计算机学报
CHINESE JOURNAL OF COMPUTERS

# 基于正则化的半监督弱标签分类方法

丁家满[1),2)]刘楠[1),2)]周蜀杰[1),2)]贾连印[1),2)]李润鑫[1),2)]

[1)](昆明理工大学信息工程与自动化学院昆明 650500)

[2)](云南省人工智能重点实验室昆明 650500)

**摘要**　针对多标签学习中实例标签的缺失补全和预测问题，本文提出一种基于正则化的半监督弱标签分类方法（简称SWCMR），方法同时兼顾实例相似性和标签相关性。SWCMR首先根据标签相关性对弱标签实例的缺失标签进行初步预估，然后利用弱标签实例和无标签实例构造邻域图，从实例相似性和标签相关性角度构建基于平滑性假设的正则化项，接下来利用预估后的弱标签实例结合无标签实例训练半监督弱标签分类模型。在多种公共多标签数据集上的实验结果表明，SWCMR提高了分类性能，尤其是标签信息较少时，分类效果提升更显著。

关键词　半监督弱标签学习；多标签分类；正则化；标签相关性

**中图法分类号**　TP181

# Semi-Supervised Weak-Label Classification Method by Regularization

DING Jia-Man[1),2)]　LIU Nan[1),2)]　ZHOU Shu-Jie[1),2)]JIA Lian-Yin[1),2)]　LI Run-Xin[1),2)]

[1)](Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming　650500)

[2)](Artificial Intelligence Key Laboratory of Yunnan Province, Kunming　650500)

**Abstract**　To solve the problem of replenishing missed labels of partially labeled instances and classifying new instances in multi-label learning,this paper proposes a semi-supervised weak-label classification method by regularization(SWCMR), which takes into account both instance similarity and label correlation. SWCMR first estimates the missing labels of weak-label instances based on label correlation, then uses weak-label instances and unlabeled instances to construct a neighborhood graph, and constructs regularization terms based on the assumption of smoothness from the perspective of instance similarity and label correlation, the next step is to train a semi-supervised weak-label linear classification model using estimated weak label instances, along with unlabeled instances. Experiments on various public multi-label datasets show that SWCMR improves the classification performance, especially when the label information is less, the classification effect is more significant.

**Key words**　semi-supervised weak-label learning; multi-label classification;regularization; label correlation

## 1 引言

多标签学习是指一个实例可以同时标注多个不同的类标签，已被广泛应用于文本分类、图像标注和基因功能分析等任务，是数据挖掘和机器学习领域中的热点问题[1]。早期的多标签学习研究大多是在训练实例相关标签集完整的假设下进行，即假设训练实例的已有标签信息是完整的。但在实际应用场景中，为每个训练实例获取其对应的完整标签信息是非常困难的，而收集标签信息不完整的弱标签实例和大量无标签实例则相对容易。

---

本课题得到国家自然科学基金（No.61562054）资助. 丁家满，硕士，副教授，计算机学会(CCF)会员，主要研究领域为数据挖掘、云计算和机器学习.E-mail: tjoman@126.com. 刘楠，硕士研究生，计算机学会(CCF)学生会员，主要研究领域为数据挖掘、机器学习.E-mail: 184208247@qq.com. 周蜀杰，硕士研究生，主要研究领域为数据挖掘、机器学习.E-mail: 550911204@qq.com. 贾连印，博士，副教授，计算机学会(CCF)会员,主要研究领域为数据库、数据挖掘、信息检索、并行计算.E-mail: jlianyin@163.com. 李润鑫（通信作者），博士，计算机学会(CCF)会员,主要研究领域为数据挖掘、优化算法.E-mail: rxli@kust.edu.cn.

