

一类基于MM方法的低秩稀疏支持矩阵机^{*}

翁 婷, 李国权

(重庆师范大学 数学科学学院, 重庆 401331)

摘要:支持矩阵机(support matrix machine, SMM)是一类直接处理矩阵类型数据的分类模型,是支持向量机的一种推广。稀疏支持矩阵机通过使用矩阵的 L_1 范数代替F范数,实现了特征选择。然而研究表明,与凸正则化项相比,非凸正则化项通常能实现更精确的近似,从而增强模型的特征选择能力。因此,提出一种新的带有非凸Geman函数诱导的低秩稀疏支持矩阵机模型。该模型在SMM的基础上,通过引入由非凸Geman函数诱导的行分组稀疏正则化项替代矩阵的核范数,在实现稀疏的同时促进矩阵的低秩性,获得更优的分类性能。为高效求解该模型,设计了一种基于majorization-minimization(MM)方法的改进优化算法。数值实验表明,所提算法比基准算法更高效,且引入非凸正则化项能够获得更好的分类效果。

关键词:支持矩阵机;Geman函数;组稀疏优化;MM方法

中图分类号:O224

文献标志码:A

文章编号:1672-6693(2026)04-0011-13

支持向量机(support vector machine, SVM)^[1]作为机器学习领域一种经典的监督学习方法,广泛应用于分类、回归等任务。然而随着数据形式的多样化,特别是矩阵数据的出现,如图像、脑电信号等二维结构数据,传统的SVM在处理这些复杂结构的数据时表现出一定的局限性。处理矩阵数据分类的一种常见策略是将矩阵数据向量化,然后应用一些成熟的基于向量的方法去处理。但这种策略存在一些缺陷:首先,它容易导致维度灾难。在向量化后,生成的向量的维数通常极高,这增加了过拟合的风险。其次,矩阵数据的向量化过程会破坏内部结构信息,例如行或列之间的相关性,而这些结构信息对于数据分类至关重要。

为保留矩阵的结构信息,研究者提出了多种矩阵分类器,例如Rank- k SVM^[2]、Low-Rank SVM^[2]、支持张量机^[3]、Bilinear SVM^[4]等。这类方法旨在通过对权重矩阵 W 施加低秩约束来捕捉结构信息。然而,它们通常会导致非凸优化问题,且矩阵 W 的秩需预先给定,从而增加了模型的调优难度。受Kobayashi等人^[5]利用核范数诱导低秩的启发,Luo等人^[6]提出了支持矩阵机(support matrix machine, SMM)。该模型联合核范数与F范数作为正则项,以同时促进模型的低秩性和控制模型复杂度。然而,在实际应用中,不是所有的特征都对分类有帮助,不排除存在特征冗余的情况。SMM未能解决特征冗余问题,冗余特征仍会干扰分类性能。

特征选择是处理高维特征矩阵的重要技术手段。该方法通过筛选关键特征变量,能够有效提升模型的可解释性、降低过拟合风险,进而增强模型的泛化能力。实现特征选择的一种常用策略是添加稀疏正则化项。2018年,Zheng等人^[7]提出了稀疏支持矩阵机(sparse support matrix machine, SSMM)模型。该方法通过同时引入核范数与 L_1 范数正则项,在保留输入矩阵结构的基础上实现特征选择,并采用广义前向-后向(generalized forward-backward, GFB)算法进行求解。SSMM在图像分类等任务中有效降低了冗余特征的影响,提升了模型的可解释性。2021年,Li等人^[8]提出了辛加权稀疏支持矩阵机模型。该方法基于辛几何变换生成辛加权系数矩阵,并引入变熵加权以增强故障特征的显著性,同样采用GFB算法求解。同年,Gu等人^[9]提出将 L_1 范数与平滑斜坡损失函数相结合,以抑制离群点对模型的影响,提出了斜坡稀疏支持矩阵机(ramp sparse SMM, RSSMM)模型。2022年,Wang等人^[10]提出了稀疏范数矩阵机模型。该模型采用 L_1 范数距离约束超平面,从而避免了矩阵求逆

* 收稿日期:2026-01-02 修回日期:2026-04-08 网络出版时间:2026-07-14T14:23

资助项目:国家自然科学基金面上项目(No. 12171063);重庆市教育委员会科技项目重点项目(No. KJZD-K202300509);重庆市自然科学基金面上项目(No. CSTB2025NSCQ-GPX1017)

第一作者简介:翁婷,女,研究方向为最优化理论与算法,E-mail: wengting_xs@163.com;通信作者简介:李国权,男,教授,博士,E-mail: ligq@cqu.edu.cn

网络出版地址:https://link.cnki.net/urlid/50.1165.n.20260713.1907.002

运算,并利用交替迭代法求解。2023 年, Li 等人^[11]提出了基于自相关函数的稀疏支持矩阵机(auto-correlation function based sparse support matrix machine, ACF-SSMM),通过自相关函数变换对输入矩阵进行时序增广,从而捕捉历史与当前时刻数据间的依赖关系,进而增强对时间序列信息的利用,最终提升模型的泛化性能,该模型同样通过 GFB 算法求解。

以上方法均是利用矩阵的 L_1 范数来诱导稀疏性。作为 L_0 范数的最佳凸近似, L_1 范数虽具有良好的性质,但所获得的稀疏解在解释性上通常不够理想。相比之下,非凸正则化项虽牺牲了问题的凸性,却能够获得比凸正则化项更好的近似效果。此外,由于 L_1 范数主要促进变量水平的稀疏,对于图像等具有先验结构(如分组结构)的矩阵数据,它并不能直接促进组级别的稀疏。在许多实际学习任务中,特征之间往往高度相关,呈现出明显的群体结构。组稀疏性已被证明在这些情况下是有效的,特别是对于图像等具有先验结构的矩阵数据,组稀疏通常效果更好^[12]。因此,如何有效刻画结构化稀疏仍是当前研究的一个关键问题。

此外,为实现低秩与稀疏的联合优化目标,现有方法大多通过引入核范数与 L_1 范数正则化项,分别促进模型的低秩性与稀疏性。例如, Zheng 等人^[13]提出的鲁棒支持矩阵机将输入矩阵 $\mathbf{X}_i$ 分解为低秩干净矩阵 $\mathbf{L}_i$ 与稀疏噪声矩阵 $\mathbf{S}_i$ 之和,即 $\mathbf{X}_i = \mathbf{L}_i + \mathbf{S}_i$,进而基于 $\mathbf{L}_i$ 进行模型训练。该方法通过对 $\mathbf{L}_i$ 与 $\mathbf{S}_i$ 分别施加核范数与 L_1 范数约束,实现了低秩与稀疏的联合优化,并采用交替方向乘子法进行求解。然而,鲁棒支持矩阵机的计算复杂度较高,且引入的多个超参数使得对参数设置较为敏感。类似地, RSSMM、SSMM 和 ACF-SSMM 等模型也通过联合使用 L_1 范数与核范数正则项实现低秩与稀疏联合优化。2024 年, Razzak 等人^[14]进一步提出了能够同步实现矩阵恢复的支持矩阵机模型,将输入矩阵分解为一个干净矩阵 $\mathbf{L}$ 和稀疏矩阵 $\mathbf{S}$ 的和,随后对权重矩阵 $\mathbf{W}$ 和干净矩阵 $\mathbf{L}$ 施加核范数约束以增强低秩性、捕捉矩阵内部结构特征,并在 $\mathbf{S}$ 上引入 $L_{2,1}$ 范数以促进列稀疏性。总体而言,现有研究通常借助矩阵的核范数来保留矩阵行列相关性(如 EEG 通道时空关联),同时利用 L_1 范数或 $L_{2,1}$ 范数诱导稀疏性,以实现特征选择或噪声分离。

