

基于动态图拉普拉斯的多标签特征选择

李永豪^{1,2}, 胡亮^{1,2}, 张平^{1,2}, 高万夫^{1,2,3}

(1. 吉林大学计算机科学与技术学院, 吉林 长春 130012; 2. 吉林大学符号计算与知识工程教育部重点实验室, 吉林 长春 130012;
3. 吉林大学化学学院, 吉林 长春 130012)

摘要: 针对基于图的多标签特征选择方法忽略图拉普拉斯矩阵的动态变化, 且利用逻辑标签来指导特征选择过程而丢失标签信息等问题, 提出了一种基于动态图拉普拉斯矩阵和实值标签的多标签特征选择方法。该方法利用特征矩阵的稳健低维空间构造动态图拉普拉斯矩阵, 并利用该稳健低维空间作为实值标签空间, 进一步使用流形约束和非负约束将逻辑标签转化为实值标签, 以此来解决上述问题。所提方法与 3 种多标签特征选择方法在 9 个多标签基准数据集上进行了对比实验, 实验结果表明, 所提多标签特征选择方法可得到高质量的特征子集, 并且能获得很好的分类表现。

关键词: 多标签特征选择; 动态图拉普拉斯矩阵; 实值标签; 分类

中图分类号: TP18

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2020244

Multi-label feature selection based on dynamic graph Laplacian

LI Yonghao^{1,2}, HU Liang^{1,2}, ZHANG Ping^{1,2}, GAO Wanfu^{1,2,3}

1. College of Computer Science and Technology, Jilin University, Changchun 130012, China

2. Key Laboratory of Symbolic Computation and Knowledge Engineering, Ministry of Education, Jilin University, Changchun 130012, China

3. College of Chemistry, Jilin University, Changchun 130012, China

Abstract: In view of the problems that graph-based multi-label feature selection methods ignore the dynamic change of graph Laplacian matrix, as well as such methods employ logical-value labels to guide feature selection process and loses label information, a multi-label feature selection method based on both dynamic graph Laplacian matrix and real-value labels was proposed. The robust low-dimensional space of feature matrix was used to construct a dynamic graph Laplacian matrix, and the robust low-dimensional space was used as the real-value label space. Furthermore, manifold and non-negative constraints were adopted to transform logical labels into real-valued labels to address the issues mentioned above. The proposed method was compared to three multi-label feature selection methods on nine multi-label benchmark data sets in experiments. The experimental results demonstrate that the proposed multi-label feature selection method can obtain the higher quality feature subset and achieve good classification performance.

Key words: multi-label feature selection, dynamic graph Laplacian matrix, real-value label, classification

1 引言

进入大数据时代后, 万物互联产生了海量的数

据, 其中高维数据导致的维度诅咒问题非常引人注意, 处理这些高维数据对现有的方法来说是一个巨大的挑战^[1-2]。更进一步地, 在这些高维数据中存在

收稿日期: 2020-07-20; 修回日期: 2020-10-18

通信作者: 高万夫, gaowf@jlu.edu.cn

基金项目: 博士后创新人才支持计划基金资助项目 (No.BX20190137); 中国博士后科学基金资助项目 (No.2020M670839); 国家重点研发计划基金资助项目 (No.2017YFA0604500); 吉林省重点科技研发基金资助项目 (No.20180201103GX)

Foundation Items: Postdoctoral Innovative Talents Support Program under Grant (No.BX20190137), China Postdoctoral Science Foundation Founded Project (No.2020M670839), The National Key Research and Development Program of China (No.2017YFA0604500), Key Scientific and Technological Research and Development Plan of Jilin Province (No.20180201103GX)

的多标签数据也越来越凸显其现实应用价值^[3-4]。与早期的单标签数据不同，多标签数据中每个样例都可能与多个不同的标签有关联^[5-7]。例如，各类音乐软件中对歌曲进行分类时，同一首歌曲可能标记不同风格的标签。如何有效地对各类多标签数据进行分类逐渐成为研究的热点^[6]。然而，高维多标签数据中存在大量特征，这些特征中包含的不相关信息严重削弱了多标签学习算法的分类性能^[1,8]。如何找到一个紧凑的、与标签相关的特征子集是一个棘手而紧迫的问题。特征选择算法可以从原始数据中获得一个最优特征子集，它不仅实现了特征降维，而且保留了原始数据的直观意义和物理解释^[9]。因此，特征选择技术成为图像、视频、文本和基因等存在大量多标签数据的领域的热门预处理方法^[10]。一般来说，多标签特征选择分为过滤式模型、包装式模型和嵌入式模型^[11-13]。过滤式模型与后续学习算法无关^[14]，而包装式模型依赖于学习算法。与过滤式和包装式模型不同，嵌入式模型将特征选择嵌入学习算法中。本文重点研究嵌入式模型。

在嵌入式模型的特征选择方法中，基于图的特征选择方法备受关注。传统的基于图的特征选择方法严格依赖于固定的图拉普拉斯矩阵，其通常采用两步策略^[15-17]：1)构造对称亲和矩阵；2)利用对称亲和矩阵指导特征选择过程，得到图拉普拉斯矩阵。然而，这种策略忽略了图拉普拉斯矩阵在算法执行过程中的动态变化。具体地，在特征选择算法执行过程中，算法的每一次更新迭代应该依赖本次迭代的图拉普拉斯矩阵。传统基于图的特征选择算法在每次算法迭代过程中，并没有选择适合本次迭代的图拉普拉斯矩阵，前一次更新造成的误差会被后续的更新不断放大。因此，特征选择方法无法获得令人满意的分类性能。另外，在有监督的多标签特征选择方法中还存在一个问题，大多数基于图的特征选择方法利用逻辑标签来指导特征选择^[18-19]，然而逻辑标签不能很好地反映相应标签的重要性，即逻辑标签无法刻画标签本身的重要程度，而且多标签数据涉及大量不同标签，这些标签之间相关性复杂，这些问题导致多标签特征选择方法无法获得令人满意的分类性能。本文针对上述问题，设计了一种动态图拉普拉斯的多标签特征选择方法。首先，本文构造了一个稳健的低维空间；其次，利用基于低维空间的动态图拉普拉斯矩阵指导特征选择过程；最后，通过在不同领域数据上的实验证明

了所提方法的分类优势。本文主要贡献如下。

1) 设计了一种基于特征矩阵的稳健低维空间的动态变化图拉普拉斯矩阵来指导特征选择。

2) 在所获得的图拉普拉斯动态更新基础上，为避免逻辑标签造成的信息丢失，将逻辑标签转化为实值标签。

3) 设计了一种基于动态图拉普拉斯矩阵和实值标签的多标签特征选择方法，针对该方法提出一种有效的优化方案，并证明了该方案的收敛性。

4) 通过在 9 个多标签基准数据集上与 3 个多标签特征选择方法的对比实验验证了该方法的分类优越性。

2 相关工作

2.1 相关符号

对于任意矩阵 $A \in \mathbb{R}^{n \times m}$ ， A_i 和 A_j 分别表示矩阵 A 的第 i 行和第 j 列。当 A 是一个方阵时， $\text{Tr}(A)$ 表示矩阵 A 的迹。 A^T 表示矩阵 A 的转置。

$\|A\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m A_{ij}^2}$ 和 $\|A\|_{2,1} = \sum_{i=1}^n \sqrt{\sum_{j=1}^m A_{ij}^2}$ 分别表示矩阵 A 的 Frobenius 范数和 $l_{2,1}$ 范数，其中 A_{ij} 表示矩阵 A 的第 i 行第 j 个元素。 $X \in \mathbb{R}^{n \times d}$ 表示特征矩阵，

$Y \in \mathbb{R}^{n \times c}$ 表示标签矩阵，其中 n 、 d 、 c 分别表示样例数目、特征数目和标签数目。当特征矩阵的第 i 行 X_i 与标签矩阵的第 j 列 Y_j 有关联时， $Y_{ij} = 1$ ；否则 $Y_{ij} = 0$ 。

2.2 相关工作

在多标签学习中，已有许多行之有效的方法处理来自不同领域的多标签数据。基于不同程度的标签相关性可以将这些方法分为一阶策略、二阶策略和高阶策略^[3]。一阶策略利用传统的单标签方法处理多标签数据，忽略了标签相关性，其中代表性方法有二元关联 (BR, binary relevance)^[20]等。由于注意到标签相关性的重要性，研究者开始考虑二阶策略，即利用标签之间的成对关系，相关多标签视频标注 (CMLVA, correlative multi-label video annotation) 和校准标签排名 (CLR, calibrated label ranking) 是二阶策略的代表方法^[21-22]。高阶策略通常考虑标签子集或所有标签的相关性，其优点是充分考虑了标签相关性，缺点是时间复杂度高且计算量大，如 LLSFC-DL (learning label-specific features

and class-dependent label)^[23]为高阶策略。本文中采用二阶策略。

近年来，图模型在数据结构挖掘方面取得显著成就，受到研究者的广泛关注。典型的图模型是流形学习，其目的是在高维空间嵌入低维空间时，保持数据的几何结构^[24]。许多方法都是从样例的角度来考虑流形结构的，即如果 \mathbf{X}_i 与 \mathbf{X}_j 具有很强的相似性，那么 \mathbf{Y}_i 与 \mathbf{Y}_j 的相似性也会很强。Ren 等^[25]提出了一种无监督特征选择方法来保持实例关联的局部和全局结构。Huang 等^[26]提出了一种考虑样例相关性的基于流形的约束拉普拉斯评分方法。Xu 等^[27]提出了一种半监督多标签特征选择方法，考虑保持特征空间与标签空间的一致性。Chen 等^[28]提出的半监督多标签学习方法考虑了样例关联和标签关联。但是，这些利用图模型的方法都严重依赖于固定的图拉普拉斯矩阵，而忽略了特征选择中图拉普拉斯矩阵的动态变化，图拉普拉斯矩阵的不同设定会对后续的更新策略产生不同的影响，尤其是前一次更新造成的误差会被后续的更新不断放大。上述方法也存在利用逻辑标签来指导标签分类的问题，逻辑标签并不能很好地反映相应标签的重要性，而且多标签数据涉及大量标签，导致标签相关性更加复杂，因此特征选择方法无法获得令人满意的分类性能^[29]。

