

直觉模糊局部保持投影最小二乘双支持向量聚类

王顺霞¹, 黄成泉^{2*}, 蔡江海¹, 杨贵燕¹, 罗森艳¹, 周丽华¹

(1. 贵州民族大学数据科学与信息工程学院, 贵州 贵阳 550025; 2. 贵州民族大学工程技术人才实践训练中心, 贵州 贵阳 550025)

摘要:针对数据样本的局部结构信息未得到充分利用以及算法对噪声的敏感特性导致聚类效果下降的问题,提出直觉模糊局部保持投影最小二乘双支持向量聚类方法。基于样本与质心的距离以及样本的异质性分配模糊得分,给样本赋予权重,充分利用训练样本的局部几何结构信息,提供样本邻域的先验信息,不仅降低了噪声和异常值对算法性能的影响,而且有效解决了数据的聚类问题。在多个数据集上进行实验,并通过统计分析检验所提算法的显著性,实验结果证实了所提算法比其它现有算法具有更好的鲁棒性能和聚类性能。

关键词:双支持向量;聚类;局部保持投影;直觉模糊;噪声

中图分类号:TP181 **文献标志码:**A

引用格式:王顺霞,黄成泉,蔡江海,等.直觉模糊局部保持投影最小二乘双支持向量聚类[J].山东大学学报(理学版),2026,61(3):124-134.

Intuitionistic fuzzy locality preserving projection least squares twin support vector clustering

WANG Shunxia¹, HUANG Chengquan^{2*}, CAI Jianghai¹, YANG Guiyan¹, LUO Senyan¹, ZHOU Lihua¹

(1. School of Data Science and Information Engineering, Guizhou Minzu University, Guiyang 550025, Guizhou, China;
2. Engineering Training Center, Guizhou Minzu University, Guiyang 550025, Guizhou, China)

Abstract: To solve the problem that the local structure information of data samples is not fully utilized and the sensitivity of the algorithm to noise leads to the decline of the clustering effect, this paper proposes the intuitionistic fuzzy local preserving projection least square twin support vector clustering method. Fuzzy scores are assigned based on the distance between samples and centroids and the heterogeneity of the samples, given weights to the samples, and the local geometric structure information of the training sample is fully utilized to provide prior information about the sample neighborhood, which not only reduces the influence of noise and outliers on the performance of the algorithm, but also effectively solves the clustering problem of data. Experiments are conducted on several datasets, and the significance of the proposed algorithm is verified by statistical analysis. These experimental results demonstrate that the proposed algorithm has better robustness and clustering performance than other existing algorithms.

Key words: twin support vector; clustering; locality preserving projection; intuitionistic fuzzy; noise

0 引言

聚类分析旨在根据数据的相似性将数据点分配到不同的类别中,广泛应用于许多研究领域^[1]。然而,现实生活中的数据集往往包含噪声或离群值,这些因素可能会降低算法的学习效率和聚类结果的准确性^[2]。

Wang 等^[3]将孪生支持向量机^[4](twin support vector machine, TWSVM)扩展到聚类问题,首次提出孪生

收稿日期:2024-04-10; 网络出版时间:2025-01-27

基金项目:国家自然科学基金资助项目(62062024);贵州省模式识别与智能系统重点实验室2022年度开放课题资助项目(GZMUKL[2022]KF03);贵州省省级科技计划资助项目(黔科合基础-ZK[2021]一般342);贵州省教育厅自然科学研究资助项目(黔教技[2022]015)

第一作者:王顺霞(1999—),女,硕士研究生,研究方向为机器学习与模式识别。E-mail:2689826749@qq.com

*通信作者:黄成泉(1976—),男,教授,硕士生导师,博士,研究方向为模式识别、机器学习与图像处理等。E-mail:hcq863@163.com

支持向量聚类 (twin support vector clustering, TWSVC) 算法。该算法通过求解一系列二次规划问题 (quadratic programming problem, QPP) 来确定 k 个聚类平面,但由于算法存在计算耗时、聚类精度不高等缺点,难以高效地解决现实中的许多问题。Khemchandani 等^[5]将 TWSVC 中的不等式约束替换为等式约束,提出最小二乘孪生支持向量聚类 (least squares TWSVC, LSTWSVC),通过求解一组线性方程而不是 QPP,缩短了算法的计算时间。Bai 等^[6]提出最小二乘孪生有界支持向量聚类 (least square twin bounded SVC, LSTBSVC),通过引入正则化项,使用等式约束条件,加快算法的学习速度。Tanveer 等^[7]通过在 TWSVC 中引入 Pinball 损失函数,提出一种基于 Pinball 损失的孪生支持向量机聚类算法 (pinball loss twin SVC, pinTSVC),该算法在包含噪声的数据集上表现出良好的泛化性能。近年来,一种基于投影轴的聚类算法展示出优秀的聚类性能。例如,Richhariya 等^[8]提出最小二乘投影双支持向量聚类 (least squares projection twin SVC, LSPTSVC)。该方法通过求解一组线性方程,减少计算时间,提高聚类准确率,但是由于该算法未能充分利用数据的局部几何结构信息,并且在解决数据样本中存在噪声和异常值的聚类问题时并不具有突出优势。在此基础上,Tanveer 等^[9]利用数据点之间的局部邻域信息,同时使用隶属度和非隶属度权重来降低噪声和异常值的影响,提出一种直觉模糊加权最小二乘双支持向量机 (intuitionistic fuzzy weighted least squares twin SVMs, IFW-LSTSVM)。Chen 等^[10]通过将局部保持投影的基本思想引入 LSPTSVC 中,提出一种局部保持投影最小二乘双支持向量机 (locality preserving projection least squares twin SVM, LPPLSTSVM)。

受 IFW-LSTSVM 和 LPPLSTSVM 的启发,本文提出一种直觉模糊局部保持投影最小二乘双支持向量聚类 (intuitionistic fuzzy locality preserving projection least squares twin support vector clustering, IFLPPLSTSVC) 算法。该算法将局部保持投影和直觉模糊数引入 LSPTSVC 中,提升算法的聚类性能和鲁棒性能,解决噪声数据的聚类问题,并通过使用凹凸过程 (concave-convex procedure, CCCP)^[11]求解一组线性方程而不是 TWSVC 算法中的 1 对 QPP 来解决算法目标函数的优化问题。将所提算法与 4 种经典算法在 11 个基准数据集上进行实验,实验结果证实了 IFLPPLSTSVC 算法的有效性。

1 理论基础

本章简要讨论 TWSVC 和 LSPTSVC 算法。样本数量用 m 表示,每个样本中的特征数量用 n 表示。这 m 个数据样本用矩阵形式表示为 $X_i = (x_1, x_2, \dots, x_m)^T$,其中 $i = 1, 2, \dots, N$, N 是集群的总数。 $X_i \in \mathbf{R}^{m \times n}$ 表示第 i 类中的所有样本, $\bar{X}_i \in \mathbf{R}^{m \times n}$ 表示不属于第 i 类的所有样本。

1.1 TWSVC 算法

线性 TWSVC 的优化问题表示为

$$\begin{aligned} \min_{\omega_i^{j+1}, b_i^{j+1}, \xi_i^{j+1}} \frac{1}{2} \| X_i \omega_i^{j+1} + b_i^{j+1} e \|^2 + c_1 e^T \xi_i^{j+1} \quad (1) \\ \text{s.t. } T(|\bar{X}_i \omega_i^{j+1} + b_i^{j+1} e|) \geq e - \xi_i^{j+1}, \quad \xi_i^{j+1} \geq 0, \end{aligned}$$

其中, $c_1 > 0$ 是惩罚参数, e 为分量全为 1 的适当维度的向量, ξ_i^{j+1} 是松弛向量, $j = 0, 1, 2, \dots, T(\cdot)$ 表示一级泰勒展开式。

