DOI: 10.11992/tis. 201809017
网络出版地址: http://kns.cnki.net/kcms/detail/23.1538.TP.20190409.0946.016.html

鲁棒的半监督多标签特征选择方法

严菲, 王晓栋
(厦门理工学院 计算机与信息工程学院, 福建 厦门 361024)

摘 要: 针对现有的半监督多标签特征选择方法利用 l_2-范数建立谱图易受到噪声影响的问题, 文中提出一种鲁棒的半监督多标签特征选择方法, 利用全局线性回归函数建立多标签特征选择模型, 结合 l_1 图获取局部描述信息提高模型准确度, 引入 $l_{2,1}$ 约束提升特征之间可区分度和回归分析的稳定性, 避免噪声干扰。在 4 种开源数据集上借助多种性能评价标准验证所提出方法, 结果表明: 本文方法能有效提高分类模型的准确性和对外界噪声的抗干扰性。

关键词: 特征选择; 半监督学习; 多标签学习; l_1 范式图; 线性回归; $l_{2,1}$ 范数; 鲁棒; 分类; 聚类
中图分类号: TP391.4 **文献标志码**: A **文章编号**: 1673−4785(2019)04−0812−08

中文引用格式: 严菲, 王晓栋. 鲁棒的半监督多标签特征选择方法 [J]. 智能系统学报, 2019, 14(4): 812–819.
英文引用格式: YAN Fei, WANG Xiaodong. A robust, semi-supervised, and multi-label feature selection method[J]. CAAI transactions on intelligent systems, 2019, 14(4): 812–819.

A robust, semi-supervised, and multi-label feature selection method

YAN Fei, WANG Xiaodong
(College of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China)

Abstract: The existing semi-supervised multi-label feature selection method constructs a spectral image based on the l_2-norm, which is sensitive to noise. To handle this problem, a robust semi-supervised multi-label feature selection method is presented in this study. A global linear regression function is utilized to construct the multi-label feature selection model, and the l_1-norm graph is combined to obtain the local discriminant information. Subsequently, the $l_{2,1}$-norm constraint is added to improve the distinguishability between these features and the stability of regression analysis to avoid noise interference. Four open source datasets are selected to verify the proposed method based on various evaluation criteria. The results demonstrate the efficiency of our method with respect to the classification accuracy and robustness.
Keywords: feature selection; semi-supervised learning; multi-label learning; l_1-norm graph; linear regression; $l_{2,1}$-norm; robust; classification; clustering

在机器学习中, 特征选择从原始数据集中提取最具代表性子集, 降低数据维度以提高后续分类、聚类处理精度, 是当前研究的热点。按已标记数据样本数量划分, 特征选择方法可分为监督、无监督和半监督学习。监督特征选择[1-2]在已知数据集和类别标签上训练学习模型。在已获取大量已知标签时, 该类方法能取得较好的识别效果。但在实际应用中获取大量已知标签较困难, 且当已知标签数量不足时该类方法性能迅速下降。无监督特征选择方法[3-4]通过对无标签数据的训练以发现其隐藏的结构性知识, 但由于缺乏可区分性的标签信息导致学习性下降。近年来, 半监督特征选择将少量已知标签数据与未标记数据结合建立学习模型, 受到学者的广泛研究。Doquire 等[5]提出以数据方差为评价准则筛选出重要特征, 但其忽略了特征间的相关性, 易陷入局

收稿日期: 2018−09−13. **网络出版日期**: 2019−04−10.
基金项目: 国家自然科学基金项目 (61871464); 福建省自然科学基金面上项目 (2017J01511); 福建省中青年教师科研项目 (JAT170417); 厦门理工学院科研攀登计划项目 (XPDKQ18012).
通信作者: 严菲. E-mail: fyan@xmut.edu.cn.

部最优。为解决此问题, 研究者提出基于谱图理论的半监督方法, 依据某准则建立 Laplacian 矩阵提取数据底层流形结构进行特征选择, 如 Liu 等[6]提出以迹比准则为评价机制, Ma 等[7]引入流形正则化的稀疏特征选择方法等。

在现实应用中, 存在某个数据样本同时隶属于一个或多个不同的类别, 如网页分类、自然场景分类等。以图像标注为例, 一幅自然图像可包含多种场景, 如树、陆地、沙漠, 即一个图像样本可属于多种类别。最简单的解决方法将多标签簇问题分解为多个独立单标签分类问题, 但其忽略了多标签间的相关性。为此, Ji 等[8]利用多标签的共享子空间建立学习框架, 文献 [9] 使用互信息和交互信息的理论方法寻找最优子集, 这些方法均属于监督类方法。但现实应用中大量训练样本中已标记数据极少。如何利用未标记数据及其之间的关系信息提高泛化性能, 给多标签特征选择方法带来了巨大的挑战。针对此问题, 研究者提出半监督多标签特征学习。Alalga 等[10]提出利用 Laplacian 得分判断多标签类间关系, 但其利用 l_2-范数建立 Laplacian 矩阵易受离群点的影响, 从而导致学习稳定性差。Chang 等[11]提出基于全局线性约束的多标签半监督方法, 无需构建 Laplacian 图和特征分解操作, 计算量少, 速度快, 但其忽略了真实数据嵌入在高维空间的底层流形结构。受启于上述研究工作, 本文提出基于 l_1 图的半监督多标签特征选择方法 SMFSL (semi-supervised multi-label feature selection method based on L_1-norm graph), 利用全局线性回归函数方法和 $l_{2,1}$ 组稀疏约束建立多标签特征选择模型, 引用 l_1 图模型以提高特征选择准确度。

1　范数图模型

在机器学习中, 基于图的学习方法通过构建近邻图, 利用样本间反映流形分布而建立问题模型, 得到广泛的研究应用。其中, 基于谱图理论的谱聚类学习方法[12], 在多种应用场景下取得较好的效果。

谱聚类根据数据样本间的相似关系建立 Laplacian 矩阵, 利用特征值和特征向量获取样本间的内在联系。给定 n 组数据集 $\boldsymbol{X}=\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\} \in \mathbf{R}^{d\times n}$, 其中 $\boldsymbol{x}_i \in \mathbf{R}^d$ 为第 i 组数据, d 为维度。定义 $\boldsymbol{G}=(\boldsymbol{V}, \boldsymbol{A})$ 为无向权重图, 其中 $\boldsymbol{V}$ 为向量集, 相似矩阵 $\boldsymbol{A}=[\boldsymbol{A}_1\ \boldsymbol{A}_2\ \cdots\ \boldsymbol{A}_n] \in \mathbf{R}^{n\times n}$, $\boldsymbol{A}_{ji}=\boldsymbol{A}_{ij}\geqslant 0$。基于高斯核函数 σ 相似矩阵 $\boldsymbol{A}$ 定义为

$$\boldsymbol{A}_{ij}=\begin{cases}\exp\left(\dfrac{-\left\|\boldsymbol{x}_i-\boldsymbol{x}_j\right\|^2}{\sigma^2}\right), & \boldsymbol{x}_i\text{为}\boldsymbol{x}_j\text{的近邻}\\ 0, & \text{其他}\end{cases} \tag{1}$$