多标签特征选择方法广泛采用了一些不同的标准，如基于互信息的方法和基于稀疏学习的方法^[8]。本文回顾了几种有代表性的多标签特征选择方法。Lee 等^[30]采用可扩展的相关性评估标准来评估条件相关性，提出了一种新的多标签特征选择方法，即大标签集的可扩展准则 (SCLS, scalable criterion for large label set)。然而，SCLS 的特征和标签组合呈指数式增长，可能导致性能下降。Jian 等^[17]设计了一种基于稀疏化的多标签信息特征选择 (MIFS, multi-label informed feature selection) 方法，利用标签的局部几何结构和低秩潜在标签矩阵来消除无关特征。MIFS 的形式为

$$\min_{\mathbf{W}, \mathbf{V}, \mathbf{B}} \|\mathbf{X}\mathbf{W} - \mathbf{V}\|_{\text{F}}^2 + \alpha \|\mathbf{Y} - \mathbf{V}\mathbf{B}\|_{\text{F}}^2 + \beta \text{Tr}(\mathbf{V}^{\text{T}}\mathbf{L}\mathbf{V}) + \gamma \|\mathbf{W}\|_{2,1} \quad (1)$$

其中， $\mathbf{X} \in \mathbb{R}^{n \times d}$ ， $\mathbf{Y} \in \mathbb{R}^{n \times l}$ ， $\mathbf{W} \in \mathbb{R}^{d \times c}$ 分别表示特征矩阵、标签矩阵和权重矩阵； $\mathbf{V} \in \mathbb{R}^{n \times c}$ 和 $\mathbf{B} \in \mathbb{R}^{c \times l}$ 分别表示潜在标签矩阵和系数矩阵； $\mathbf{L} \in \mathbb{R}^{n \times n}$ 表示

拉普拉斯矩阵； α 、 β 和 γ 表示 MIFS 方法的 3 个正则化超参数； c 表示标签的聚类簇数。

Cai 等^[31]提出了一种基于稀疏学习的特征选择方法，该方法被称为稳健的增光拉格朗日乘子特征选择 (RALM-FS, robust augmented Lagrange multiplier for feature selection)。RALM-FS 对权重矩阵施加 $l_{2,0}$ 范数，从而获得目标函数如式(2)所示。

$$\min_{\mathbf{W}, \mathbf{V}, \mathbf{B}} \|\mathbf{Y} - \mathbf{X}\mathbf{W} - \mathbf{1}\mathbf{b}^{\text{T}}\|_{2,1} \quad (2)$$

$$\text{s.t. } \|\mathbf{W}\|_{2,0} = q$$

其中， $\mathbf{1}$ 表示元素全为 1 的列向量， \mathbf{b} 表示偏置列向量， q 表示所选特征数目。

3 动态图拉普拉斯多标签特征选择方法描述

3.1 设计方法

本节提出一种新的多标签特征选择算法，考虑到图拉普拉斯的动态变化能够提供更有指导，并且为避免逻辑标签造成的性能退化，将逻辑标签转化为实值标签，加强挖掘特征选择过程中的标签相关性，使用式(3)所示的学习框架。

$$\mathbf{W}^*, \mathbf{F}^* = \underset{\mathbf{W}, \mathbf{F}}{\text{argmin}} Q(\mathbf{X}, \mathbf{W}, \mathbf{F}) + \lambda \mathcal{M}(\mathbf{F}, \mathbf{Y}) + \gamma \mathcal{Q}(\mathbf{W}) \quad (3)$$

其中，第一项表示损失函数；第二项和第三项表示对该损失函数进行正则化处理，帮助减少损失函数造成的损失； $\mathbf{F} \in \mathbb{R}^{n \times c}$ 表示重构的标签矩阵。

通常，损失函数利用最小二乘回归模型学习从特征空间到标签空间的映射矩阵 \mathbf{W} ，但这种模型对于数据中存在的异常值非常敏感，特别是基于图模型的学习模型对异常值的抗干扰能力非常弱。为获得一个更加稳健的损失函数，本文设计了如式(4)所示的形式。

$$\Theta(\mathbf{W}, \mathbf{F}) = \text{Tr}(\mathbf{W}\mathbf{L}_F\mathbf{W}^{\text{T}}) + \|\mathbf{X}\mathbf{W} - \mathbf{F}\|_{2,1} \quad (4)$$

其中， $\Theta(\mathbf{W}, \mathbf{F})$ 表示关于 \mathbf{W} 和 \mathbf{F} 的函数； $\mathbf{W} \in \mathbb{R}^{d \times c}$ 表示特征权重矩阵，用于度量特征矩阵 \mathbf{X} 中每一个特征的重要性，即 $\|\mathbf{W}_i\|_2$ 值越大，第 i 个特征的影响越大； $\|\cdot\|_{2,1}$ 表示 $l_{2,1}$ 范数，可以有效减少异常值的干扰^[18]； $\text{Tr}(\mathbf{W}\mathbf{L}_F\mathbf{W}^{\text{T}})$ 的设计受文献[32]启发，即在多标签学习中保持标签的局部几何结构是至关重要的，本文利用上述图模型来保持标签的局部几何结构。与传统的图模型不同，本文设计的拉普拉斯矩阵 \mathbf{L}_F 与重构标签矩阵 \mathbf{F} 紧密相关， \mathbf{L}_F 随 \mathbf{F} 的更新

而变化。 $\text{Tr}(\mathbf{W}\mathbf{L}_F\mathbf{W}^T)$ 的构造过程如下

$$\begin{aligned} & \frac{1}{2} \sum_{i=1}^c \sum_{j=1}^c S_{Fij} \|\mathbf{W}_{\cdot i} - \mathbf{W}_{\cdot j}\|_2^2 = \\ & \frac{1}{2} \sum_{i=1}^c \sum_{j=1}^c S_{Fij} (\mathbf{W}_{\cdot i} - \mathbf{W}_{\cdot j})(\mathbf{W}_{\cdot i} - \mathbf{W}_{\cdot j})^T = \\ & \sum_{i=1}^c \mathbf{W}_{\cdot i} \mathbf{W}_{\cdot i}^T A_{Fii} - \sum_{i=1}^c \sum_{j=1}^c \mathbf{W}_{\cdot i} \mathbf{W}_{\cdot j}^T S_{Fij} = \\ & \text{Tr}(\mathbf{W}(\mathbf{A}_F - \mathbf{S}_F)\mathbf{W}^T) = \text{Tr}(\mathbf{W}\mathbf{L}_F\mathbf{W}^T) \end{aligned} \quad (5)$$

其中, $\mathbf{L}_F = \mathbf{A}_F - \mathbf{S}_F$ 表示基于重构标签矩阵的图拉普拉斯矩阵, \mathbf{A}_F 和 \mathbf{S}_F 分别表示度矩阵和近邻矩阵, $A_{Fii} = \sum_{j=1}^c S_{Fij}$ 。近邻矩阵 \mathbf{S}_F 采用式(6)所示的热核函数来构建。

$$S_{Fij} = \begin{cases} \frac{\|\mathbf{F}_{\cdot i} - \mathbf{F}_{\cdot j}\|_2^2}{\sigma}, & \mathbf{F}_{\cdot i} \in (\mathcal{N})_p(\mathbf{F}_{\cdot j}) \text{ 或 } \mathbf{F}_{\cdot j} \in (\mathcal{N})_p(\mathbf{F}_{\cdot i}) \\ 0, & \text{其他} \end{cases} \quad (6)$$

其中, $(\mathcal{N})_p(\mathbf{F}_{\cdot j})$ 表示 $\mathbf{F}_{\cdot j}$ 的 p 个最近邻集合, σ 表示热核函数的带宽参数。根据文献[29]可知, 传统的有监督多标签方法利用逻辑标签评价输入数据与输出数据之间的相关性, 不能很好地反映相应标签的重要性, 因此将式(3)中的第二项设计为式(7)所示形式。

$$\alpha \text{Tr}(\mathbf{F}^T \mathbf{L}_X \mathbf{F}) + \beta \|\mathbf{F} - \mathbf{Y}\|_F^2 \quad (7)$$