Zhao 等人^[15]通过引入非凸函数诱导的行分组非凸正则化项,实现了降秩回归模型的组稀疏。受此启发,为提升矩阵分类模型的特征选择能力和泛化性能,本文在 SMM 模型基础上,利用由非凸 Geman 函数诱导的行分组稀疏正则化项 $R(\mathbf{W})$ 替换核范数正则化项,提出了低秩稀疏支持矩阵机(low-rank and sparse support matrix machine, LoRa-SSMM)。针对由此产生的非凸优化问题,设计了一种基于 majorization-minimization(MM)方法^[16]及拉格朗日对偶理论的求解算法。为实现更高效的求解,将非凸性重分配方法^[17]与 MM 方法相结合,将非凸优化问题转化为凸优化问题求解,从而降低了计算复杂度。为验证所提模型与算法的有效性,在图像等具有先验结构的数据集上进行了评估。

1 预备知识

本节将简要介绍 SVM、SMM 和 SSMM。首先介绍本文将用到的主要符号的含义。对于矩阵 $\mathbf{X} \in \mathbf{R}^{pq}$,定义它的秩为 $\text{Rank}(\mathbf{X}) = r$,其中 $r \leq \min\{p, q\}$ 。 $\text{vec}(\cdot)$ 为向量化算子,表示按列优先顺序将矩阵堆叠为列向量,即 $\mathbf{x} = \text{vec}(\mathbf{X}^T) = (X_{11}, X_{12}, \dots, X_{1q}, X_{21}, \dots, X_{pq})^T \in \mathbf{R}^{pq}$ 。矩阵 $\mathbf{X}$ 的 F 范数定义为 $\|\mathbf{X}\|_F = \sqrt{\sum_{i,j} X_{ij}^2}$, L_1 范数定义为 $\|\mathbf{X}\|_1 = \sum_{i,j} |X_{ij}|$,核范数定义为矩阵的奇异值的和,即 $\|\mathbf{X}\|_* = \sum_{i=1}^r \sigma_i(\mathbf{X})$,其中 $\sigma_i(\mathbf{X}) (i=1, \dots, r)$ 为矩阵 $\mathbf{X}$ 的奇异值。合页损失函数定义为 $(1-u)_+ = \max\{1-u, 0\}$ 。

1.1 SVM

SVM 是一种基于结构风险最小化原则的经典监督学习算法,核心思想是通过构造最大间隔超平面实现二分类问题的建模。给定训练样本集 $\{\mathbf{x}_i, y_i\}_{i=1}^n$,其中 $\mathbf{x}_i \in \mathbf{R}^{pq}$ 表示第 i 个向量形式的输入样本,对应的标签 $y_i \in \{1, -1\}$ 。SVM 旨在找到一个超平面 $\mathbf{w}^T \mathbf{x} + b = 0$,尽可能将 2 类样本分隔开,优化模型为:

$$\min_{\mathbf{w}, b} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n [1 - y_i (\mathbf{w}^T \mathbf{x}_i + b)]_+, \quad (1)$$

其中: $\mathbf{w}$ 和 b 分别表示权重向量和偏置, C 是罚参数。通过引入松弛变量,上述问题的对偶问题为:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j),$$

$$\text{s. t. } \sum_{i=1}^n \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i = 1, \dots, n. \quad (2)$$

其中: α_i 是拉格朗日乘子。由式(2)求解出 α_i^* , 可得到能判断未知样本 $\tilde{\mathbf{x}}$ 的标签 $\tilde{y}$ 的决策函数:

$$\tilde{y} = \text{sign}\left(\sum_{i=1}^n \alpha_i y_i (\mathbf{x}_i \cdot \tilde{\mathbf{x}}) + b\right).$$

此外, 当 $\mathbf{w} = \text{vec}(\mathbf{W}^T)$, $\mathbf{x}_i = \text{vec}(\mathbf{X}_i^T)$, 式(1)可写成:

$$\min_{\mathbf{w}, b} \frac{1}{2} \text{tr}(\mathbf{W}^T \mathbf{W}) + C \sum_{i=1}^n \{1 - y_i [\text{tr}(\mathbf{W}^T \mathbf{X}_i) + b]\}_+. \quad (3)$$

其中: $\mathbf{W}$ 是权重矩阵, $\mathbf{X}_i$ 为输入矩阵, $\text{tr}(\cdot)$ 表示矩阵的迹。从计算角度来看, 式(1)与式(3)本质上是等价的, 但当输入数据为矩阵时, 式(1)不能捕捉矩阵的结构信息。

1.2 SMM

SMM 是面向矩阵型数据的分类器, 是 SVM 的一种推广形式。给定训练数据集 $\{\mathbf{X}_i, y_i\}_{i=1}^n$, 其中 $\mathbf{X}_i \in \mathbf{R}^{p \times q}$ 表示第 i 个输入样本, $y_i \in \{1, -1\}$ 表示第 i 个样本的标签。SMM 遵循 SVM 的“最大间隔原则”, 旨在寻找一个超平面 $\text{tr}(\mathbf{W}^T \mathbf{X}_i) + b = 0$ 尽可能将属于不同类的数据分隔开。最初的 SMM 模型是由 Luo 等人^[6]提出, 优化模型为:

$$\min_{\mathbf{W}, b} \frac{1}{2} \text{tr}(\mathbf{W}^T \mathbf{W}) + \tau \|\mathbf{W}\|_* + C \sum_{i=1}^n \{1 - y_i [\text{tr}(\mathbf{W}^T \mathbf{X}_i) + b]\}_+, \quad (4)$$

其中: $\mathbf{W}$ 和 b 分别表示分类超平面的权重矩阵和偏置, $\text{tr}(\cdot)$ 表示矩阵的迹, $\|\mathbf{W}\|_*$ 表示为 $\mathbf{W}$ 的核范数, τ 是低秩参数, C 是罚参数, 最后一项表示合页损失函数。在式(4)中, $\frac{1}{2} \text{tr}(\mathbf{W}^T \mathbf{W}) + \tau \|\mathbf{W}\|_*$ 被统称为谱弹性网正则, F 范数和核范数分别控制矩阵 $\mathbf{W}$ 的复杂度和促进矩阵 $\mathbf{W}$ 的低秩性。SMM 将合页损失函数和谱弹性网惩罚相结合, 可直接作用于矩阵空间, 在保留矩阵结构信息的同时实现对矩阵数据的直接分类。

虽然式(4)是凸问题, 但合页损失和核范数的非光滑性使得传统的光滑优化算法不能直接用于求解该问题。文献[6]使用交替方向乘子法求解该问题。因此, 通过引入辅助变量矩阵 $\mathbf{Z}$, 式(4)可等价地写成:

$$\begin{aligned} \min_{\mathbf{W}, b} \quad & \frac{1}{2} \text{tr}(\mathbf{W}^T \mathbf{W}) + \tau \|\mathbf{Z}\|_* + C \sum_{i=1}^n \{1 - y_i [\text{tr}(\mathbf{W}^T \mathbf{X}_i) + b]\}_+, \\ \text{s. t. } \quad & \mathbf{W} - \mathbf{Z} = \mathbf{0}. \end{aligned}$$

上述问题的增广拉格朗日函数为:

$$\frac{1}{2} \text{tr}(\mathbf{W}^T \mathbf{W}) + \tau \|\mathbf{Z}\|_* + C \sum_{i=1}^n \{1 - y_i [\text{tr}(\mathbf{W}^T \mathbf{X}_i) + b]\}_+ + \text{tr}(\mathbf{A}^T (\mathbf{Z} - \mathbf{W})) + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{W}\|_F^2. \quad (5)$$

其中: $\mathbf{A}$ 为拉格朗日乘子, 是惩罚参数且 $\mu > 0$ 。从式(5)可知, 需求解 3 个矩阵变量: $\mathbf{Z}$ 、 $(\mathbf{W}, b)$ 和 $\mathbf{A}$ 。求解 SMM 的核心步骤是在每次迭代中计算 $\mathbf{Z}^{(k)}$ 和 $(\mathbf{W}^{(k)}, b^{(k)})$, 即交替执行以下步骤。

