

基于自适应图学习的半监督特征选择

江兵兵¹,何文达¹,吴兴宇²,项俊浩¹,洪立斌¹,盛伟国¹

(1. 杭州师范大学信息科学与技术学院,浙江杭州 311121; 2. 中国科学技术大学计算机科学与技术学院,安徽合肥 230027)

摘要: 随着数据特征维数的增加,如何在少量有标签和大量无标签高维样本的情况下选择相关的特征子集已成为特征选择领域的热点问题. 针对现有半监督特征选择算法直接忽略特征选择与局部结构学习之间的相互作用,从而难以有效获取样本分布结构的问题,本文提出了一种基于自适应图学习的半监督特征选择(Semi-supervised Feature Selection with Adaptive Graph learning, SFSAG)算法. 利用标签传播将特征空间的稀疏投影学习和近邻图的构建有效地结合起来,实现在选择相关特征的同时还能学习样本的局部结构;自适应地利用样本在投影特征空间中的相似性信息构建可靠的近邻图,从而有效降低噪声特征的干扰并选择更具判别性的特征子集. 多种数据集上的实验验证了SFSAG的有效性及其相对于现有半监督特征选择算法的优越性.

关键词: 特征选择; 自适应图学习; 半监督学习; 标签传播; $L_{2,1}$ 稀疏正则化

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112(2022)07-1643-10

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20210415

Semi-Supervised Feature Selection with Adaptive Graph Learning

JIANG Bing-bing¹, HE Wen-da¹, WU Xing-yu², XIANG Jun-hao¹, HONG Li-bin¹, SHENG Wei-guo¹

(1. School of Information Science and Technology, Hangzhou Normal University, Hangzhou, Zhejiang 311121, China;

2. School of Computer Science and Technology, University of Science and Technology, Hefei, Anhui 230027, China)

Abstract: With the increasing feature dimensionality, how to select a relevant feature subset in the case of a few labeled and large amount of unlabeled high-dimensional samples has become a hot issue in feature selection. However, existing semi-supervised feature selection algorithms directly ignore the interaction between feature selection and local structure learning, making it difficult to obtain the distribution structure information. To these ends, a semi-supervised feature selection algorithm with adaptive graph learning(SFSAG) is developed in this paper. Firstly, the label propagation is used to link the tasks of sparse projection learning on the original feature space and construction of affinity graph, such that the feature selection and local structure learning can be performed simultaneously. Then, a reliable neighbor graph is adaptively constructed by using the similarity information of samples in the projected feature space, which largely alleviates the adverse effects of noisy dimensions and facilitates selecting more discriminative features. Extensive experiments are conducted on various datasets, and the results demonstrate the effectiveness of the proposed SFSAG and its superiority in comparison with the state-of-the-art feature selection algorithms.

Key words: feature selection; adaptive graph learning; semi-supervised learning; label propagation; $L_{2,1}$ sparse regularization

1 引言

随着信息技术的发展,实际应用中产生的数据具有很高的特征维数. 这些高维数据中包含大量冗余和噪声特征,直接使用这些高维数据不仅需要消耗大量的计算资源和存储空间,还可能导致算法性能的下降. 通过特征抽取或特征选择技术,可以实现高维数据的

降维. 其中特征抽取^[1]是把数据从原始特征空间映射到一个低维特征空间;而特征选择^[2]能够根据特定的规则从原始特征中挑选出少量的相关特征. 由于没有改变数据的原始特征空间,特征选择对高维数据具有更好的可解释性,已广泛应用于图像分类^[3]、人脸识别^[4]等机器学习应用中.

收稿日期:2021-03-30;修回日期:2021-11-25;责任编辑:王天慧

基金项目:国家自然科学基金(No.62006065, No.61873082);杭州师范大学科研启动项目(No.20204003)

根据训练样本中是否包含标签信息,可将现有特征选择算法分成3类:监督特征选择、半监督特征选择和无监督特征选择. 监督特征选择算法利用特征与标签之间的联系来评估特征的重要性. 无监督特征选择算法通过数据中蕴含的特定信息(如数据的方差、分布结构等)来评估特征的相关性. 监督特征选择的性能主要依赖充足的有标签训练样本^[5];而无监督特征选择由于缺少标签信息的指导,其性能通常不理想^[6]. 值得注意的是,实际应用中产生的数据往往包含大量的无标签样本和少量的有标签样本. 因此,研究能够同时利用有标签样本和无标签样本的半监督特征选择算法具有重要的意义. 半监督特征选择的主要思想是:充分利用训练样本的标签信息和分布结构去评估特征的相关性,进而选出具有判别性的特征.

基于半监督学习框架^[7],研究者们相继提出了许多半监督特征选择算法. 这些研究中,早期的方法主要以基于图的过滤式算法为主,此类算法基于特征保持数据分布结构的能力或特定信息准则来选择相关特征. 例如,文献[8]提出了局部敏感的半监督特征选择(Locality Sensitive Discriminant Feature, LSDF)算法,其利用标签信息最大化有标签样本间的类间距,并基于训练样本的局部结构进行特征选择. 文献[9]提出了基于迹比准则的半监督特征选择(Trace Ratio Criterion for Feature Selection, TRCFS)算法. TRCFS首先通过标签传播计算出无标签样本的临时标签,然后根据这些临时标签在训练样本上构建类内矩阵和类间矩阵,最后基于迹比准则选择相关特征子集. 然而,上述过滤式算法没有充分考虑特征间的相关性,同时也忽略了选择特征和训练分类器之间的相互作用^[5]. 最近,研究者们将特征选择作为训练分类器的一部分,提出了能够在训练过程中自动进行特征选择的嵌入式半监督特征选择算法. 例如,Chang等人提出的半监督特征选择(Convex Semi-supervised Feature Selection, CSFS)算法通过最小化预测标签与样本投影的损失来训练特征投影矩阵,并利用特征投影矩阵上的稀疏约束实现特征选择^[10]. Chen等人提出了重新调节线性回归半监督特征选择算法,其利用标度因子调节回归系数,最后选择相应特征子集^[11].

