题的大小约为 l , 所以其计算复杂度大约是 $O(2 \times l^3)$. 因此,MD-ANPSVM 和 NPSVM 之间的时间成本比率约为 $[O(2l^3)/O((27l^3/4)]=8/27$. 换句话说,MD-ANPSVM 比 NPSVM 快 27/8 倍. 但是,考虑到 MD-ANPSVM 需要计算训练样本的间隔分布信息,且需要计算矩阵逆运算,这导致额外的时间被需要. 因此,对于 MD-ANPSVM 而言,其计算复杂度通常要高于 $O(2 \times l^3)$.

4 实验分析

4.1 实验设置

为了验证 MD-ANPSVM 的有效性,本节研究了 MD-ANPSVM 在 UCI 数据集^[16]和真实应用数据集上的性能. 在实验中,重点比较了 MD-ANPSVM 和四种最先进的方法,包括 SVM、FSVM、LDM 和 NPSVM. 所有实验均在 MATLAB 2016b^[17]内使用一台装有 Intel i7 处理器,内存为 32 GB 的台式电脑上完成. 公平起见,SVM、FSVM、LDM、NPSVM 和 MD-ANPSVM 都使用 Matlab 中的“quadprog”函数求解优化问题.

在本文中,对于线性情况,所有算法均采用线性核函数 $\psi(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$. 对于非线性情况,采用径向基核函数 $\psi(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / \sigma^2)$, 其中, σ 是核半径参数. 为了降低非平行分类器参数设置的复杂性,本实验令 NPSVM 和 MD-ANPSVM 中的参数 $c_1 = c_3, c_2 = c_4, \nu_1 = \nu_3, \nu_2 = \nu_4$ 和 $t_1 = t_2$. 在本实验中,所有算法的惩罚参数和核半径参数 σ 均从集合 $\{2^i \mid i = -7, -6, \dots, +7\}$ 中选取, NPSVM 中的调节参数 ε 从集合 $\{0, 0.1, 0.5, 0.9\}$ 中选择.

在训练分类器之前,事先通常不知道哪些参数是最优的. 因此,该实验在训练集上进行五折交叉验证和网格搜索进行参数选择,以分类精度作为交叉验证选参数的标准,通过参数选择后的最优参数作为模型训练的参数. 一旦最优参数被确定,将利用它们来训练最终的决策函数.

4.2 UCI 数据集

为了评估所提出的 MD-ANPSVM 的强泛化性能,在本节中,从 UCI 数据库中选择了 10 个数据集进行比较实验,在实验之前,对所有数据集进行归一化处理,使特征在区间 $[0, 1]$ 内进行缩放.

表 1 列出了每个线性分类器在上述 UCI 数据集上获得的分类结果. 其中,“精度”表示平均测试精度和方差,“时间”代表平均训练时间. 在表 1 中,最高的分类精度用粗体表示. 此外,在表 1 的底部还列出了基于双边 t 检验的胜率/平率/负率 (W/T/L). 对于非线性情况,可获得相似的实验结果.

从表 1 可以观察到,在分类精度方面,MD-ANPSVM 在大多数数据集上都优于其他分类器. 这是因为 MD-ANPSVM 考虑了数据结构信息. 此外,可以看出,与其他四种分类器相比,MD-ANPSVM 的胜率/平率/负率往往优于它们或者至少持平. 至于模型的训练时间,可以发现在这些分类器中,除了 NPSVM 分类器之外,MD-ANPSVM 是最慢的. 事实上,这是因为 MD-ANPSVM 需要解决两个与 SVM 具有相同规模的对偶问题,且需要计算样本的间隔分布信息和矩阵逆运算,这导致其学

习速度较慢.

为了从统计学上来比较本章提出的 MD-ANPSVM 算法和其他四种算法所得的实验结果,本实验采用了弗里德曼检验^[18],来验证 MD-ANPSVM 和其他分类器在整个数据集上是否存在显著差异.表 1 的最后一行列出了分类器的平均秩.