步骤 1, 固定 $\mathbf{Z} = \mathbf{Z}^{(k-1)}$, 求解如下问题:

$$\begin{aligned} & (\mathbf{W}^{(k)}, b^{(k)}) = \\ & \arg \min_{\mathbf{W}, b} \left\{ \frac{1}{2} \text{tr}(\mathbf{W}^T \mathbf{W}) + C \sum_{i=1}^n \{1 - y_i [\text{tr}(\mathbf{W}^T \mathbf{X}_i) + b]\}_+ - \text{tr}(\mathbf{A}^{(k-1)T} \mathbf{W}) + \frac{\mu}{2} \|\mathbf{Z}^{(k-1)} - \mathbf{W}\|_F^2 \right\}; \end{aligned} \quad (6)$$

步骤 2, 固定 $\mathbf{W} = \mathbf{W}^{(k)}$, 求解如下问题:

$$\mathbf{Z}^{(k)} = \arg \min_{\mathbf{Z}} \left\{ \tau \|\mathbf{Z}\|_* + \text{tr}(\mathbf{A}^{(k-1)T} \mathbf{Z}) + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{W}^{(k)}\|_F^2 \right\}; \quad (7)$$

步骤 3, 更新 $\mathbf{A}^{(k)}$:

$$\mathbf{A}^{(k)} = \mathbf{A}^{(k-1)} + \mu (\mathbf{Z}^{(k)} - \mathbf{W}^{(k)}).$$

式(6)、(7)的解分别为:

$$\begin{aligned} \hat{\mathbf{W}} &= \frac{1}{\mu + 1} (\mathbf{A} + \mu \mathbf{Z}^{(k-1)} + \sum_{i=1}^n \alpha_i^* y_i \mathbf{X}_i), \\ \hat{b} &= \frac{1}{|\Omega|} \sum_{i \in \Omega} \{y_i - \text{tr}(\hat{\mathbf{W}}^T \mathbf{X}_i)\}, \end{aligned}$$

$$\hat{\mathbf{Z}} = \frac{1}{\mu} D_{\tau}(\mu \mathbf{W}^{(k)} - \mathbf{A}).$$

其中: $\Omega = \{i: 0 < \alpha_i^* < C\}$, $|\Omega|$ 表示集合 Ω 中的元素数量; D_{τ} 表示为奇异值阈值算子^[18]。

通过交替方向乘子法得到问题的最优解 $\mathbf{Z}^*$, $\mathbf{A}^*$ 和 $(\mathbf{W}^*, b^*)$, 从而找到分类超平面为:

$$\text{tr}(\mathbf{W}^{*\top} \mathbf{X}_i) + b^* = 0.$$

1.3 SMM

SMM 假设每个输入矩阵中的所有元素都可以作为标签的解释特征。然而, 在实际应用中, 许多特征对于某些分类任务是冗余的, 因此通过特征选择过滤掉冗余特征来提升模型的可解释性, 这是一种常用的策略。基于此, Zheng 等人^[7]在文献[6]的基础上, 用 L_1 范数正则化项代替 F 范数, 提出 SSMM 模型:

$$\min_{\mathbf{W}, b} \gamma \|\mathbf{W}\|_1 + \tau \|\mathbf{W}\|_* + \sum_{i=1}^n \{1 - y_i [\text{tr}(\mathbf{W}^{\top} \mathbf{X}_i) + b]\}_+, \quad (8)$$

其中: $\mathbf{W}$ 和 b 分别表示分类超平面的权重矩阵和偏置, γ, τ 分别是稀疏参数、低秩参数。模型以矩阵 $\mathbf{W}$ 的核范数和 L_1 范数作为正则项, 分别促进 $\mathbf{W}$ 的低秩和稀疏特性, 同时实现结构信息的提取和重要特征的选择, 以更好地对矩阵形式的数据进行分类。

式(8)是一个凸优化问题, 但合页损失、 L_1 范数和核范数的组合使得该问题难以求解。为了解决这一问题, 本文提出了基于 GFB^[19]算法框架的迭代算法, 求解过程如下。

首先, 将合页损失函数用广义合页损失函数 h_a 代替, 得到新的优化问题。

然后, 引入辅助变量 $\mathbf{Z}_1, \mathbf{Z}_2$, 利用 GFB 算法求解该问题, 即交替执行下列步骤:

步骤 1, 更新 $\mathbf{Z}_1, \mathbf{Z}_2$:

$$\begin{aligned} \mathbf{Z}_1 &\leftarrow \mathbf{Z}_1 + \lambda_i (\text{prox}_{\frac{\theta}{\omega_1} \gamma \|\cdot\|_1} (2\mathbf{W} - \mathbf{Z}_1 - \theta \nabla_{\mathbf{W}} h_a) - \mathbf{W}), \\ \mathbf{Z}_2 &\leftarrow \mathbf{Z}_2 + \lambda_i (\text{prox}_{\frac{\theta}{\omega_2} \tau \|\cdot\|_*} (2\mathbf{W} - \mathbf{Z}_2 - \theta \nabla_{\mathbf{W}} h_a) - \mathbf{W}); \end{aligned}$$

步骤 2, 更新 $\mathbf{W}, b$:

$$\begin{aligned} \mathbf{W} &\leftarrow \sum_{k=1}^2 \omega_k \mathbf{Z}_k, \\ b &\leftarrow b - \theta \nabla_b h_a. \end{aligned}$$

令 $\mathbf{Y} = 2\mathbf{W} - \mathbf{Z}_1 - \theta \nabla_{\mathbf{W}} h_a$, $\bar{\mathbf{Y}} = 2\mathbf{W} - \mathbf{Z}_2 - \theta \nabla_{\mathbf{W}} h_a$, 则:

$$\text{prox}_{\frac{\theta}{\omega_1} \gamma \|\cdot\|_1} (Y_{ij}) = \text{sign}(Y_{ij}) \cdot \max\left(|Y_{ij}| - \frac{\theta}{\omega_1} \gamma, 0\right),$$

其中: Y_{ij} 为矩阵 $\mathbf{Y}$ 的第 i 行第 j 列的元素; $\text{prox}_{\frac{\theta}{\omega_2} \tau \|\cdot\|_*} (\bar{\mathbf{Y}}) = D_{\frac{\theta}{\omega_2} \tau}(\bar{\mathbf{Y}})$, $D_{\frac{\theta}{\omega_2} \tau}$ 为奇异值阈值算子^[18]。由该算法求得式(8)的最优解 $\mathbf{W}^*$ 和 b^* , 则分类超平面为:

$$\text{tr}(\mathbf{W}^{*\top} \mathbf{X}_i) + b^* = 0;$$

对未知样本 $\tilde{\mathbf{X}}$, 预测标签 $\tilde{y}$ 可通过 $\tilde{y} = \text{sign}(\text{tr}(\mathbf{W}^{*\top} \tilde{\mathbf{X}}) + b^*)$ 确定。

2 LoRa-SSMM

在式(4)和式(8)中均采用凸正则化项(核范数)来促使系数矩阵 $\mathbf{W}$ 低秩。与 SMM 和 SSMM 不同, 本文考虑利用非凸正则化项 $R(\mathbf{W})$ 替代传统的核范数 $\|\mathbf{W}\|_*$, 以期获得更精确的低秩与稀疏近似。下文将详细介绍本文所提出的模型及求解算法。

2.1 模型

通过引入由非凸 Geman 函数定义的行分组稀疏正则化项 $R(\mathbf{W})$, 本文提出 LoRa-SSMM:

$$\min_{\mathbf{W}, b} F(\mathbf{W}, b) = \frac{1}{2} \text{tr}(\mathbf{W}^{\top} \mathbf{W}) + \tau R(\mathbf{W}) + C \sum_{i=1}^n \{1 - y_i [\text{tr}(\mathbf{W}^{\top} \mathbf{X}_i) + b]\}_+, \quad (9)$$