值得注意的是,上述嵌入式半监督特征选择算法直接利用训练样本的预测标签去训练特征选择模型,同时忽略了样本的局部结构信息^[4]. 然而,半监督场景下预测标签中包含的真实标签信息非常有限,直接基于预测标签训练得到的特征选择模型很难选出完全有效的特征子集. 为了充分利用样本中蕴含的局部结构信息,Ma等人将流形正则化^[12]和 $L_{2,p}$ 稀疏约束结合起来,提出了一种基于图嵌入的半监督结构化稀疏特征

选择(Structural Feature Selection with Sparsity, SFSS)算法^[13]. Luo等人使用 $L_{2,p}$ ($0 < p \leq 1$)代替 $L_{2,1}$ 约束,提出基于稀疏回归的半监督特征选择算法^[14]. 此外,在SFSS算法的基础上,文献[15]将 $L_{2,p}$ ($0 < p \leq 1$)约束同时应用到投影损失函数和特征投影矩阵上. 上述基于图嵌入的半监督特征选择算法在一些实际应用中取得了较好的性能,但是它们忽略了特征选择与局部结构学习之间的相互作用,难以有效地利用数据的局部结构信息,进而影响所选特征子集的有效性^[16]. 具体来说,这些算法直接基于原始特征空间中数据间的欧式距离去构建用于刻画数据局部分布结构的近邻图,并在特征选择的过程中使用固定不变的近邻图. 然而,原始特征空间中通常包含大量的噪声和冗余特征,这会使构建的近邻图陷入局部最优. 因此,学习高质量的近邻图对获取数据的局部结构及提升所选特征子集的有效性具有重要作用.

为解决上述问题,本文提出了一种基于自适应图学习的半监督特征选择(Semi-supervised Feature Selection with Adaptive Graph learning, SFSAG)算法. SFSAG首先通过学习样本间的近邻结构,有效利用训练样本中蕴含的局部分布信息;其次,充分利用样本在投影特征空间中的近邻关系;最后,通过稀疏投影学习和标签传播^[17]有效地预测无标签样本的标签,并将学习到的特征投影矩阵用于选择判别性强的特征. 本文的主要贡献总结如下:

- (1) 所提算法能够充分利用样本在投影特征空间中的近邻信息,从而有效地学习样本的局部分布结构,并提高所选特征的有效性和算法的性能;
- (2) 所提算法在选择相关特征的同时还能自适应地学习样本的局部近邻结构,并通过两者间的信息交互降低噪声特征对算法的不利影响;
- (3) 提出了一种有效的优化方法来求解目标函数,并通过合成数据集和高维数据集上的实验验证了所提算法的有效性及其相对其他算法的优越性.

2 自适应图学习的半监督特征选择

本文所提基于自适应图学习的半监督特征选择(SFSAG)算法包含稀疏特征投影学习、自适应图学习和标签传播. 在描述算法模型之前,首先给出算法中的符号说明. 令 $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_n]$ 表示训练样本,其中, $\mathbf{x}_i \in \mathbb{R}^d$ 表示第*i*个样本,*n*为训练样本数,*d*为特征数. $\mathbf{Y} = [\mathbf{Y}_l \quad \mathbf{Y}_u]^T \in \mathbb{R}^{n \times c}$ 为标签矩阵,*c*表示样本的类别标签数, \mathbf{Y}_l 表示有标签样本的标签,如果样本 \mathbf{x}_i 属于第*j*类,那么对应类标签 $\mathbf{Y}_{ij} = 1$,否则 $\mathbf{Y}_{ij} = 0$; \mathbf{Y}_u 表示无标签样本的真实标签. 由于在训练过程中无标签样本的真实标签 \mathbf{Y}_u 是未知的,所以在训练时将其设置为 $\mathbf{0}$ 矩阵^[14].

2.1 算法描述

SFSAG 算法首先在半监督框架下进行稀疏特征投影学习,其公式定义如下

$$\min_{\mathbf{W}, \mathbf{b}, \mathbf{F}} \left\| \mathbf{X}^T \mathbf{W} + \mathbf{1} \mathbf{b}^T - \mathbf{F} \right\|_{2,1} + \gamma \left\| \mathbf{W} \right\|_{2,1} \quad (1)$$

其中, $\mathbf{W} \in \mathbb{R}^{d \times c}$ 表示特征投影矩阵, $\mathbf{b} \in \mathbb{R}^c$ 为偏置向量, $\mathbf{F} = [\mathbf{F}_l \ \mathbf{F}_u]^T \in \mathbb{R}^{n \times c}$ 表示训练样本的预测标签矩阵, γ 是正则化参数. 式(1)将 $L_{2,1}$ 稀疏约束施加在线性投影损失和投影矩阵 \mathbf{W} 上,不仅可以使模型更加鲁棒,还能够使学习到的 \mathbf{W} 具有行稀疏的特点. 将有标签和无标签样本输入到线性回归模型中,然后通过最小化特征投影子空间与预测标签 \mathbf{F} 之间的损失简化特征空间,从而实现特征选择. 基于式(1),研究者们相继提出了一些嵌入式半监督特征选择算法,如文献[10, 11]. 然而预测标签 \mathbf{F} 中包含的真实标签信息相对有限,完全基于 \mathbf{F} 训练的特征选择模型很难选出判别性强的特征子集. 研究表明,在样本标签信息较少时,训练样本(主要是无标签样本)中蕴含的局部分布信息对提升特征选择效力非常重要^[18].

为了充分利用无标签样本,研究者们通过构建样本间的近邻图来获取样本的分布信息(主要是局部结构信息). 传统的方法^[19]使用高斯核函数去构造样本间的近邻图 \mathbf{S} ,这种方式主要存在两个问题:首先,直接通过高斯核构建的近邻图没有考虑原始特征空间中存在冗余和噪声特征,容易产生不可靠的近邻图,进而影响所选特征保持数据分布结构的能力及特征选择性能. 其次,需要提前设定高斯核宽度参数,影响算法的实用性和通用性. 最近, Nie 等人提出了一种自适应近邻图构造方法^[20]:

$$\min_{\mathbf{A}} \sum_{i,j=1}^n \left\| \mathbf{x}_i - \mathbf{x}_j \right\|_2^2 A_{ij} + \alpha A_{ij}^2 \quad (2)$$

其中, A_{ij} 度量了第 i 个样本 \mathbf{x}_i 和第 j 个样本 \mathbf{x}_j 在原始特征空间中的相似性,正则化参数 α 能够避免学习到的相似性矩阵 \mathbf{A} 出现平凡解和均匀分布. 具体来说,如果 $\alpha \rightarrow 0$, 样本将选择最近邻的样本作为其邻居; 如果 $\alpha \rightarrow +\infty$, 任意样本间的相似性 $A_{ij} = 1/n$. 由于 \mathbf{A} 中的每一行 $\mathbf{A}_i (i=1, \dots, n)$ 相互独立,可单独求解行向量 \mathbf{A}_i . 因此式(2)简化为

$$\min_{\mathbf{A}} \sum_{i=1}^n \left\| \mathbf{A}_i + \frac{1}{2\alpha} \mathbf{e}_i \right\|_2^2 \quad (3)$$