其中, α 和 β 表示正则超参数; $\mathbf{F} \in \mathbb{R}^{n \times c}$ 与逻辑标签矩阵 \mathbf{Y} 同型, 但 \mathbf{F} 显然是连续数值型的, 这种连续性可以较好地刻画标签的重要程度。将 \mathbf{F} 作为一个更新变量时, 为了保证 \mathbf{F} 和 \mathbf{Y} 之间的结构一致性, 本文采用常规的图模型对 \mathbf{F} 进行约束。根据 \mathbf{W} 可度量特征矩阵 \mathbf{X} 中每一个特征的重要性这一特性, 本文方法能够有效实现特征选择。式(3)中的第三项被设计为 $\|\mathbf{W}\|_{2,1}$, 可有效实现稀疏化的特征筛选。最终目标函数被构造为式(8)所示形式。

$$\begin{aligned} \Theta(\mathbf{W}, \mathbf{F}) = & \text{Tr}(\mathbf{W}\mathbf{L}_F\mathbf{W}^T) + \|\mathbf{X}\mathbf{W} - \mathbf{F}\|_{2,1} + \\ & \alpha \text{Tr}(\mathbf{F}^T \mathbf{L}_X \mathbf{F}) + \beta \|\mathbf{F} - \mathbf{Y}\|_F^2 + \gamma \|\mathbf{W}\|_{2,1} \end{aligned} \quad (8)$$

其中, γ 表示稀疏化正则超参数, 用于调整目标函数的稀疏程度; $\|\cdot\|_{2,1}$ 控制 \mathbf{W} 的行稀疏性^[18]。但是行稀疏性并不能总被保证^[33], 同时, \mathbf{Y} 中仅包含 0 和 1 这 2 种非负元素, 因此需要避免 \mathbf{F} 的元素负值化。

根据上述原因, 本文对 \mathbf{W} 和 \mathbf{F} 实施了非负约束, 最终的目标函数构造如下

$$\begin{aligned} \Theta(\mathbf{W}, \mathbf{F}) = & \text{Tr}(\mathbf{W}\mathbf{L}_F\mathbf{W}^T) + \|\mathbf{X}\mathbf{W} - \mathbf{F}\|_{2,1} + \\ & \alpha \text{Tr}(\mathbf{F}^T \mathbf{L}_X \mathbf{F}) + \beta \|\mathbf{F} - \mathbf{Y}\|_F^2 + \gamma \|\mathbf{W}\|_{2,1} \\ \text{s.t. } & \{\mathbf{W}, \mathbf{F}\} \geq 0 \end{aligned} \quad (9)$$

3.2 优化方案

本节设计了一套针对式(9)所示目标函数的简单有效的优化方案。根据分析可以得出, 目标函数关于 \mathbf{W} 和 \mathbf{F} 是联合非凸的。由于 $l_{2,1}$ 范数的存在, 导致目标函数存在非光滑性问题。因此, 本节提出了一种交替迭代的方法来解决非凸问题, 同时引入了一种松弛化方法来处理非光滑问题^[18], 获得了拉格朗日函数, 如式(10)所示。

$$\begin{aligned} \mathcal{L}(\mathbf{W}, \mathbf{F}) = & \text{Tr}(\mathbf{W}\mathbf{L}_F\mathbf{W}^T) + \\ & 2 \text{Tr}((\mathbf{X}\mathbf{W} - \mathbf{F})^T \mathbf{D}_1 (\mathbf{X}\mathbf{W} - \mathbf{F})) + \\ & \alpha \text{Tr}(\mathbf{F}^T \mathbf{L}_X \mathbf{F}) + \beta \|\mathbf{F} - \mathbf{Y}\|_F^2 + \\ & 2\gamma \text{Tr}(\mathbf{W}^T \mathbf{D}_2 \mathbf{W}) - \text{Tr}(\Phi \mathbf{W}^T) - \text{Tr}(\Psi \mathbf{F}^T) \end{aligned} \quad (10)$$

其中, $\mathcal{L}(\mathbf{W}, \mathbf{F})$ 表示关于 \mathbf{W} 和 \mathbf{F} 的拉格朗日函数; $\Phi \in \mathbb{R}_+^{d \times c}$ 和 $\Psi \in \mathbb{R}_+^{n \times c}$ 表示 2 个与 \mathbf{W} 和 \mathbf{F} 同型的拉格朗日乘子, 这 2 项同时将非负约束条件整合到目标函数中, 从而方便了优化方案的设计; \mathbf{D}_1 和 \mathbf{D}_2 是 2 个对角矩阵, 其第 i 个对角元素分别为

$$\begin{cases} D_{1_{ii}} = \frac{1}{2\|\mathbf{X}\mathbf{W} - \mathbf{F}\|_{2,1} + \epsilon} \\ D_{2_{ii}} = \frac{1}{2\|\mathbf{W}_{\cdot i}\|_2 + \epsilon} \end{cases} \quad (11)$$

其中, ϵ 是一个非负的极小常数。对式(10)分别求 \mathbf{W} 和 \mathbf{F} 的偏导数, 可得

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \mathbf{W}} = 2\mathbf{W}\mathbf{L}_F + 2\mathbf{X}^T \mathbf{D}_1 \mathbf{X}\mathbf{W} - 2\mathbf{X}^T \mathbf{D}_1 \mathbf{F} + 2\gamma \mathbf{D}_2 \mathbf{W} - \Phi \\ \frac{\partial \mathcal{L}}{\partial \mathbf{F}} = 4\mathbf{F}\mathbf{M} + 2\mathbf{D}_1 \mathbf{F} - 2\mathbf{D}_1 \mathbf{X}\mathbf{W} + 2\alpha \mathbf{L}_X \mathbf{F} + 2\beta \mathbf{F} - 2\beta \mathbf{Y} - \Psi \end{cases} \quad (12)$$

其中, $\mathbf{M} = \text{Diag}\left(-\frac{1}{2\sigma} \mathbf{S} \circ (\text{diag}(\mathbf{W}^T \mathbf{W}) \mathbf{1}^T + \mathbf{1} \text{diag}(\mathbf{W}^T \mathbf{W}) - 2\mathbf{W}^T \mathbf{W})\right)$, $\text{diag}()$ 和 $\text{Diag}()$ 分别表示矩阵对角向量化和向量对角矩阵化, \circ 表示哈达马积。根

据 KKT 的互补松弛化条件, $\Phi_{ij}W_{ij} = 0$ 和 $\Psi_{ij}F_{ij} = 0$, 可得

$$\begin{cases} (2\mathbf{W}\mathbf{L}_F + 2\mathbf{X}^T\mathbf{D}_1\mathbf{X}\mathbf{W} - 2\mathbf{X}^T\mathbf{D}_1\mathbf{F} + \\ 2\gamma\mathbf{D}_2\mathbf{W})_{ij}W_{ij} = 0 \\ (4\mathbf{F}\mathbf{M} + 2\mathbf{D}_1\mathbf{F} - 2\mathbf{D}_1\mathbf{X}\mathbf{W} + 2\alpha\mathbf{L}_X\mathbf{F} + \\ 2\beta\mathbf{F} - 2\beta\mathbf{Y})_{ij}F_{ij} = 0 \end{cases} \quad (13)$$

根据式(10), 获得最终的更新规则为

$$\begin{cases} W_{ij}^{t+1} \leftarrow W_{ij}^t \frac{(\mathbf{X}^T\mathbf{D}_1\mathbf{F} + \mathbf{W}\mathbf{S}_F)_{ij}}{(\mathbf{X}^T\mathbf{D}_1\mathbf{X}\mathbf{W} + \mathbf{W}\mathbf{A}_F + \gamma\mathbf{D}_2\mathbf{W})_{ij}} \\ F_{ij}^{t+1} \leftarrow F_{ij}^t \frac{(-2\mathbf{F}\text{diag}(\mathbf{R}\mathbf{1}) + \mathbf{D}_1\mathbf{X}\mathbf{W} + \alpha\mathbf{S}_X\mathbf{F} + \beta\mathbf{Y})_{ij}}{(-2\mathbf{F}\mathbf{R} + \mathbf{D}_1\mathbf{F} + \alpha\mathbf{A}_X\mathbf{F} + \beta\mathbf{F})_{ij}} \end{cases} \quad (14)$$

其中, t 表示迭代次数, $\mathbf{R} = -\frac{1}{2\sigma}\mathbf{S} \circ (\text{diag}(\mathbf{W}^T\mathbf{W})\mathbf{1}^T + \mathbf{1}\text{diag}(\mathbf{W}^T\mathbf{W}) - 2\mathbf{W}^T\mathbf{W})$ 。然而, 式(14)中分母上的一些元素在更新过程中可能变为 0。为避免这种情况的发生, 本文在分母上添加了一个极小的常数。该方法的伪代码如算法 1 所示。

算法 1 动态图拉普拉斯多标签特征选择算法

输入 特征矩阵 $\mathbf{X} \in \mathbb{R}^{n \times d}$, 标签矩阵 $\mathbf{Y} \in \mathbb{R}^{n \times c}$, 正则化超参数 α 、 β 、 γ

输出 选择的前 k 个特征的索引集

- 1) 随机初始化 \mathbf{W} 和 \mathbf{F}
- 2) $\mathbf{W} \in \mathbb{R}_+^{d \times c}$
- 3) $\mathbf{F} \in \mathbb{R}_+^{n \times c}$
- 4) $t = 0$
- 5) 根据式(5)计算 \mathbf{A}_F 、 \mathbf{S}_F 、 \mathbf{A}_X 和 \mathbf{S}_X
- 6) while 满足收敛条件 do
- 7) 更新 $\begin{cases} D_{1i} = \frac{1}{2\|(\mathbf{X}\mathbf{W} - \mathbf{F})_i\|_2 + \epsilon} \\ D_{2ii} = \frac{1}{2\|\mathbf{W}_i\|_2 + \epsilon} \end{cases}$, 计算 \mathbf{D}_1 和 \mathbf{D}_2
- 8) 更新

$$\begin{cases} W_{ij}^{t+1} \leftarrow W_{ij}^t \frac{(\mathbf{X}^T\mathbf{D}_1\mathbf{F} + \mathbf{W}\mathbf{S}_F)_{ij}}{(\mathbf{X}^T\mathbf{D}_1\mathbf{X}\mathbf{W} + \mathbf{W}\mathbf{A}_F + \gamma\mathbf{D}_2\mathbf{W})_{ij}} \\ F_{ij}^{t+1} \leftarrow F_{ij}^t \frac{(-2\mathbf{F}\text{diag}(\mathbf{R}\mathbf{1}) + \mathbf{D}_1\mathbf{X}\mathbf{W} + \alpha\mathbf{S}_X\mathbf{F} + \beta\mathbf{Y})_{ij}}{(-2\mathbf{F}\mathbf{R} + \mathbf{D}_1\mathbf{F} + \alpha\mathbf{A}_X\mathbf{F} + \beta\mathbf{F})_{ij}} \end{cases}$$