标签信息的不完整性显著影响多标签学习的性能，学者们提出了改进的多标签学习方法。弱标签学习方法考虑到已有标签信息的不完整性，半监督多标签学习方法则关注到大量非常有用的无标签实例。其中，直推式半监督学习方法只处理样本空间内给定的训练数据，同时利用训练数据中的有标签实例和无标签实例进行学习，预测训练数据中无标签实例的标签，取得了良好的预测性能，但由于其无法对训练数据以外未知新实例的标签进行预测，存在一定局限性。

针对上述问题，本文提出一种基于正则化的半监督弱标签分类方法（简称 SWCMR）。SWCMR 考虑到多标签实例的已有标签信息可能存在缺失这一特点，可以增量标注实例的缺失标签，也可以基于弱标签实例和无标签实例对未知新实例的标签进行预测，为归纳式方法。SWCMR 首先根据标签相关性对弱标签实例的缺失标签进行初步预估，度量预估后弱标签实例的经验损失；然后利用所有训练实例构造邻域图，从实例相似性和标签相关性角度构建基于平滑性假设的正则化项；接下来利用预估后的弱标签实例结合无标签实例训练半监督弱标签分类模型；最后进行标签预测和模型评估。

本文其余部分组织结构：第 2 节介绍了相关工作；第 3 节给出了本文方法的原理和流程；第 4 节通过算例对算法流程进行推演；第 5 节汇报实验结果及对比分析；第 6 节总结全文。

## 2 相关工作

处理多标签学习问题的常见策略是将任务分解成 $n$ 个独立的二分类问题，如标签传播[2]和基于支持向量机的多标签分类[3]等，但这种方式忽略了标签间的相关性。合理利用标签间的相关性可以带来更加理想的预测性能，这也是近年来多标签学习研究的关键之一。比如，Zhang 等人[4]将特征集作为所有标签的共同父类，利用贝叶斯网络结构对标签和特征集的条件依赖进行有效编码。Tsoumakas 等人[5]将初始标签集分成若干个随机子集，并为每个子集训练一个分类器，然后集成这些分类器进行预测。Huang 等人[6]利用标签间高阶依赖关系，从不完整标签矩阵中学习新的补充标签矩阵。Wang 等人[7]结合相似度约束和排名约束，利用特权信息和标签间的依赖关系构造最大边缘分类器。Huang 等人[8]利用局部标签相关性增强每个实例的特征表示，并将全局判别拟合和局部相关灵敏度结合到统一框架中，提出一种交替求解的优化方法。

虽然以上方法较好的利用标签相关性并提升了分类预测性能，但是在早期的多标签学习研究中，训练样本的基本假设是每个实例的相关标签集是完整的，而实际应用场景中这种假设受到很大的挑战，为此许多学者提出了基于弱标签学习的解决思路。弱标签学习关注标签信息不完整的实例，解决实例标签集不完整的问题。比如，Bucak 等人[9]提出了使用分组最小角回归对训练错误进行正则化的多标签学习方法 MLR-GL。Wu 等人[10]提出一种基于混合图的弱标签分类方法 MLML。Durand 等人[11]基于深度卷积网络提出一种弱标签分类损失算法。Tan 等人[12]从带有弱标签的不完整视图中学习共享子空间，并提出预测模型 iMVWL。Duan 等人[13]对每个训练实例只给出一个相关标签，提出一种极值弱标签学习算法。Wang 等人[14]采用级联森林结构，提出基于树集成的弱标签深度学习方法。弱标签学习在解决实例标签的不完整性问题上取得了良好效果。

在实际应用中，如何有效利用大量无标签实例来提高分类性能成为了另一个关键。半监督多标签学习同时关注有标签实例和无标签实例，在监督多标签学习的基础上解决实例标签集为空的问题。比如，Liu 等人[15]针对训练样本数量少和标签类别多的问题，提出一种半监督多标签学习框架，并通过有约束的非负矩阵分解进行模型求解。Tang 等人[16]提出一种基于超图的自适应学习方法，通过高阶方式保留数据的局部几何结构，并将特征投影集成获得特征空间。Kong 等人[17]将方法分成概念学习和预测两步，对实例特征构建近邻图，采用直推式方法预测。Zhan 等人[18]提出了一种归纳式的协同训练方法,通过对特征空间进行二分生成两个分类模型来处理多标签实例，然后对无标签实例进行成对排序预测。Guo 等人[19]提出了一种半监督多标签特征学习方法，通过学习一个变换矩阵来进行特征学习，且合并放大后的多标签信息来减少维数影响。半监督学习在有效利用无标签实例的问题上获得广泛应用。

为了兼有弱标签学习和半监督学习的优点，半监督弱标签学习方法应运而生，它同时关注标签信息不完整的弱标签实例和无标签实例，现已成为国内外的研究热点，涌现出许多相关方法。比如，Yu 等人[20]提出一种基于标签和特征依赖最大化的半