其中, 向量 \mathbf{e}_i 的第 j 个分量 $e_{ij} = \left\| \mathbf{x}_i - \mathbf{x}_j \right\|_2^2$. 假设每一个样本有 k 个近邻样本,则可以自动确定正则化参数 α 的值^[21], $\alpha = \sum_{i=1}^n \left(k \hat{e}_{i,k+1} - \sum_{j=1}^k \hat{e}_{i,j} \right) / (2n)$, 其中 $\hat{e}_{i1}, \dots, \hat{e}_{im}$ 表示向量 \mathbf{e}_i 中的元素按照从小到大的顺序排列后得到的

序列. 因此 \mathbf{A} 的计算公式如下

$$A_{ij} = \frac{\hat{e}_{i,k+1} - \hat{e}_{i,j}}{k \hat{e}_{i,k+1} - \sum_{j=1}^k \hat{e}_{i,j}} \quad (4)$$

相比于传统的构图方法(如高斯核方法),上述构图方法可以为每个样本选择 k 个近邻样本,并自适应地确定样本与近邻样本之间的相似性,且不需要使用额外的参数. 虽然上述构图方法能够根据原始特征空间中样本间的近邻信息自适应地构建近邻图,但是在原始特征空间中样本间的近邻信息容易被冗余和噪声特征干扰. 由于特征投影能够有效减轻冗余和噪声特征的不利影响^[16],因此样本在投影特征空间中的近邻关系能够更加准确地描述样本的局部分布结构. 因此,本文提出的自适应图学习模型将特征投影 \mathbf{W} 直接引入式(2)中:

$$\min_{\mathbf{S}} \sum_{i,j=1}^n \left\| \mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j \right\|_2^2 S_{ij} + \theta \left\| \mathbf{S} \right\|_F^2 \quad (5)$$

其中,第一项度量了样本在投影特征空间中的近邻结构. 具体来说,假如 \mathbf{x}_i 和 \mathbf{x}_j 在投影特征空间中较接近,即 $\left\| \mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j \right\|_2^2$ 较小,那么此时 \mathbf{x}_i 和 \mathbf{x}_j 的相似性 S_{ij} 应该较大.

在算法优化的初始阶段,可令 \mathbf{S} 等于 \mathbf{A} ,这样可以使所提构图模型能够在原始特征空间中样本间近邻信息的辅助下,通过学习投影特征空间中的样本近邻关系,更加有效地获取数据的真实局部结构. 正则化项 $\theta \left\| \mathbf{S} \right\|_F^2$ 能够避免 \mathbf{S} 出现平凡解,其中参数 θ 和 α 类似,可根据邻居样本数 k 自动确定:

$$\theta = \frac{1}{n} \sum_{i=1}^n \left(\frac{k}{2} \hat{d}_{i,k+1}^w - \frac{1}{2} \sum_{j=1}^k \hat{d}_{i,j}^w \right) \quad (6)$$

其中, $\hat{d}_{i1}^w, \hat{d}_{i2}^w, \dots, \hat{d}_{im}^w$ 是向量 \mathbf{d}_i^w 中的元素按照从小到大的顺序排列后的序列 ($\mathbf{d}_{ij}^w = \left\| \mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j \right\|_2^2$).

为了有效地结合所提自适应图学习模型和稀疏投影学习,实现在选择相关特征子集的同时还能自适应地学习样本在投影特征空间中的局部结构, SFSAG 算法通过标签传播对预测标签 \mathbf{F} 进行平滑约束,其目标函数为

$$\begin{aligned} & \min_{\mathbf{S}} \sum_{i,j=1}^n \left\| \mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j \right\|_2^2 S_{ij} + \theta \left\| \mathbf{S} \right\|_F^2 \\ & + \lambda \text{tr} \left(\mathbf{F}^T \mathbf{L}_s \mathbf{F} \right) + \text{tr} \left[(\mathbf{F} - \mathbf{Y})^T \mathbf{B} (\mathbf{F} - \mathbf{Y}) \right] \\ & + \beta \left(\left\| \mathbf{X}^T \mathbf{W} + \mathbf{1} \mathbf{b}^T - \mathbf{F} \right\|_{2,1} + \gamma \left\| \mathbf{W} \right\|_{2,1} \right) \end{aligned} \quad (7)$$

其中,第一项表示投影特征空间中的样本近邻结构对图的约束;第二项能够避免 \mathbf{S} 出现平凡解, θ 是正则化

参数;第三项为定义在预测标签 F 上的平滑约束,使近邻样本具有相似的预测标签, L_S 表示定义在 S 上的拉普拉斯矩阵, λ 是正则化参数;第四项能够使有标签样本的预测标签 F_i 与其真实标签 Y_i 尽量保持一致, B 为对角矩阵,通过将其前 l 个对角元素设置为一个较大的常数,充分发挥有标签样本对算法的引导作用.

2.2 算法优化

SFSAG 算法的目标函数中包含目标近邻图 S 、预测标签 F 、投影矩阵 W 和偏差向量 b , 以及投影矩阵 W 的 $L_{2,1}$ 范数约束, 这导致无法直接优化目标函数及求解参数. 因此, 使用交替迭代优化方法来求解目标函数.

首先, 固定 F 和 W , 目标函数可简化为

$$\min \|X^T W + \mathbf{1} b^T - F\|_{2,1} \quad (8)$$

计算式(8)关于 b 的导数, 并令导数等于 $\mathbf{0}$, 可直接得到 b 的最优解

$$b^T = (\mathbf{1}^T F - \mathbf{1}^T X^T W) / n \quad (9)$$

然后, 固定 S 和 W , 并将式(9)中的 b 代入到目标函数中, 可得

$$\min \lambda \text{tr}(F^T L_S F) + \text{tr}[(F - Y)^T B (F - Y)] + \beta \|HX^T W - HF\|_{2,1} \quad (10)$$

其中, $H = I - \mathbf{1}\mathbf{1}^T/n$, I 是单位阵. 式(10)对 F 求导并令导数等于 $\mathbf{0}$, 可得

$$\lambda L_S F + B(F - Y) + 2\beta ZF - 2\beta ZX^T W = \mathbf{0} \quad (11)$$

其中, $Z = HD^v H$, D^v 为对角矩阵, 其第 i 个对角元素 $D_{ii}^v = 1/2 \left(\|x_i^T W + b^T - F_i\|_2 + \varepsilon \right)$.