通过使用 ω_i^j 和 b_i^j 的 $|\bar{X}_i \omega_i^j + b_i^j e|$ 次梯度和泰勒级数展开式,TWSVC 优化问题可以写为

$$\begin{aligned} \min_{\omega_i^{j+1}, b_i^{j+1}, \xi_i^{j+1}} \frac{1}{2} \| X_i \omega_i^{j+1} + b_i^{j+1} e \|^2 + c_1 e^T \xi_i^{j+1} \quad (2) \\ \text{s.t. } \text{diag}(\text{sign}(|\bar{X}_i \omega_i^j + b_i^j e|)) (|\bar{X}_i \omega_i^{j+1} + b_i^{j+1} e|) \geq e - \xi_i^{j+1}, \quad \xi_i^{j+1} \geq 0. \end{aligned}$$

将式(2)转化为其对偶问题为

$$\begin{aligned} \min_{\alpha} \frac{1}{2} \alpha^T F (H^T H)^{-1} F^T \alpha - e^T \alpha \quad (3) \\ \text{s.t. } 0 \leq \alpha \leq c_1 e, \end{aligned}$$

其中, $F = \text{diag}(\text{sign}(\bar{X}_i \omega_i^j + b_i^j e)) [\bar{X}_i \quad e]$, $H = [X_i \quad e]$, 并且 $\alpha \in \mathbf{R}^{m \times m}$ 是拉格朗日乘子向量。

每个聚类超平面表示为

$$[\boldsymbol{\omega}_i^{j+1}; b_i^{j+1}]^T = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{F}^T \boldsymbol{\alpha}. \quad (4)$$

1.2 LSPTSVC 算法

线性 LSPTSVC 算法的优化问题表示为

$$\min_{\boldsymbol{\omega}_i^{j+1}} \frac{1}{2} (\boldsymbol{\omega}_i^{j+1})^T \mathbf{S}_i \boldsymbol{\omega}_i^{j+1} + \frac{c_1}{2} \sum_{q=1}^{m_i} (\boldsymbol{\xi}_{iq}^{j+1})^2 + \frac{c_2}{2} \|\boldsymbol{\omega}_i^{j+1}\|^2 \quad (5)$$

$$\text{s.t. } T(|\bar{X}_i \boldsymbol{\omega}_i^{j+1} - \frac{1}{m_i} \bar{\mathbf{e}}_i \mathbf{e}_i^T X_i \boldsymbol{\omega}_i^{j+1}|) + \boldsymbol{\xi}_{iq}^{j+1} = \bar{\mathbf{e}}_i,$$

其中, $c_1, c_2 > 0$ 是惩罚参数, $\boldsymbol{\omega}_i^{j+1}$ 表示第 $j+1$ 次迭代的权重向量. \mathbf{e}_i 和 $\bar{\mathbf{e}}_i$ 分别表示维度为 p 和 q 的分量全为 1 的向量. 矩阵 \mathbf{S}_i 可以表示为

$$\mathbf{S}_i = \sum_{p=1}^{m_i} \left(\mathbf{x}_p^{(i)} - \frac{1}{m_i} \sum_{p=1}^{m_i} \mathbf{x}_p^{(i)} \right) \left(\mathbf{x}_p^{(i)} - \frac{1}{m_i} \sum_{p=1}^{m_i} \mathbf{x}_p^{(i)} \right)^T. \quad (6)$$

通过使用 $\boldsymbol{\omega}_i^j$ 的 $|\bar{X}_i \boldsymbol{\omega}_i^j - \frac{1}{m_i} \bar{\mathbf{e}}_i \mathbf{e}_i^T X_i \boldsymbol{\omega}_i^j|$ 次梯度和泰勒级数展开式, 可得

$$L = \frac{1}{2} (\boldsymbol{\omega}_i^{j+1})^T \mathbf{S}_i \boldsymbol{\omega}_i^{j+1} + \frac{c_2}{2} \|\boldsymbol{\omega}_i^{j+1}\|^2 + \frac{c_1}{2} \left\| -\text{diag} \left(\text{sign} \left(\bar{X}_i \boldsymbol{\omega}_i^j - \frac{1}{m_i} \bar{\mathbf{e}}_i \mathbf{e}_i^T X_i \boldsymbol{\omega}_i^j \right) \right) \left(\bar{X}_i \boldsymbol{\omega}_i^{j+1} - \frac{1}{m_i} \bar{\mathbf{e}}_i \mathbf{e}_i^T X_i \boldsymbol{\omega}_i^{j+1} \right) + \bar{\mathbf{e}}_i \right\|^2. \quad (7)$$

求解式(7)关于 $\boldsymbol{\omega}_i^{j+1}$ 梯度并令其为 0, 可得

$$\mathbf{S}_i \boldsymbol{\omega}_i^{j+1} + c_2 \boldsymbol{\omega}_i^{j+1} + c_1 \mathbf{D}_i^T (\mathbf{D}_i \boldsymbol{\omega}_i^{j+1} - \bar{\mathbf{e}}_i) = 0, \quad (8)$$

其中 $\mathbf{D}_i = \text{diag} \left(\text{sign} \left(\bar{X}_i \boldsymbol{\omega}_i^j - \frac{1}{m_i} \bar{\mathbf{e}}_i \mathbf{e}_i^T X_i \boldsymbol{\omega}_i^j \right) \right) \left(\bar{X}_i - \frac{1}{m_i} \bar{\mathbf{e}}_i \mathbf{e}_i^T X_i \right)$.

求解式(8)得

$$\boldsymbol{\omega}_i^{j+1} = \left(\mathbf{D}_i^T \mathbf{D}_i + \frac{c_2}{c_1} \mathbf{I}_i + \frac{\mathbf{S}_i}{c_1} \right)^{-1} \mathbf{D}_i^T \bar{\mathbf{e}}_i. \quad (9)$$

2 本文所提算法 IFLPPLSTSVC

本章分别描述直觉模糊集^[12]和局部保持类内散射矩阵^[13]的基本概念, 并讨论所提算法 IFLPPLSTSVC 在线性和非线性情况下的算法推导过程及相关分析.

2.1 直觉模糊集

考虑一个非空集 χ , 一个模糊集在 χ 中定义为

$$\mathbf{Z} = \{ (z, \mu_{\mathbf{Z}}(z)) \mid z \in \chi \}, \quad (10)$$

其中, $\mu_{\mathbf{Z}}: \chi \rightarrow [0, 1]$ 且 $\mu_{\mathbf{Z}}(z)$ 是 $z \in \chi$ 的隶属度. 由于每个样本的隶属度的模糊集不同, 因此直觉模糊集同时考虑了每个样本的隶属度和非隶属度. 在数学上, 一个直觉模糊集^[9]表示为

$$\bar{\mathbf{Z}} = \{ (z, \mu_{\bar{\mathbf{Z}}}(z), \nu_{\bar{\mathbf{Z}}}(z)) \mid z \in \chi \}, \quad (11)$$

其中, $\mu_{\bar{\mathbf{Z}}}(z)$ 和 $\nu_{\bar{\mathbf{Z}}}(z)$ 分别是 $z \in \chi$ 的隶属度和非隶属度, $\mu_{\bar{\mathbf{Z}}}: \chi \rightarrow [0, 1]$, $\nu_{\bar{\mathbf{Z}}}: \chi \rightarrow [0, 1]$, 且 $0 \leq \mu_{\bar{\mathbf{Z}}}(z) + \nu_{\bar{\mathbf{Z}}}(z) \leq 1$. $z \in \chi$ 的犹豫度为

$$\pi_{\bar{\mathbf{Z}}}(z) = 1 - \mu_{\bar{\mathbf{Z}}}(z) - \nu_{\bar{\mathbf{Z}}}(z). \quad (12)$$