其中: $R(\mathbf{W}) = \sum_{t=1}^p \xi_t \rho(\|\mathbf{w}_t\|_2^2)$, $\mathbf{w}_t$ 为 $\mathbf{W}$ 的第 t 行, $\rho(|x|) = \frac{|x|}{\theta + |x|}$ ($\theta > 0$) 为非凸 Geman 函数^[20], τ, C 分别是稀疏参数和罚参数。并利用矩阵的 F 范数和行分组稀疏正则化项分别控制 $\mathbf{W}$ 的复杂度和促进 $\mathbf{W}$ 的稀疏性与低

秩性。 $R(W)$ 是由非凸 Geman 函数诱导而来,能够实现特征的组稀疏,保留矩阵内部的分组结构信息。特别地,当 $\rho(|x|) = |x|$, $\xi_t \equiv 1 (1 \leq t \leq p)$ 时, $R(W) = \|W\|_{2,1}$; 当 $R(W) = \|W\|_*$ 时,式(9)退化为 SMM。因此正则项 $R(W)$ 既可以提升 W 的稀疏度,还能有效地保持 W 的低秩性。

2.2 基于 MM 方法的迭代算法

非凸正则化项虽能有效提升模型的稀疏性并避免有偏估计,但对应的非凸优化问题通常具有较高的计算复杂度,求解比较困难。为此,根据文献[15],本文提出一种基于 MM 方法^[16]的改进优化算法。MM 方法是一类用于求解复杂优化问题的迭代框架,核心思想是通过在每次迭代中构造一个易于优化的近似函数来替代原始目标函数,并通过连续最小化该近似函数来实现对原目标函数的单调下降。该方法包含 2 个核心步骤:一是在当前迭代点构造目标函数的上界近似;二是通过最小化该上界函数来更新迭代点。考虑如下形式的优化问题:

$$\min_{x \in \mathcal{X}} f(x), \quad (10)$$

其中: $\mathcal{X} \subseteq \mathbf{R}^n$ 是非空闭集, $f: \mathcal{X} \rightarrow \mathbf{R}$ 为连续函数。算法 1 总结了求解该问题的 MM 方法的基本框架。

算法 1(MM 方法) 输入: 目标函数 $f(x)$, 初始点 $x^{(0)}$, 收敛阈值 $\epsilon > 0$ 。

过程:

1) 初始化: 设置 $k = 0$ 。

2) 循环:

a) 构造上界: 在当前点 $x^{(k)}$ 构造满足以下条件的近似函数 $g(x | x^{(k)})$:

$$\begin{aligned} g(x | x^{(k)}) &\geq f(x), \text{ 对所有 } x; \\ g(x^{(k)} | x^{(k)}) &= f(x^{(k)}). \end{aligned}$$

b) 最小化上界: 更新迭代点 $x^{(k+1)} = \arg \min_x g(x | x^{(k)})$ 。

c) 更新: $k \leftarrow k + 1$ 。

直到满足收敛条件 ($\|x^{(k)} - x^{(k-1)}\| < \epsilon$)。

输出: 近似最优解 $x^{(k)}$ 。

根据文献[17]有:

$$R(W) = R^+(W) + R^-(W),$$

其中: $R^+(W) = \sigma \sum_{t=1}^p \xi_t \|w_t\|_2^2 (\sigma = 1/\theta)$, $R^-(W) = R(W) - R^+(W)$ 。

将 $R^-(W)$ 在 $W^{(k)}$ 处一阶泰勒展开有:

$$\overline{R}^-(W) \simeq \text{tr}(G^{(k)T} W),$$

其中: $G^{(k)}$ 表示函数 $R^-(W)$ 在 $W^{(k)}$ 处的梯度, $g_t^{(k)}$ 表示 $G^{(k)}$ 的第 t 行, 且:

$$g_t^{(k)} \triangleq 2\xi_t [\rho'(\|w_t^{(k)}\|_2^2) - \sigma] w_t^{(k)}, \quad (11)$$

则 $R(W)$ 的近似函数为:

$$\overline{R}(W) \simeq R^+(W) + \text{tr}(G^{(k)T} W)。$$

故 $F(W, b)$ 在 $W^{(k)}$ 处的近似函数为:

$$\overline{F}(W, b) = \frac{1}{2} \text{tr}(W^T W) + \tau [R^+(W) + \text{tr}(G^{(k)T} W)] + C \sum_{i=1}^n \{1 - y_i [\text{tr}(W^T X_i) + b]\}_+, \quad (12)$$

从而将式(9)转化为迭代求解下列凸优化子问题:

$$\min_{W, b} \frac{1}{2} \text{tr}(W^T W) + \tau [R^+(W) + \text{tr}(G^{(k)T} W)] + C \sum_{i=1}^n \{1 - y_i [\text{tr}(W^T X_i) + b]\}_+, \quad (13)$$

式(13)可等价地写成带有线性约束的凸二次规划问题,并通过求该式的对偶问题来获得问题的最优解。

定理 1 对 $\tau, \sigma, \{\xi_t\}_{t=1}^p > 0$, 令 $\overline{D} = 2\sigma\tau D$, $D = \text{diag}\{\xi_1, \xi_2, \dots, \xi_p\}$ 为对角矩阵, 则式(13)的最优解为:

$$\begin{aligned} \hat{W} &= (I + \overline{D})^{-1} \left(\sum_{i=1}^n \alpha_i^* y_i X_i - \tau G^{(k)} \right), \\ \hat{b} &= \frac{1}{|\Omega|} \sum_{i \in \Omega} \{y_i - \text{tr}(\hat{W}^T X_i)\}. \end{aligned} \quad (14)$$

其中: $\alpha^* \in \mathbf{R}^n$ 是下列约束二次规划问题的解:

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \alpha^T K \alpha + q^T \alpha, \\ \text{s. t.} \quad & 0 \leq \alpha_i \leq C, \sum_{i=1}^n \alpha_i y_i = 0. \end{aligned}$$

其中: $K \in \mathbf{R}^{n \times n}$, $q \in \mathbf{R}^n$, K, q 中的元素 K_{ij}, q_i 分别定义为:

$$\begin{aligned} K_{ij} &= y_i y_j \operatorname{tr}[\mathbf{X}_i \mathbf{X}_j^T (\mathbf{I} + \overline{\mathbf{D}})^{-1}], \\ q_i &= 1 + \tau y_i \operatorname{tr}[\mathbf{X}_i \mathbf{G}^{(k)T} (\mathbf{I} + \overline{\mathbf{D}})^{-1}]. \end{aligned}$$

证明 引入松弛变量 η_i , 式(13)等价于:

$$\begin{aligned} \min_{\mathbf{W}, b} \quad & \frac{1}{2} \operatorname{tr}(\mathbf{W}^T \mathbf{W}) + \tau [R^+(\mathbf{W}) + \operatorname{tr}(\mathbf{G}^{(k)T} \mathbf{W})] + C \sum_{i=1}^n \eta_i, \\ \text{s. t.} \quad & 1 - y_i [\operatorname{tr}(\mathbf{W}^T \mathbf{X}_i) + b] \leq \eta_i, \eta_i \geq 0, \forall i. \end{aligned} \quad (15)$$

式(15)的拉格朗日函数为:

$$L = \frac{1}{2} \operatorname{tr}(\mathbf{W}^T \mathbf{W}) + \tau [R^+(\mathbf{W}) + \operatorname{tr}(\mathbf{G}^{(k)T} \mathbf{W})] + C \sum_{i=1}^n \eta_i - \sum_{i=1}^n \alpha_i \{y_i [\operatorname{tr}(\mathbf{W}^T \mathbf{X}_i) + b] - 1 + \eta_i\} - \sum_{i=1}^n \gamma_i \eta_i, \quad (16)$$

其中: α, γ 分别为拉格朗日乘子, 且 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$, $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)$ 。对式(16)分别关于 $\mathbf{W}, b, \eta$ 求偏导:

$$\frac{\partial L}{\partial \mathbf{W}} = 0, \frac{\partial L}{\partial b} = 0, \frac{\partial L}{\partial \eta} = 0,$$

有:

$$\begin{aligned} \mathbf{W} &= (\mathbf{I} + \overline{\mathbf{D}})^{-1} \left(\sum_{i=1}^n \alpha_i y_i \mathbf{X}_i - \tau \mathbf{G}^{(k)} \right), \\ \sum_{i=1}^n \alpha_i y_i &= 0, \\ C &= \alpha_i + \gamma_i. \end{aligned} \quad (17)$$