- 9) $t = t + 1$
- 10) end while
- 11) return \mathbf{W}

12) 通过 $\|\mathbf{W}_i\|_2, i = 1, 2, \dots, d$, 选取前 k 个特征序号

算法 1 中核心步骤为第 7)~8)行, 这 2 个步骤促使算法逐步收敛于最终状态。第 12)行中 k 的取值主要依据文献参考经验值, 如 MIFS 中所采纳的 k 值。

3.3 收敛性证明

本节以 \mathbf{W} 的证明为例, 给出相关的定义和概念^[34-35]。

定义 1 若 $\mathcal{G}(W_{ij}, W_{ij}^t) \geq \mathcal{F}(W_{ij})$, $\mathcal{G}(W_{ij}, W_{ij}) = \mathcal{F}(W_{ij})$, 则 $\mathcal{G}(W_{ij}, W_{ij}^t)$ 被认为是 $\mathcal{F}(W_{ij})$ 的辅助函数。

引理 1 若 $\mathcal{G}(W_{ij}, W_{ij}^t)$ 是 $\mathcal{F}(W_{ij})$ 的辅助函数, 则 $\mathcal{F}(W_{ij})$ 在式(15)所示条件下单调递减。

$$W_{ij}^{t+1} = \arg \min_{W_{ij}} \mathcal{G}(W_{ij}, W_{ij}^t) \quad (15)$$

其中, W_{ij} 表示矩阵 \mathbf{W} 中的任意元素。考虑到上述更新规则是基于矩阵元素更新的, 本文使用 W_{ij} 作为变量。 $\mathcal{F}_{ij}(W_{ij})$ 表示与 W_{ij} 相关的 $\mathcal{O}(\mathbf{W})$ 的部分函数。 $\mathcal{F}_{ij}(W_{ij})$ 的一阶导数和二阶导数分别为

$$\begin{cases} \mathcal{F}'_{ij} = 2(\mathbf{W}\mathbf{L}_F + \mathbf{X}^T\mathbf{D}_1\mathbf{X}\mathbf{W} - \mathbf{X}^T\mathbf{D}_1\mathbf{F} + \gamma\mathbf{D}_2\mathbf{W})_{ij} \\ \mathcal{F}''_{ij} = (\mathbf{L}_F)_{jj} + (\mathbf{X}^T\mathbf{D}_1\mathbf{X})_{ii} + \gamma(\mathbf{D}_2)_{ii} \end{cases} \quad (16)$$

因此, $\mathcal{F}_{ij}(W_{ij})$ 的泰勒展开式为

$$\begin{aligned} \mathcal{F}_{ij}(W_{ij}) &= \mathcal{F}_{ij}(W_{ij}^t) + \mathcal{F}'_{ij}(W_{ij}^t)(W_{ij} - W_{ij}^t) + \\ &\quad \mathcal{F}''_{ij}(W_{ij}^t)(W_{ij} - W_{ij}^t)^2 \end{aligned} \quad (17)$$

受文献[36]的启发, 本文设置了关于 $\mathcal{F}_{ij}(W_{ij})$ 的辅助函数, 如式(18)所示。

$$\begin{aligned} \mathcal{G}(W_{ij}, W_{ij}^t) &= \mathcal{F}_{ij}(W_{ij}^t) + \mathcal{F}'_{ij}(W_{ij}^t)(W_{ij} - W_{ij}^t) + \\ &\quad \frac{(\mathbf{W}\mathbf{A}_F + \mathbf{X}^T\mathbf{D}_1\mathbf{X}\mathbf{W} - \mathbf{X}^T\mathbf{D}_1\mathbf{F} + \gamma\mathbf{D}_2\mathbf{W})_{ij}}{W_{ij}^t} (W_{ij} - W_{ij}^t)^2 \end{aligned} \quad (18)$$

证明 只需要证明所设的辅助函数满足定义 1 便可证明更新规则的收敛性。显然, 式(18)中当 $W_{ij} = W_{ij}^t$ 时, $\mathcal{G}(W_{ij}, W_{ij}^t) = \mathcal{F}(W_{ij})$ 成立, 只需要在证明另一条件 $\mathcal{G}(W_{ij}, W_{ij}^t) \geq \mathcal{F}(W_{ij})$ 即可证明收敛性, 所以, 需要证明式(19)成立。

$$\begin{aligned} \frac{(\mathbf{W}\mathbf{A}_F + \mathbf{X}^T\mathbf{D}_1\mathbf{X}\mathbf{W} - \mathbf{X}^T\mathbf{D}_1\mathbf{F} + \gamma\mathbf{D}_2\mathbf{W})_{ij}}{W_{ij}^t} &\geq \\ (\mathbf{L}_F)_{jj} + (\mathbf{X}^T\mathbf{D}_1\mathbf{X})_{ii} + \gamma(\mathbf{D}_2)_{ii} &\quad (19) \end{aligned}$$

显然，可以推导出式(20)。

$$\left\{ \begin{aligned} (X^T D_1 X W + \gamma D_2 W)_{ij} &= \sum_{l=1}^d (X^T D_1 X W + \gamma D_2)_{il} W_{lj}^t \geq \\ (X^T D_1 X W + \gamma D_2)_{ii} W_{ij}^t & \\ (W A_F)_{ij} = \sum_{l=1}^c W_{il}^t (A_F)_{lj} &\geq W_{ij}^t (A_F)_{jj} \geq \\ W_{ij}^t (A_F - S_F)_{jj} &= W_{ij}^t (L_F)_{jj} \end{aligned} \right. \quad (20)$$

至此，证明所设置的函数是 $\mathcal{F}(W_{ij})$ 的辅助函数。

根据式(15)，通过 $\frac{\partial \mathcal{G}(W_{ij}, W_{ij}^t)}{\partial W_{ij}} = 0$ 可得关于 W

的更新公式。 F 的收敛性证明类似于 W ，本节不再列出。证毕。

4 实验分析

为了验证所提方法的分类效果，本文在 9 个多标签基准数据集上与 3 个先进的多标签特征选择算法进行比较。所有实验均在内存为 16 GB 的 3.4 GHz 的英特尔酷睿 i7-6700 计算机上进行。

4.1 数据集描述及实验设置

本文实验所用数据均来自 MulanLibrary^[37]。这些属于不同领域的数据集已经被众多文献使用^[11-15]。例如，Birds 数据集是野外条件下采集的鸟类声音数据，其中包括 645 个音频记录，与 19 种未压缩 WAV 格式的鸟类声音相关。Yeast 数据集来自生物领域，该数据集包含 2 417 个数据样例，每个样例有 103 个特征和 14 个标签。Enron 数据集属于文本领域，是安然电子邮件语料库的一个子集。数据集的参数如表 1 所示。

为了证明所提方法的有效性，将其与 MIFS、RALM-FS 和 SCLS 这 3 种先进的方法进行比较。

此外，一些参数需要提前设定。首先，在构造近邻矩阵的热核函数中，参数 p 和 σ 分别被设置为 $c-1$ 和 1。为了方便，涉及超参数的各个对比方法中的参数统一在网格 {0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0} 范围内进行搜索。然后，在五折交叉验证过程中记录参数的最佳值。根据文献，使用 BR 模型^[20]将多标签问题转化为几个二进制问题，使用线性支持向量机 (SVM, support vector machines) 分类器和 K 最近邻 (KNN, K-nearest neighbor) 分类器 ($K=3$) 进行分类处理，本文采用相同的方式以确保公平性。所有方法的分类性能由 2 个评价标准来评估，即基于 F1 度量的 Micro-F1 和 Macro-F1^[38]。

$$\left\{ \begin{aligned} \text{Micro-F1} &= \frac{\sum_{i=1}^z 2TP^i}{\sum_{i=1}^z (2TP^i + FP^i + FN^i)} \\ \text{Macro-F1} &= \frac{1}{z} \sum_{i=1}^z \frac{2TP^i}{(2TP^i + FP^i + FN^i)} \end{aligned} \right. \quad (21)$$

其中， z 和 i 分别表示标签数和第 i 个标签，TP、FP 和 FN 分别表示真阳性、假阳性和假阴性。Micro-F1 和 Macro-F1 均是值越大表示相应的方法分类性能越好。为评估所提方法的分类性能，本文在 9 个不同的多标签数据集上进行实验。根据一些经验方法^[17]，本文使用每个数据集中总特征的前 20% 来计算不同方法的平均结果和标准偏差。

4.2 实验结果及分析

表 2~表 5 记录了 4 个多标签特征选择方法在 9 个数据集上的实验结果。表中每一行最优值用黑色粗体表示。最后一行计算每个特征选择方法下所有数据集的平均值。从表 2~表 5 可以看出，与其