直觉模糊数(intuitionistic fuzzy number, IFN)定义为 $\theta = (\mu_{\theta}, \nu_{\theta})$, 其中 $\mu_{\theta} \in [0, 1]$, $\nu_{\theta} \in [0, 1]$, 且 $0 \leq \mu_{\theta} + \nu_{\theta} \leq 1$, $\theta^- = [0, 1]$ 和 $\theta^+ = [1, 0]$ 分别是最小和最大的 IFN. 对于给定的 $\theta = (\mu_{\theta}, \nu_{\theta})$, IFN 的得分值表示为

$$S(\theta) = \mu_{\theta} - \nu_{\theta}, \quad (13)$$

对于一些 IFN 来说, 分数函数难以确定. 为了避免这个问题, 使用函数

$$h(\theta) = \mu_{\theta} + \nu_{\theta}. \quad (14)$$

通过式(12)和式(14)可得

$$h(\theta) + \pi(\theta) = 1, \tag{15}$$

如果 $S(\theta_1) = S(\theta_2)$ 且 $h(\theta_1) < h(\theta_2)$, $\theta_1 < \theta_2$ 。使用式(13),得分函数

$$H(\theta) = \frac{1 - v(\theta)}{2 - \mu(\theta) - v(\theta)}. \tag{16}$$

隶属度和非隶属度之间的关系表示为

$$S(\theta_1) < S(\theta_2) \Rightarrow H(\theta_1) < H(\theta_2), \tag{17}$$

$$S(\theta_1) = S(\theta_2), h(\theta_1) = h(\theta_2) \Rightarrow H(\theta_1) = H(\theta_2). \tag{18}$$

2.2 局部保持类内散射矩阵

定义 1 设 L 是数据集 X 的拉普拉斯矩阵,矩阵 $Z = X^T L X = X^T (D - W) X$ 称为局部保持散射矩阵。

定义 2 假设数据集 X 被分成 i 个不同的类,矩阵 $Z_w = \sum_{i=1}^N Z_i$ 称为局部保持类内散射矩阵, Z_i 是第 i 类 X_i 的局部保持散射矩阵,即 $Z_i = X_i^T (D^i - W^i) X_i$,其中 D^i 是对角矩阵, $D^i_{kk} = \sum_j W^i_{kj}$, W^i 是第 i 类 X_i 的权重矩阵,定义为

$$W^i_{ij} = \begin{cases} \exp\left(\frac{-\|x_{ik} - x_{ij}\|^2}{\mu}\right), & x_{ik} \in N_i(x_{ij}) \text{ 或 } x_{ij} \in N_i(x_{ik}) \\ 0, & \text{其他,} \end{cases} \tag{19}$$

其中, x_{ij} 指第 i 类 ($i=1, 2, \dots, N$) 中的第 i 个样本点, μ 是内核参数。

局部保持类内散射矩阵 Z_w 是对称且半正定的,并且形式上类似于类内散射矩阵 S_w ,反映了数据的内在几何形状和局部结构。

2.3 线性情况

线性 IFLPPLSTSVC 算法的优化问题为

$$\min_{\omega_i} \frac{1}{2} \omega_i^T Z_i \omega_i + \frac{c_1}{2} \|\omega_i\|^2 + \frac{c_2}{2} \|\bar{S}_i \xi_i\|^2 \tag{20}$$

$$\text{s.t. } T\left(\left|\bar{X}_i \omega_i - \frac{1}{m_i} \bar{e}_i e_i^T X_i \omega_i\right|\right) + \xi_i = \bar{e}_i,$$

其中, $c_1, c_2 > 0$ 是惩罚参数, ξ_i 表示松弛向量, \bar{e}_i 表示不属于第 i 类数据样本且分量全为 1 的适当维度的向量, $T(\cdot)$ 表示一级泰勒展开式, Z_i 表示局部保持类内散射矩阵, \bar{S}_i 表示不属于第 i 类数据样本的模糊数得分值。使用 CCCP 求解式(20)的优化问题,可以写为

$$\min_{\omega_i} \frac{1}{2} (\omega_i^{j+1})^T Z_i \omega_i^{j+1} + \frac{c_1}{2} \|\omega_i^{j+1}\|^2 + \frac{c_2}{2} \|\bar{S}_i \xi_i^{j+1}\|^2 \tag{21}$$

$$\text{s.t. } T\left(\left|\bar{X}_i \omega_i^{j+1} - \frac{1}{m_i} \bar{e}_i e_i^T X_i \omega_i^{j+1}\right|\right) + \xi_i^{j+1} = \bar{e}_i.$$

通过引入 $|\bar{X}_i \omega_i^j - \frac{1}{m_i} \bar{e}_i e_i^T X_i \omega_i^j|$ 关于 ω_i^j 的次梯度,可得

$$T\left(\bar{X}_i \omega_i^{j+1} - \frac{1}{m_i} \bar{e}_i e_i^T X_i \omega_i^{j+1}\right) = \text{diag}\left(\text{sign}\left(\bar{X}_i \omega_i^j - \frac{1}{m_i} \bar{e}_i e_i^T X_i \omega_i^j\right)\right) \left(\bar{X}_i \omega_i^{j+1} - \frac{1}{m_i} \bar{e}_i e_i^T X_i \omega_i^{j+1}\right). \tag{22}$$

将式(22)代入式(21)的约束条件中,并将约束条件代入目标函数,可以写为

$$L = \frac{1}{2} (\omega_i^{j+1})^T Z_i \omega_i^{j+1} + \frac{c_1}{2} \|\omega_i^{j+1}\|^2 + \frac{c_2}{2} \left\| \bar{S}_i \left(\bar{e}_i - \text{diag}\left(\text{sign}\left(\bar{X}_i \omega_i^j - \frac{1}{m_i} \bar{e}_i e_i^T X_i \omega_i^j\right)\right) \left(\bar{X}_i \omega_i^{j+1} - \frac{1}{m_i} \bar{e}_i e_i^T X_i \omega_i^{j+1}\right) \right) \right\|^2. \tag{23}$$

求解式(23)关于 ω_i^{j+1} 梯度并令其为 0,可得

$$Z_i \omega_i^{j+1} + c_1 \omega_i^{j+1} + c_2 G_i^T \bar{S}_i^T (\bar{S}_i (G_i \omega_i^{j+1} - \bar{e}_i)) = 0, \tag{24}$$

其中 $G_i = \text{diag}\left(\text{sign}\left(\bar{X}_i \omega_i^j - \frac{1}{m_i} \bar{e}_i e_i^T X_i \omega_i^j\right)\right) \left(\bar{X}_i - \frac{1}{m_i} \bar{e}_i e_i^T X_i\right)$ 。

对式(24)求解 ω_i^{j+1} 得到最终聚类平面为

$$\omega_i^{j+1} = \left(\mathbf{G}_i^T \bar{\mathbf{S}}_i^T \mathbf{S}_i \mathbf{G}_i + \frac{c_1}{c_2} \mathbf{I}_i + \frac{\mathbf{Z}_i}{c_2} \right)^{-1} \mathbf{G}_i^T \bar{\mathbf{S}}_i^T \mathbf{S}_i \bar{\mathbf{e}}_i. \quad (25)$$

连续迭代直到满足终止条件,获得最终解,对于测试样本 \mathbf{x}_t ,根据式(26)进行聚类:

$$y(\mathbf{x}_t) = \arg \min_{i=1,2,\dots,N} \left| \omega_i^T \mathbf{x}_t - \frac{1}{m_i} \mathbf{e}_i^T \mathbf{X}_i \omega_i \right|. \quad (26)$$

综上,线性 IFLPPLSTSVC 算法步骤如下。

算法 1 线性 IFLPPLSTSVC。

输入 数据 $\mathbf{X} \in \mathbf{R}^n$, 惩罚参数 c_1, c_2 。

输出 聚类标签 \mathbf{Y} 和投影权重向量 ω_i 。

(1) 初始化标签:根据最近邻图(nearest neighbor graph, NNG)进行样本标签初始化,并计算聚类 i 的初始权重向量 ω_i^0 。

(2) 直觉模糊得分值与局部保持类内散射矩阵的计算:

① 根据式(16)计算第 i 类的直觉模糊得分值;