令 $P = (\lambda L_S + B + 2\beta Z)^{-1}$, $Q = BY + 2\beta ZX^T W$, 可计算出预测标签 F 的最优解:

$$F = PQ \quad (12)$$

固定 b 和 S , 并将式(12)中的 F 代入到目标函数中, 可得

$$\begin{aligned} \min \sum_{i,j=1}^n \|W^T x_i - W^T x_j\|_2^2 S_{ij} + \lambda \text{tr}((PQ)^T L_S PQ) \\ + \text{tr}[(PQ - Y)^T B (PQ - Y)] \\ + \beta \left(\|HX^T W - HPQ\|_{2,1} + \gamma \|W\|_{2,1} \right) \end{aligned} \quad (13)$$

由矩阵迹的定义可将式(13)化简成

$$\begin{aligned} \min 2\text{tr}(W^T X L_S X^T W) + \text{tr}(2\beta W^T X Z X^T W) \\ - \text{tr}(Q^T P^T Q) + \beta \gamma \|W\|_{2,1} \end{aligned} \quad (14)$$

将 $Q = BY + 2\beta ZX^T W$ 代入到式(14)中, 并计算式(14)关于 W 的导数, 可得 W 的最优解:

$$W = (E + \gamma K)^{-1} C \quad (15)$$

其中, $E = X(L_S/\beta + Z - 2\beta ZPZ)X^T$, $C = XZPB Y$, 对角矩阵 K 的第 i 个对角元素 $K_{ii} = 1/2(\|w^i\| + \varepsilon)$.

最后, 固定 W 和 F , 目标函数简化为

$$\min \sum_{i,j=1}^n \|W^T x_i - W^T x_j\|_2^2 S_{ij} + \theta \|S\|_F^2 + \lambda \text{tr}(F^T L_S F), \quad (16)$$

$$\text{s.t. } \mathbf{S}\mathbf{1} = \mathbf{1}, 0 \leq S_{ij} \leq 1$$

其中, 矩阵 S 的每一行都是相互独立的, 因此, 可以单独求解 S 中的行向量 S_i :

$$\theta S_i^2 + S_i^T d_i^w + \lambda S_i^T d_i^f / 2 \quad (17)$$

其中, 向量 d_i^f 和 d_i^w 的第 j 个分量为 $d_{ij}^f = \|f_i - f_j\|_2^2$, $d_{ij}^w = \|W^T x_i - W^T x_j\|_2^2$. 令标量 $d_{ij} = d_{ij}^w + \lambda d_{ij}^f / 2$, 式(17)可化简为

$$\min_{S_i^T \mathbf{1} = \mathbf{1}, 0 \leq S_{ij} \leq 1} \left\| S_i + \frac{1}{2\theta} d_i \right\|_2^2 \quad (18)$$

式(18)可利用文献[22]中提到的方法直接求解. 假设每个样本有 k 个近邻样本, 那么参数 θ 的值可以通过式(6)自动确定.

综上所述, SFSAG 目标函数的迭代求解过程如下: 首先, 初始化 $S = A$ 和 $F = Y$; 其次, 利用式(9)和式(12)分别计算 b 和 F ; 然后利用初始化的对角矩阵 K 和式(15)更新特征投影矩阵 W , 并基于当前的 W 更新 K , 通过式(18)自适应地更新近邻图 S , 重复上述优化过程, 直到目标函数满足收敛条件. 所提算法通过稀疏投影学习可以将特征投影矩阵 W 的行向量 $w^i (i = 1, \dots, d)$ 与训练样本 X 的对应特征关联起来. 因此, W 的第 i 个行向量直接度量了样本 X 中第 i 个特征的重要程度^[12] ($\|w^i\|_2$ 越大表明对应特征越重要). 通过计算特征投影 W 的行向量模长 $\{\|w^1\|_2, \dots, \|w^d\|_2\}$, 并将其从大到小排序, 最终选择对应前 r 个特征作为特征选择的输出.

3 实验结果与分析

本节将在合成数据集和高维实际数据集上测试 SFSAG 算法, 并将其与一些先进的半监督特征选择算法进行对比, 以验证 SFSAG 的有效性.

3.1 合成数据集上的实验

本实验参照文献[4]生成 Two-Moon 数据集来验证 SFSAG 算法获取样本局部结构的能力及其对噪声特征的鲁棒性. Two-Moon 数据集的原始数据分布如图 1(a) 所示: 数据包含 2 个类簇, 每个类簇约有 100 个样本点, 分别是图中的红色和黑色样本点, 其中蓝色圆圈表示有标签样本. 使用式(2)构造的初始化近邻图 A 如图 1

(b). 从中可以看出,虽然初始化图 A 能大致刻画数据的分布结构,但是 2 类簇间仍存在一些相互连接的边,使类簇无法完全分离,这表明基于原始特征空间的样本间距离信息构建的近邻图 A 并不完全可靠. 同时,当在 Two-Moon 的特征空间中增加 8 个随机噪声特征(范围在 $-0.3 \sim 0.3$) 时,此时构造的近邻图如图 1(c) 所示,可以看出,增加的噪声特征使不同类簇间连接边数增多,不仅降低了近邻图可靠性,还增加了分离不同类簇的难度. 图 1(d) 展示了 SFSAG 基于初始化图 A 和数据在投影特征空间中的近邻结构学习到的近邻图 S , 可以看出,在 S 中,两类簇间不存在相互连接的边,说明 S 不仅准确获取数据的分布结构,还能有效分离不同的类簇. 实验结果表明: SFSAG 在进行特征选择的同时还能自适应地利用数据在投影特征空间中的近邻信息,进而降低噪声对数据局部分布结构的干扰,提高算法的性能.

3.2 高维实际数据集上的实验

为了进一步验证 SFSAG 算法在高维数据集上的性能,本节实验使用了 8 个常用的高维实际数据集,其详细信息如表 1 所示.

本实验在每个数据集中随机选择 70% 的样本作为训练集,剩余样本为测试集. 实验首先利用特征选择算法在训练集上选出相关的特征子集,然后在所选特征子集表示的有标签样本上训练线性 SVM 分类器(正则化参数固定为 0.1),最后通过计算分类器在由所选特征表示的测试样本上的分类准确率来验证特征选择算法的有效性. 实验选择 4 种半监督特征选择算法与 SFSAG 算法进行对比,分别是 CSFS, LSDF, SFSS 和 TRCFS. 为了降低实验结果的统计误差,每组实验都会在不同的训练集和测试集分组上独立运行 20 次,使用平均分类准确率作为实验结果. 所有实验算法的正则化参数或平衡参数使用网格搜索法在 $\{10^{-3}, \dots, 10^3\}$ 的范围内调整. 所提算法希望有标签样本上的预测标签

表 1 实验数据集描述

数据集名称	样本数	特征数	类别数
Binalpha ^[16]	1 404	320	36
Coil20 ^[16]	1 440	1 024	20
Flower-17	1 360	1 239	17
Isolet	1 560	617	26
JAFFE ^[22]	213	1 024	10
MSRA25 ^[20]	1 799	256	12
MSRC ^[23]	210	512	7
USPST ^[16]	2 007	256	10

1. Flower-17 来自: <https://www.robots.ox.ac.uk/~vgg/data/flowers/17/>

2. Isolet 来自: <https://archive.ics.uci.edu/ml/datasets/isolet>

F_i 能够充分接近它们的真实标签 Y_i , 由于无标签样本的真实标签 Y_u 未知,所以在算法训练过程中将其设置为 $\mathbf{0}$ 矩阵,因此 SFSAG 算法只考虑有标签样本的真实标签 Y_l 和预测标签 F_l 的关系. 于是将对角矩阵 B 的前 l 个对角元素设置成较大的常数 (10^4); 将 B 的后 u 个对角元素设置成较小的常数 (10^{-4}). 实验从训练集中随机选择不同比率的有标签样本,并通过改变所选特征数量来分析特征选择算法的性能. 图 2、图 3 分别展示了使用 10% 和 30% 有标样本时算法分类准确率与所选特征数量的变化曲线.