式中: σ 为宽度参数, 控制函数的径向作用范围。

基于式 (1) 构造加权图, 谱聚类算法将聚类问题转化为图划分问题, 但其最优解为 NP 问题。传统解决方法以借助松弛方法得到连续的类别标签, 进而转换为率切 (ratio cut) 问题:

$$\min_{\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\boldsymbol{I}} \mathrm{Tr}\left(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{Q}\right) \tag{2}$$

式中: $\boldsymbol{Q}=[\boldsymbol{q}_1\ \boldsymbol{q}_2\ \cdots\ \boldsymbol{q}_n]^{\mathrm{T}} \in \mathbf{R}^{n\times c}$ 为聚类指标矩阵, c 为类别标签数, $\boldsymbol{q}_k \in \mathbf{R}^c$ 为 $\boldsymbol{Q}$ 矩阵第 k 列; $\boldsymbol{L}$ 为谱图 Laplacian 矩阵, 其定义为 $\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{A}$, $\boldsymbol{D}$ 为对角矩阵, 其每个 i 对角元素 $\boldsymbol{D}_{ii}=\sum_{j=1}^{n}\boldsymbol{A}_{ij}$。为获取最终的类别标签, 必须进一步借助聚类算法将连续值矩阵 $\boldsymbol{Q}$ 进行离散化, 如采用 K-means 算法等。Nie 等[13]提出基于 l_1 范数图模型来获取更清晰的流形结构, 上述式 (2) 转换为

$$\min_{\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\boldsymbol{I}} \sum_{i,j=1}^{n}\boldsymbol{A}_{ij}\left\|\boldsymbol{q}_i-\boldsymbol{q}_j\right\|_2 \tag{3}$$

2　基于 l_1 图的半监督多标签特征选择方法

2.1　问题定义

给定数据集 $\boldsymbol{X}=[\boldsymbol{x}_1\ \boldsymbol{x}_2\ \cdots\ \boldsymbol{x}_l\ \boldsymbol{x}_{l+1}\cdots\ \boldsymbol{x}_n] \in \mathbf{R}^{d\times n}$, $\boldsymbol{x}_i \in \mathbf{R}^d$ 为第 i 组数据, d 为维度, l 为已标记样本数 ($l \ll n$)。设 $\boldsymbol{Y}_l=[\boldsymbol{y}_1\ \boldsymbol{y}_2\ \cdots\ \boldsymbol{y}_l]^{\mathrm{T}} \in \mathbf{R}^{l\times c}$ 为已标记数据集的标签矩阵, $\boldsymbol{y}_i \in \{0,1\}^c$ 为数据 $\boldsymbol{x}_i$ 的类别标签。若 $\boldsymbol{x}_i$ 被标识为 j 类, 则 $\boldsymbol{Y}_{ij}=1$, 否则 $\boldsymbol{Y}_{ij}=0$。为获取未标签数据的类别标签信息, 定义预测标签矩阵 $\boldsymbol{F}=[\boldsymbol{f}_1\ \boldsymbol{f}_2\ \cdots\ \boldsymbol{f}_n]^{\mathrm{T}}=\begin{bmatrix}\boldsymbol{F}_l\\ \boldsymbol{F}_u\end{bmatrix} \in \mathbf{R}^{n\times c}$, 其中 $\boldsymbol{F}_l$ 初始化为 $\boldsymbol{Y}_l$, $\boldsymbol{F}_u \in \mathbf{R}^{(n-l)\times c}$ 则为未标签数据的标签矩阵, 且初始化 $\boldsymbol{F}_u=\boldsymbol{O}$, $\boldsymbol{O}$ 为所有元素为 0 的矩阵。定义 $\boldsymbol{W}=[\boldsymbol{w}_1\ \boldsymbol{w}_2\ \cdots\ \boldsymbol{w}_d]^{\mathrm{T}} \in \mathbf{R}^{d\times c}$ 为特征选择分类器, 半监督多标签特征选择学习模型定义为

$$\min_{\boldsymbol{W},\boldsymbol{F},\boldsymbol{F}_l=\boldsymbol{Y}_l} \sum_{i=1}^{n}\mathrm{loss}(\boldsymbol{W},\boldsymbol{f}_i)+\gamma\Omega(\boldsymbol{W}) \tag{4}$$

在式 (4) 中, $\Omega(\cdot)$ 为正则化项 (可以选择不同的正则化模型, 如 l_1 范数、$l_{2,1}$ 范数等), 参数 γ 为正则化参数, loss$(\cdot)$ 为损失函数。从模型的简单性、高效性角度进行考虑, 本文选择最小二乘法作为损失函数, 式 (4) 可表示为

$$\min_{\boldsymbol{W},\boldsymbol{F},\boldsymbol{b},\boldsymbol{F}_l=\boldsymbol{Y}_l} \left\|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}+\boldsymbol{1}\boldsymbol{b}^{\mathrm{T}}-\boldsymbol{F}\right\|_F^2+\gamma\Omega(\boldsymbol{W}) \tag{5}$$

式中: $\boldsymbol{b} \in \mathbf{R}^c$ 为偏置量; $\boldsymbol{1} \in \mathbf{R}^n$ 为元素值全是 1 的列向量。

从式 (5) 可以看出，利用线性回归函数逐渐逼近可找出全局最优，但却忽略了其局部数据之间相关性。为提高特征选择准确度，文献 [7] 提出建立相似矩阵以获取局部流形结构，建立的学习模型如下：

$$\min_{\boldsymbol{W},\boldsymbol{F},\boldsymbol{b},\boldsymbol{F}_l=\boldsymbol{Y}_l}\sum_{i,j=1}^{n}\boldsymbol{A}_{ij}\left\|\boldsymbol{f}_i-\boldsymbol{f}_j\right\|_2^2+\alpha\left\|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}+\boldsymbol{1}\boldsymbol{b}^{\mathrm{T}}-\boldsymbol{F}\right\|_F^2+\gamma\Omega(\boldsymbol{W}) \tag{6}$$

式中 α 为平衡参数。

半监督学习应用中，已标签的数据集往往只占据小部分，无标签数据集非常庞大，而离群点一般存在无标签数据集中。Nie 等[13] 研究证明，采用 l_1 范数有效减少噪音的影响，从而获取更清晰的谱聚类结构，因此本文提出将 l_1 范数引入半监督学习模型中。同时为减少外界噪声点的干扰，本文提出采用 $l_{2,1}$ 范数[3] 来约束最小二乘损失函数。给定任意矩阵 $\boldsymbol{M}\in\mathbf{R}^{d\times c}$，其定义为 $\boldsymbol{M}_{2,1}=\sum_{i=1}^{d}\sqrt{\sum_{j=1}^{c}\boldsymbol{M}_{ij}^2}$。结合式 (3)，式 (6) 转换为

$$\min_{\boldsymbol{W},\boldsymbol{F},\boldsymbol{b},\boldsymbol{F}_l=\boldsymbol{Y}_l}\sum_{i,j=1}^{n}\boldsymbol{A}_{ij}\left\|\boldsymbol{f}_i-\boldsymbol{f}_j\right\|_2+\alpha\left\|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}+\boldsymbol{1}\boldsymbol{b}^{\mathrm{T}}-\boldsymbol{F}\right\|_{2,1}+\gamma\Omega(\boldsymbol{W})$$