根据五种算法的平均秩,并在所有算法都等价的零假设下,弗里德曼统计参数计算如下:

$$\chi_F^2 = \frac{12n}{k(k+1)} \left[\sum_j \text{Rank}_j^2 - \frac{k(k+1)^2}{4} \right] \quad (28)$$

其中,Rank_j表示第 j 个算法的平均排名, n 是数据集个数, k 是比较算法个数.考虑到原始弗里德曼检验过于保守,现在通常采用服从自由度为 (k-1) 和 (k-1)(n-1) 的 F 分布:

$$F_F = \frac{(n-1)\chi_F^2}{n(k-1) - \chi_F^2} \quad (29)$$

根据表 1 可知 k=5 和 n=10. 因此,对于表 1,根据式(28)和式(29),计算可得到 $\chi_F^2=26.06$ 和 $F_F=16.82$, 其中, F_F 是服从自由度为 (4, 36) 的 F 分布. 当显著性水平为 $\alpha=0.05$ 时, $F(4, 36)$ 的临界值为 2.633. 由于 $F_F=16.82$ 比 2.633 大得多. 因此,可以得出五种算法之间存在显著差异的结论. 此外,从表 1 可以看出,所提出的 MD-ANPSVM 的分类精度排名第一. 这进一步表明 MD-ANPSVM 比其他四种算法更有效.

接下来,为了验证 MD-ANPSVM 的抗噪声特性,在上述 UCI 数据集上添加了零均值和 0.05 方差的高斯白噪声,测试结果见表 2 所示. 可以发现,在大多数数据集上, MD-ANPSVM 比其他四种分类器具有更好的鲁棒性. 此外,从表 2 最后一行的平均秩可以看出, MD-ANPSVM 的优势更加突出,这表明 MD-ANPSVM 具有更好的噪声不敏感性.

表 1 UCI 数据集上线性分类器的测试结果

数据集	MD-ANPSVM		NPSVM		LDM		FSVM		SVM	
	精度/%	时间/s	精度/%	时间/s	精度/%	时间/s	精度/%	时间/s	精度/%	时间/s
australian	87.39±2.28	0.534 6	85.51±1.64	2.573 8	85.51±1.63	0.369 1	85.51±1.64	0.463 2	85.51±1.64	0.357 8
breast	74.39±4.38	0.072 2	71.88±3.90	0.516 7	72.93±4.68	0.091 7	72.97±4.23	0.082 9	71.88±3.90	0.106 0
diabetes	75.90±2.26	0.955 8	75.78±1.92	3.898 4	65.75±1.65	0.176 7	74.34±4.76	0.563 7	65.23±2.78	0.484 5
haberman	73.87±1.62	0.105 7	75.83±2.03	0.164 7	73.53±0.48	0.041 8	73.53±0.48	0.038 9	73.53±0.48	0.055 6
heart	82.96±4.60	0.112 5	82.22±5.81	0.395 9	81.48±3.70	0.055 6	79.63±3.51	0.060 6	80.37±3.81	0.069 2
ionosphere	89.45±2.97	0.322 2	88.31±2.95	1.256 4	81.76±1.48	0.092 4	87.46±2.32	0.182 6	80.33±1.96	0.194 9
iris	98.67±1.63	0.085 9	98.00±2.67	0.167 9	66.67±0.00	0.055 3	96.00±3.27	0.055 1	66.67±0.00	0.052 9
sonar	79.83±3.12	0.065 7	78.39±2.87	0.131 6	75.92±4.23	0.034 5	76.46±2.64	0.048 9	75.41±6.94	0.043 7
vote	95.86±1.72	0.120 1	94.70±1.74	0.588 2	94.27±2.96	0.173 9	93.80±2.74	0.163 3	93.34±2.63	0.156 1
wdbc	97.54±0.35	0.700 4	97.01±1.19	1.137 2	93.84±3.23	0.223 6	97.71±0.70	0.330 3	93.31±4.30	0.340 3
W/T/L	—		9/0/1		10/0/0		9/0/1		10/0/0	
平均秩	1.20		2.40		3.70		3.15		4.55	