表 1 数据集参数

| 数据集 | 样例数/个 | 特征数/个 | 标签数/个 | 训练样例数/个 | 测试样例数/个 | 领域 |
|-----------|-------|-------|-------|---------|---------|---------|
| Arts | 5 000 | 462 | 26 | 2 000 | 3 000 | 文本(Web) |
| Birds | 645 | 260 | 19 | 322 | 323 | 声音 |
| Yeast | 2 417 | 103 | 14 | 1 500 | 917 | 生物 |
| Education | 5 000 | 550 | 33 | 2 000 | 3 000 | 文本(Web) |
| Enron | 1 702 | 1 001 | 53 | 1 123 | 579 | 文本 |
| Social | 5 000 | 1 047 | 39 | 2 000 | 3 000 | 文本(Web) |
| Science | 5 000 | 743 | 40 | 2 000 | 3 000 | 文本(Web) |
| Entertain | 5 000 | 636 | 27 | 2 000 | 3 000 | 文本(Web) |
| Society | 5 000 | 640 | 21 | 2 000 | 3 000 | 文本(Web) |

他方法相比，所提方法在所有数据集下分类效果更好。在 SVM 分类器的基础上，分别使用评估指标 *Micro-F1* 和 *Macro-F1* 获得所有算法的分类结果，如表 2 和表 3 所示。从表 2 和表 3 可以看出，与其他方法相比，所提方法的 *Micro-F1* 和 *Macro-F1* 在所有数据集平均值均为最优值，分别为 0.315 和 0.097。

表 2 特征选择方法在 SVM 分类器上的 *Micro-F1* 结果

| 数据集 | 所提方法 | MIFS | RALM-FS | SCLS |
|-----------|--------------------|-------------|-------------|-------------|
| Arts | 0.158±0.076 | 0.139±0.078 | 0.106±0.046 | 0.102±0.061 |
| Birds | 0.123±0.058 | 0.116±0.059 | 0.060±0.040 | 0.096±0.046 |
| Yeast | 0.564±0.036 | 0.547±0.035 | 0.532±0.008 | 0.552±0.027 |
| Education | 0.261±0.061 | 0.073±0.059 | 0.073±0.059 | 0.193±0.056 |
| Enron | 0.524±0.051 | 0.372±0.027 | 0.488±0.031 | 0.389±0.059 |
| Social | 0.468±0.118 | 0.276±0.136 | 0.363±0.120 | 0.149±0.112 |
| Science | 0.141±0.053 | 0.129±0.057 | 0.037±0.035 | 0.097±0.054 |
| Entertain | 0.283±0.101 | 0.228±0.112 | 0.113±0.072 | 0.214±0.100 |
| Society | 0.317±0.042 | 0.300±0.042 | 0.216±0.028 | 0.223±0.059 |
| 平均值 | 0.315 | 0.242 | 0.221 | 0.224 |

表 3 特征选择方法在 SVM 分类器上的 *Macro-F1* 结果

| 数据集 | 所提方法 | MIFS | RALM-FS | SCLS |
|-----------|--------------------|-------------|-------------|-------------|
| Arts | 0.067±0.038 | 0.055±0.034 | 0.038±0.016 | 0.039±0.026 |
| Birds | 0.077±0.037 | 0.075±0.036 | 0.039±0.026 | 0.058±0.024 |
| Yeast | 0.243±0.048 | 0.219±0.048 | 0.207±0.014 | 0.229±0.036 |
| Education | 0.068±0.015 | 0.019±0.017 | 0.019±0.015 | 0.052±0.014 |
| Enron | 0.135±0.036 | 0.074±0.017 | 0.119±0.031 | 0.074±0.027 |
| Social | 0.054±0.019 | 0.031±0.016 | 0.036±0.014 | 0.014±0.012 |
| Science | 0.056±0.025 | 0.034±0.016 | 0.007±0.007 | 0.036±0.020 |
| Entertain | 0.119±0.039 | 0.097±0.047 | 0.044±0.030 | 0.081±0.038 |
| Society | 0.060±0.016 | 0.055±0.020 | 0.021±0.003 | 0.032±0.012 |
| 平均值 | 0.097 | 0.073 | 0.059 | 0.068 |

表 4 征选择方法在 3NN 分类器上的 *Micro-F1* 结果

| 数据集 | 所提方法 | MIFS | RALM-FS | SCLS |
|-----------|--------------------|-------------|-------------|--------------------|
| Arts | 0.212±0.052 | 0.202±0.052 | 0.164±0.024 | 0.182±0.043 |
| Birds | 0.135±0.061 | 0.167±0.061 | 0.144±0.043 | 0.171±0.066 |
| Yeast | 0.532±0.06 | 0.525±0.053 | 0.518±0.035 | 0.529±0.019 |
| Education | 0.289±0.043 | 0.183±0.055 | 0.176±0.037 | 0.260±0.050 |
| Enron | 0.488±0.033 | 0.410±0.024 | 0.437±0.026 | 0.365±0.073 |
| Social | 0.456±0.061 | 0.338±0.081 | 0.376±0.057 | 0.315±0.054 |
| Science | 0.188±0.028 | 0.171±0.037 | 0.115±0.016 | 0.160±0.048 |
| Entertain | 0.314±0.062 | 0.276±0.065 | 0.234±0.033 | 0.273±0.065 |
| Society | 0.313±0.046 | 0.305±0.043 | 0.245±0.021 | 0.255±0.050 |
| 平均值 | 0.325 | 0.286 | 0.268 | 0.279 |

表 5 特征选择方法在 3NN 分类器上的 *Macro-F1* 结果

| 数据集 | 所提方法 | MIFS | RALM-FS | SCLS |
|-----------|--------------------|--------------------|-------------|-------------|
| Arts | 0.096±0.043 | 0.095±0.033 | 0.074±0.014 | 0.071±0.025 |
| Birds | 0.085±0.038 | 0.106±0.042 | 0.078±0.028 | 0.093±0.036 |
| Yeast | 0.322±0.055 | 0.282±0.057 | 0.301±0.026 | 0.300±0.027 |
| Education | 0.094±0.018 | 0.043±0.018 | 0.059±0.021 | 0.084±0.023 |
| Enron | 0.129±0.018 | 0.087±0.014 | 0.111±0.013 | 0.081±0.026 |
| Social | 0.067±0.020 | 0.051±0.017 | 0.049±0.013 | 0.038±0.012 |
| Science | 0.079±0.021 | 0.062±0.015 | 0.037±0.010 | 0.057±0.021 |
| Entertain | 0.155±0.039 | 0.138±0.042 | 0.112±0.020 | 0.131±0.040 |
| Society | 0.088±0.020 | 0.089±0.021 | 0.055±0.011 | 0.053±0.016 |
| 平均值 | 0.124 | 0.106 | 0.097 | 0.101 |

在 3NN 分类器上所有方法的 *Micro-F1* 和 *Macro-F1* 如表 4 和表 5 所示。从表 4 可以看出，所提方法的 *Micro-F1* 平均值为 0.325，相对于 MIFS、RALM-FS 和 SCLS，分别提升了 13.6%、21.3% 和 16.5%。从表 5 可以看出，所提方法的 *Macro-F1* 平均值为 0.124，相对于 MIFS、RALM-FS 和 SCLS，分别提升了 17.0%、27.8% 和 22.8%。综上所述，所提方法在不同评估条件下均取得优异的分类表现。为了进一步展示所提方法的分类优势，本文选取 6 个代表性的数据集（Arts、Yeast、Enron、Science、Education 和 Social）绘制折线分析图，如图 1~图 4 所示。

通过分析图 1~图 4，可以直观地看到所提方法相比其他 3 个多标签特征选择方法具有最佳分类表现。随着所选特征数目的增加，不同方法的分类性能都总体先增加，后趋于稳定。图 1~图 4 的曲线都是振荡上升的，产生这种现象的原因如下。所提方法属于序列前向搜索方式的嵌入式特征选择，这种策略通过一定的标准对所有特征进行排序，然后选择 k 个排名靠前的特征。图 1~图 4 中，横轴表示所选排名靠前的特征的占比。举例说明如下，通常选择前 k 个特征导致的分类性能可能会高于选择 $k-1$ 个特征，而低于 $k+1$ 个特征的性能，这是因为不同的特征子集的组合导致的分类性能是不同的，即前 k 个单独排名靠前的特征联合导致的分类性能可能低于相互有关联的 k 个特征组成的子集的分类性能。因此，前 k 个特征导致的分类性能可能会低于 $k+1$ 个特征的性能，但随着特征数目的增加，依然可以导致整体分类性能的提升。这也就导致了曲线振荡上升的现象。同时，可以观察到，图 1~图 4 中