为有效地去除原数据集的冗余特征，本文对特征选择矩阵 $\boldsymbol{W}$ 引入正则化模型 $l_{2,1}$ 范数，最后提出的目标函数定义如下：

$$\min_{\boldsymbol{W},\boldsymbol{F},\boldsymbol{b},\boldsymbol{F}_l=\boldsymbol{Y}_l}\sum_{i,j=1}^{n}\boldsymbol{A}_{ij}\left\|\boldsymbol{f}_i-\boldsymbol{f}_j\right\|_2+\alpha\left\|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}+\boldsymbol{1}\boldsymbol{b}^{\mathrm{T}}-\boldsymbol{F}\right\|_{2,1}+r\|\boldsymbol{W}\|_{2,1} \tag{7}$$

2.2 学习模型求解

为获取最终选择特征子集，将对多标签特征选择的目标函数进行模型求解。由于目标函数式 (7) 引入的 $l_{2,1}$ 范数具有非光滑特征，无法直接进行求解，因此本文提出一种迭代优化方法来解决。

首先，将目标函数 (7) 进行转换。定义对角矩阵 $\boldsymbol{S}$，其元素 $\boldsymbol{S}_{ii}=\dfrac{1}{2\|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}+\boldsymbol{1}\boldsymbol{b}^{\mathrm{T}}-\boldsymbol{F}\|_2+\delta}$。其中，为防止 $\boldsymbol{S}_{ii}$ 除数为 0，δ 的值为接近于 0 的正常量。式 (7) 转换后形式如下：

$$\min_{\boldsymbol{W},\boldsymbol{S},\boldsymbol{F},\boldsymbol{b},\boldsymbol{F}_l=\boldsymbol{Y}_l}\operatorname{tr}\left(\boldsymbol{F}^{\mathrm{T}}\tilde{\boldsymbol{L}}\boldsymbol{F}+\alpha(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}+\boldsymbol{1}\boldsymbol{b}^{\mathrm{T}}-\boldsymbol{F})^{\mathrm{T}}\boldsymbol{S}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}+\boldsymbol{1}\boldsymbol{b}^{\mathrm{T}}-\boldsymbol{F})\right)+r\|\boldsymbol{W}\|_{2,1} \tag{8}$$

式中：$\tilde{\boldsymbol{L}}$ 为对角矩阵且定义为 $\tilde{\boldsymbol{L}}=\tilde{\boldsymbol{D}}-\tilde{\boldsymbol{A}}$；矩阵 $\tilde{\boldsymbol{A}}$ 元素定义为 $\tilde{\boldsymbol{A}}_{ij}=\dfrac{\boldsymbol{A}_{ij}}{2\left\|\boldsymbol{f}_i-\boldsymbol{f}_j\right\|_2}$；$\tilde{\boldsymbol{D}}$ 为对角矩阵，元素值为 $\tilde{\boldsymbol{D}}_{ii}=\sum_j\tilde{\boldsymbol{A}}_{ij}$。给定任意矩阵 $\boldsymbol{B}\in\mathbf{R}^{n\times n}$，对矩阵 $\boldsymbol{B}$ 迹函数定义 $\operatorname{tr}(\boldsymbol{B})=\sum_{i=1}^{n}\boldsymbol{B}_{ii}$，$\boldsymbol{B}_{ii}$ 为矩阵 $\boldsymbol{B}$ 的第 i 个对角元素。

保持 $\boldsymbol{W}$, $\boldsymbol{S}$, $\boldsymbol{F}$ 不变，将式 (8) 对 $\boldsymbol{b}$ 求导，令求导结果为 0，得到：

$$\boldsymbol{b}=\frac{1}{\boldsymbol{1}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{1}}\left(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{S}-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{S}\right) \tag{9}$$

将式 (9) 代入式 (8)，为简化公式，令 $\boldsymbol{H}=\left(\boldsymbol{I}-\dfrac{1}{\boldsymbol{1}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{1}}\boldsymbol{1}\boldsymbol{1}^{\mathrm{T}}\boldsymbol{S}\right)$，其中 $\boldsymbol{H}$ 为中心化矩阵，$\boldsymbol{I}\in\mathbf{R}^{n\times n}$ 为单位矩阵。式 (8) 将转换为：

$$\min_{\boldsymbol{W},\boldsymbol{S},\boldsymbol{F}}\operatorname{tr}(\boldsymbol{F}^{\mathrm{T}}\tilde{\boldsymbol{L}}\boldsymbol{F}+\alpha((\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}-\boldsymbol{H}\boldsymbol{F})^{\mathrm{T}}\boldsymbol{S}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}-\boldsymbol{H}\boldsymbol{F}))+\gamma\|\boldsymbol{W}\|_{2,1} \tag{10}$$

将式 (10) 对 $\boldsymbol{W}$ 求导，并令求导结果为 0，得到

$$\alpha\boldsymbol{X}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}+\gamma\boldsymbol{D}_{\mathrm{W}}\boldsymbol{W}=\alpha\boldsymbol{X}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{F} \tag{11}$$

式中 $\boldsymbol{D}_W$ 为对角矩阵，可表示为

$$\boldsymbol{D}_W=\begin{bmatrix}\dfrac{1}{2\|\boldsymbol{w}_1\|_2+\delta} & & \\ & \ddots & \\ & & \dfrac{1}{2\|\boldsymbol{w}_d\|_2+\delta}\end{bmatrix} \tag{12}$$

由于矩阵 $\boldsymbol{D}_W$ 与 $\boldsymbol{W}$ 相关，无法直接求解上式。为解决此问题，将 $\boldsymbol{W}$ 随机初始化以获取矩阵 $\boldsymbol{D}_W$，转换式 (11)，推导出：

$$\boldsymbol{W}=\alpha\left(\alpha\boldsymbol{X}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}+\gamma\boldsymbol{D}_W\right)^{-1}\boldsymbol{X}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{F} \tag{13}$$

为求解 $\boldsymbol{F}$，将式 (10) 进行变换，得到：

$$\min_{\boldsymbol{W},\boldsymbol{F}}\operatorname{tr}(\boldsymbol{F}^{\mathrm{T}}\tilde{\boldsymbol{L}}\boldsymbol{F})+\alpha\operatorname{tr}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{F})+\gamma\operatorname{tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{D}_{\mathrm{w}}\boldsymbol{W})+\alpha\operatorname{tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W})-2\alpha\operatorname{tr}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}) \tag{14}$$

转换式 (14)，得到：

$$\min_{\boldsymbol{W},\boldsymbol{F}}\operatorname{tr}(\boldsymbol{F}^{\mathrm{T}}\tilde{\boldsymbol{L}}\boldsymbol{F})+\alpha\operatorname{tr}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{F})+\operatorname{tr}(\boldsymbol{W}^{\mathrm{T}}(\alpha\boldsymbol{X}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}+\gamma\boldsymbol{D}_{\mathrm{w}})\boldsymbol{W})-2\alpha\operatorname{tr}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}) \tag{15}$$

将 $\boldsymbol{W}$ 的求导结果式 (13) 代入式 (15)，得到：

$$\min_{\boldsymbol{F}}\operatorname{tr}(\boldsymbol{F}^{\mathrm{T}}\tilde{\boldsymbol{L}}\boldsymbol{F})+\alpha\operatorname{tr}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{F})-\alpha^2\operatorname{tr}\left(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}(\alpha\boldsymbol{X}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}+\gamma\boldsymbol{D}_W)^{-1}\boldsymbol{X}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{F}\right) \tag{16}$$