代入式(16)有:

$$\begin{aligned} L &= \frac{1}{2} \operatorname{tr}(\mathbf{W}^T \mathbf{W}) + \tau [R^+(\mathbf{W}) + \operatorname{tr}(\mathbf{G}^{(k)T} \mathbf{W})] + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i y_i \operatorname{tr}(\mathbf{W}^T \mathbf{X}_i) = \\ &= \frac{1}{2} \operatorname{tr}(\mathbf{W}^T \mathbf{W}) + \tau [\sigma \operatorname{tr}(\mathbf{D} \mathbf{W} \mathbf{W}^T) + \operatorname{tr}(\mathbf{G}^{(k)T} \mathbf{W})] + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i y_i \operatorname{tr}(\mathbf{W}^T \mathbf{X}_i) = \\ &= \frac{1}{2} \operatorname{tr}[(\mathbf{I} + \overline{\mathbf{D}}) \mathbf{W} \mathbf{W}^T] + \tau \operatorname{tr}(\mathbf{G}^{(k)T} \mathbf{W}) - \sum_{i=1}^n \alpha_i y_i \operatorname{tr}(\mathbf{W}^T \mathbf{X}_i) + \sum_{i=1}^n \alpha_i = \\ &= -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \operatorname{tr}[\mathbf{X}_i \mathbf{X}_j^T (\mathbf{I} + \overline{\mathbf{D}})^{-1}] \cdot \alpha_i \alpha_j + \sum_{i=1}^n \{1 + \tau y_i \operatorname{tr}[\mathbf{X}_i \mathbf{G}^{(k)T} (\mathbf{I} + \overline{\mathbf{D}})^{-1}]\} \cdot \alpha_i + \text{const.} = \\ &= -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n K_{ij} \alpha_i \alpha_j + \sum_{i=1}^n q_i \alpha_i + \text{const.} = -\frac{1}{2} \alpha^T K \alpha + q^T \alpha + \text{const.} \end{aligned}$$

其中: $\text{const.} = -\frac{1}{2} \tau^2 \operatorname{tr}[\mathbf{G}^{(k)} \mathbf{G}^{(k)T} (\mathbf{I} + \overline{\mathbf{D}})^{-1}]$, 为常数。

令:

$$K_{ij} = y_i y_j \operatorname{tr}[\mathbf{X}_i \mathbf{X}_j^T (\mathbf{I} + \overline{\mathbf{D}})^{-1}], q_i = 1 + \tau y_i \operatorname{tr}[\mathbf{X}_i \mathbf{G}^{(k)T} (\mathbf{I} + \overline{\mathbf{D}})^{-1}].$$

得到对偶问题:

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \alpha^T K \alpha + q^T \alpha, \\ \text{s. t.} \quad & 0 \leq \alpha_i \leq C, \sum_{i=1}^n \alpha_i y_i = 0. \end{aligned} \quad (18)$$

求解式(18)得到最优解 α^* 。将 α^* 代入式(17)可得:

$$\hat{\mathbf{W}} = (\mathbf{I} + \overline{\mathbf{D}})^{-1} \left(\sum_{i=1}^n \alpha_i^* y_i \mathbf{X}_i - \tau \mathbf{G}^{(k)} \right).$$

由 KKT 条件中的互补松弛条件,有:

$$\alpha_i^* \{y_i [\text{tr}(\hat{\mathbf{W}}^T \mathbf{X}_i) + b] - 1 + \eta_i\} = 0, \gamma_i \eta_i = 0.$$

则对 $0 < \alpha_i^* < C, \gamma_i > 0$ 有:

$$b = y_i - \text{tr}(\hat{\mathbf{W}}^T \mathbf{X}_i),$$

取平均有:

$$\hat{b} = \frac{1}{|\Omega|} \sum_{i \in \Omega} \{y_i - \text{tr}(\hat{\mathbf{W}}^T \mathbf{X}_i)\}.$$

其中: $\Omega = \{i: 0 < \alpha_i^* < C\}$, $|\Omega|$ 表示集合 Ω 中元素的数量。

证毕

基于上述讨论,给出求解式(9)的算法框架,即算法 2。

算法 2 (LoRa-SSMM) 步 0 初始化: $\mathbf{W}^{(0)} \in \mathbf{R}^{p \times q}, b^{(0)} \in \mathbf{R}$;

步 1 通过式(11)计算 $\mathbf{G}^{(k)}$;

步 2 求解对偶问题式(18)计算 α^* ;

步 3 计算 $\mathbf{W}^{(k+1)}, b^{(k+1)}$, 通过

$$\begin{aligned} \mathbf{W}^{(k)} &= (\mathbf{I} + \overline{\mathbf{D}})^{-1} \left(\sum_{i=1}^n \alpha_i^* y_i \mathbf{X}_i - \tau \mathbf{G}^{(k)} \right) \\ b^{(k)} &= \frac{1}{|\Omega|} \sum_{i \in \Omega} \{y_i - \text{tr}(\mathbf{W}^{(k)T} \mathbf{X}_i)\} \end{aligned}$$

步 4 若满足收敛条件,算法终止;否则令 $k \leftarrow k+1$, 转步 1。

算法 2 是基于经典 MM 方法的算法框架,全局收敛性结果已由文献[21-22]给出,具体见定理 2。

定理 2 假设 $\{(\mathbf{W}^{(k)}, b^{(k)})\}$ 是算法 2 产生的无穷点列,式(9)的目标函数 $F(\mathbf{W}, b)$ 是强制的且 MM 方法映射 M 是连续映射^[22], 则 $\lim_{k \rightarrow \infty} \|\mathbf{W}^{(k+1)} - \mathbf{W}^{(k)}\| = 0$ 且 $\lim_{k \rightarrow \infty} \|b^{(k+1)} - b^{(k)}\| = 0$ 。

注 在算法 2 中,更新 $\mathbf{G}^{(k)}$ 的计算复杂度 $O(pq)$ 。在第 2 步中,首先需计算 $\mathbf{K}$ 和 $\mathbf{q}$, 因为:

$$K_{ij} = y_i y_j \text{tr}[\mathbf{X}_i \mathbf{X}_j^T (\mathbf{I} + \overline{\mathbf{D}})^{-1}],$$

其中 $(\mathbf{I} + \overline{\mathbf{D}})^{-1}$ 为对角矩阵,所以只需计算 $\mathbf{X}_i \mathbf{X}_j^T$ 的主对角元,计算复杂度为 $O(pq)$,从而 $\mathbf{K}$ 更新的计算复杂度为 $O(n^2 pq)$, $\mathbf{q}$ 的计算复杂度为 $O(npq)$;其次,求解对偶变量 α^* 的计算复杂度为 $O(n^3)$ 。在第 3 步中,更新 $\mathbf{W}$ 的复杂度为 $O(npq)$,更新 b 的复杂度为 $O(pq |\Omega|)$, $|\Omega|$ 表示非支持向量的样本个数。假设 T 为算法 2 循环迭代次数,则算法 2 的计算复杂度为 $O(T(n^3 + n^2 pq))$ 。

3 数值实验

为了说明本文所提 LoRa-SSMM 的有效性,将 LoRa-SSMM 与 SVM、 L_1 -SVM、SMM、SSMM 这 4 种方法进行比较。实验过程中,使用 Matlab 统计与机器学习工具箱中内置的 fitcsvm 和 fitclinear 函数分别求解 SVM 和 L_1 -SVM 模型。所有实验均以 Matlab 2021b 为软件平台,以 Intel Core i5-8265U CPU @ 1.60 GHz 为处理器,内存 16.0 GB 的笔记本电脑为硬件平台。当连续 2 次迭代的值之差满足 $\|\mathbf{W}^{(k+1)} - \mathbf{W}^{(k)}\| \leq 10^{-4}$ 或迭代次数大于 500 时,迭代过程终止。当 $|\mathbf{W}_{ij}| \leq 10^{-6}$ 时,令 $\mathbf{W}_{ij} = 0$,去除冗余特征,即此时认为样本的第 (i, j) 个特征为冗余特征。所有方法的超参数均由五折交叉验证选取,并以准确率作为交叉验证的性能评价指标。准确率定义为正确分类样本数占总样本数的比例,记为 x_{Acc} ,计算公式如下:

$$x_{\text{Acc}} = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{TN}} + N_{\text{FP}} + N_{\text{FN}}},$$