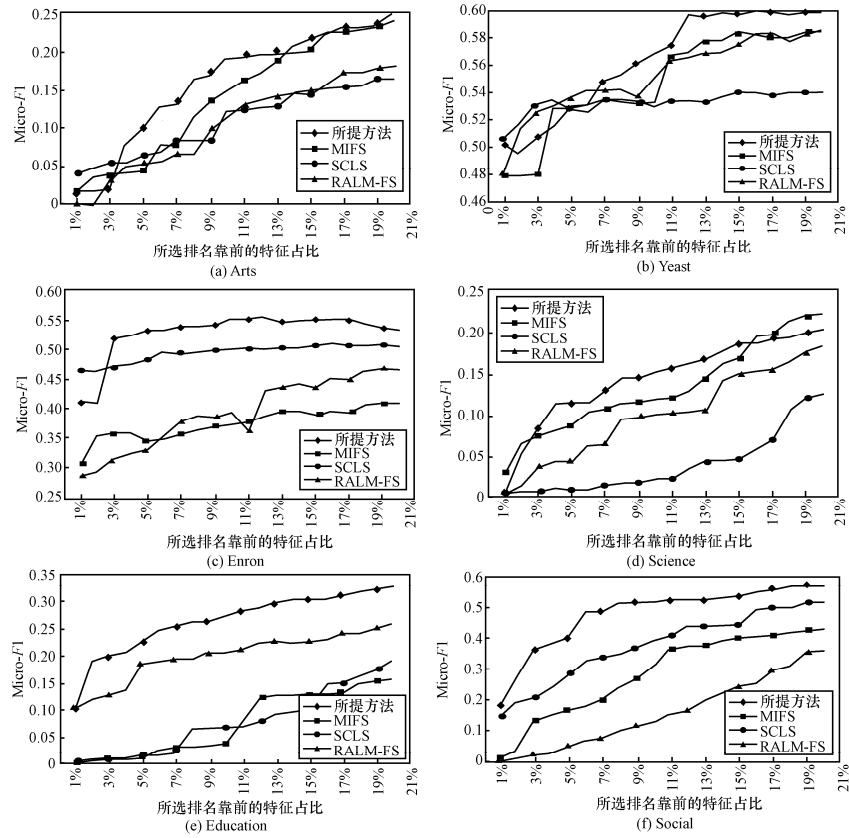


图 1 6 个数据集在 Micro-F1(SVM)上的实验结果

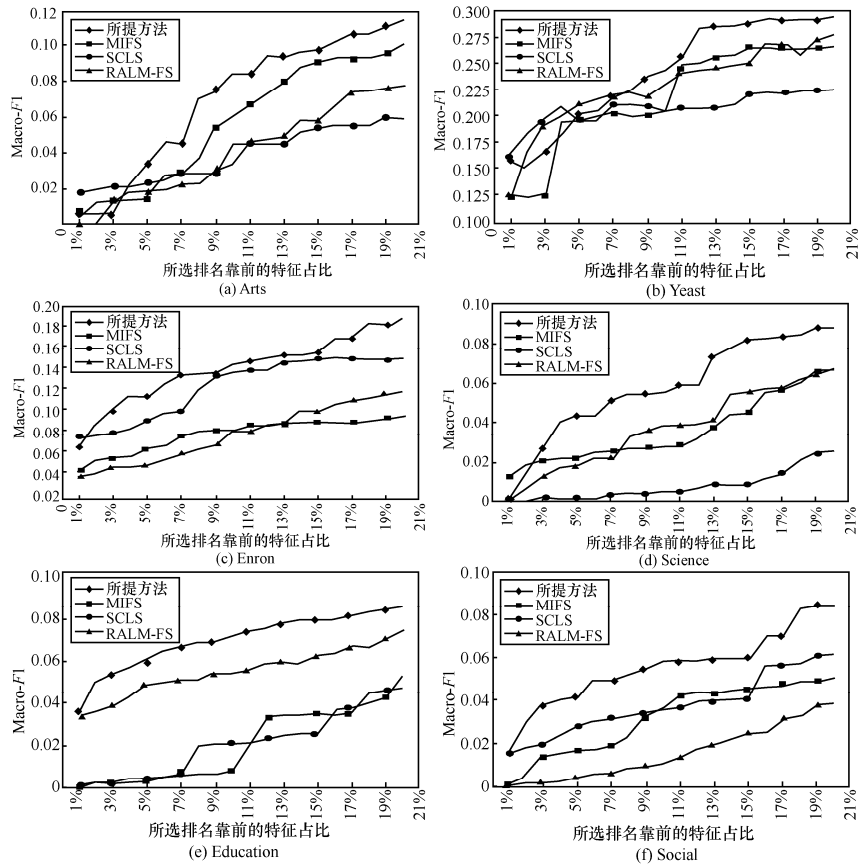


图 2 6 个数据集在 Macro-F1(SVM)上的实验结果

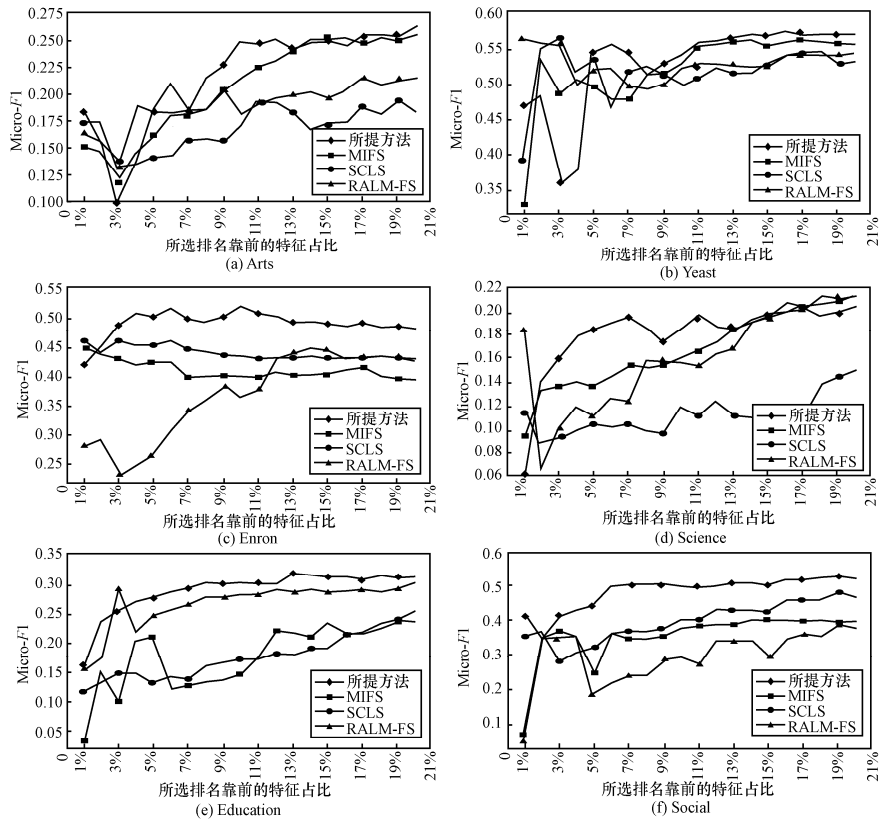


图 3 6 个数据集在 Micro-F1(3NN)上的实验结果

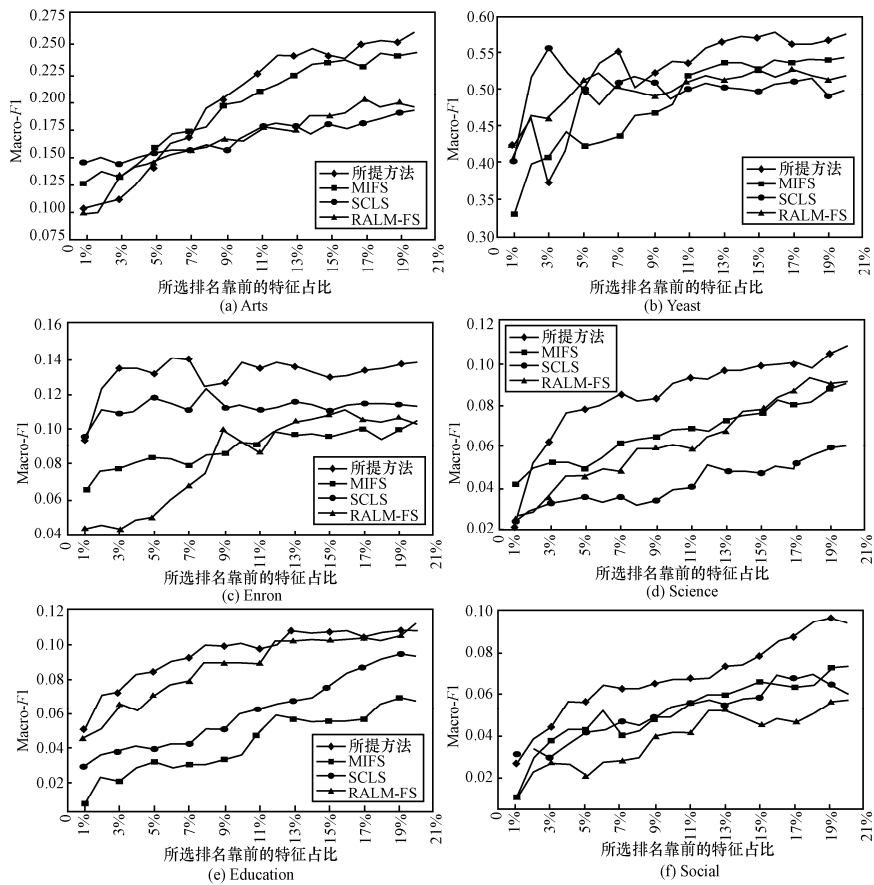


图 4 6 个数据集在 Macro-F1(3NN)上的实验结果

所提方法对应折线总是在最上部,说明所提方法取得了更优异的性能。总体来说,所提方法取得优于对比方法的分类性能,原因是其考虑了特征选择过程中图拉普拉斯矩阵的动态变化,保证每次更新过程中所利用的图拉普拉斯优于上一次的更新,并考虑数值标签刻画标签的重要程度,以便更好地选择特征。