监督弱标签分类方法 ProDM，并应用到蛋白质功能预测领域。Wu 等人[21]考虑到标签间的不平衡问题，提出一种基于类别不平衡的半监督弱标签学习方法。Dong 等人[22]模型假设相似的实例在其标签集合中有相似的概念组合，并结合集成学习思想，提出一种半监督弱标签分类模型 SSWL。针对各自的实际情形，以上半监督弱标签学习方法都取得了良好的效果，但它们大多是直推式半监督学习方法，仅对学习过程中观察到的无标签实例进行预测。

基于上述研究现状，本文提出一种基于正则化的半监督弱标签分类方法（简称 SWCMR），下文将对 SWCMR 的具体原理和流程进行详细描述。

## 3 基于正则化的半监督弱标签分类方法

### 3.1 问题定义

假设 $X=[x_1,x_2,...,x_n]\in R^{d\times n}$ 为实例特征矩阵，其中 $x_i\in R^d$ 为第 $i$ 个实例的特征向量，$n$ 为实例数，$d$ 为特征类别数。$Y=[y_1,y_2,...,y_n]\in R^{c\times n}$ 为初始标签矩阵，其中 $y_i\in R^c$ 为第 $i$ 个实例的标签向量，$c$ 为标签类别数。$Y\in\{0,1\}^{n\times c}$，$Y_{\mathrm{ij}}=1$ 表示第 i 个实例有第 j 个标签，$Y_{\mathrm{ij}}=0$ 表示第 i 个实例没有第 j 个标签。为不失一般性，假设 $n=l+\mu$，前 $l$ 个为带有部分标签信息的弱标签实例，后 $\mu$ 个为标签信息完全未知的无标签实例。本文的目标是利用 $X$ 中所有实例训练半监督弱标签分类模型，对训练集以外新实例的多个标签进行预测。

### 3.2 分类模型

综合利用少量弱标签实例和大量无标签实例，训练半监督弱标签分类模型，最小化的目标方程形式如下：

$$\Phi(f)=\Omega_1(f)+\alpha\Omega_2(f)+\beta\Omega_3(f) \tag{1}$$

式（1）由三项构成，第一项 $\Omega_1(f)$ 是基于一致性假设的，保证预测标签矩阵和初始标签矩阵的一致性，度量前 $l$ 个弱标签实例的经验损失，惩罚实例的预测标签不等于给定标签的情况；第二项 $\Omega_2(f)$ 和第三项 $\Omega_3(f)$ 是基于平滑性假设的，保证分类模型预测的平滑性，使近邻点的预测标签相同，其中第二项 $\Omega_2(f)$ 是从实例相似性角度综合利用前 $l$ 个弱标签实例和后 $\mu$ 个无标签实例的正则化；第三项 $\Omega_3(f)$ 是从标签相关性角度利用预测的实例标签的正则化；参数 $\alpha\geq 0$，$\beta\geq 0$ 平衡三项的相对重要性。

分类模型的一般形式定义为：

$$f(x)=W^T x+b \tag{2}$$

其中，$W=[w_1,w_2,...,w_c]\in R^{d\times c}$ 是预测矩阵，$b\in R^c$ 是标签偏差，$f(x)\in R^c$ 是实例 $x$ 的预测标签向量。