其中: N_{TP} 、 N_{TN} 、 N_{FP} 和 N_{FN} 分别表示真正例、真负例、假正例与假负例。

针对 LoRa-SSMM,参数选择范围设置如下:

$$C \in \{0.001, 0.01, 0.1, 1\}, \tau \in \{0.1, 1, 10, 100, 1\,000\}, \theta \in \{0.005, 0.05\}, \xi_i (i=1, \dots, p) = 1;$$

其余对比方法中, SMM 的参数设置参考文献[13]的相关设置,即:

$$C \in \{0.01, 0.1, 1, 10, 100\}, \tau \in \{0.001, 0.01, 0.1, 1, 10\};$$

SSMM 的正则化参数设置为:

$$\gamma \in \{0.001, 0.01, 0.1, 1\}, \tau \in \{0.001, 0.01, 0.1, 1, 10\};$$

SVM 和 SSVM 的正则化参数 C, λ 均取自集合 $\tau \in \{10^{-6}, 10^{-5}, 10^{-4}, \dots, 10^4, 10^5\}$ 。

实验中共采用 3 个可天然表示为矩阵结构的数据集: INRIA Person、Caltech faces 以及 Caltech 101。为确保实验的公平性与可复现性,对于各对比方法,本文均采用随机抽样策略,将每个数据集的 70% 划分为训练集,剩余 30% 划分为测试集。各数据集的详细描述如下。

Caltech faces 数据集由加州理工学院收集,包含 26 位受试者的 450 张彩色人脸图像。原始图像分辨率为 896×592 像素,为减少计算复杂度,将原始图像统一采样至 448×296 像素。所有样本按性别属性被划分为 2 类(男性为 277 张,女性为 173 张)。图 1 给出了该数据集的部分样本示例。

INRIA Person 数据集是行人检测领域的基准数据集,本实验使用其中的正负样本进行二分类任务。该数据集包含 902 张行人图像(正样本)与 1 212 张背景图像(负样本)。所有图像均已统一处理为 64×128 像素的灰度图。图 2 给出了该数据集的部分样本示例。



图 1 Caltech faces 性别分类数据集样本示例

Fig. 1 Sample examples of Caltech faces gender classification dataset



图 2 INRIA Person 数据集中的行人和背景图像样本示例

Fig. 2 Pedestrian and background samples from the INRIA Person dataset

Caltech 101 数据集包含 101 个类别的物体图像,由李飞飞等人于 2003 年创建。数据集中各类别图像数量介于 30~800 张之间(多数约为 50 张)。为验证所提算法在图像二分类任务上的泛化能力,从该数据集中选取了 6 组具有较高语义相关性且在视觉特征上具有一定可比性的类别对,构建了 6 个二分类评估子集(例如“飞机”与“摩托车”同属交通工具,且在视觉上均具有流线型轮廓;“火烈鸟”与“鸽子”同属鸟类,在形态结构上存在共性),从而构成更具挑战性的分类任务。最终选定的 6 组类别对分别为:(Wheelchair, Windsor chair)、(Cup, Ewer)、(Flamingo, Pigeon)、(Emu, Ibis)、(Bass, Dolphin)以及(Airplanes, Motorbikes),将每一对中的前者标记为正类,后者标记为负类,所有图像均已统一处理为 100×100 像素的灰度图。各类别对的原始样本量汇总于表 1,详细的数据划分情况见表 2。

为全面评估算法性能,在每个数据集上进行 5 次独立实验,每次实验均重新划分训练集与测试集。最终报告的各项性能指标为 5 次实验结果的算术平均值,以确保结果的统计可靠性。采用以下 4 项核心指标:准确率,用于衡量模型的整体分类精度;稀疏度,定义为模型参数中零元素的比例,该值越高表明特征选择能力越强;模型有效秩,用于衡量模型对矩阵结构信息的利用能力,值越低表明低秩约束效果越好,向量型分类器的秩通过将得到的权重向量重塑为矩阵进而计算有效秩;训练时间,即算法完成 1 次训练所消耗的时间,用于评估计算效

率。基于上述指标,表 3 详细列出了各对比方法在所有数据集上的平均性能。

表 1 从 Caltech 101 数据集中选取的类别对

Tab. 1 Category pairs selected from the Caltech 101 dataset

数据集	数据维度/像素	样本数/个	数据集	数据维度/像素	样本数/个
Airplanes	100×100	800	Wheelchair	100×100	59
Motorbikes	100×100	798	Windsor chair	100×100	56
Flamingo	100×100	67	Bass	100×100	54
Pigeon	100×100	45	Dolphin	100×100	65
Cup	100×100	57	Emu	100×100	53
Ewer	100×100	85	Ibis	100×100	80

表 2 实验中使用的数据集

Tab. 2 The dataset used in the experiment

数据集	数据维度/像素	训练集(正样本/负样本)	测试集(正样本/负样本)	样本数/个
Caltech faces	448×296	315(197/118)	135(80/55)	450
INRIA Person	64×128	1 526(614/912)	588(288/300)	2 114
Wheelchair&Windsor chair	100×100	80(39/41)	35(17/18)	115
Cup&Ewer	100×100	98(59/39)	44(26/18)	142
Flamingo&Pigeon	100×100	77(46/31)	35(14/21)	112
Emu&Ibis	100×100	93(56/37)	40(24/16)	133
Bass&Dolphin	100×100	76(45/31)	37(20/17)	113
Airplanes&Motorbikes	100×100	1 118(558/560)	480(240/240)	1 598

实验结果表明,LoRa-SSMM 与其他方法相比,具有一定优势。

在分类准确率方面,LoRa-SSMM 在 8 个数据集集中的 6 个数据集上取得了最高准确率,而其他方法最多在 2 个数据集上达到最优。具体而言,在 Caltech faces、INRIA Person、Cup&Ewer、Emu&Ibis、Bass&Dolphin 和 Airplane&Motorbike 数据集上,LoRa-SSMM 均取得了最高的分类准确率;在 Wheelchair&Windsor chair 和 Flamingo&Pigeon 数据集上,LoRa-SSMM 虽分别以 0.58% 和 0.57% 的差距略低于 SSMM,但仍明显优于其他对比方法。从平均准确率来看,LoRa-SSMM 达到 76.43%,明显高于 SVM 的 68.38%、 L_1 -SVM 的 58.70%、SMM 的 67.25% 和 SSMM 的 73.77%,充分验证了本文方法在矩阵数据分类任务上的有效性。

在模型稀疏度和有效秩方面,LoRa-SSMM 在多数数据集上保持了合理的稀疏水平(如 Caltech faces: 43.11%, INRIA Person: 19.96%),既避免了过拟合,又保留了足够的模型表达能力。相比之下,SVM 和 SSMM 在某些数据集上稀疏度接近 0,存在过拟合风险;而 L_1 -SVM 在部分数据集上稀疏度过高(如 Airplane&Motorbike: 100%),虽然具有较强的特征选择能力,但可能损失有价值的判别信息,以至于分类准确率普遍偏低。在模型的有效秩方面,LoRa-SSMM 在多个数据集上取得了较低的有效秩,证明能更有效地捕捉矩阵数据的内部结构。