② 根据式(19)计算第 i 类的局部保持类内散射矩阵 $\mathbf{Z}_i = \mathbf{X}_i^T (\mathbf{D}^i - \mathbf{W}^i) \mathbf{X}_i$, 其中 \mathbf{D}^i 是对角矩阵, $\mathbf{D}_{kk}^i = \sum_j \mathbf{W}_{kj}^i$, \mathbf{W}_i 是第 i 类样本 \mathbf{X}_i 的权重矩阵。

(3) CCCP 过程:对于每个聚类 i ,当 $j=0$ 时使用初始权重向量 ω_i^0 和式(25)计算 ω_i^{j+1} ,如果 $\|\omega_i^{j+1} - \omega_i^j\|$ 小于容差值 t_{ol} ,则 $j=j+1$,进行(3),否则进行(4)。

(4) 分配聚类标签:将新标签分配给每个聚类 i 的数据点,初始 $k=0$,使用式(26)计算 \mathbf{Y}_{k+1} ,如果 $(\|\mathbf{Y}_{k+1} - \mathbf{Y}_k\| \neq 0)$,则 $k=k+1$,转向(3),否则,停止迭代,输出聚类标签 \mathbf{Y} 和投影权重向量 ω_i 。

2.4 非线性情况

通过核函数将线性 IFLPPLSTSVC 算法扩展到非线性情况,得到非线性 IFLPPLSTSVC 算法的优化问题为

$$\begin{aligned} \min_{\omega_i} & \frac{1}{2} \omega_i^T \tilde{\mathbf{Z}}_i \omega_i + \frac{c_1}{2} \|\omega_i\|^2 + \frac{c_2}{2} \|\bar{\mathbf{S}}_i \xi_i\|^2 \\ \text{s.t.} & \quad T \left(\left| \mathbf{K}(\bar{\mathbf{X}}_i, \mathbf{C}^T) \omega_i - \frac{1}{m_i} \bar{\mathbf{e}}_i \mathbf{e}_i^T \mathbf{K}(\mathbf{X}_i, \mathbf{C}^T) \omega_i \right| \right) + \xi_i = \bar{\mathbf{e}}_i, \end{aligned} \quad (27)$$

其中, $\mathbf{K}(\cdot, \mathbf{C}^T)$ 是内核函数, $\tilde{\mathbf{Z}}_i$ 表示引入核函数后的局部保持散射矩阵,并且 $\mathbf{C} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N]$, 与线性情况类似,使用 CCCP 求解式(27),可得

$$\begin{aligned} \min_{\omega_i} & \frac{1}{2} (\omega_i^{j+1})^T \tilde{\mathbf{Z}}_i \omega_i^{j+1} + \frac{c_1}{2} \|\omega_i^{j+1}\|^2 + \frac{c_2}{2} \|\bar{\mathbf{S}}_i \xi_i^{j+1}\|^2 \\ \text{s.t.} & \quad T \left(\left| \mathbf{K}(\bar{\mathbf{X}}_i, \mathbf{C}^T) \omega_i^{j+1} - \frac{1}{m_i} \bar{\mathbf{e}}_i \mathbf{e}_i^T \mathbf{K}(\mathbf{X}_i, \mathbf{C}^T) \omega_i^{j+1} \right| \right) + \xi_i^{j+1} = \bar{\mathbf{e}}_i. \end{aligned} \quad (28)$$

引入 $\left| \mathbf{K}(\bar{\mathbf{X}}_i, \mathbf{C}^T) \omega_i^j - \frac{1}{m_i} \bar{\mathbf{e}}_i \mathbf{e}_i^T \mathbf{K}(\mathbf{X}_i, \mathbf{C}^T) \omega_i^j \right|$ 关于 ω_i^j 的次梯度,并将约束条件代入目标函数可以写为

$$\begin{aligned} L = & \frac{1}{2} (\omega_i^{j+1})^T \tilde{\mathbf{Z}}_i \omega_i^{j+1} + \frac{c_1}{2} \|\omega_i^{j+1}\|^2 \\ & + \frac{c_2}{2} \left\| \bar{\mathbf{S}}_i \left(\bar{\mathbf{e}}_i - \text{diag} \left(\text{sign} \left(\mathbf{K}(\bar{\mathbf{X}}_i, \mathbf{C}^T) \omega_i^j - \frac{1}{m_i} \bar{\mathbf{e}}_i \mathbf{e}_i^T \mathbf{K}(\mathbf{X}_i, \mathbf{C}^T) \omega_i^j \right) \right) \left(\mathbf{K}(\bar{\mathbf{X}}_i, \mathbf{C}^T) \omega_i^{j+1} - \frac{1}{m_i} \bar{\mathbf{e}}_i \mathbf{e}_i^T \mathbf{K}(\mathbf{X}_i, \mathbf{C}^T) \omega_i^{j+1} \right) \right) \right\|^2. \end{aligned} \quad (29)$$

求解式(29)关于 ω_i^{j+1} 的梯度并令其为0,可得

$$\tilde{\mathbf{Z}}_i \omega_i^{j+1} + c_1 \omega_i^{j+1} + c_2 \mathbf{U}_i^T \bar{\mathbf{S}}_i^T (\bar{\mathbf{S}}_i (\mathbf{U}_i \omega_i^{j+1} - \bar{\mathbf{e}}_i)) = 0, \quad (30)$$

其中 $\mathbf{U}_i = \text{diag} \left(\text{sign} \left(\mathbf{K}(\bar{\mathbf{X}}_i, \mathbf{C}^T) \omega_i^j - \frac{1}{m_i} \bar{\mathbf{e}}_i \mathbf{e}_i^T \mathbf{K}(\mathbf{X}_i, \mathbf{C}^T) \omega_i^j \right) \right) \left((\bar{\mathbf{X}}_i, \mathbf{C}^T) - \frac{1}{m_i} \bar{\mathbf{e}}_i \mathbf{e}_i^T \mathbf{K}(\mathbf{X}_i, \mathbf{C}^T) \right)$ 。

对式(30)求解 ω_i^{j+1} ,得到最终聚类平面为

$$\omega_i^{j+1} = \left(U_i^T S_i^T S_i U_i + \frac{c_1}{c_2} I_i + \frac{\tilde{Z}_i}{c_2} \right)^{-1} U_i^T S_i^T S_i \tilde{e}_i. \quad (31)$$

连续迭代直到满足终止条件,获得最终解。对于测试样本 x_i ,根据式(32)进行聚类,可得

$$y(x_i) = \arg \min_{i=1,2,\dots,N} \left| \omega_i^T K(x_i, C^T) - \frac{1}{m_i} e_i^T K(X_i, C^T) \omega_i \right|. \quad (32)$$

综上,非线性 IFLPPLSTSVC 算法步骤如下。

算法 2 非线性 IFLPPLSTSVC。

输入 数据 $X \in \mathbf{R}^n$, 惩罚参数 c_1, c_2 和内核参数 μ 。

输出 聚类标签 Y 和投影权重向量 ω_i 。

(1) 初始化标签:根据 NNG 进行样本标签初始化,并计算聚类 i 的初始权重向量 ω_i^0 。

(2) 直觉模糊得分值与局部保持类内散射矩阵的计算:

① 根据式(16)计算第 i 类的直觉模糊得分值;

② 根据式(19)计算第 i 类的局部保持类内散射矩阵 $\tilde{Z}_i = K(X_i^T, C^T) (D^i - W^i) K(X_i, C^T)$, 其中 D^i 是对角矩阵, $D_{kk}^i = \sum_j W_{kj}^i$, W_i 是第 i 类 X_i 的权重矩阵。

(3) CCCP 过程:对于每个聚类 i ,当 $j=0$ 时使用初始权重向量 ω_i^0 计算式(31)计算 ω_i^{j+1} ,如果 $\|\omega_i^{j+1} - \omega_i^j\|$ 小于容差值 t_{ol} ,则 $j=j+1$,继续进行(3),否则进行(4)。