从图 2、图 3 可以看出: (1) 在实验数据集上,特征选择算法的分类性能随着选择特征数量的增加而提高,而且 SFSAG 算法几乎在所有数据集上的分类准确率均高于对比算法,验证了所提算法在高维数据集上的有效性; (2) 在大多数数据集上,基于图的过滤式算法 (LSDF 和 TRCFS) 的性能表现均不及基于图嵌入式的 SFSS 算法,且 SFSS 算法的性能不如 SFSAG 算法,这表明特征空间的稀疏投影学习和自适应的局部结构学习是影响特征选择的重要因素; (3) 当标签样本的比率为 10% 时,CSFS 在所有实验数据集上的性能均不如基于

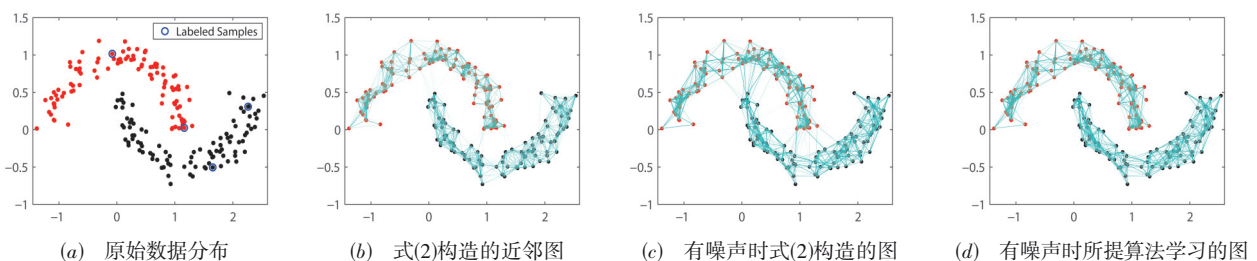


图 1 公式(2)和所提算法在 Two-Moon 数据集上构建的近邻图

图嵌入的算法,甚至在多个数据集(如 Binalpha、Coil20 和 JAFFE)上的性能不如基于图的过滤式算法,这表明在有标签样本较少时,利用样本的局部结构信息对提升特征选择的性能具有重要的意义.

为了进一步分析 SFSAG 与现有半监督特征选择算

法在不同有标签样本比率下的分类性能,下面实验将训练样本中有标签样本的比率设定在 $\{10\%, 20\%, 30\%, 40\%, 50\%\}$ 范围内,并将所选特征数量分别固定为 $\{80, 200, 200, 80, 200, 80, 80, 80\}$. 实验结果如表 2 所示,其中加粗项为相同标签比率下算法在对应数据集