表 2 五种分类器在方差为 0.05 的噪声数据集上的实验结果

数据集	MD-ANPSVM		NPSVM		LDM		FSVM		SVM	
	精度/%	时间/s	精度/%	时间/s	精度/%	时间/s	精度/%	时间/s	精度/%	时间/s
australian	85.93±1.77	1.067 1	85.51±1.50	5.222 3	85.36±1.63	0.312 0	85.51±1.50	0.342 2	85.22±1.76	0.457 4
breast	74.74±3.17	0.071 8	73.67±4.03	0.414 6	74.04±4.63	0.051 2	71.83±1.17	0.070 5	73.32±3.73	0.083 1
diabetes	75.52±1.91	2.433 7	74.47±4.38	8.546 2	67.19±2.02	0.567 1	75.12±2.98	2.922 0	67.19±1.87	2.744 6
haberman	73.87±1.62	0.156 8	73.53±0.48	0.138 5	73.86±0.90	0.058 1	73.53±0.48	0.039 9	73.86±0.90	0.057 7
heart	81.11±3.59	0.132 6	80.74±3.43	0.216 1	80.37±3.43	0.069 7	79.26±3.78	0.054 8	78.89±3.43	0.067 1
ionosphere	88.60±4.78	0.295 1	87.74±3.70	1.216 3	79.76±2.37	0.077 3	86.33±3.16	0.144 7	79.19±2.88	0.202 8
iris	89.33±2.49	0.114 7	89.33±2.49	0.302 6	67.33±2.49	0.056 2	84.67±4.52	0.052 5	66.67±0.00	0.077 0
sonar	77.90±5.03	0.086 6	77.40±7.50	0.159 8	75.01±7.66	0.056 6	75.52±5.80	0.044 7	74.53±8.08	0.063 7
vote	93.57±2.74	0.476 2	93.44±2.31	1.166 0	93.11±2.59	0.164 4	92.19±3.56	0.161 6	82.19±1.81	0.128 6
wdbc	96.48±0.97	0.633 7	95.96±1.53	1.431 9	87.17±3.67	0.247 3	96.48±1.85	0.301 5	86.29±3.24	0.358 1
平均秩	1.10		2.55		3.50		3.25		4.60	

4.3 应用数据集

为了进一步说明 MD-ANPSVM 的有效性,在本小节中,使用 MD-ANPSVM 进行了钢板表面缺陷分类应用,并将其与 SVM、LDM 和 NPSVM 进行了比较.

在本实验中,采用了东北大学表面缺陷数据库的六种典型缺陷数据集^[19],包含裂纹缺陷(CZ)、夹杂缺陷(IN)、斑块缺陷(PH)、麻点缺陷(PE)、氧化铁皮缺陷(RD)和划痕缺陷(SH).实验之前,通过采用一系列的图像处理技术,这些缺陷图像被提取为实验所需要的缺陷样本,共计 3 781 个,其中,特征描述包括了 3 个几何特征、7 个不变矩特征、9 个灰度特征和 16 个纹理特征.实验之前,将数据集随机划分为两个子集,一个用于训练,占有数据的 75%,另一个用于测试,占有数据的 25%.

考虑到钢板表面缺陷属于多分类问题,本文基于二叉树结构建立了多分类模型框架.图 1 记录了四个模型在缺陷数据集上的总分类精度和总训练时间.从图 1 可以看出,MD-ANPSVM 模型的总精度明显高于另外三个分类器.此外,相比于 NPSVM 模型,MD-ANPSVM 的运行效率更占据明显优势.