(4) 分配聚类标签:将新标签分配给每个聚类 i 的样本,初始 $k=0$,使用式(32)计算 Y_{k+1} ,如果 $(\|Y_{k+1} - Y_k\| \neq 0)$,则 $k=k+1$,转向(3),否则,停止迭代,输出聚类标签 Y 和投影权重向量 ω_i 。

2.5 时间复杂度

聚类算法的计算时间主要用于解决优化问题,本节使用 big-O 表示法分析所提 IFLPPLSTSVC 算法优化问题的时间复杂度^[14]。设 $m = m_i + \bar{m}_i$ 是数据样本的总数, n 是矩阵的维数, N 是聚类总数。IFLPPLSTSVC 算法的时间复杂度主要由以下 3 个部分组成:

(1) 构造局部保持类内散射矩阵 Z_i 。用于查找所有数据点的 k 最近邻的时间复杂度为 $O(N(m_i^2 \log(m_i) + \bar{m}_i^2 \log(\bar{m}_i)))$ 。

(2) 隶属度的计算。在计算隶属度的过程中包括类中心、类半径、每个样本与类中心的距离以及每个样本的隶属度的计算,时间复杂度为 $O(N) + O(N) + O(Nm) + O(Nm)$, 而计算非隶属度的时间复杂度为 $O(Nm) + O(Nm)$ 。因此,计算样本得分函数的总时间复杂度为 $O(N) + O(N) + O(Nm) + O(Nm) + O(Nm)$, 而当 m 扩展到无穷大时,样本得分函数计算的总复杂度为 $O(Nm)$ 。

(3) 求解 2 个线性方程组。线性情况下的时间复杂度为 $O(Nn^3)$,非线性情况下的时间复杂度为 $O(N(m_i^3 + \bar{m}_i^3))$ 。

3 实验结果及分析

为验证本文所提算法 IFLPPLSTSVC 的聚类性能以及对噪声的鲁棒性能,共使用 11 个基准数据集进行实验,并与 TWSVC、LSTBSVC、LSPTSVC 和 pinTSVC 算法的实验结果进行比较分析。分别通过添加不同标准差的零均值高斯噪声来引入不同的特征噪声,特征噪声的标准差 σ 分别取 0(无噪声)、0.05 和 0.10。实验使用数据集的详细信息如表 1 所示。

3.1 实验设置

本文的实验过程在 64 位 Windows10 操作系统, Intel® Core™ i5-4590 CPU@ 3.30 GHz 处理器, 8GB 内存, MATLAB R2020a 环境的 PC 上进

表 1 实验数据集的详细信息

Table 1 Details of the experimental datasets			
数据集	样本数	属性数	类别数
Ds2c2sc13	588	2	13
Spherical_5_2	250	2	5
Elliptical_10_2	500	2	10
2d-4c-no4	863	2	4
2d-4c-no9	876	2	4
Longsquare	900	2	6
R15	600	2	15
Aggregation	788	2	7
Haberman	306	3	2
Seeds	210	7	3
Glass	214	9	6

行。本文所有非线性算法都使用高斯核函数 $K(x, y) = \exp(-\|x - y\|^2 / (2\mu^2))$ 。总样本的 50% 用于训练, 其余用于测试。所提算法的惩罚参数 c_1, c_2 从集合 $\{10^{-5}, 10^{-4}, \dots, 10^5\}$ 中选择, 内核参数 μ 从集合 $\{2^{-5}, 2^{-4}, \dots, 2^5\}$ 中选择。对于所有算法, CCCP 过程的 $t_{ol} = 0.001$ 。

使用如下相似性矩阵^[15]来度量具有聚类标签 y 的 m 个数据样本的聚类精度

$$P(i, j) = \begin{cases} 1, & y_i = y_j, \\ 0, & \text{其他。} \end{cases} \quad (33)$$

设 P_t 是预测聚类标签的相似性矩阵, P_a 是实际标签的相似矩阵, n_0 是 P_a 和 P_t 中值为 0 的元素个数, n_1 是 P_a 和 P_t 中值为 1 的元素个数。将聚类精度的计算定义为

$$\text{accuracy} = \frac{n_0 + n_1 - m}{m^2 - m} \times 100\%。 \quad (34)$$

3.2 噪声实验结果与分析

非线性算法 IFLPPLSTSVC 与其他方法(即 TWSVC、LSTBSVC、LSPTSVC 和 pinTSVC)在无噪声(标准差 σ 为 0)和有噪声(σ 分别为 0.05 和 0.10)情况下的实验结果比较如表 2 所示, 表中加粗字体表示对应算法的最佳准确率。

表 2 非线性算法在不同数据集上的聚类准确率
Table 2 Clustering accuracy of nonlinear algorithms on different datasets

单位: %

数据集	TWSVC			LSTBSVC			LSPTSVC			pinTSVC			IFLPPLSTSVC		
	$\sigma=0$	$\sigma=0.05$	$\sigma=0.10$	$\sigma=0$	$\sigma=0.05$	$\sigma=0.10$	$\sigma=0$	$\sigma=0.05$	$\sigma=0.10$	$\sigma=0$	$\sigma=0.05$	$\sigma=0.10$	$\sigma=0$	$\sigma=0.05$	$\sigma=0.10$
Ds2c2sc13	93.28	83.69	83.91	95.64	83.72	83.11	92.18	84.03	84.42	98.34	84.40	84.25	94.32	85.01	84.46
Spherical_5_2	94.99	89.48	80.05	85.32	82.98	83.17	95.92	88.31	85.50	95.85	90.76	87.00	97.95	88.57	86.37
Elliptical_10_2	93.94	90.85	88.16	95.42	93.47	95.22	97.82	98.31	96.20	99.05	98.06	96.66	99.64	98.59	97.69
2d-4c-no4	58.14	56.24	51.08	63.52	87.31	85.01	89.91	89.97	84.22	99.69	99.15	99.00	99.79	99.15	98.73
2d-4c-no9	96.61	95.99	94.85	97.72	95.22	94.14	90.00	94.04	90.27	99.53	98.89	97.63	98.35	98.04	97.75
Longsquare	77.14	73.84	83.39	81.02	80.88	82.61	95.36	94.57	93.81	99.06	98.15	97.22	99.70	98.84	97.94
R15	99.51	94.29	84.47	99.00	93.88	92.39	96.44	88.21	88.20	99.16	94.65	92.32	99.87	95.38	92.97
Aggregation	79.88	84.89	81.89	75.56	78.86	73.84	71.39	71.03	71.10	99.40	97.73	97.47	99.81	99.52	94.18
Haberman	61.76	62.42	63.78	63.09	63.78	63.78	64.48	63.78	63.09	62.93	60.54	61.26	64.48	63.78	63.78
Seeds	74.87	76.63	76.92	86.15	84.69	83.24	86.37	83.77	83.92	92.24	84.60	80.95	86.37	82.12	83.17
Glass	64.72	61.12	64.08	68.91	65.02	65.97	68.51	67.50	65.21	75.40	67.65	68.08	66.92	68.52	65.95

由表 2 可以看到, 在 $\sigma=0, \sigma=0.05, \sigma=0.10$ 情况下, 所提算法在大多数数据集上的聚类精度优于其他算法。例如, 对于 Elliptical_10_2 数据集, 在标准差 $\sigma=0$ 的无噪声情况下, 所提算法的聚类准确率为 99.64%, TWSVC、LSTBSVC、LSPTSVC 和 pinTSVC 的聚类准确率分别为 93.94%、95.42%、97.82%、99.05%。在标准差 $\sigma=0.05$ 的噪声情况下, 所提算法的聚类准确率为 98.59%, 而 TWSVC、LSTBSVC、LSPTSVC 和 pinTSVC 分别为 90.85%、93.47%、98.31%、98.06%。在标准差 $\sigma=0.10$ 的噪声情况下, 所提算法的聚类准确率为 97.69%, 而 TWSVC、LSTBSVC、LSPTSVC 和 pinTSVC 分别为 88.16%、95.22%、96.20%、96.66%。此外, 在 11 个实验数据集中, 非线性 IFLPPLSTSVC 算法在标准差 σ 为 0 的无噪声情况下以及 σ 分别为 0.05 和 0.10 的噪声情况下, 分别在 7、8 和 6 个数据集上取得最高的聚类精度。由表 2 中可以看到非线性 IFLPPLSTSVC 算法在有无噪声的数据集上均获得良好的聚类性能。