定义 $\boldsymbol{M}$ 为：

$$\boldsymbol{M}=\tilde{\boldsymbol{L}}+\alpha\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}-(\alpha\boldsymbol{X}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}+\gamma\boldsymbol{D}_W)^{-1}\boldsymbol{X}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H} \tag{17}$$

将式 (16) 按分块矩阵形式转换为：

$$\min_{\boldsymbol{F}_u}\operatorname{tr}\left(\begin{bmatrix}\boldsymbol{F}_l\\ \boldsymbol{F}_u\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\boldsymbol{M}_{ll} & \boldsymbol{M}_{lu}\\ \boldsymbol{M}_{ul} & \boldsymbol{M}_{uu}\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_l\\ \boldsymbol{F}_u\end{bmatrix}\right)$$

将上式对 $\boldsymbol{F}_u$ 求导，并令求导结果为 0，得到：

$$\boldsymbol{F}_u=-\boldsymbol{M}_{uu}^{-1}\boldsymbol{M}_{ul}\boldsymbol{F}_l$$

基于以上推导过程求解学习模型，本文方法具体描述如下：

算法 1 SMFSL

输入 数据集 $\boldsymbol{X}\in\mathbf{R}^{d\times n}$，类别标签 $\boldsymbol{Y}_l\in\mathbf{R}^{l\times c}$，相似矩阵 $\boldsymbol{A}$，参数 α,γ；

输出 特征选择矩阵 $\boldsymbol{W}$，预测标签矩阵 $\boldsymbol{F}$。

1) t=0，随机初始化 $\boldsymbol{W}^t\epsilon\mathbf{R}^{d\times c}$，$\boldsymbol{b}^t\epsilon\mathbf{R}^c$，$\boldsymbol{F}^t\epsilon\mathbf{R}^{n\times c}$

2) repeat

3) 计算 $\tilde{\boldsymbol{L}}=\tilde{\boldsymbol{D}}-\tilde{\boldsymbol{A}}$, 其中 $\tilde{\boldsymbol{A}}_{ij}=\dfrac{\boldsymbol{A}_{ij}}{2\left\|f_i-f_j\right\|_2}$, $\tilde{\boldsymbol{D}}$ 为对角矩阵, 其元素值为 $\tilde{\boldsymbol{D}}_{ii}=\sum_j\tilde{\boldsymbol{A}}_{ij}$

4) 计算对象矩阵 $\boldsymbol{S}^t$, 其第 i 个元素

$$\boldsymbol{S}_{ii}^{\ t}=\frac{1}{2\left\|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}^t+\boldsymbol{1}(\boldsymbol{b}^t)^{\mathrm{T}}-\boldsymbol{F}^t\right\|_2+\delta}$$

5) 计算 $\boldsymbol{H}$, $\boldsymbol{H}=\left(\boldsymbol{I}-\dfrac{1}{\boldsymbol{1}^{\mathrm{T}}\boldsymbol{S}^t\boldsymbol{1}}\boldsymbol{1}\boldsymbol{1}^{\mathrm{T}}\boldsymbol{S}^t\right)$

6) 根据式 (12) 计算对角矩阵 $\boldsymbol{D}_w^t$

7) 根据式 (17) 计算 $\boldsymbol{M}^{t+1}=\tilde{\boldsymbol{L}}+\alpha\boldsymbol{H}\boldsymbol{S}^t\boldsymbol{H}-\alpha^2\boldsymbol{H}\boldsymbol{S}^t\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}(\alpha\boldsymbol{X}\boldsymbol{H}\boldsymbol{S}^t\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}+\gamma\boldsymbol{D}_w^t)^{-1}\boldsymbol{X}\boldsymbol{H}\boldsymbol{S}^t\boldsymbol{H}$

8) 计算 $\boldsymbol{F}_u^{t+1}=-(\boldsymbol{M}_{uu}^{t+1})^{-1}\boldsymbol{M}_{ul}^{t+1}\boldsymbol{F}_l$, 并且更新 $\boldsymbol{F}^{t+1}=\begin{bmatrix}\boldsymbol{F}_l\\ \boldsymbol{F}_u^{t+1}\end{bmatrix}$

9) 更新

$$\boldsymbol{W}^{t+1}=\alpha\left(\alpha\boldsymbol{X}\boldsymbol{H}\boldsymbol{S}^t\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}+\gamma\boldsymbol{D}_w^t\right)^{-1}\boldsymbol{X}\boldsymbol{H}\boldsymbol{S}^t\boldsymbol{H}\boldsymbol{F}^{t+1}$$

10) 更新

$$\boldsymbol{b}^{t+1}=\frac{1}{\boldsymbol{1}^{\mathrm{T}}\boldsymbol{S}^t\boldsymbol{1}}\left(\left(\boldsymbol{F}^{t+1}\right)^{\mathrm{T}}\boldsymbol{S}^t\boldsymbol{1}-\left(\boldsymbol{W}^{t+1}\right)^{\mathrm{T}}\boldsymbol{X}\boldsymbol{S}^t\boldsymbol{1}\right)$$

11) t=t+1

12) until 收敛

在得到特征选择矩阵 $\boldsymbol{W}$ 后, 按照 $\left\|\boldsymbol{w}_i^t\right\|_2$ $(i=1,2,\cdots,d)$ 的值, 对输入数据的特征进行降序排列, 其中最前面的 p 个特征即为所选择的特征。

2.3　收敛性分析

本小节将对上一小节提出的 SMFSL 算法收敛性进行证明。

引理 1　对于任意非零的向量 $\boldsymbol{p},\boldsymbol{q}\in\mathbf{R}^c$, 有以下不等式成立:

$$\|\boldsymbol{p}\|_2-\frac{\|\boldsymbol{p}\|_2^2}{2\|\boldsymbol{q}\|_2}\leqslant\|\boldsymbol{q}\|_2-\frac{\|\boldsymbol{q}\|_2^2}{2\|\boldsymbol{q}\|_2}\tag{18}$$

根据以上引理, 本文提出以下定理:

定理 1　算法 SMFSL 在每次迭代过程中最小化目标函数。

证明　为方便起见, 首先定义 $g(\boldsymbol{W},\boldsymbol{F},\boldsymbol{b}^{\mathrm{T}},\boldsymbol{S})=\alpha\mathrm{tr}((\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}+\boldsymbol{1}\boldsymbol{b}^{\mathrm{T}}-\boldsymbol{F})^{\mathrm{T}}\boldsymbol{S}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}+\boldsymbol{1}\boldsymbol{b}^{\mathrm{T}}-\boldsymbol{F}))$。目标函数可写为:

$$\min_{\boldsymbol{W},\boldsymbol{F},\boldsymbol{b}}\sum_{i,j=1}^{n}\tilde{\boldsymbol{A}}_{ij}\left\|f_i-f_j\right\|_2^2+\gamma\|\boldsymbol{W}\|_{2,1}+g\left(\boldsymbol{W},\boldsymbol{F},\boldsymbol{b}^{\mathrm{T}},\boldsymbol{S}\right)\tag{19}$$