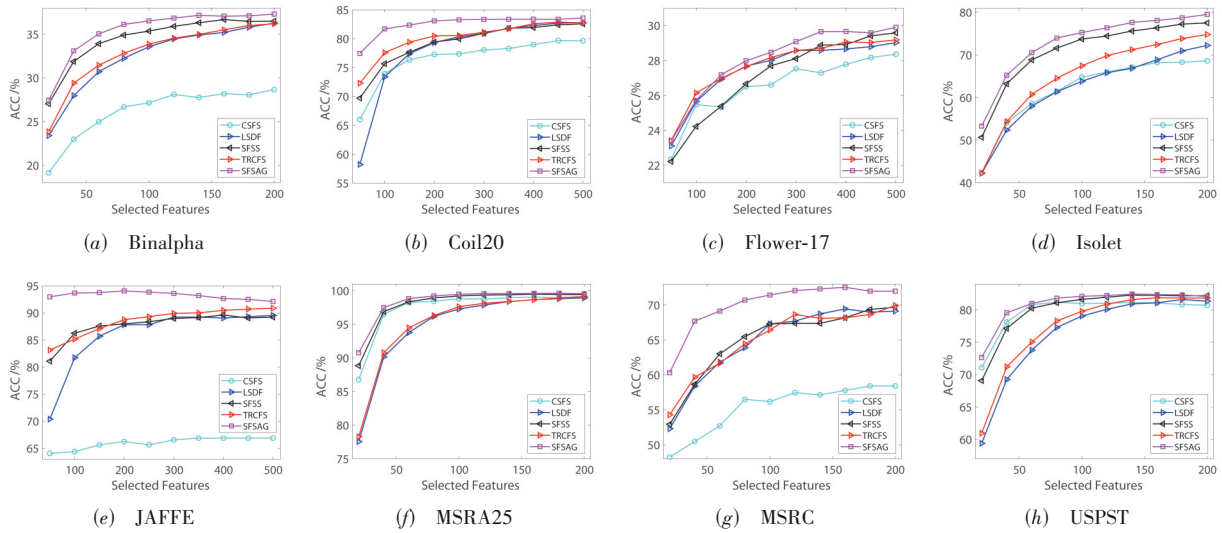


图2 使用10%的有标签样本,SFSAG和对比较法在实验数据集上的分类准确率与所选特征数量的变化曲线

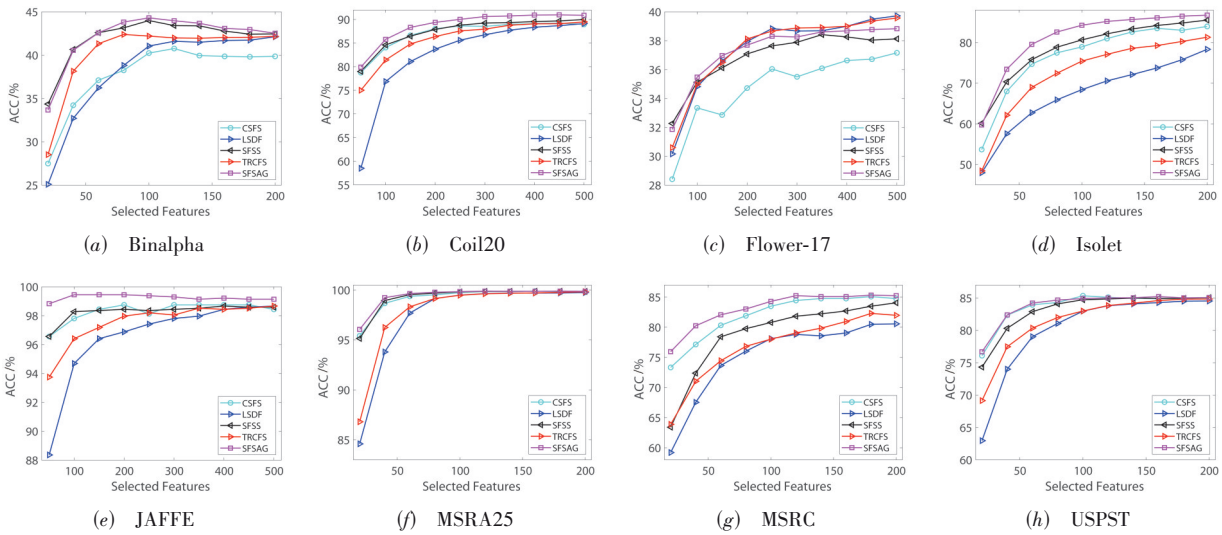


图3 使用30%的有标签样本,SFSAG和对比较法在实验数据集上的分类准确率与所选特征数量的变化曲线

上的最优结果.从表2中可以看出,特征选择算法的分类准确率随着样本标签比率的增加不断增长,这表明在有标签样本充足时,特征选择算法能够选择出判别性更强的特征子集.在不同的标签比率下,SFSAG算法在实验数据集上都能够取得较好甚至最好的分类准确率.此外,在MSRC数据集上,当有标签样本比率为10%和20%时,基于预测标签的CSFS算法的性能明显低于LSDF、TRCFS和SFSS算法.当标签比率达到30%甚至更高时,CSFS算法的性能仅仅低于SFSAG算法,这表明在样本的标签比率较低时,SFSAG算法能够更充分地利用样本的局部结构信息,在性能上较其他算法更具优势.值得注意的是,SFSAG算法在实验数据集上的分类准确率均高于SFSS算法,进一步验证了SFSAG算法的有效性.

3.3 算法的参数敏感性和收敛性分析

SFSAG算法包含3个正则化参数 λ 、 β 和 γ .本实验分析参数在 $\{10^{-3}, \dots, 10^3\}$ 范围内变化对SFSAG性能的影响.图4给出了 β 、 γ 和 λ 在Isolet和Coil20上对SFSAG性能的影响.实验结果表明:SFSAG算法对较宽取值范围内的参数 β 、 γ 和 λ 的敏感程度不同.从图4可以看出,在给定所选特征数量时,参数 β 、 γ 和 λ 的变化对算法性能的影响较大.具体来说,在所选特征数量不变时,算法性能会随着 β 、 γ 和 λ 取值的变化而明显变化,表明SFSAG算法性能对特征空间中的稀疏投影学习和预测标签上的平滑约束更敏感;同时当所选特征数量较少时,SFSAG算法的性能对参数 β 、 γ 的变化更敏感.

表 2 使用不同比率的有标签样本, SFSAG 以及其他半监督特征选择算法在测试集上的平均分类准确率(均值±标准差%)

数据集名称	模型	10%	20%	30%	40%	50%
Binalpha	CSFS	26.68±2.20	32.61±1.96	38.25±2.75	40.81±3.70	42.13±3.01
	LSDF	32.25±2.02	36.13±2.62	38.82±3.17	40.45±2.16	42.93±2.02
	SFSS	34.91±2.46	39.09±2.15	43.15±2.18	45.20±2.34	47.48±3.30
	TRCFS	32.82±1.79	37.42±2.67	42.39±2.76	44.87±1.84	45.81±3.02
	SFSAG	35.90±2.66	40.11±2.21	43.82±2.29	45.95±2.54	47.19±2.87
Coil20	CSFS	77.27±1.73	83.74±2.39	88.04±1.92	89.42±3.62	88.82±0.35
	LSDF	79.33±3.13	82.47±3.00	83.72±2.15	84.26±2.14	84.86±1.82
	SFSS	79.42±3.37	84.82±2.86	87.90±1.76	89.71±1.56	90.21±1.47
	TRCFS	80.48±2.81	84.72±2.96	86.39±2.27	88.08±1.89	86.93±2.31
	SFSAG	81.80±3.08	87.46±2.29	89.39±1.12	90.70±1.19	91.44±1.61
Flower-17	CSFS	26.50±4.21	32.18±1.81	34.72±1.26	38.97±1.43	40.44±3.06
	LSDF	27.68±2.29	33.13±2.37	37.95±1.88	40.02±2.21	41.70±2.09
	SFSS	26.66±2.69	32.13±3.05	37.08±1.86	38.99±2.58	41.75±2.28
	TRCFS	27.65±2.12	33.14±2.32	38.12±2.02	40.05±2.27	41.65±2.11
	SFSAG	27.98±2.25	33.73±2.71	37.69±2.08	39.84±2.05	42.22±2.33
Isolet	CSFS	61.45±2.70	74.57±3.88	77.48±1.87	78.85±3.25	81.24±1.08
	LSDF	61.36±4.20	64.65±3.15	65.90±2.59	66.51±2.53	66.71±1.73
	SFSS	71.60±3.06	76.92±2.08	78.86±2.30	77.99±2.22	79.08±2.30
	TRCFS	64.52±3.40	70.57±2.53	72.43±2.68	72.79±3.47	75.10±3.82
	SFSAG	73.96±3.11	79.74±1.93	82.59±2.29	83.80±1.95	83.68±2.38
JAFEE	CSFS	66.25±9.73	91.88±9.84	98.75±2.04	100.00±0.00	98.75±1.71
	LSDF	87.81±6.89	94.06±4.17	96.88±1.96	96.80±1.79	97.50±1.47
	SFSS	87.97±5.53	97.11±2.11	98.44±1.76	99.30±1.56	99.77±0.76
	TRCFS	88.75±4.80	95.31±3.13	97.97±2.10	98.44±1.83	98.67±1.46
	SFSAG	94.06±2.80	98.67±1.37	99.45±0.76	99.61±0.86	99.92±0.35
MSRA25	CSFS	98.44±0.75	99.52±0.21	99.52±0.31	99.89±0.17	99.96±0.08
	LSDF	96.19±2.69	98.66±0.75	99.15±0.05	99.25±0.52	99.41±0.32
	SFSS	98.93±0.91	99.51±0.29	99.69±0.23	99.81±0.18	99.81±0.17
	TRCFS	96.31±2.19	98.71±0.84	99.15±0.46	99.24±0.60	99.44±0.35
	SFSAG	99.23±0.53	99.65±0.26	99.78±0.22	99.83±0.16	99.84±0.17
MSRC	CSFS	56.51±9.30	70.48±14.70	81.90±4.14	83.49±5.79	87.62±2.84
	LSDF	63.89±6.07	73.25±4.64	76.03±5.82	75.48±4.79	78.89±5.60
	SFSS	65.48±5.87	74.76±8.30	79.76±5.07	80.16±5.39	83.33±4.14
	TRCFS	64.44±5.84	73.49±4.72	76.83±5.51	78.57±3.80	79.05±4.42
	SFSAG	70.71±7.59	79.29±6.40	83.02±4.81	83.41±4.23	86.03±4.23
USPST	CSFS	81.16±2.09	83.75±1.21	84.31±2.50	85.31±1.10	85.80±0.99
	LSDF	77.30±2.82	80.16±2.44	81.07±2.44	82.76±1.66	82.60±1.94
	SFSS	81.11±1.37	83.02±2.17	84.09±1.81	84.89±1.65	84.62±1.45
	TRCFS	78.31±2.33	81.10±2.11	81.99±2.20	82.40±2.22	82.59±1.96
	SFSAG	81.82±1.98	83.76±1.63	84.68±1.53	85.62±1.19	85.50±1.60