SWCMR 的模型框架图如图 1 所示：

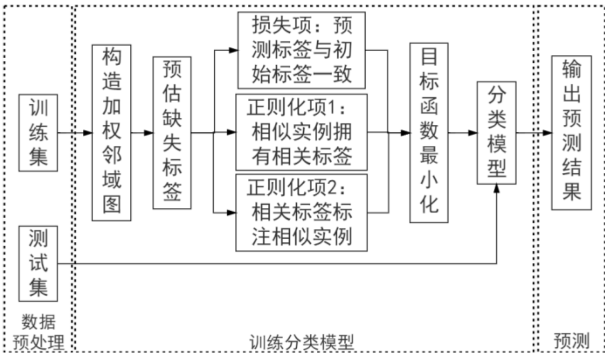


图 1 模型框架图

#### 3.2.1 标签一致性

基于一致性假设，最小化前 $l$ 个弱标签实例的经验损失，即弱标签实例的标签预测结果应与其已知标签一致。式（1）的第一项 $\Omega_1(f)$ 定义为：

$$\Omega_1(f)=\sum_{i=1}^{l}\|f(x_i)-y_i\| \tag{3}$$

由于前 $l$ 个实例为弱标签实例，标签信息存在一定程度上的缺失，式（3）忽视了这一点。为克服标签缺失，本文利用标签相关性对弱标签实例的缺失标签进行初步预估，以提高分类模型的性能。现有的度量标签相关性的策略分为“一阶策略”、“二阶策略”和“高阶策略”：“一阶策略”，即逐一学习单个标签而忽略标签之间的相关性；“二阶策略”，即学习两两标签之间的相关性，在一定程度上利用了标签相关性，具有良好的泛化性能；“高阶策略”，即学习标签间的高阶相关性，具有更强的相关性建模能力，但在计算上要求更高且可伸缩性较差。由于“二阶策略”的简单性和有效性，本文采用“二阶策略”。

基于余弦相似度，标签相关矩阵 $L\in R^{c\times c}$ 定义如下：

$$L(c_1,c_2)=\frac{<Y_{c_1\cdot},Y_{c_2\cdot}>}{\|Y_{c_1\cdot}\|\cdot\|Y_{c_2\cdot}\|} \tag{4}$$

其中，$L(c_1,c_2)$ 是标签 $c_1$，$c_2$ 的相关度，$Y_{c_1\cdot}$ 是 $Y$ 的第 $c_1$ 行。两个标签同时标注的样本数越多，他们之间的相关度越高。利用标签相关矩阵对前 $l$ 个弱标签实例 $x_i$ 的缺失标签进行初步预估，方法如下：

$$\tilde{y}_{ic}=\begin{cases} y_i^T L(\cdot,c), & if\ \ y_{ic}=0 \\ 1, & otherwise \end{cases} \tag{5}$$

当 $y_{ic}=0$ 时，为保证 $\tilde{y}_{ic}\in[0,1]$，将 $\tilde{y}_{ic}$ 归一化为 $\tilde{y}_{ic}/\|\tilde{y}_i\|$，式（5）是通过实例的弱标签以及标签间

的相关度对缺失标签进行初步预估。如果 $y_{ic}=0$，也就是实例 $i$ 没有用标签 $c$ 标记，但标签 $c$ 与实例 $i$ 已经标注的标签间有较大相关度，那么标签 $c$ 很可能就是缺失标签， $\tilde{y}_{ic}$ 将被赋予一个较大的概率值。在考虑前 $l$ 个弱标签实例标签信息缺失的基础上，式（3）可重新定义为：

$$\begin{aligned}\Omega_1(f)&=\sum_{i=1}^{l}\|f(x_i)-\tilde{y}_i\|^2\\&=\sum_{i=1}^{N}(f(x_i)-\tilde{y}_i)V_{ii}(f(x_i)-\tilde{y}_i)^T\\&=tr((F-\tilde{Y})V(F-\tilde{Y})^T)\end{aligned}\tag{6}$$

其中， $tr()$ 是矩阵的迹运算， $F=[f(x_1),f(x_2),...f(x_n)]=W^TX+be^T$， $\tilde{Y}=\left[\tilde{y}_1,\tilde{y}_2,...\tilde{y}_n\right]$， $e\in R^n$ 是所有元素为1的向量。$V\in R^{n\times n}$ 是对角矩阵，前 $l$ 个对角元素为 1，后 $\mu$ 个对角元素为 0。

### 3.2.2 实例相似性

基于 $n$ 个实例样本构造加权邻域图 $G$， $G$ 中每一个顶点对应一个实例 $x_i$， $x_i$ 和 $x_p$ 之间的边表示 $x_i$ 是 $x_p$ 的 $k$ 近邻或 $x_p$ 是 $x_i$ 的 $k$ 近邻。定义实例相似矩阵 $S\in R^{n\times n}$ 表示相邻实例间的相似性：

$$S_{ij}=\begin{cases}\dfrac{1}{\mathrm{z_i}}\exp(-\dfrac{\left\|x_i-x_j\right\|_2^2}{2\sigma^2}), & if\ j\in N_i\\ 0, & otherwise\end{cases}\tag{7}$$