非线性算法在 $\sigma=0, \sigma=0.05, \sigma=0.10$ 情况下在不同基准数据集上的平均学习时间如图 1 所示。IFLPPLSTSVC 算法的平均学习时间明显比 TWSVC 和 pinTSVC 算法短, 这主要是因为这 2 种聚类算法都需要通过求解 1 对 QPP 来获得聚类超平面, 而所提算法只需通过求解线性方程组获得投影轴对数据样本进行聚类。在大多数情况下与 LSTBSVC 算法的学习效率相当, 比 LSPTSVC 耗时稍短一些, 但是相差并不大, 这主要是因为所提算法需要耗费一定的时间计算数据样本点的直觉模糊得分值和数据样本之间的局部保持类内散射矩阵。

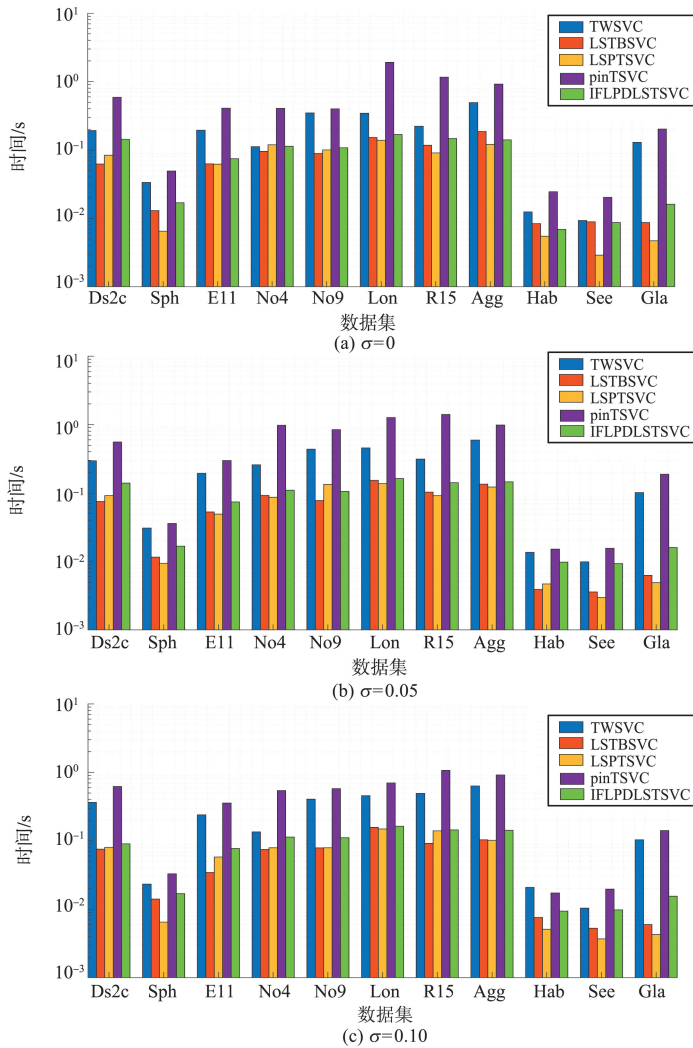


图1 非线性算法在不同数据集上的平均学习时间

Fig.1 Average learning time of the nonlinear algorithms on different datasets

Longsquare 数据集在不同噪声水平下的聚类如图 2 所示。图 2(a) 表示 Longsquare 数据集在 σ 为 0 的无噪声情况下的原始聚类图,图 2(b)、(c) 分别表示在 Longsquare 数据集中添加标准差 σ 为 0.05 和 0.10 高斯噪声的真实聚类图。

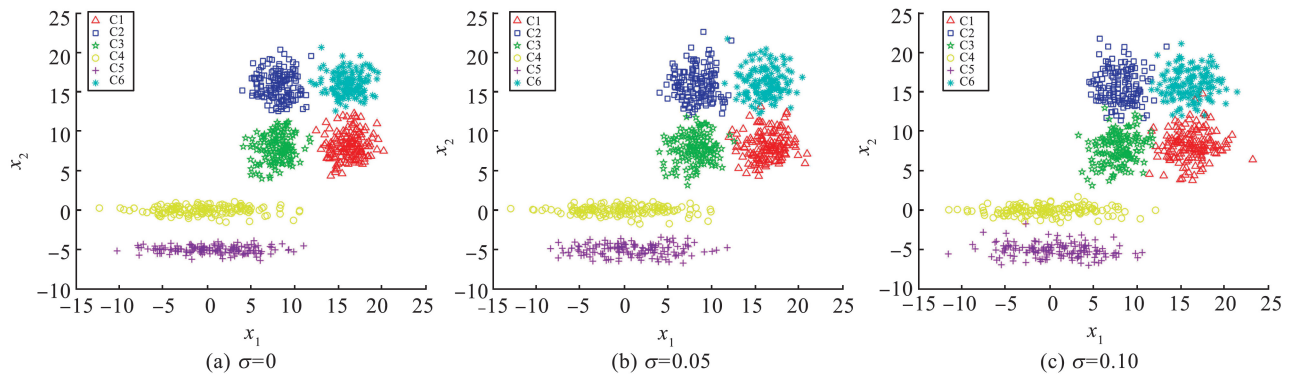


图2 Longsquare 数据集在不同噪声水平下的原始聚类

Fig.2 Primitive clustering of Longsquare dataset under different noise levels

非线性 IFLPLDSTSVC 算法在 Longsquare 数据集的不同噪声水平下的聚类效果如图 3 所示。图 3(a)、(b)和(c) 分别表示在标准差 σ 为 0 的无噪声情况下,以及在 Longsquare 数据集中分别添加标准差 σ 为 0.05 和 0.10 的高斯噪声后非线性 IFLPLDSTSVC 算法的聚类效果图。由图 3(a) 可以看到,在标准差 σ 为 0 的无

噪声情况下,所提算法对 Longsquare 数据集的聚类准确率达到 99.70%,聚类效果图与原始聚类几乎相同,仅仅只有几个数据点被误分。由图 3(b)、(c)可以看到,所提算法在标准差 σ 分别为 0.05 和 0.10 的噪声情况下对 Longsquare 数据集进行聚类的聚类准确率分别达到 98.84% 和 97.94%。综上所述,本文所提算法 IFLPPLSTSVC 对噪声具有良好的鲁棒性。

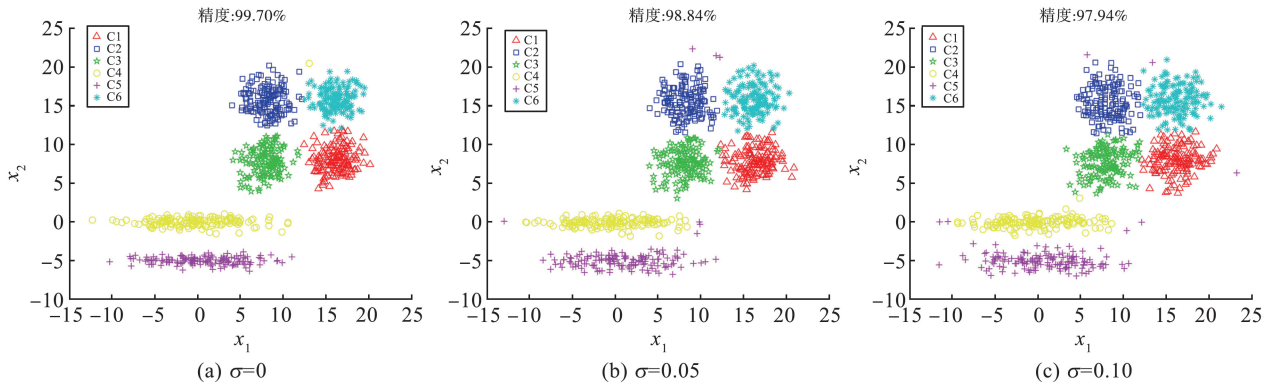


图 3 不同噪声水平下的 Longsquare 数据集在非线性 IFLPPLSTSVC 算法下的聚类效果图

Fig.3 Clustering graphs of nonlinear IFLPPLSTSVC algorithm under different noise levels of Longsquare dataset