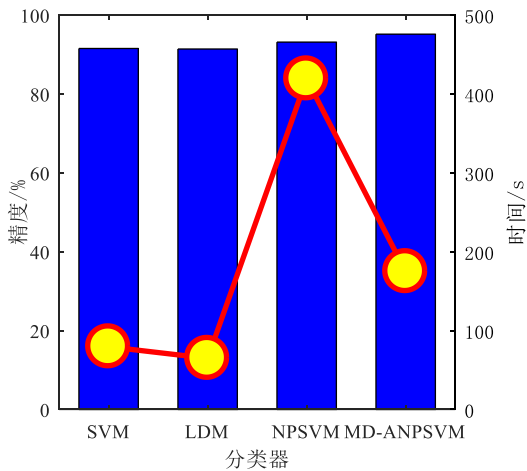


图 1 缺陷数据集中四个分类器的分类性能

为了直观地验证分类结果的有效性,本实验使用混淆矩阵对测试结果进行评估,如图 2 所示,其中每类缺陷正确的分类结果在对角线上.通过混淆矩阵可以看出,除了“PH”、“PE”和“CZ”缺陷数据集之外,MD-ANPSVM 在三种缺陷上都获得了最好的分类性能.此外,尽管 MD-ANPSVM 在其他三个数据集上的精度不是最优的,但相差无几.这进一步验证了 MD-ANPSVM 算法具有可行性和适用性.

SH	80.74	0.00	14.81	0.00	4.44	0.00
IN	0.00	89.93	0.00	0.72	3.60	5.76
PH	0.00	0.00	100.00	0.00	0.00	0.00
RD	0.00	5.26	0.00	78.95	12.63	3.16
PE	0.00	0.82	0.00	0.00	99.18	0.00
CZ	0.00	3.74	0.00	0.00	0.93	95.33
	SH	IN	PH	RD	PE	CZ

(a) SVM

SH	80.74	0.00	14.81	0.00	4.44	0.00
IN	0.00	89.93	0.00	1.44	3.60	5.04
PH	0.00	0.00	100.00	0.00	0.00	0.00
RD	0.00	5.26	0.00	80.00	11.58	3.16
PE	0.00	0.00	0.00	0.00	100.00	0.00
CZ	0.00	3.74	0.00	0.00	0.93	95.33
	SH	IN	PH	RD	PE	CZ

(b) LDM

SH	90.37	0.00	8.15	0.00	1.48	0.00
IN	0.00	91.37	1.44	3.60	0.00	3.60
PH	0.68	0.00	99.32	0.00	0.00	0.00
RD	0.00	3.16	0.00	85.26	4.21	7.37
PE	0.00	0.00	0.00	0.00	97.54	2.46
CZ	0.00	6.54	0.00	0.93	0.00	92.52
	SH	IN	PH	RD	PE	CZ

(c) NPSVM

SH	95.56	0.00	3.70	0.00	0.74	0.00
IN	0.00	92.09	0.00	4.32	1.44	2.16
PH	1.37	0.00	98.63	0.00	0.00	0.00
RD	0.00	3.16	0.00	90.53	6.32	0.00
PE	0.82	0.00	0.00	0.00	99.18	0.00
CZ	0.00	5.61	0.00	0.93	0.00	93.46
	SH	IN	PH	RD	PE	CZ

(d) MD-ANPSVM

图 2 四种算法的混淆矩阵

5 结论

在本文,主要关注于提高NPSVM模型的分类精度和抗噪声能力.受大间隔分布研究的启发,同时考虑损失函数对生成最优分类模型的影响,本文提出一种新的非平行超平面支持向量机,称为MD-NPSVM.与LDM一样,MD-ANPSVM在分类前通过一阶和二阶统计量来挖掘数据的结构分布信息,然后将优化间隔分布集成到NPSVM目标函数中.由于NPSVM和结构信息被有效地结合起来,与其他算法相比,MD-ANPSVM在泛化性能上具有显著优势.此外,由于使用了基于L1范数的绝对损失和改进的铰链损失,MD-ANPSVM在具有不同方差的噪声数据集上具有较强的鲁棒性.实验结果表明,MD-ANPSVM算法能够更充分地利用数据的先验结构信息,而且具有更强的噪声不敏感性.

在本文中,MD-ANPSVM主要用于处理中小型规模数据的分类问题.这是基于SVM类型的分类器的一个常见问题,主要归因于求解大型凸规划问题增加了模型的计算复杂度.因此,未来工作可能集中于如何更有效地求解凸规划问题.此外,用于选择最优参数的网格搜索法是耗时的.因此,在未来工作中,设计更有效的参数选择方法也是一个需要解决的关键问题.