其中， $N_i$ 是实例 $i$ 的 $k$ 近邻实例集， $\sigma$ 是实例特征矩阵 $X$ 中任意两点间欧氏距离的平均值。$z_i=\sum_{j\in N_i}\exp(-\dfrac{\left\|x_i-x_j\right\|_2^2}{2\sigma^2})$，因此 $\sum_{j\in N_i}S_{ij}=1$。为了降低 $k$ 近邻搜索在所有实例中的计算成本，本文使用 kd-tree 有效搜索每个实例的 $k$ 近邻。

基于平滑性假设，从实例相似性的角度，实例的标签集可由其最近邻实例的标签集派生，即相似实例拥有相似标签。式（1）的第二项 $\Omega_2(f)$ 定义为：

$$\begin{aligned}\Omega_2(f)&=\frac{1}{2}\sum_{i,j=1}^{n}\left\|W^Tx_i-W^Tx_j\right\|^2S_{ij}\\&=tr(W^T\sum_{i=1}^{n}(x_iS_{ii}x_i{}^T)W\\&\quad-W^T\sum_{i,j=1}^{n}(x_iS_{ij}x_j{}^T)W)\\&=tr(W^TX(\Lambda-S)X^TW)\\&=tr(W^TXM_SX^TW)\end{aligned}\tag{8}$$

其中，$\Lambda$ 是对角矩阵，$\Lambda_{ii}=\sum_{j=1}^{n}S_{ij}$， $M_S=\Lambda-S$ 为 $S$ 对应近邻图上的规范化 Laplacian 矩阵。

### 3.2.3 标签相关性

基于平滑性假设，从标签相关性的角度，标签在实例上的赋值可以通过它邻近标签的赋值中衍生出来，即相似标签（相关度高的标签）标注相似实例。式（1）的第三项 $\Omega_3(f)$ 定义为：

$$\begin{aligned}\Omega_3(f)&=\frac{1}{2}\sum_{i,j=1}^{c}\left\|X^Tw_i-X^Tw_j\right\|^2L_{ij}\\&=tr(X^T\sum_{i=1}^{c}(w_iL_{ii}w_i{}^T)X\\&\quad-X^T\sum_{i,j=1}^{c}(w_iL_{ij}w_j{}^T)X)\\&=tr(X^TW(\Lambda'-L)W^TX)\\&=tr(X^TWM_LW^TX)\end{aligned}\tag{9}$$

其中，$\Lambda'$ 是对角矩阵，$\Lambda'_{ii}=\sum_{j=1}^{c}L_{ij}$， $M_L=\Lambda'-L$ 为 $L$ 对应近邻图上的规范化 Laplacian 矩阵。

由式（6）、式（8）和式（9），将式（1）写成

$$\begin{aligned}\Phi(f)=tr((F-\tilde{Y})V(F^T-\tilde{Y})^T)&\\+\alpha tr(W^TXM_SX^TW)&\\+\beta tr(X^TWM_LW^TX)&\\ s.t.\ F=W^TX+be^T\end{aligned}\tag{10}$$

对式（10）分别求关于 $W$ 和 $b$ 的偏导

$$\begin{aligned}\frac{\partial\Phi}{\partial W}&=2XV(X^TW+eb^T-\tilde{Y})\\&\quad+2\alpha XM_SX^TW+2\beta XX^TWM_L\end{aligned}\tag{11}$$

$$\frac{\partial\Phi}{\partial b}=2(W^TX+be^T-\tilde{Y}^T)Ve\tag{12}$$

令 $\dfrac{\partial\Phi}{\partial W}=0$， $\dfrac{\partial\Phi}{\partial b}=0$，可得：

$$\begin{aligned}vec(W)=(kron(1,(XV_cX^T+\alpha XM_SX^T))&\\+kron(M_L,\beta XX^T))^{-1}vec(XV_c\tilde{Y}^T)&\end{aligned}\tag{13}$$

$$b=\frac{(\tilde{Y}^T-W^TX)Ve}{n}\tag{14}$$

**其中**，$V_c=V-\dfrac{Vee^TV^T}{n}$ 。

**算法 1**. 基于正则化的半监督弱标签分类方法.

```
输入：
    $X$：实例特征矩阵， $X=\left[x_1,x_2,\cdots,x_n\right]$
    $Y$：初始标签矩阵， $Y=\left[y_1,y_2,\cdots,y_n\right]$
    $\alpha$，$\beta$：式（1）中的参数
    $k$：式（7）中的 $k$ 近邻数
    $x$：无标签新实例
输出：
    $f(x)$：$x$ 的预测标签似然向量
训练：
1. 根据式（4）计算标签相关矩阵 $L$
2. 根据式（5）对弱标签实例的缺失标签进行初步预估
3. 根据式（7）计算实例相似矩阵 $S$
4. 计算 $S$ 对应近邻图上的规范化 Laplacian 矩阵 $M_S$
5. 计算 $L$ 对应近邻图上的规范化 Laplacian 矩阵 $M_L$
```

```
6. 根据式（13）、式（14）解出 $W$ 和 $b$
7. 根据式（2）返回 $f(x)=W^T x+b$
```