4.3 参数敏感性分析

为了研究 3 个超参数(α, β, γ)在多标签特征选择过程中产生的影响,本文通过搜索网格{0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0}来调整这些超参数。然而,网格搜索策略时间成本过高,为此本文参考文献[17]中的策略,即固定其他超参数,仅调整其中一个超参数。本文设定被固定的超参数值为 0.5, 选择 Education

数据集通过 SVM 分类器进行超参数敏感性分析,分析结果如图 5 所示。图 5 中,超参数 α 在选择特征数目相同的情况下,在网格范围内波动幅度较小,仅在 $\alpha=0.1 \sim 0.5$ 时会对模型的性能产生影响,根据文献[15, 17],这种程度的影响在实验中是可以接受的,即算法对超参数 α 的变化不敏感,而且随着选择的特征数目的增加,影响更小。超参数 β 在选择特征数目相同的情况下,在网格范围内波动幅度较大,即非常敏感,这种敏感程度对算法性能的影响是不能忽略的。因此,超参数 β 在实际应用中可使用更大范围的网格进行搜索以获得令人满意的性能。相对于超参数 α 和 β ,超参数 γ 选择的特征数目相同情况下,在给定的网格范围内波动

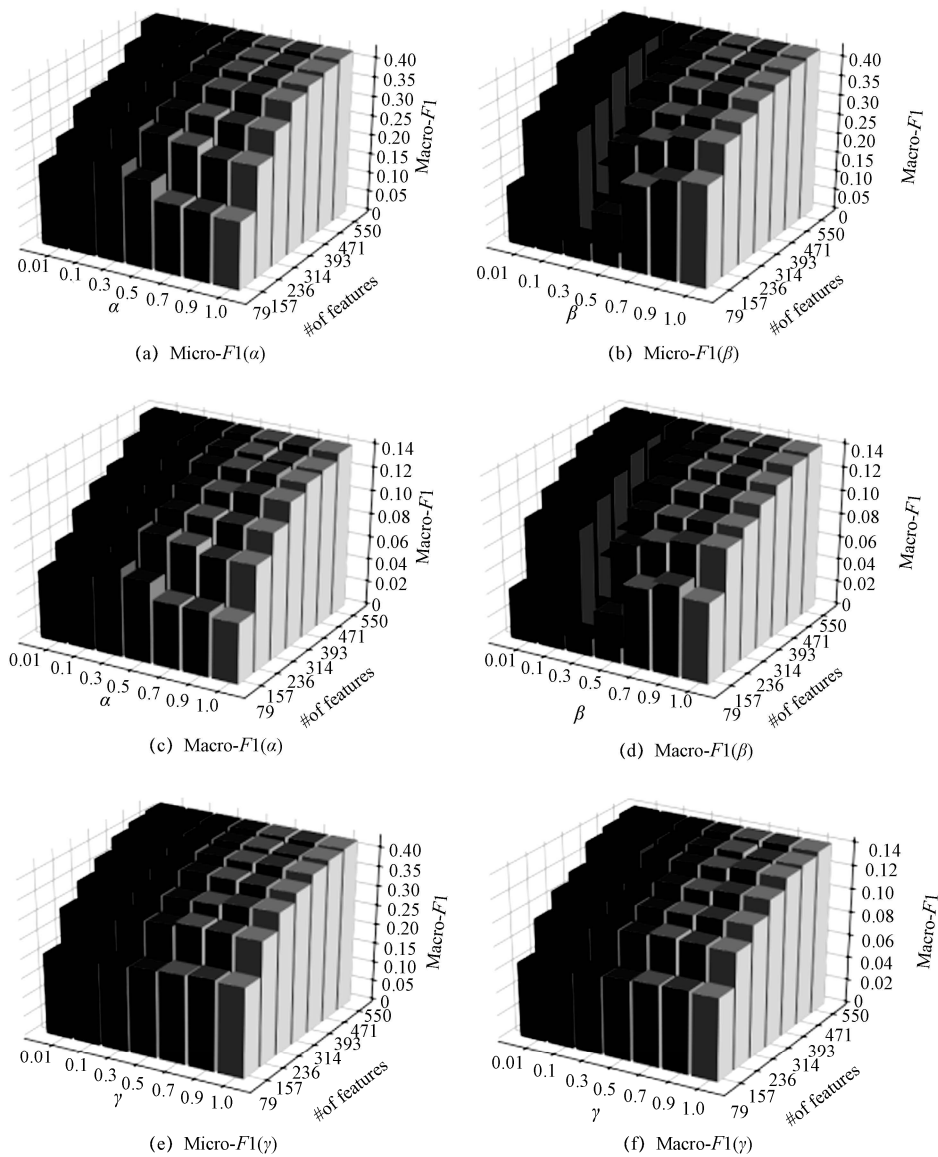


图 5 在 Education 数据集上所提算法关于 α, β 和 γ 的 Micro-F1 和 Macro-F1 (SVM)

幅度最小，因此与超参数 α 一样，算法对超参数 γ 的变化不敏感。

4.4 收敛性分析与时间复杂度

本节对所提方法的收敛性和时间复杂度进行分析。首先，通过 6 个代表性的数据集 (Arts、Yeast、Enron、Science、Education 和 Social) 对所提方法的收敛性进行验证，实验结果如图 6 展示。从图 6 可以看出，所提方法的迭代收敛速度很快。在前 2~3 次迭代中目标函数的损失值快速下降，然后下降速度开始变缓。特别是数据集 Yeast 和 Enron，仅迭代 2 次之后，已无法直接观察到目标函数损失值的变化，但

根据分析可知，后面的迭代结果依然接近给定的迭代停止触发条件。同样地，数据集 Arts、Science、Education 和 Social 上的目标函数损失值也随着迭代次数的增加迅速减小，并最终趋于稳定。实验结果证明所提方法在 3.2 节中所设计的优化方案下可有效收敛，同时验证了 3.3 节理论证明的正确性。

下面分析所提方法和对比方法的时间复杂度。设 p 、 d 、 n 和 c 分别表示已选特征数量、特征总数、样例数和标签总数。SCLS 的时间复杂度为 $O(dc + pd)$ ；MIFS 在每次迭代的时间复杂度为 $O(ndl + n^2)$ ；由于涉及矩阵的逆运算，RALM-FS