参考文献

- [1] VAPNIK V N. The Nature of Statistical Learning Theory [M]. Berlin: Springer Science & Business Media, 1999.
- [2] DENG N, TIAN Y, ZHANG C. Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions [M]. Boca Raton: CRC Press, 2012.
- [3] 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.
- [4] DING S, HUA X, YU J. An overview on nonparallel hyperplane support vector machine algorithms[J]. Neural Computing and Applications, 2014, 25(5): 975-982.
- [5] TIAN Y, QI Z, JU X, et al. Nonparallel support vector machines for pattern classification[J]. IEEE Transactions on Cybernetics, 2013, 44(7): 1067-1079.
- [6] HUANG X, SHI L, SUYKENS J A K. Support vector machine classifier with pinball loss[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 36(5): 984-997.
- [7] 李凯, 李慧. 基于 pinball 损失的结构模糊孪生支持向量机[J]. 电子学报, 2019, 47(10): 2221-2227.
LI Kai, LI Hui. Structural fuzzy twin support vector machine with pinball loss[J]. Acta Electronica Sinica, 2019, 47(10): 2221-2227. (in Chinese)
- [8] YEUNG D S, WANG D, NG W W Y, et al. Structured large margin machines: Sensitive to data distributions[J]. Machine Learning, 2007, 68(2): 171-200.
- [9] TANG L, TIAN Y, LI W, et al. Structural improved regular simplex support vector machine for multiclass classification[J]. Applied Soft Computing, 2020, 91: 106235.
- [10] ZHANG T, ZHOU Z H. Large margin distribution machine[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. NY: ACM, 2014: 313-322.
- [11] PENG X, XU D, KONG L, et al. L1-norm loss based twin support vector machine for data recognition[J]. Information Sciences, 2016, 340: 86-103.
- [12] 袁亚湘, 孙文瑜. 最优化理论与方法[M]. 北京: 科学出版社, 1997.
- [13] DENG N, TIAN Y, ZHANG C. Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions[M]. Boca Raton: CRC Press, 2012.
- [14] SCHOLKOPF B, SMOLA A J. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond[M]. Cambridge: MIT Press, 2018.
- [15] LIN C F, WANG S D. Fuzzy support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2): 464-471.
- [16] DUA D, TANISKIDOU E K. UCI machine learning repository, 2017[OL]. [2022-10-16]. <http://archive.ics.uci.edu/ml/>.
- [17] The Math Works (MATLAB 2016b), Inc[OL]. [2022-10-16]. <http://www.mathworks.com>.
- [18] DEMSAR J. Statistical comparisons of classifiers over multiple data sets[J]. The Journal of Machine Learning Research, 2006, 7: 1-30.
- [19] HE Y, SONG K, MENG Q, et al. An end-to-end steel sur-

face defect detection approach via fusing multiple hierarchical features[J]. IEEE Transactions on Instrumentation and Measurement, 2019, 69(4): 1493-1504.

作者简介



刘历铭 女, 1994年6月出生, 辽宁本溪人. 2019年获得辽宁科技大学控制科学与工程硕士学位. 现为博士研究生, 研究方向为模式识别和机器学习.

E-mail: llm06101021@hotmail.com



李平 男, 1964年3月出生. 湖南涟源人. 1995年获得浙江大学工业自动化博士学位. 现任辽宁科技大学电子与信息工程学院教授. 研究方向为工业过程的先进控制和优化.



储茂祥 男, 1978年出生. 安徽桐城人. 2015年获得东北大学的博士学位. 现为辽宁科技大学电子与信息工程学院的教授. 研究方向包括模式识别、机器学习、图像处理和智能控制.



蔡宏斌 男, 1987年出生. 辽宁铁岭人. 2019年获得西北工业大学控制科学与工程博士学位. 现任辽宁科技大学电子与信息工程学院讲师. 研究方向为工业过程的先进控制和优化.