## 4 算例

为更直观的理解算法 SWCMR，本文通过具体的实例进行进一步说明。但由于实际数据维度过高，在此仅选取实例中的部分主要特征和标签进行推演。

### 4.1 训练样例

本文选取 5 个训练样例，其中，实例 1、实例 2 和实例 3 为弱标签样例，实例 4 和实例 5 为无标签样例。构造如图 2~3 所示的实例特征矩阵 $X$ 和初始标签矩阵 $Y$，设置模型参数 $\alpha=0.8$，$\beta$=0.12，近邻数 $k=2$，并将其作为输入数据。

$$\begin{array}{c|ccccc|} & 实例1 & 实例2 & \cdots & & 实例5 \\ 特征1 & 0.8 & 0.7 & 0.3 & 0.2 & 0.9 \\ 特征2 & 0.2 & 0.2 & 0.8 & 0.7 & 0.1 \\ & 0.6 & 0.5 & 0.7 & 0.6 & 0.3 \\ \vdots & 0.9 & 0.8 & 0.2 & 0.1 & 0.9 \\ & 0.5 & 0.3 & 0.8 & 0.8 & 0.4 \\ 特征6 & 0.4 & 0.5 & 0.1 & 0.1 & 0.6 \end{array}$$

图 2 实例特征矩阵 $X$ 示意

$$\begin{array}{c|ccccc|} & 实例1 & 实例2 & \cdots & & 实例5 \\ 标签1 & 1 & 0 & 1 & 0 & 0 \\ 标签2 & 1 & 1 & 0 & 0 & 0 \\ & 0 & 0 & 1 & 0 & 0 \\ \vdots & 0 & 0 & 1 & 0 & 0 \\ 标签5 & 1 & 1 & 0 & 0 & 0 \end{array}$$

图 3 初始标签矩阵 $Y$ 示意

### 4.2 训练过程

步骤一：根据初始标签矩阵 $Y$ 和式（4）计算标签相关矩阵 $L$，如图 4 所示：

$$\begin{array}{c|ccccc|} & 标签1 & 标签2 & \cdots & & 标签5 \\ 标签1 & 1 & 0.25 & 0.5 & 0.5 & 0.25 \\ 标签2 & 0.25 & 1 & 0 & 0 & 0.5 \\ & 0.5 & 0 & 1 & 1 & 0 \\ \vdots & 0.5 & 0 & 1 & 1 & 0 \\ 标签5 & 0.25 & 0.5 & 0 & 0 & 1 \end{array}$$

图 4 标签相关矩阵 $L$ 示意

观察标签相关矩阵 $L$ 可知，它显示了标签 1~5 间彼此的相关性。以标签 2 为例，它与标签 1 和标签 5 间都存在相关性，但与标签 3 和标签 4 间的相关性则为 0，原因是由标签 2 标注的实例（实例 1 和实例 2）拥有的其它标签分别是标签 1 和标签 5，而无标签 3 和标签 4。

步骤二：考虑到初始标签矩阵 $Y$ 中的标签信息可能存在缺失，根据标签相关矩阵 $L$ 和式（5）计算 $\tilde{y}_{ic}$，得到如图 5 所示的预估标签矩阵 $\tilde{Y}$；

$$\begin{array}{c|ccccc|} & 实例1 & 实例2 & \cdots & & 实例5 \\ 标签1 & 1 & 0.2 & 1 & 0 & 0 \\ 标签2 & 1 & 1 & 0.07 & 0 & 0 \\ & 0.13 & 0 & 1 & 0 & 0 \\ \vdots & 0.13 & 0 & 1 & 0 & 0 \\ 标签5 & 1 & 1 & 0.07 & 0 & 0 \end{array}$$