假设在第 t 次迭代后获得了 $\boldsymbol{W}$, $\boldsymbol{F}$, $\boldsymbol{b}^{\mathrm{T}}$ 和 $\boldsymbol{S}$。在下一次迭代中, 固定 $\boldsymbol{W}$ 为 $\boldsymbol{W}^t$、$\boldsymbol{b}$ 为 $\boldsymbol{b}^t$、$\boldsymbol{S}$ 为 $\boldsymbol{S}^t$ 来求解 $\boldsymbol{F}^{t+1}$, 参考文献 [14] 中的方法可得出:

$$\sum_{i,j=1}^{n}(\tilde{\boldsymbol{A}}^t)_{ij}\left\|\boldsymbol{f}_i^{t+1}-\boldsymbol{f}_j^{t+1}\right\|_2^2+\gamma\|\boldsymbol{W}^t\|_{2,1}+g\left(\boldsymbol{W}^t,\boldsymbol{F}^{t+1},(\boldsymbol{b}^t)^{\mathrm{T}},\boldsymbol{S}^t\right)\leqslant\sum_{i,j=1}^{n}(\tilde{\boldsymbol{A}}^t)_{ij}\left\|\boldsymbol{f}_i^t-\boldsymbol{f}_j^t\right\|_2^2+\gamma\|\boldsymbol{W}^t\|_{2,1}+g\left(\boldsymbol{W}^t,\boldsymbol{F}^t,(\boldsymbol{b}^t)^{\mathrm{T}},\boldsymbol{S}^t\right)\tag{20}$$

将 $(\tilde{\boldsymbol{A}}^t)_{ij}=\dfrac{\boldsymbol{A}_{ij}}{2\left\|\boldsymbol{f}_i^t-\boldsymbol{f}_j^t\right\|_2}$ 代入式 (19), 得到:

$$\sum_{i,j=1}^{n}\frac{\boldsymbol{A}_{ij}\left\|\boldsymbol{f}_i^{t+1}-\boldsymbol{f}_j^{t+1}\right\|_2^2}{2\left\|\boldsymbol{f}_i^t-\boldsymbol{f}_j^t\right\|_2}+\gamma\left\|\boldsymbol{W}^t\right\|_{2,1}+g\left(\boldsymbol{W}^t,\boldsymbol{F}^{t+1},(\boldsymbol{b}^t)^{\mathrm{T}},\boldsymbol{S}^t\right)\leqslant\sum_{i,j=1}^{n}\frac{\boldsymbol{A}_{ij}\left\|\boldsymbol{f}_i^t-\boldsymbol{f}_j^t\right\|_2^2}{2\left\|\boldsymbol{f}_i^t-\boldsymbol{f}_j^t\right\|_2}+\gamma\left\|\boldsymbol{W}^t\right\|_{2,1}+g\left(\boldsymbol{W}^t,\boldsymbol{F}^t,(\boldsymbol{b}^t)^{\mathrm{T}},\boldsymbol{S}^t\right)\tag{21}$$

结合引理 1 的式 (18), 有以下公式:

$$\sum_{i,j=1}^{n}\boldsymbol{A}_{ij}\left(\left\|\boldsymbol{f}_i^{t+1}-\boldsymbol{f}_j^{t+1}\right\|_2-\frac{\left\|\boldsymbol{f}_i^{t+1}-\boldsymbol{f}_j^{t+1}\right\|_2^2}{2\left\|\boldsymbol{f}_i^t-\boldsymbol{f}_j^t\right\|_2}\right)\leqslant\sum_{i,j=1}^{n}\boldsymbol{A}_{ij}\left(\left\|\boldsymbol{f}_i^t-\boldsymbol{f}_j^t\right\|_2-\frac{\left\|\boldsymbol{f}_i^t-\boldsymbol{f}_j^t\right\|_2^2}{2\left\|\boldsymbol{f}_i^t-\boldsymbol{f}_j^t\right\|_2}\right)\tag{22}$$

结合式 (20) 和 (21), 得到

$$\sum_{i,j=1}^{n}\boldsymbol{A}_{ij}\left\|\boldsymbol{f}_i^{t+1}-\boldsymbol{f}_j^{t+1}\right\|_2+\gamma\|\boldsymbol{W}^t\|_{2,1}+g\left(\boldsymbol{W}^t,\boldsymbol{F}^{t+1},(\boldsymbol{b}^t)^{\mathrm{T}},\boldsymbol{S}^t\right)\leqslant\sum_{i,j=1}^{n}\boldsymbol{A}_{ij}\left\|\boldsymbol{f}_i^t-\boldsymbol{f}_j^t\right\|_2+\gamma\|\boldsymbol{W}^t\|_{2,1}+g\left(\boldsymbol{W}^t,\boldsymbol{F}^t,(\boldsymbol{b}^t)^{\mathrm{T}},\boldsymbol{S}^t\right)\tag{23}$$

接下来, 固定 $\boldsymbol{F}$ 为 $\boldsymbol{F}^{t+1}$、$\boldsymbol{b}$ 为 $\boldsymbol{b}^t$、$\boldsymbol{S}$ 为 $\boldsymbol{S}^t$ 来求解 $\boldsymbol{W}^{t+1}$, 根据算法 1 可得出:

$$\boldsymbol{W}^{t+1}=\arg\min_{\boldsymbol{W}}\sum_{i,j=1}^{n}(\tilde{\boldsymbol{A}}^t)_{ij}\left\|\boldsymbol{f}_i^{t+1}-\boldsymbol{f}_j^{t+1}\right\|_2^2+\gamma\|\boldsymbol{W}^t\|_{2,1}+g\left(\boldsymbol{W}^t,\boldsymbol{F}^{t+1},(\boldsymbol{b}^t)^{\mathrm{T}},\boldsymbol{S}^t\right)\tag{24}$$

同样, 参考文献 [14] 中的方法, 可得出:

$$\sum_{i,j=1}^{n}\boldsymbol{A}_{ij}\left\|\boldsymbol{f}_i^{t+1}-\boldsymbol{f}_j^{t+1}\right\|_2+\gamma\left\|\boldsymbol{W}^{t+1}\right\|_{2,1}+g\left(\boldsymbol{W}^{t+1},\boldsymbol{F}^{t+1},(\boldsymbol{b}^t)^{\mathrm{T}},\boldsymbol{S}^t\right)\leqslant\sum_{i,j=1}^{n}\boldsymbol{A}_{ij}\left\|\boldsymbol{f}_i^{t+1}-\boldsymbol{f}_j^{t+1}\right\|_2+\gamma\left\|\boldsymbol{W}^t\right\|_{2,1}+g\left(\boldsymbol{W}^t,\boldsymbol{F}^{t+1},(\boldsymbol{b}^t)^{\mathrm{T}},\boldsymbol{S}^t\right)\tag{25}$$

接下来, 固定 $\boldsymbol{F}$ 为 $\boldsymbol{F}^{t+1}$、$\boldsymbol{W}$ 为 $\boldsymbol{W}^{t+1}$ 求解 $\boldsymbol{b}^{t+1}$、$\boldsymbol{S}^{t+1}$, 同样, 参考文献 [14] 中的方法, 可得出:

$$\sum_{i,j=1}^{n}\boldsymbol{A}_{ij}\left\|\boldsymbol{f}_i^{t+1}-\boldsymbol{f}_j^{t+1}\right\|_2+\gamma\left\|\boldsymbol{W}^{t+1}\right\|_{2,1}+g\left(\boldsymbol{W}^{t+1},\boldsymbol{F}^{t+1},(\boldsymbol{b}^{t+1})^{\mathrm{T}},\boldsymbol{S}^{t+1}\right)\leqslant\sum_{i,j=1}^{n}\boldsymbol{A}_{ij}\left\|\boldsymbol{f}_i^{t+1}-\boldsymbol{f}_j^{t+1}\right\|_2+\gamma\left\|\boldsymbol{W}^t\right\|_{2,1}+g\left(\boldsymbol{W}^t,\boldsymbol{F}^{t+1},(\boldsymbol{b}^t)^{\mathrm{T}},\boldsymbol{S}^t\right)\tag{26}$$