3.3 统计分析

本文使用 Friedman 检验和 Nemenyi 检验^[16-17]来验证非线性 IFLPPLSTSVC 算法与非线性 TWSVC、LSTBSVC、LSPTSVC 和 pinTSVC 算法之间的统计显著性。首先根据聚类精度计算每个数据集上每种算法在不同噪声水平下的秩排名(表 3)。非线性 TWSVC、LSTBSVC、LSPTSVC、pinTSVC 算法和非线性 IFLPPLSTSVC 算法的平均秩分别为 4.181 8、3.515 2、3.454 5、2.136 4 和 1.712 1,非线性 IFLPPLSTSVC 算法具有最低平均秩。最初,假设这两种方法之间没有差异,并将其作为零假设。

表 3 非线性算法在不同数据集上的秩排名

Table 3 Rank ranking of nonlinear algorithms on different datasets

数据集	TWSVC			LSTBSVC			LSPTSVC			pinTSVC			IFLPPLSTSVC		
	$\sigma=0$	$\sigma=0.05$	$\sigma=0.10$	$\sigma=0$	$\sigma=0.05$	$\sigma=0.10$	$\sigma=0$	$\sigma=0.05$	$\sigma=0.10$	$\sigma=0$	$\sigma=0.05$	$\sigma=0.10$	$\sigma=0$	$\sigma=0.05$	$\sigma=0.10$
Ds2c2sc13	4.0	5.0	4.0	2.0	4.0	5.0	5.0	3.0	2.0	1.0	2.0	3.0	3.0	1.0	1.0
Spherical_5_2	4.0	2.0	5.0	5.0	5.0	4.0	2.0	4.0	3.0	3.0	1.0	1.0	1.0	3.0	2.0
Elliptical_10_2	5.0	5.0	5.0	4.0	4.0	4.0	3.0	2.0	3.0	2.0	3.0	2.0	1.0	1.0	1.0
2d-4c-no4	5.0	5.0	5.0	4.0	4.0	3.0	3.0	3.0	4.0	2.0	1.5	1.0	1.0	1.5	2.0
2d-4c-no9	4.0	3.0	3.0	3.0	4.0	4.0	5.0	5.0	5.0	1.0	1.0	2.0	2.0	2.0	1.0
Longsquare	5.0	5.0	4.0	4.0	4.0	5.0	3.0	3.0	3.0	2.0	2.0	2.0	1.0	1.0	1.0
R15	2.0	3.0	5.0	4.0	4.0	2.0	5.0	5.0	4.0	3.0	2.0	3.0	1.0	1.0	1.0
Aggregation	3.0	3.0	3.0	4.0	4.0	4.0	5.0	5.0	5.0	2.0	2.0	1.0	1.0	1.0	2.0
Haberman	5.0	4.0	2.0	3.0	2.0	2.0	1.5	2.0	4.0	4.0	5.0	5.0	1.5	2.0	2.0
Seeds	5.0	5.0	5.0	4.0	1.0	2.0	2.5	3.0	1.0	1.0	2.0	4.0	2.5	4.0	3.0
Glass	5.0	5.0	5.0	2.0	4.0	2.0	3.0	3.0	4.0	1.0	2.0	1.0	4.0	1.0	3.0

通过表 3 计算 Friedman 检验的 χ_F^2 值表示为

$$\chi_F^2 = \frac{12N}{v(v+1)} \left[\sum_{i=1}^v R_i^2 - \frac{v(v+1)^2}{4} \right], \tag{35}$$

$$F_F = \frac{(N-1)\chi_F^2}{N(v-1) - \chi_F^2}, \tag{36}$$

其中, N 是所使用的数据集的数量, v 表示算法的数量, R_i 是第 i 种算法在 N 个数据集上的平均秩。将对应值代入式(35)和(36)中,得到 $\chi_F^2 = 56.405 5$, $F_F = 23.877 1$ 。这里 F 分布自由度为 $(5-1, (5-1)(33-1)) = (4, 128)$ 。在 $\alpha = 0.05$ 的显著性水平下,临界值 $F(4, 128) = 2.442$, 因为 $F_F = 23.877 1 > 2.442$, 所以拒绝零假设。

为了检验所提算法和其他方法之间的成对统计差异,使用 Nemenyi 检验计算

$$CD = t_\alpha \sqrt{\frac{v(v+1)}{6N}}, \tag{37}$$

其中, t_α 是 α 显著性水平的临界值, CD 是 N 个数据集上的 ν 算法的临界差。对于 $\nu = 5, N = 33$, 得到 $t_\alpha = 2.728$, 将其代入式 (37), 得到 $CD = 1.0619$ 。如果 2 种算法的平均秩大于 CD 值, 则 2 种方法之间存在显著差异。非线性 IFLPPLSTSVC 算法与非线性 TWSVC、LSTBSVC、LSPTSVC 和 pinTSVC 算法之间的平均秩差分别为: $4.1818 - 1.7121 = 2.4697, 3.5152 - 1.7121 = 1.8030, 3.4545 - 1.7121 = 1.7424$ 和 $2.1364 - 1.7121 = 0.4242$ 。对于 $\alpha = 0.05$ 的显著性水平, 非线性 IFLPPLSTSVC 算法与非线性 TWSVC、LSTBSVC 和 LSPTSVC 算法之间的平均秩差均大于 CD 值, 因此, 非线性 IFLPPLSTSVC 算法与非线性 TWSVC、LSTBSVC 和 LSPTSVC 算法之间存在显著差异, 非线性 IFLPPLSTSVC 算法的性能优于非线性 TWSVC、LSTBSVC 和 LSPTSVC 算法; 由于非线性 IFLPPLSTSVC 算法与非线性 pinTSVC 算法之间的平均秩差均小于 CD 值, 所以非线性 IFLPPLSTSVC 算法与非线性 pinTSVC 算法之间的差异不够显著。

3.4 参数敏感性分析

为进一步分析参数对所提算法性能的影响, 本文对参数 c_1, c_2 和内核参数 μ 进行敏感性分析。图 4 显示在 Spherical_5_2 和 Aggregation 数据集上参数 $c_1 = c_2$ 和 μ 对非线性 IFLPPLSTSVC 算法性能的影响, 所提算法的性能对所选参数存在一定的敏感性。在图 4(a)、(b) 中, μ 的取值偏大时, 改变参数 c_1, c_2 的值, 所提算法能够获得更好的聚类准确率。以上分析均表明, 在非线性情况下, 核映射起着非常重要的作用, 内核参数 μ 决定核函数的值, 导致数据的非线性变换。因此, 为了获得更好的聚类性能, 需要仔细选择所提算法的参数。