在训练时间方面,LoRa-SSMM 在保持优异性能的同时,在多数情况下展现了较高的计算效率。在 Wheelchair&Windsor chair、Cup&Ewer、Flamingo&Pigeon、Bass&Dolphin 和 Airplane&Motorbikes 数据集上,LoRa-SSMM 的训练时间明显低于其他矩阵型方法,例如 SMM、SSMM,甚至还优于部分向量型方法。特别是在 Caltech faces 数据集上,LoRa-SSMM 的训练时间仅需 1.75 s,而 SSMM 耗时则超过 10 min,因此,LoRa-SSMM 展现出较强的效率优势。图 3、图 4 展示了 SSMM、SMM 以及 LoRa-SSMM 在 Emu&Ibis 和 Cup&Ewer 数据集上的收敛性能对比。以图 3 为例,图 3a 为归一化目标函数值的收敛轨迹,函数数值已归一化至 $[0,1]$ 区间以便于比较;图 3b 量化了各算法达到预定收敛阈值所需的迭代次数,其中收敛阈值设定为归一化目标函数值的

10%和5%;图3c展示了LoRa-SSMM和SMM算法在前100次迭代中的收敛细节。收敛曲线(图3a)显示,LoRa-SSMM具有最快的收敛特性,SSMM收敛得最慢。量化分析(图3b)表明,达到10%目标函数值时,LoRa-SSMM、SMM和SSMM分别需要5、9和361次迭代;达到5%目标函数值时,则分别需要8、15和427次迭代。

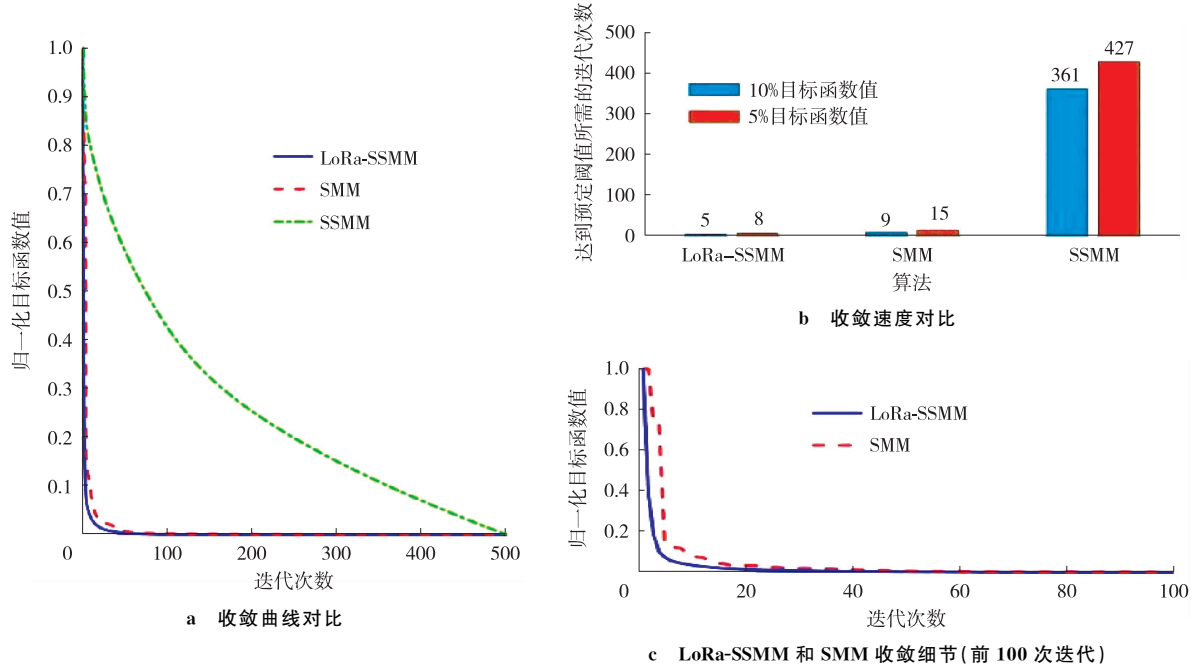
表 3 实验结果

Tab. 3 Experimental results

数据集	参数	SVM	L_1 -SVM	SMM	SSMM	LoRa-SSMM
Caltech faces	准确率	(57.04±0.00)%	(59.26±0.00)%	(62.22±11.82)%		(74.96±3.25)%
	稀疏度	100.00%	0.00%	26.53%		43.11%
	有效秩	28	296	281		296
	时间/s	1.36	1.01	56.33	>10 min	1.75
INRIA Person	准确率	(80.41±0.19)%	(79.76±0.70)%	(67.79±15.47)%	(80.20±0.19)%	(80.61±0.24)%
	稀疏度	0.00%	61.39%	0.01%	0.13%	19.96%
	有效秩	64	58	64	33	52
	时间/s	6.78	9.13	10.90	195.63	125.64
Wheelchair& Windsor chair	准确率	(74.29±0.00)%	(50.86±1.28)%	(66.28±2.39)%	(83.44±1.27)%	(82.86±0.00)%
	稀疏度	5.24%	91.08%	4.36%	1.02%	10.49%
	有效秩	97	18	80	98	95
	时间/s	0.05	0.11	1.59	7.78	0.03
Cup&Ewer	准确率	(57.27±3.73)%	(58.64±2.49)%	(60.91±4.37)%	(55.46±1.24)%	(63.18±1.02)%
	稀疏度	0.03%	78.61%	0.05%	4.90%	0.08%
	有效秩	100	35	100	98	99
	时间/s	0.20	0.18	2.70	8.69	0.05
Flamingo&Pigeon	准确率	(57.14±0.00)%	(60.00±0.00)%	(64.00±4.78)%	(72.57±1.57)%	(72.00±5.11)%
	稀疏度	3.99%	37.99%	4.69%	7.03%	10.19%
	有效秩	99	72	98	100	97
	时间/s	0.08	0.03	2.19	7.59	0.05
Emu&Ibis	准确率	(57.50±0.00)%	(57.00±2.09)%	(59.00±1.37)%	(60.50±1.12)%	(64.50±2.74)%
	稀疏度	0.02%	19.16%	0.01%	11.71%	0.01%
	有效秩	100	90	98	71	100
	时间/s	0.10	0.46	2.15	9.41	2.59
Bass&Dolphin	准确率	(67.17±0.89)%	(54.05±0.00)%	(61.08±2.41)%	(68.11±6.45)%	(75.68±0.00)%
	稀疏度	3.89%	4.97%	3.06%	0.41%	10.05%
	有效秩	98	94	76	99	95
	时间/s	0.04	0.03	2.06	8.88	0.03
Airplane& Motorbike	准确率	(96.25±0.00)%	(50.00±0.00)%	(96.71±0.88)%	(96.08±0.35)%	(97.63±0.12)%
	稀疏度	3.32%	100.00%	2.82%	0.00%	4.28%
	有效秩	98	0	73	100	98
	时间/s	1.46	0.25	21.44	128.00	0.82
平均准确率		68.38%	58.70%	67.25%	73.77%	76.43%

注:加粗的数据表示“最高平均准确率±标准差”。

总体而言,LoRa-SSMM 在分类精度、模型复杂度和计算效率之间实现了良好平衡。与现有矩阵分类器相比,它不仅实现了更高的平均分类准确率,而且具备更合理的模型复杂度和更有竞争力的训练速度,展现出优异的综合性能。



注:图 3b 中 10%目标函数值和 5%目标函数值分别指归一化目标函数值的 10%和 5%。下同。

图 3 Emu&Ibis 数据集上的算法收敛性能综合分析

Fig. 3 Comprehensive analysis of algorithm convergence performance on Emu&Ibis dataset