最后, 结合式 (22)、(23) 和 (24), 可得出:

$$\sum_{i,j=1}^{n}\boldsymbol{A}_{ij}\left\|\boldsymbol{f}_i^{t+1}-\boldsymbol{f}_j^{t+1}\right\|_2+\gamma\left\|\boldsymbol{W}^{t+1}\right\|_{2,1}+g\left(\boldsymbol{W}^{t+1},\boldsymbol{F}^{t+1},(\boldsymbol{b}^{t+1})^{\mathrm{T}},\boldsymbol{S}^{t+1}\right)\leqslant\sum_{i,j=1}^{n}\boldsymbol{A}_{ij}\left\|\boldsymbol{f}_i^t-\boldsymbol{f}_j^t\right\|_2+\gamma\left\|\boldsymbol{W}^t\right\|_{2,1}+g\left(\boldsymbol{W}^t,\boldsymbol{F}^t,(\boldsymbol{b}^t)^{\mathrm{T}},\boldsymbol{S}^t\right)\tag{27}$$

从式 (25) 中可证明出算法 1 是收敛的。

3　实验分析

3.1　对比方法及实验数据

为验证本文方法的有效性, 对相关方法进行

描述。

All-feature: 其数据未采用特征选择, 本次实验以该分类结果作为基准线。

TRCFS(trace ratio criterion for feature selection)[6]: 采用谱图的半监督学习方法, 其通过引入具有抗噪声的率切准则提高特征选择的效率。

CSFS(convex semi-supervised multi-label feature selection)[12]: 该将线性回归模型引入特征选择模型中, 是一种快速的半监督特征选择方法。

FSNM(feature selection via joint $l_{2,1}$-norms minimization)[1]: 监督学习方法, 其在损失函数和正则化方面采用 $l_{2,1}$ 范数模型进行特征选择。

本次实验将所提出方法应用到各种场景, 包括自然场景分类、网页分类和基因功能分类。同时本文将各方法应用到多种开源数据库, 包括 MIML[15]、YEAST[16]、Education[17] 和 Entertainment[17], 数据集的相关属性描述如表 1 所示。

本文对于每一种方法所有涉及到的参数 (如果有的话) 的范围设定为 $\{10^{-4}、10^{-2}、10^{0}、10^{2}、10^{4}\}$。对于每种数据集, 随机选择 n 个样本作为训练集, 其中分别选择 10%、20% 和 40% 的数据为已标记数据集, 其余为未标记数据。实验独立重复 5 次, 最后取其平均值。本次实验选择 MLKNN 作为多标签分类器, 其中 MLKNN 的优化参数参照文献 [18]。

表 1　实验数据集

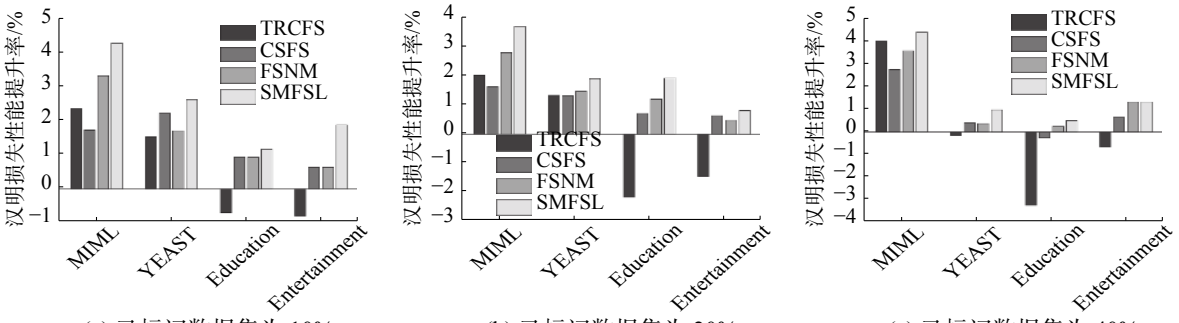
Table 1　Experimental datasets

数据集	样本数	特征数	类别数	特征数
MIML	2 000	135	5	{20, 40, 60, 80, 100, 120}
YEAST	2 417	103	14	{20, 40, 60, 80, 100}
Education	5 000	550	33	{100, 200, 300, 400, 500}
Entertainment	5 000	640	21	{100, 200, 300, 400, 500, 600}

3.2　性能对比分析

本次实验选择 Hamming loss(HL, 汉明损失)、One-Error(OE, 单错误) 作为评价指标[19] 用来评价方法的分类性能。其中, Hamming loss 和 One Error 值越低代表性能越好。图 1、2 列出了以 All-feature 方法的分类结果为基准线, TRCFS、CSFS、FSNM 以及本文提出的 SMFSL 方法在各种数据集的性能对比分析。其中图 1 分别为 HL 评价标准的性能提升百分比, 以及图 2 分别为 OE 评价标准的性能提升百分比。从图中可以看出:

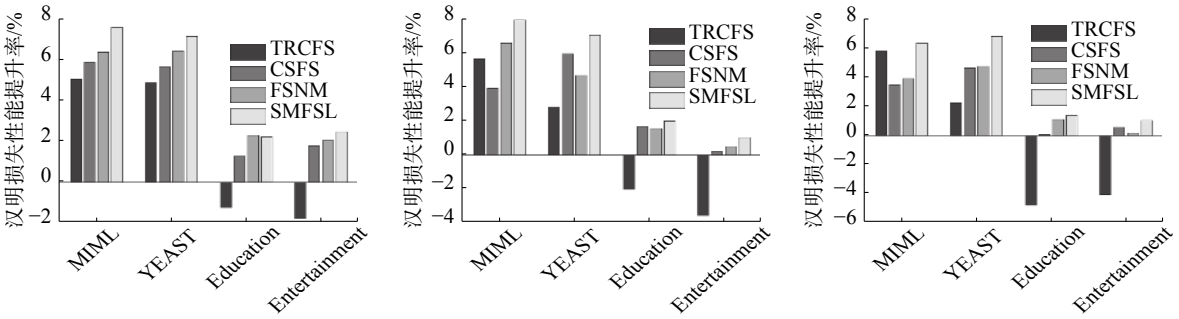
1) 大部分特征选择方法要优于未采用特征选择的 All-feature, 由此可证明特征选择有助于提高多标记学习性能。但在 YEAST、Education 和 Entertainment 数据集中, TRCFS 学习性能整体略低于 All-feature, 但该方法经过特征选择后维度会有所降低, 从而能有效地节省后续分类时间。