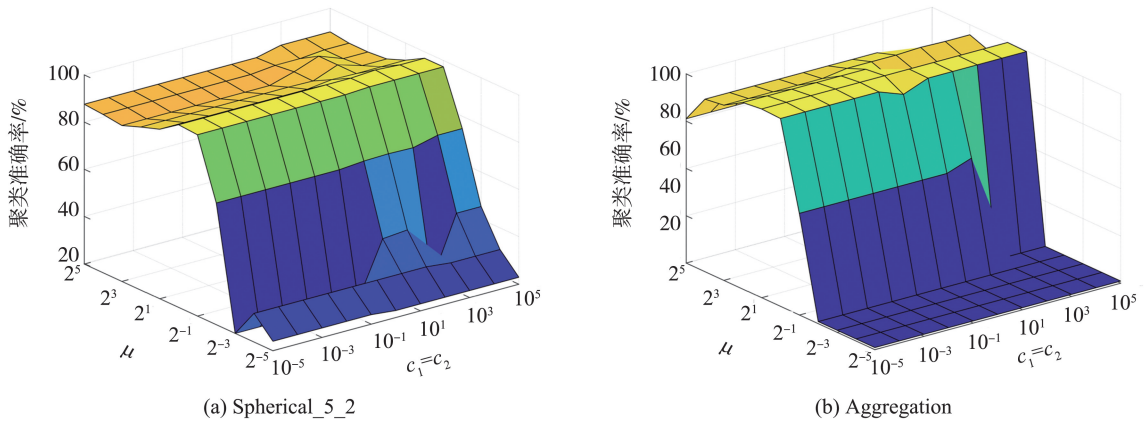


图 4 参数 c_1, c_2 和 μ 对非线性 IFLPPLSTSVC 算法聚类性能的影响

Fig.4 Influence of parameters c_1, c_2 and μ on clustering performance of nonlinear IFLPPLSTSVC algorithm

4 结论

本文提出的直觉模糊局部保持投影最小二乘双支持向量聚类算法, 将直觉模糊数和局部保持投影引入 LSPTSVC 算法中, 不仅继承 TWSVC、LSTWSVC 和 LSPTSVC 算法的优点, 而且充分考虑样本之间的局部几何结构, 有效提取数据局部潜在的判别信息, 同时通过直觉模糊得分值为每个样本分配权重, 减少噪声和异常值带来的负面影响。实验结果证实 IFLPPLSTSVC 算法聚类性能更好, 鲁棒性能更强。在未来, 改进直觉模糊数和构建更高效的最优参数选择策略是后续研究的主要方向。

参考文献:

[1] 郑晨颖, 陈颖悦, 侯贤宇, 等. 一种邻域粒的模糊 C 均值聚类算法[J]. 山东大学学报(理学版), 2024, 59(5):35-44.
ZHENG Chenying, CHEN Yingyue, HOU Xianyu, et al. A fuzzy C-means clustering algorithm for neighborhood particles[J]. Journal of Shandong University (Natural Science), 2024, 59(5):35-44.

[2] 邢璐, 魏毅强. 基于鲁棒块对角表示的子空间聚类[J]. 计算机应用研究, 2020, 37(S2):102-104.
XING Lu, WEI Yiqiang. Subspace clustering based on diagonal representation of robust blocks[J]. Application Research of Computers, 2020, 37(S2):102-104.

[3] WANG Zhen, SHAO Yuanhai, BAI Lan, et al. Twin support vector machine for clustering[J]. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(10):2583-2588.

[4] 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.

- [5] KHEMCHANDANI R, PAL A, CHANDRA S. Fuzzy least squares twin support vector clustering[J]. *Neural Computing and Applications*, 2018, 29(2):553-563.
- [6] BAI Lan, SHAO Yuanhai, WANG Zhen, et al. Clustering by twin support vector machine and least square twin support vector classifier with uniform output coding[J]. *Knowledge-Based Systems*, 2019, 163:227-240.
- [7] TANVEER M, GUPTA T, SHAH M, et al. Pinball loss twin support vector clustering[J]. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 2021, 17(2s):1-23.
- [8] RICHHARIYA B, TANVEER M. Least squares projection twin support vector clustering (LSPTSVC) [J]. *Information Sciences*, 2020, 533:1-23.
- [9] TANVEER M, GANAIE M A, BHATTACHARJEE A, et al. Intuitionistic fuzzy weighted least squares twin SVMs[J]. *IEEE Transactions on Cybernetics*, 2023, 53(7):4400-4409.
- [10] CHEN Sugeng, WU Xiaojun, XU Juan. Locality preserving projection least squares twin support vector machine for pattern classification[J]. *Pattern Analysis and Applications*, 2020, 23:1-13.
- [11] ZHU Jiao, CHEN Sugeng, LIU Yufei, et al. Energy-based structural least squares twin support vector clustering [J]. *Engineering Applications of Artificial Intelligence*, 2024, 128:107467.
- [12] REZVANI S, WANG X, POURPANAH F. Intuitionistic fuzzy twin support vector machines[J]. *IEEE Transactions on Fuzzy Systems*, 2019, 27(11):2140-2151.
- [13] HUA Xiaopeng, DING Shifei. Weighted least squares projection twin support vector machines with local information[J]. *Neurocomputing*, 2015, 160:228-237.
- [14] PHALKE S, VAIDYA Y, METKAR S. Big-O time complexity analysis of algorithm[C]//2022 International Conference on Signal and Information Processing (IConSIP). Pune: IEEE, 2022:1-5.
- [15] 陈素根,刘玉菲. 改进的 Ramp 孪生支持向量机聚类[J]. *计算机科学与探索*, 2023, 17(11):2767-2776.
CHEN Sugeng, LIU Yufei. Improved Ramp twin support vector machine clustering[J]. *Exploration of Computer Science and Technology*, 2023, 17(11):2767-2776.
- [16] RICHHARIYA B, TANVEER M. A fuzzy universum least squares twin support vector machine (FULSTSVM) [J]. *Neural Computing and Applications*, 2022, 34(14):11411-11422.
- [17] RICHHARIYA B, TANVEER M. An efficient angle-based universum least squares twin support vector machine for classification[J]. *ACM Transactions on Internet Technology (TOIT)*, 2021, 21(3):59.

(编辑:唐秋)

(上接第123页)

- [18] RAN Guangtao, LIU Jian, LI Chuanjiang, et al. Event-based finite-time consensus control of second-order delayed multi-agent systems[J]. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2020, 68(1):276-280.
- [19] ZHOU Ding, ZHANG An, YANG Pan. Fixed-time event-triggered consensus of second-order multi-agent systems with fully continuous communication free[J]. *IET Control Theory & Applications*, 2020, 14(16):2385-2394.
- [20] 刘剑. 固定时间多智能体系统事件触发协同控制[D]. 北京:北京科技大学, 2020:86-94.
LIU Jian. Fixed-time event-triggered cooperative control of multi-agent systems[D]. Beijing: University of Science and Technology Beijing, 2020: 86-94.
- [21] LIU Jian, RAN Guangtao, WU Yongbao, et al. Dynamic event-triggered practical fixed-time consensus for nonlinear multiagent systems[J]. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2021, 69(4):2156-2160.
- [22] 孙梦薇,任璐,刘剑,等. 切换拓扑下动态事件触发多智能体系统固定时间一致性[J]. *自动化学报*, 2023, 49(6):1295-1305.
SUN Mengwei, REN Lu, LIU Jian, et al. Fixed-time consensus of dynamic event-triggered multi-agent systems under switching topologies[J]. *Acta Automatica Sinica*, 2023, 49(6):1295-1305.
- [23] 邓甲,王付永,刘忠信,等. 动态事件触发机制下二阶多智能体系统完全分布式控制[J]. *控制理论与应用*, 2024, 41(1):11-20.
DENG Jia, WANG Fuyong, LIU Zhongxin, et al. Fully distributed control for second-order multi-agent systems under dynamic event-triggered mechanism[J]. *Control Theory & Applications*, 2024, 41(1):11-20.
- [24] 宁小刚,李宗刚,陈引娟,等. 时滞和干扰下多智能体系统动态事件触发固定时间一致性[J]. *控制理论与应用*, 2025, 42(2):323-334.
NING Xiaogang, LI Zonggang, CHEN Yinjuan. Dynamic event-triggered fixed-time consensus for multi-agent systems with input delay and unknown disturbances[J]. *Control Theory & Applications*, 2025, 42(2):323-334.

(编辑:唐秋)