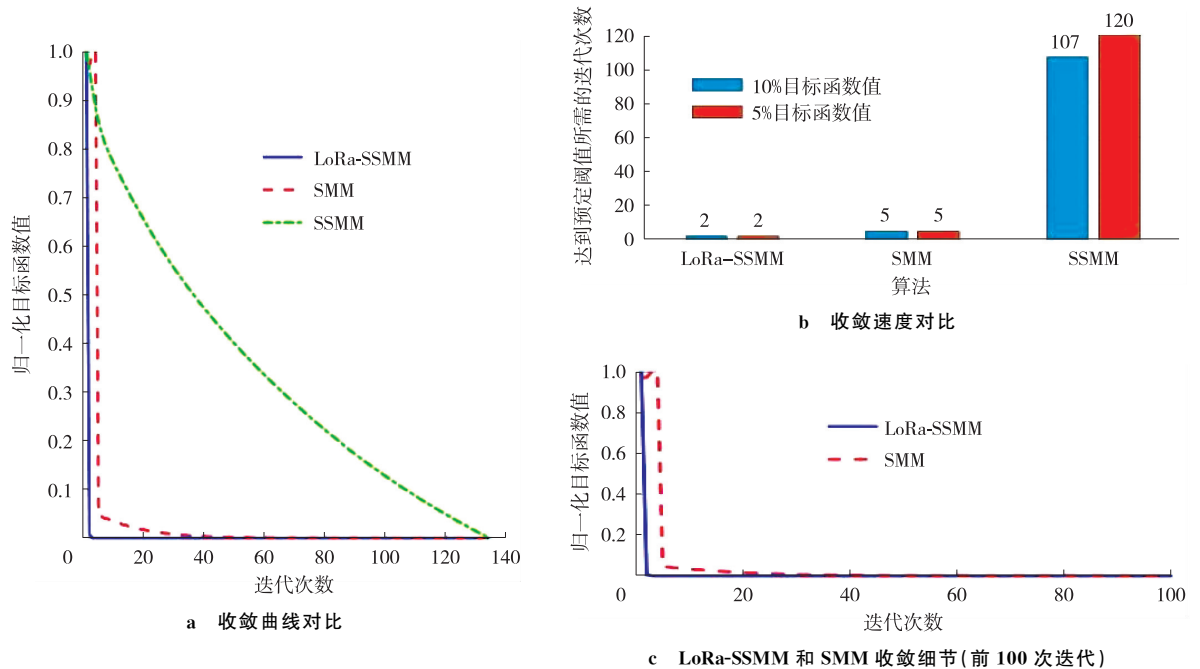


图 4 Cup & Ewer 数据集上的算法收敛性能综合分析

Fig. 4 Comprehensive analysis of algorithm convergence performance on Cup&Ewer dataset

4 结论

本文针对矩阵分类问题,提出了 LoRa-SSMM 及对应的高效求解算法。数值实验表明,通过以非凸正则项

替代经典 SMM 中的核范数,新模型能更有效地实现特征选择与低秩约束,从而实现对矩阵先验结构信息更充分的利用。所设计的基于 MM 方法的求解算法,通过将原问题转化为一系列对偶子问题求解,使得每次迭代仅需处理 1 个子问题,明显降低了计算复杂度,在效率上优于基准方法。数值实验表明,本文所提方法在矩阵结构信息的利用以及特征选择、泛化性能等方面具有明显优势。但本文所提算法中仍涉及求解对偶二次规划问题,计算效率有待进一步提升。

参考文献:

- [1] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20: 273-297.
- [2] Wolf L, Jhuang H, Hazan T. Modeling appearances with low-rank SVM[C]//2007 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2007: 1-6.
- [3] Cai D, He X F, Wen J R, et al. Support tensor machines for text categorization[J]. International Journal of Academic Research in Business & Social Science, 2006, 2 (12): 2222-6990.
- [4] Pirsiavash H, Ramanan D, Fowlkes C C. Bilinear classifiers for visual recognition[C]//Advances in Neural Information Processing Systems. 2009: 1482-1490.
- [5] Kobayashi T, Otsu N. Efficient optimization for low-rank integrated bilinear classifiers[C]//Fitzgibbon A, Lazebnik S, Perona P, et al. Computer vision-ECCV 2012. Berlin: Springer, 2012: 474-487.
- [6] Luo L, Xie Y B, Zhang Z H, et al. Support matrix machines[C]//Proceedings of the 32nd International Conference on Machine Learning (ICML 2015). 2015: 938-947.
- [7] Zheng Q Q, Zhu F Y, Qin J, et al. Sparse support matrix machine[J]. Pattern Recognition, 2018, 76: 715-726.
- [8] Li X, Yang Y, Shao H D, et al. Symplectic weighted sparse support matrix machine for gear fault diagnosis[J]. Measurement, 2021, 168: 108392.
- [9] Gu M G, Zheng J D, Pan H Y, et al. Ramp sparse support matrix machine and its application in roller bearing fault diagnosis[J]. Applied Soft Computing, 2021, 113: 107928.
- [10] Wang M, Xu H F, Pan H Y, et al. Sparse norm matrix machine and its application in roller bearing fault diagnosis[J]. Measurement Science and Technology, 2022, 33 (11): 115114.
- [11] Li Y X, Wang D Q, Liu F. The auto-correlation function aided sparse support matrix machine for EEG-based fatigue detection[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2023, 70 (2): 836-840.
- [12] Qin Z W, Goldfarb D. Structured sparsity *via* alternating direction methods[J]. Journal of Machine Learning Research, 2012, 13 (48): 1435-1468.
- [13] Zheng Q Q, Zhu F Y, Heng P A. Robust support matrix machine for single trial EEG classification[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2018, 26 (3): 551-562.
- [14] Razzak I, Bouadjenek M R, Saris R A, et al. Support matrix machine *via* joint $\ell_{2,1}$ and nuclear norm minimization under matrix completion framework for classification of corrupted data[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35 (11): 16341-16352.
- [15] Zhao Z P, Palomar D P. Sparse reduced rank regression with nonconvex regularization[C]//2018 IEEE Statistical Signal Processing Workshop (SSP). Piscataway: IEEE, 2018: 811-815.
- [16] Sun Y, Babu P, Palomar D P. Majorization-minimization algorithms in signal processing, communications, and machine learning[J]. IEEE Transactions on Signal Processing, 2017, 65 (3): 794-816.
- [17] Yao Q M, Kwok J T. Efficient learning with a family of nonconvex regularizers by redistributing nonconvexity[J]. Journal of Machine Learning Research, 2018, 18 (1): 1-52.
- [18] Cai J F, Candès E J, Shen Z W. A singular value thresholding algorithm for matrix completion[J]. SIAM Journal on Optimization, 2010, 20 (4): 1956-1982.
- [19] Raguet H, Fadili J, Peyré G. A generalized forward-backward splitting[J]. SIAM Journal on Imaging Sciences, 2013, 6 (3): 1199-1226.
- [20] Geman D, Reynolds G. Constrained restoration and the recovery of discontinuities[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, 14 (3): 367-383.
- [21] Lange K. Numerical analysis for statisticians[M]. New York: Springer, 2010.
- [22] Lange K, Won J H, Landeros A, et al. Nonconvex optimization *via* MM algorithms: convergence theory[PP/OL]. arXiv (2021-06-05) [2026-01-02]. <https://doi.org/10.48550/arXiv.2106.02805>.

Operations Research and Cybernetics

A Low-Rank and Sparse Support Matrix Machine Framework Based on the Majorization-Minimization Method

WENG Ting, LI Guoquan

(School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China)

Abstract: Support matrix machine (SMM) is a type of classification model that directly processes matrix-type data and serves as a generalization of support vector machine. Sparse support matrix machine achieves feature selection by using the L_1 norm of the matrix instead of the F norm. However, studies have shown that nonconvex regularization terms generally yield more accurate approximations, thereby enhancing the model's feature selection capability. To this end, this paper proposes a novel low-rank and sparse support matrix machine model incorporating a nonconvex regularization term. Built upon SMM, this approach replaces the nuclear norm of the matrix with a row-wise group sparsity regularizer induced by a nonconvex Geman function. This achieves sparsity while simultaneously fostering low-rank properties in the matrix, leading to improved classification performance. An improved optimization algorithm, built upon the Majorization-Minimization method, is proposed to solve this nonconvex optimization problem. Numerical experiments demonstrate that the proposed algorithm is more efficient than benchmark algorithms, and the introduction of a nonconvex regularizer can achieve better classification performance.

Keywords: SMM; Geman function; group sparsity; Majorization-Minimization method

(责任编辑 黄 颖)