(a) 已标记数据集为 10%　(b) 已标记数据集为 20%　(c) 已标记数据集为 40%

图 1　各种方法在各数据集上的汉明损失

Fig. 1　Hamming loss comparison of various methods on various datasets



(a) 已标记数据集为 10%　(b) 已标记数据集为 20%　(c) 已标记数据集为 40%

图 2　各种方法在各数据集上的单错误

Fig. 2　One-Error comparison of various methods on various datasets

2) CSFS 在部分数据集上表现略差于 FSNM, 这是由于 CSFS 方法采用 l_2 范式损失函数, 对噪声较敏感。当无标签数据集比例较大时, 噪声较多, CSFS 的分类性能受噪声影响较大。而当已标记数据集比例增加时, 其与 FSNM 差距越小。

3) 本文提出的 SMFSL 方法优于其他方法, 由此可证明采用 $l_{2,1}$ 范式约束全局回归函数能有效减少外界噪声影响, 同时结合 l_1 图建立相似矩阵能有效获取底层局部流形结构, 提高分类性能。

3.3　模型鲁棒性分析

为验证本文所提出模型的鲁棒性, 本次实验针对不同类型相似矩阵进行对比分析。实验采用如图 3(a) 所示的双月 (two-moon) 数据集。该数据集可分为上半月形和下半月形 2 个类别数据, 具有明显的流形结构。基于 l_2 范数的相似矩阵如图 3(b) 所示, 其中任意 2 个数据间的连线代表其具有相似性。从图中可以看出, 该相似矩阵存在过多冗余连接, 无法提取清晰的流形结构信息, 很难直接应用于后续对数据的分类任务。本文模型输出的相似矩阵如图 3(c) 所示, 可明显看出, 在 l_1 范数稀疏性约束下, 该相似矩阵可有效剔除数据间的无关连接, 提取更加清晰的流形结构信息, 进而有助于提高分类模型的准确性和对外界噪声的抗干扰性。

3.4　特征数对分类性能的影响

本次实验挑选了 MIML、YEAST 和 Education 数据集在已标记样本集为 20% 时, 选择不同特征数的 Average Precision(AP, 平均查准率) 性能[19] 对比效果, 具体如图 4 所示。据文献 [19] 得知, AP 值越高, 表示方法性能越好。从图中可以看出, MIML、YEAST、Education 数据集在选择特征数分别为 40、80 和 400 时 AP 值最高。这意味着选择最大特征数并不一定能产生最高的 AP 值。在给定不同的特征数量时, 本文所提方法普遍高于其他方法, 尤其是在 MIML 和 Education 数据集优势更加明显。另外, 从图中看出不同方法的结果曲线出现交叉, 这是由于不同方法所选择出的最优特征子集不同, 其对应的分类准确度也会有所不同。

3.5　参数敏感性分析

本次实验将对学习模型中涉及的参数进行具体分析。为节省篇幅, 本次实验挑选 YEAST、Education 和 Entertainment 在已标记数据为 40%、评价标准为 One Error 的性能分析。首先, 固定 α 值为 1, 即参数调节范围的中位数, 调整 γ 和特征数进行分析, 结果如图 5 所示。同样, 固定参数 γ 的值为 1, 对 α 和特征数的变化进行分析, 具体如图 6 所示。从图 5、图 6 可以看出, 参数 α、γ 的选择依赖于所选数据集, 如 Entertainment 数据集在固定 α、特征数为 600 时, 选择 $\gamma=10^{-4}$ 时 One Error 性能最佳, 而当 $\gamma=1$ 该性能表现最差。因此在实验测试时需对不同的数据集设置不同的参数。

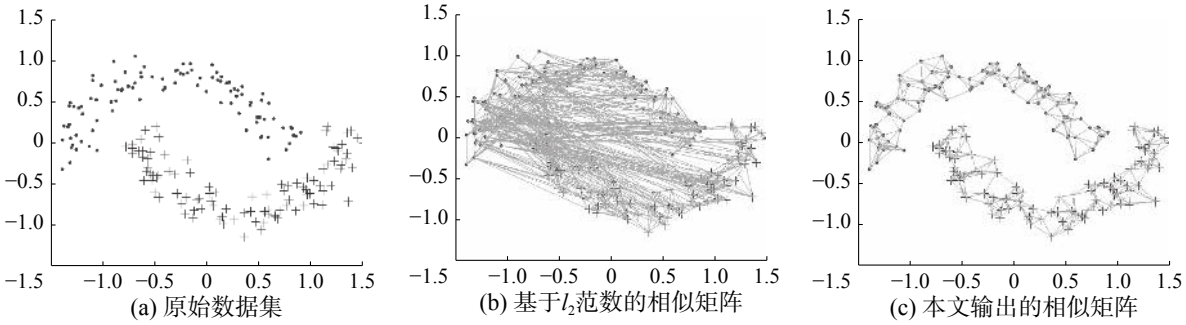


图 3　l_1 范数对相似矩阵的影响分析

Fig. 3　Influence analysis of l_1-norm on similarity matrix

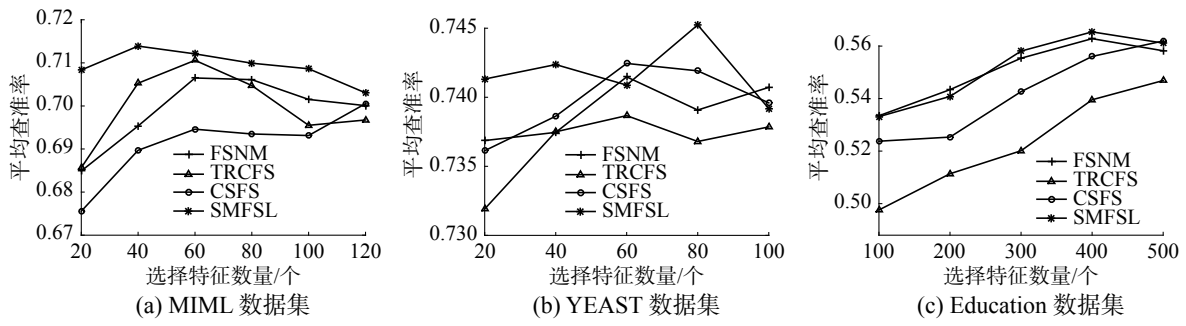
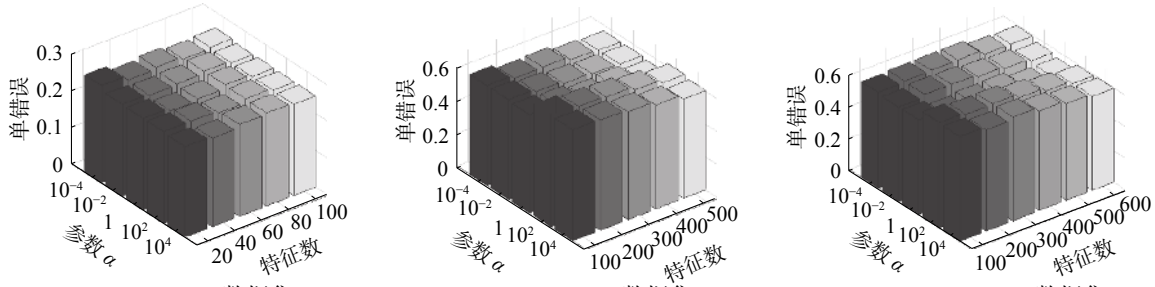