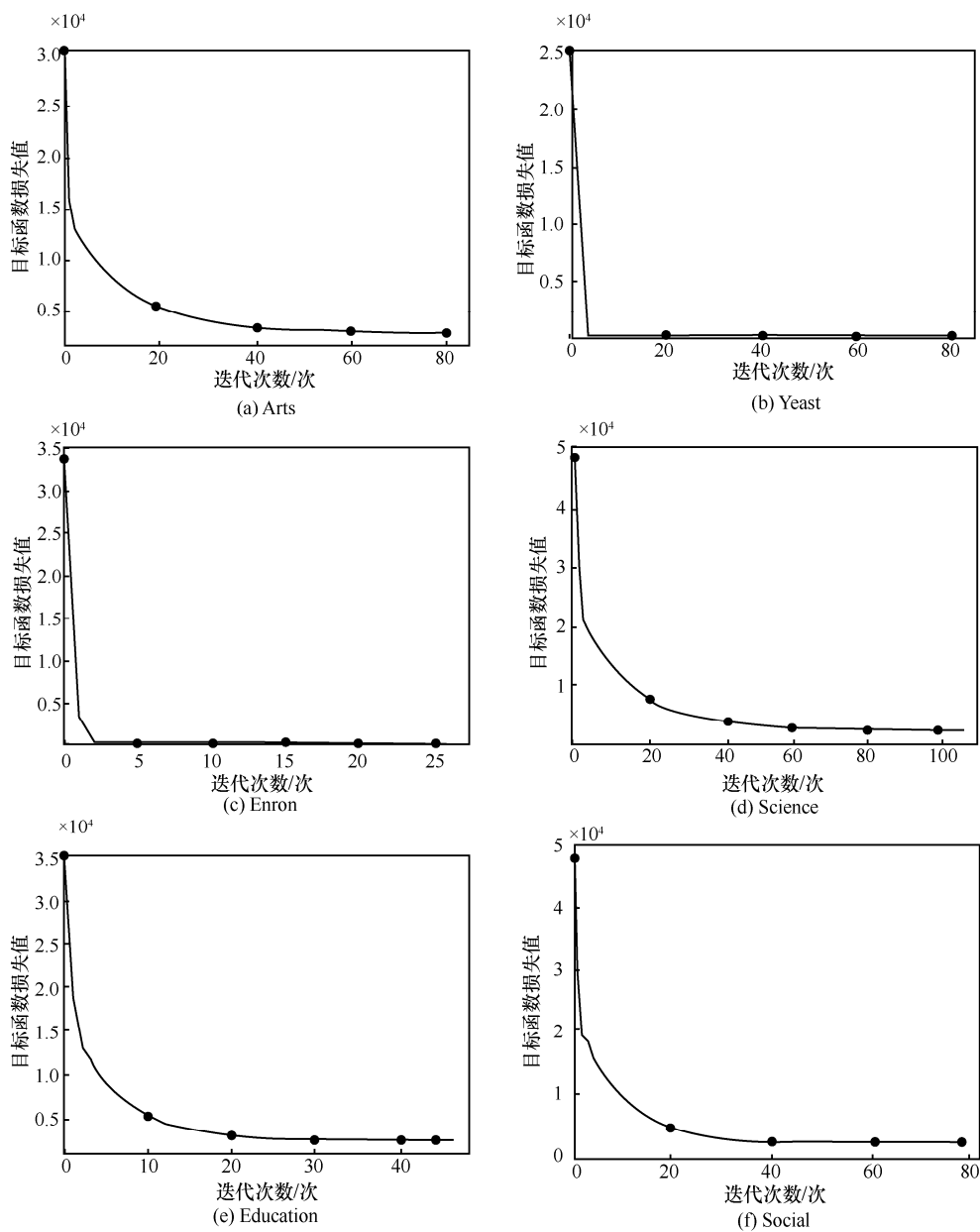


图 6 所提方法的收敛曲线

的时间复杂度为 $O(d^3)$ ；所提方法的时间复杂度为 $O(dn^2 + d^2n)$ 。

5 结束语

针对多标签分类和特征选择的结合这一开放性问题，本文提出了一种基于动态图拉普拉斯矩阵的多标签特征选择方法。该方法不同于以往基于图的多标签特征选择方法依赖于固定的图拉普拉斯矩阵，而是利用特征选择过程中可以动态变化的图拉普拉斯矩阵。在图拉普拉斯矩阵的动态变化过程中，由于逻辑标签导致标签信息丢失，而其对应的实值标签能够更好地反映相应标签的重要性，因此在新设计的动态图拉普拉斯矩阵变化下，本文将逻辑标签重构为实数值标签，同时，利用 $l_{2,1}$ 范数减少动态构造拉普拉斯矩阵时异常值产生的影响。最后，本文设计了一套针对所提方法的简单有效的优化方案。为了验证所提方法的优越性，将其与 3 个多标签特征选择方法在 9 个不同领域的多标签数据集上进行实验对比。实验结果表明，所提方法性能显著优于对比方法，且可得到高质量的特征子集。

下一阶段，将进一步研究在非凸优化问题下的多标签特征选择方法。由于非凸优化问题和多标签问题在现实生活中广泛存在，因此多标签特征选择具有巨大的研究价值。

参考文献:

- [1] GUI J, SUN Z N, JIS W, et al. Feature selection based on structured sparsity: a comprehensive study[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 28(7):1-18.
- [2] BOLON C N, SANCHEZ M N, ALONSO B A, et al. A review of microarray datasets and applied feature selection methods[J]. Information Sciences, 2014, 282:111-135.
- [3] ZHANG M L, ZHOU Z H. A review on multi-label learning algorithms [J]. IEEE Transactions on Knowledge and data Engineering, 2014, 26(8): 1819-1837.
- [4] TSOU MAKAS G, KATAKIS I, VLAHAVAS I. Mining multi-label data[M]. Berlin: Springer, 2009: 667-685.
- [5] TSOU MAKAS G, KATAKIS I. Multi-label classification: an overview [J]. International Journal of Data Warehousing and Mining, 2007, 3(3): 1-13.
- [6] KASHEF S, NEZAMABADI-POUR H, NIKPOUR B. Multilabel feature selection: a comprehensive review and guiding experiments[J]. Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery, 2018, 8(2): 12-40.
- [7] 刘慧婷, 冷新杨, 王利利, 等. 联合嵌入式多标签分类算法[J]. 自动化学报, 2019, 45(10): 1969-1982.
- LIU H T, LENG X Y, WANG L L, et al. A joint embedded multi-label classification algorithm[J]. Acta Automatica Sinica, 2019, 45(10): 1969-1982.
- [8] LI J, CHENG K, WANG S, et al. Feature selection: a data perspective [J]. ACM Computing Surveys, 2018, 50(6): 1-45.
- [9] SAEYS Y, INZA I, LARRAÑAGA P. A review of feature selection techniques in bioinformatics [J]. Bioinformatics, 2007, 23(19): 2507-2517.
- [10] 李占山, 刘兆庚. 基于 XGBoost 的特征选择算法[J]. 通信学报, 2019, 40(10): 101-108.
- LI Z S, LIU Z G. Feature selection algorithm based on XGBoost[J]. Journal on Communications, 2019, 40(10): 101-108.
- [11] ZHANG P, LIU G X, GAO W F. Distinguishing two types of labels for multi-label feature selection[J]. Pattern Recognition, 2019, 95(1): 72-82.
- [12] CAI Z, ZHU W. Multi-label feature selection via feature manifold learning and sparsity regularization[J]. International Journal of Machine Learning & Cybernetics, 2018, 9(8): 1321-1334.
- [13] RODRIGUES D, PEREIRA L A, NAKAMURA R Y, et al. A wrapper approach for feature selection based on bat algorithm and optimum-path forest[J]. Expert Systems with Applications, 2014, 41(5): 2250-2258.
- [14] 张俐, 王枫. 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):115-126.
- ZHANG L, WANG C. Multi-label feature selection algorithm based on joint mutual information of max-relevance and min-redundancy[J]. Journal on Communications, 2018, 39(5):115-126.
- [15] ZHANG J, LUO Z, LI C, et al. Manifold regularized discriminative feature selection for multi-label learning[J]. Pattern Recognition, 2019, 95(1): 136-150.
- [16] HUANG J, NIE F P, HUANG H, et al. Robust manifold nonnegative matrix factorization[J]. ACM Transactions on Knowledge Discovery from Data, 2014, 8(3):1-21.
- [17] JIAN L, LI J, SHU K, et al. Multi-label informed feature selection[C]// Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2016:1627-1633.
- [18] NIE F P, HUANG H, CAI X, et al. Efficient and robust feature selection via joint $l_{2,1}$ -norms minimization[C]//Proceedings of the Advances in neural Information Processing Systems. Massachusetts: MIT Press, 2010:1813-1821.
- [19] CHANG X, NIE F P, YANG Y, et al. A convex formulation for semi-supervised multi-label feature selection[C]//Proceedings of the National Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2014:1171-1177.
- [20] BOUTELL M R, LUO J, SHEN X, et al. Learning multi-label scene classification[J]. Pattern Recognition, 2004, 37(9): 1757-1771.
- [21] FÜRNKRANZ J, HÜLLERMEIER E, MENCÍA E L, et al. Multilabel classification via calibrated label ranking[J]. Machine Learning, 2008, 73(2): 133-153.
- [22] QI G J, HUA X S, RUI Y, et al. Correlative multi-label video annotation[C]//Proceedings of the 15th International Conference on Multimedia. New York: ACM Press, 2007: 24-29.
- [23] HUANG J, LI G, HUANG Q, et al. Learning label-specific features and class-dependent labels for multi-label classification [J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(12): 3309-3323.
- [24] ZHANG Y, YANG Y, LI T, et al. A multitask multiview clustering algorithm in heterogeneous situations based on LLE and LE[J].

- Knowledge-Based Systems, 2019, 163(1): 776-786.
- [25] REN Y, ZHANG G, YU G, et al. Local and global structure preserving based feature selection[J]. Neurocomputing, 2012, 89(1): 147-157.
- [26] HUANG R, JIANG W, SUN G. Manifold-based constraint Laplacian score for multi-label feature selection[J]. Pattern Recognition Letters, 2018, 112(1): 346-352.
- [27] XU 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. New York: ACM Press, 2018: 783-792.
- [28] CHEN G, SONG Y, WANG F, et al. Semi-supervised multi-label learning by solving a Sylvester equation[C]//Proceedings of the 2008 SIAM International Conference on Data Mining. Saarland: DBLP, 2008: 410-419.
- [29] SHAO R, XU N, GENG X. Multi-label learning with label enhancement [C]//Proceedings of the 2018 IEEE International Conference on Data Mining. Piscataway: IEEE Press, 2018: 437-446.
- [30] LEE J, KIM D W. SCLS: multi-label feature selection based on scalable criterion for large label set[J]. Pattern Recognition, 2017, 66(1): 342-352.
- [31] CAI X, NIE F P, HUANG H. Exact top-k feature selection via $l_2, 0$ -norm constraint[C]//Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence. Palo Alto: AAAI Press 2013: 1240-1246.
- [32] ZHU Y, KWOK J T, ZHOU Z H. Multi-label learning with global and local label correlation[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(6): 1081-1094.
- [33] YAN H, YANG J, YANG J. Robust joint feature weights learning framework[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(5): 1327-1339.
- [34] DEMPSTER A P, LAIRD N M, RUBIN D B. Maximum likelihood from incomplete data via the EM algorithm[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1977, 39(1): 1-22.
- [35] LEE D D, SEUNG H S. Algorithms for non-negative matrix factorization[C]//Proceedings of the Advances in Neural Information Processing Systems. Massachusetts: MIT Press, 2001: 556-562.
- [36] CAI D, HE X, HAN J, et al. Graph regularized nonnegative matrix factorization for data representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(8): 1548-1560.
- [37] TSOUMAKAS G, SPYROMITROS-XIOUFIS E, VILCEK J, et al. Mulan: a Java library for multi-label learning[J]. Journal of Machine Learning Research, 2011, 12(7): 2411-2414.

- [38] YU K, YU S, TRESP V. Multi-label informed latent semantic indexing[C]//Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2005: 258-265.

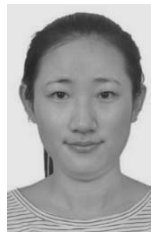
[作者简介]



李永豪（1992- ），男，河南安阳人，吉林大学博士生，主要研究方向为多标签学习、特征选择。



胡亮（1968- ），男，吉林长春人，博士，吉林大学教授、博士生导师，主要研究方向为人工智能和分布式计算。



张平（1991- ），女，河北石家庄人，吉林大学博士生，主要研究方向为多标签学习、特征选择。



高万夫（1990- ），男，吉林辽源人，博士，吉林大学讲师，主要研究方向为机器学习、特征选择、多标签学习。