本文使用交替迭代优化方法求解 SFSAG 算法的目标函数。为了验证 SFSAG 算法的收敛性,图 5 展示了 SFSAG 算法的目标函数与迭代次数的变化曲线,从图

中可以看出:在实验数据集上, SFSAG 算法的目标函数只需迭代 5 次就能达到收敛状态,验证了交替迭代优化求解方法的快速性和有效性。

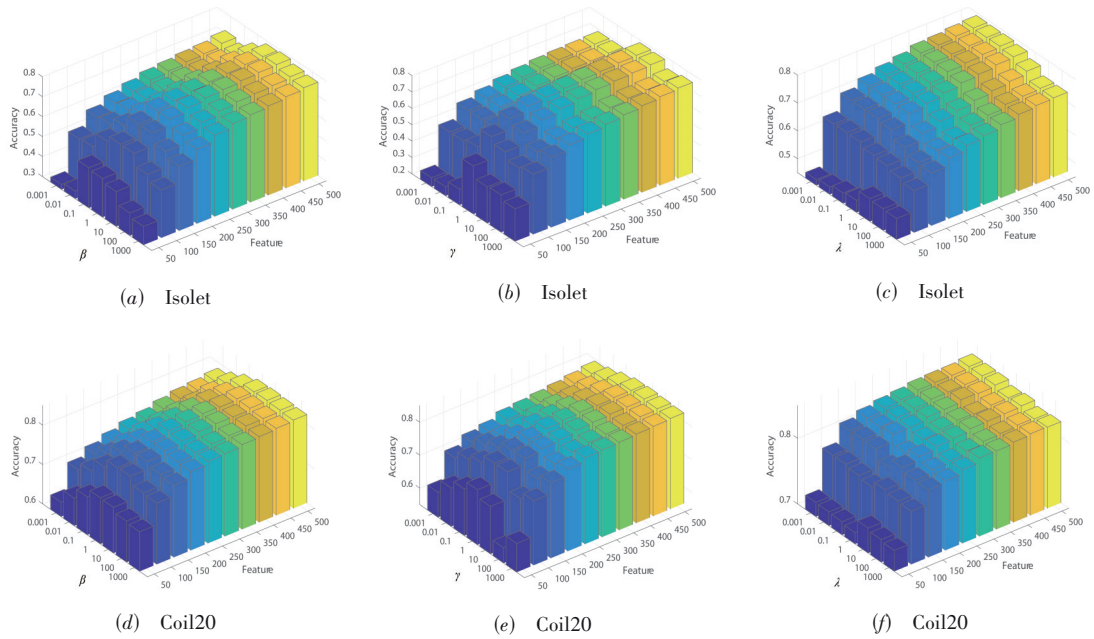


图4 在 Isolet 和 Coil20 数据集上,参数 β , γ 和 λ 对算法性能的影响

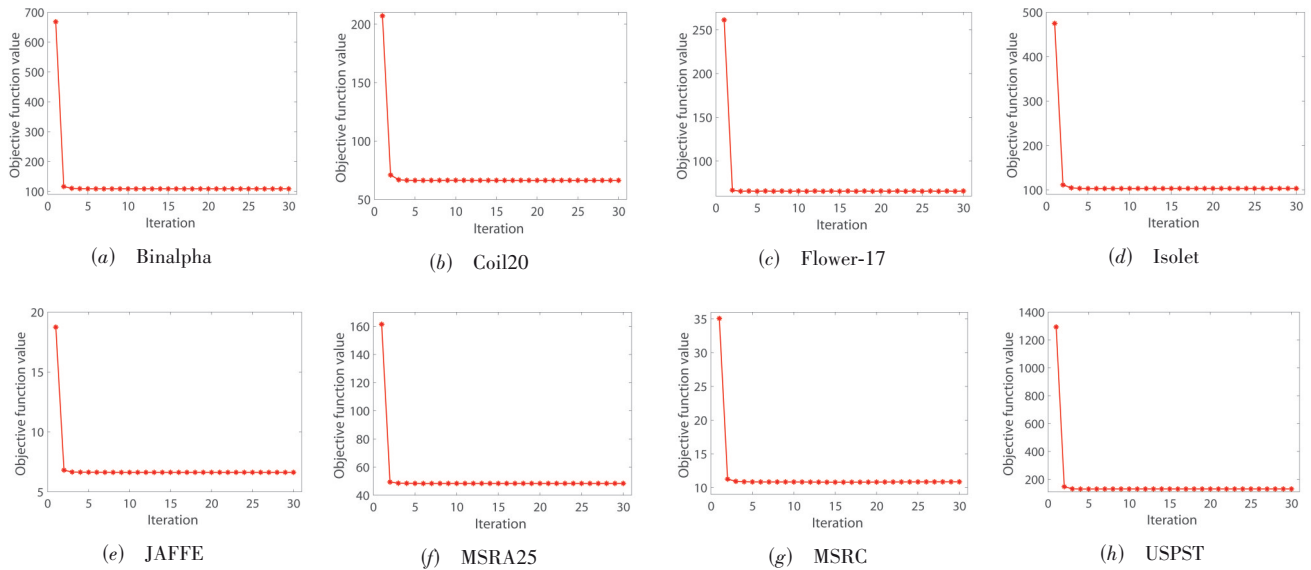


图5 在实验数据集上 SFSAG 算法的目标函数与迭代次数的变化曲线

4 结论

本文提出了一种基于自适应图学习的半监督特征选择(SFSAG)算法. 首先,SFSAG算法充分考虑了样本在投影特征空间中的近邻结构,能够自适应学习样本间的最优近邻图,从而充分利用样本中蕴含的数据分布信息. 其次,SFSAG算法能够同时进行特征选择和样本的局部结构学习,并通过两者间的交互来提升模型对噪声特征的鲁棒性,从而有利于选择判别性更强的特征子集. 然后,使用交替优化方法来求解参数的最优

解,并通过实验验证了SFSAG算法的收敛性. 最后,合成数据集和高维实际数据集上的大量实验充分验证了所提构图方法的有效性以及SFSAG算法的优越性.