图 4　平均查准率与选择的特征数量关系

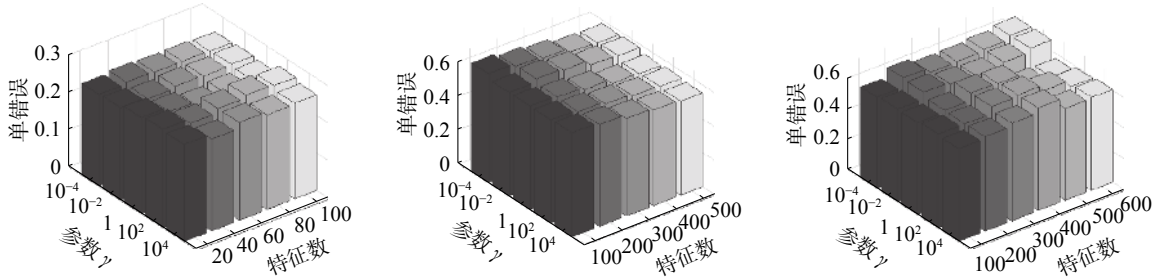
Fig. 4　Relation between average precision and the number of selected features



(a) YEAST 数据集　(b) Education 数据集　(c) Entertainment 数据集

图 5　本文方法的参数敏感性分析 (a=1)

Fig. 5　Parameter sensitivity analysis of the proposed method (a=1)



(a) YEAST 数据集　(b) Education 数据集　(c) Entertainment 数据集

图 6　本文方法的参数敏感性分析 (γ=1)

Fig. 6　Parameter sensitivity analysis of the proposed method (γ=1)

4　结束语

本文提出一种鲁棒的半监督多标签特征选择方法 SMFSL。不同于传统基于谱图的特征选择方法，本文方法利用 $l_{2,1}$ 范数约束全局线性回归函数，减少外界噪声干扰，还借助 l_1 图获取更清晰的数据底层流形结构，有效提取局部特征，以提高学习效率。为提高特征选择准确度，本文引入 $l_{2,1}$ 范数约束特征选择过程，有效利用特征间相关信息，进而过滤冗余特征。文中所提出的模型涉及 $l_{2,1}$ 范数具有非光滑特征，无法直接对其求闭合解，因此提出一套快速有效迭代方法求解学习模型。最后通过多个开源数据集实验证明了本文方法的有效性。结合自适应学习及采用鲁棒性更好的损失函数以进一步提高特征选择的准确度，为本文的下一步研究目标。

参考文献：

[1] YU Lei, LIU Huan. Feature selection for high-dimensional data: a fast correlation-based filter solution[C]//Proceedings of the 20th International Conference on International Conference on Machine Learning. Washington DC, USA, 2003: 856-863.

[2] 胡敏杰, 林耀进, 杨红和, 等. 基于特征相关的谱特征选择算法 [J]. 智能系统学报, 2017, 12(4): 519–525.
HU Minjie, LIN Yaojin, YANG Honghe, et al. Spectral feature selection based on feature correlation[J]. CAAI transactions on intelligent systems, 2017, 12(4): 519–525.

[3] YANG Yi, SHEN Hengtao, MA Zhigang, et al. $\ell_{2,1}$-norm regularized discriminative feature selection for unsupervised learning[C]//Proceedings of the 22th International Joint Conference on Artificial Intelligence. Barcelona, Spain, 2011: 1589-1594.

[4] WANG Xiaodong, ZHANG Xu, ZENG Zhiqiang, et al. Unsupervised spectral feature selection with l_1 norm graph[J]. Neurocomputing, 2016, 200: 47–54.

[5] DOQUIRE G, VERLEYSEN M. A graph Laplacian based approach to semi-supervised feature selection for regression problems[J]. Neurocomputing, 2013, 121: 5–13.

[6] LIU Yun, NIE Feiping, WU Jigang, et al. Efficient semi-supervised feature selection with noise insensitive trace ratio criterion[J]. Neurocomputing, 2013, 105: 12–18.

[7] MA Zhigang, NIE Feiping, YANG Yi, et al. Discriminating joint feature analysis for multimedia data understanding[J]. IEEE transactions on multimedia, 2012, 14(6): 1662–1672.

[8] JI Shuiwang, TANG Lei, YU Shipeng, et al. A shared-subspace learning framework for multi-label classification[J]. ACM transactions on knowledge discovery from data, 2010, 4(2): 8.

[9] 张俐, 王枞. 基于最大相关最小冗余联合互信息的多标签特征选择算法 [J]. 通信学报, 2018, 39(5): 111–122.
ZHANG Li, WANG Cong. Multi-label feature selection algorithm based on joint mutual information of max-relev-

ance and min-redundancy[J]. Journal on communications, 2018, 39(5): 111–122.

[10] ALALGA A, BENABDESLEM K, TALEB N. Soft-constrained Laplacian score for semi-supervised multi-label feature selection[J]. Knowledge and information systems, 2016, 47(1): 75–98.

[11] CHANG Xiaojun, NIE Feiping, YANG Yi, et al. A convex formulation for semi-supervised multi-label feature selection[C]//Proceedings of the 28th AAAI Conference on Artificial Intelligence. Québec City, Québec, Canada, 2014: 1171-1177.

[12] ZHOU Sihang, LIU Xinwang, ZHU Chengzhang, et al. Spectral clustering-based local and global structure preservation for feature selection[C]//Proceedings of 2014 International Joint Conference on Neural Networks. Beijing, China, 2014: 550-557.

[13] NIE Feiping, WANG Hua, HUANG Heng, et al. Unsupervised and semi-supervised learning via ℓ_1 norm graph[C]//Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain, 2011: 2268-2273.

[14] LIU Yun, GUO Yiming, WANG Hua, et al. Semi-supervised classifications via elastic and robust embedding[C]// Proceedings of the 31th AAAI Conference on Artificial Intelligence. San Francisco, USA, 2017: 2294-2300.

[15] SCHÖLKOPF B, PLATT J, HOFMANN T. Multi-instance multi-label learning with application to scene classification[C]//Proceedings of the 13th International Conference on Neural Information Processing Systems. Hong Kong, China, 2006: 1609-1616.

[16] ELISSEEFF A, WESTON J. A kernel method for multi-labelled classification[C]//Proceedings of the 14th International Conference on Neural Information Processing Systems. Vancouver, Canada, 2001: 681-687.

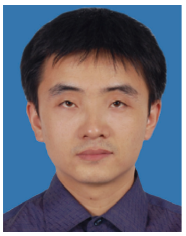
[17] UEDA N, SAITO K. Parametric mixture models for multi-labeled text[C]//Proceedings of the 15th International Conference on Neural Information Processing Systems. Cambridge, MA, USA, 2002: 737-744.

[18] ZHANG Minling, ZHOU Zhihua. ML-KNN: a lazy learning approach to multi-label learning[J]. Pattern recognition, 2007, 40(7): 2038–2048.

[19] MARON O, RATAN A L. Multiple-instance learning for natural scene classification[C]//Proceedings of 15th International Conference on Machine Learning. San Francisco, CA, USA, 1998: 341-349.

作者简介:

严菲. 女, 1985 年生, 实验师, 主要研究方向为特征选择、机器学习。主持福建省教育厅中青年教师项目 1 项。发表学术论文 5 篇。



王晓栋, 男, 1983 年生, 副教授, 博士, 主要研究方向为机器学习、图像处理。主持福建省自然科学基金面上项目 1 项, 福建教育厅中青年教师项目 1 项。发表学术论文 10 篇。