参考文献

[1] LI H F, JIANG T, ZHANG K S. Efficient and robust feature extraction by maximum margin criterion[J]. IEEE Transactions on Neural Networks, 2006, 17(1): 157-165.
 [2] 蒋盛益, 郑琪, 张倩生. 基于聚类的特征选择方法[J]. 电子学报, 2008, 36(S1): 157-160.

- JIANG S Y, ZHENG Q, ZHANG Q S. Clustering-based feature selection[J]. *Acta Electronica Sinica*, 2008, 36(S1): 157-160. (in Chinese)
- [3] SONG X N, ZANG J G, HAN Y H, et al. Semi-supervised feature selection via hierarchical regression for web image classification[J]. *Multimedia Systems*, 2016, 22(1): 41-49.
- [4] 刘艳芳, 李文斌, 高阳. 基于自适应邻域嵌入的无监督特征选择算法[J]. *计算机研究与发展*, 2020, 57(8): 1639-1649.
- LIU Y F, LI W B, GAO Y. Adaptive neighborhood embedding based unsupervised feature selection[J]. *Journal of Computer Research and Development*, 2020, 57(8): 1639-1649. (in Chinese)
- [5] SHEIKHPOUR R, SARRAM M A, GHARAGHANI S, et al. A Survey on semi-supervised feature selection methods [J]. *Pattern Recognition*, 2017, 64: 141-158.
- [6] ZHAO H F, LI Q, WANG Z, et al. Joint adaptive graph learning and discriminative analysis for unsupervised feature selection[J]. *Cognitive Computation*, 2021: 1-11.
- [7] 余游, 冯林, 王格格, 等. 一种基于伪标签的半监督少样本学习模型[J]. *电子学报*, 2019, 47(11): 2284-2291.
- YU Y, FENG L, WANG G G, et al. A few-shot learning model based on semi-supervised with pseudo label[J]. *Acta Electronica Sinica*, 2019, 47(11): 2284-2291. (in Chinese)
- [8] ZHAO J D, LU K, HE X F. Locality sensitive semi-supervised feature selection[J]. *Neurocomputing*, 2008, 71(10/11/12): 1842-1849.
- [9] LIU Y, NIE F P, WU J G, et al. Efficient semi-supervised feature selection with noise insensitive trace ratio criterion [J]. *Neurocomputing*, 2013, 105: 12-18.
- [10] CHANG X J, NIE F P, YANG Y, et al. A convex formulation for semi-supervised multi-label feature selection [C]//*Proceedings of the 28th AAAI Conference on Artificial Intelligence*. Québec City: AAAI Press, 2014: 1171-1177.
- [11] CHEN X J, YUAN G W, NIE F P, et al. Semi-supervised feature selection via rescaled linear regression[C]//*Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. Melbourne: IJCAI, 2017: 1525-1531.
- [12] BELKIN M, NIYOGI P, SINDHWANI V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples[J]. *Journal of Machine Learning Research*, 2006, 7: 2399-2434.
- [13] MA Z G, NIE F P, YANG Y, et al. Discriminating joint feature analysis for multimedia data understanding[J]. *IEEE Transactions on Multimedia*, 2012, 14(6): 1662-1672.
- [14] LUO T J, HOU C P, NIE F P, et al. Semi-supervised feature selection via insensitive sparse regression with application to video semantic recognition[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30(10): 1943-1956.
- [15] SHEIKHPOUR R, SARRAM M A, GHARAGHANI S, et al. A robust graph-based semi-supervised sparse feature selection method[J]. *Information Sciences*, 2020, 531: 13-30.
- [16] NIE F P, ZHU W, LI X L. Structured graph optimization for unsupervised feature selection[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2021, 33(3): 1210-1222.
- [17] ZHANG Z, ZHAO M B, CHOW T W S. Graph based constrained semi-supervised learning framework via label propagation over adaptive neighborhood[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(9): 2362-2376.
- [18] JIANG B B, WU X Y, YU K, et al. Joint semi-supervised feature selection and classification through Bayesian approach[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 3983-3990.
- [19] 陈兵飞, 江兵兵, 周熙人, 等. 基于稀疏贝叶斯的流形学习[J]. *电子学报*, 2018, 46(1): 98-103.
- CHEN B F, JIANG B B, ZHOU X R, et al. Manifold learning based on sparse Bayesian approach[J]. *Acta Electronica Sinica*, 2018, 46(1): 98-103. (in Chinese)
- [20] NIE F P, WANG X Q, HUANG H. Clustering and projected clustering with adaptive neighbors[C]//*Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York: ACM, 2014: 977-986.
- [21] XU Y Y, WANG J, AN S, et al. Semi-supervised multi-label feature selection by preserving feature-label space consistency[C]//*Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. Torino: ACM, 2018: 783-792.
- [22] HUANG J, NIE F P, HUANG H. A new simplex sparse learning model to measure data similarity for clustering [C]//*Proceedings of the 24th International Joint Conference on Artificial Intelligence*. Buenos Aires: AAAI Press, 2015: 3569-3575
- [23] HOU C P, NIE F P, TAO H, et al. Multi-view unsupervised feature selection with adaptive similarity and view

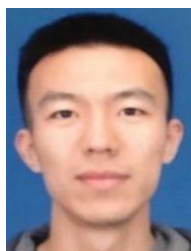
weight[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(9): 1998-2011.

作者简介



江兵兵 男,1991年生,安徽颍上人.2019年毕业于中国科学技术大学计算机应用技术专业,获工学博士学位.现为杭州师范大学信息科学与技术学院硕士生导师,讲师.主要研究方向为半监督学习、特征选择、多视图学习和贝叶斯学习.

E-mail: jiangbb@hznu.edu.cn



何文达 男,1996年生,浙江海盐人.杭州师范大学信息科学与技术学院在读硕士研究生,主要研究方向为半监督学习、特征选择等.

E-mail: wendahe1996@stu.hznu.edu.cn



盛伟国(通讯作者) 男,1977年生,浙江瑞安人.2002年毕业于英国诺丁汉大学,获信息技术硕士学位;2005年毕业于英国布鲁耐尔大学,获计算机科学博士学位.现为杭州师范大学信息科学与技术学院教授,博士生导师.主要研究方向为智能算法理论和设计及其在数据挖掘、模式识别和信息安全等.

E-mail: weiguok@hotmail.com