图 5 预估标签矩阵 $\tilde{Y}$ 示意

观察预估标签矩阵 $\tilde{Y}$ 可知，相对于初始标签矩阵 $Y$，它标注了弱标签实例的未标注标签缺失概率值。以实例 2 为例，它缺失标签 1 的概率值被补充标注为 0.2，而缺失标签 4 和标签 5 的概率值仍为 0，原因是实例 2 已标注的标签（标签 2 和标签 5）与标签 1 间存在一定程度上的相关度，而与标签 3 和标签 4 间的相关度为 0。

步骤三：根据实例特征矩阵 $X$ 和式（7）计算实例相似矩阵 $S$，如图 6 所示：

$$\begin{array}{c|ccccc|} & 实例1 & 实例2 & \cdots & & 实例5 \\ 实例1 & 0 & 0.52 & 0 & 0 & 0.48 \\ 实例2 & 0.51 & 0 & 0 & 0 & 0.49 \\ & 0.22 & 0 & 0 & 0.78 & 0 \\ \vdots & 0.19 & 0 & 0.81 & 0 & 0 \\ 实例5 & 0.49 & 0.51 & 0 & 0 & 0 \end{array}$$

图 6 实例相似矩阵 $S$ 示意

观察实例相似矩阵 $S$ 可知，它显示了实例 1~5 间彼此的相似性。以实例 1 为例，实例 1 与实例 2 和实例 5 间的相似性较高，而与实例 3 和实例 4 间的相似性较低。（注：实例相似矩阵 $S$ 不是对称矩阵，原因是式（7）是基于 $k$ 近邻算法度量实例间的相似性，而不同实例间可能不互为近邻实例。）

步骤四：计算对角矩阵 $\Lambda$ 和实例相似矩阵 $S$ 对应近邻图上的规范化 Laplacian 矩阵 $M_S$，如图 7~8 所示：

$$\begin{array}{c|ccccc|} & 实例1 & 实例2 & \cdots & & 实例5 \\ 实例1 & 1 & 0 & 0 & 0 & 0 \\ 实例2 & 0 & 1 & 0 & 0 & 0 \\ & 0 & 0 & 1 & 0 & 0 \\ \vdots & 0 & 0 & 0 & 1 & 0 \\ 实例5 & 0 & 0 & 0 & 0 & 1 \end{array}$$

图 7 对角矩阵 $\Lambda$ 示意

$$\begin{array}{cc} & \begin{array}{ccccc}\text{实例1} & \text{实例2} & \cdots & & \text{实例5}\end{array} \\ \begin{array}{c}\text{实例1}\\ \text{实例2}\\ \\ \vdots \\ \text{实例5}\end{array} & \begin{bmatrix} 1 & -0.52 & 0 & 0 & -0.48 \\ -0.51 & 1 & 0 & 0 & -0.49 \\ -0.22 & 0 & 1 & -0.78 & 0 \\ -0.19 & 0 & -0.81 & 1 & 0 \\ -0.49 & -0.51 & 0 & 0 & 1 \end{bmatrix}\end{array}$$

图 8 规范化 Laplacian 矩阵 $M_S$ 示意

观察对角矩阵 $\Lambda$ 和规范化 Laplacian 矩阵 $M_S$ 可知，对角矩阵 $\Lambda$ 中元素的值是实例相似矩阵 $S$ 对应行所有元素的总和，且由于实例相似矩阵 $S$ 每行中的所有元素和为 1，则对角矩阵 $\Lambda$ 的所有元素都为 1；规范化 Laplacian 矩阵 $M_S$ 是对角矩阵 $\Lambda$ 与实例相似矩阵 $S$ 的差值。

步骤五：计算对角矩阵 $\Lambda'$ 和标签相关矩阵 $L$ 对应近邻图上的规范化 Laplacian 矩阵 $M_L$，如图 9~10 所示：

$$\begin{array}{cc} & \begin{array}{ccccc}\text{标签1} & \text{标签2} & \cdots & & \text{标签5}\end{array} \\ \begin{array}{c}\text{标签1}\\ \text{标签2}\\ \\ \vdots \\ \text{标签5}\end{array} & \begin{bmatrix} 2.5 & 0 & 0 & 0 & 0 \\ 0 & 1.75 & 0 & 0 & 0 \\ 0 & 0 & 2.5 & 0 & 0 \\ 0 & 0 & 0 & 2.5 & 0 \\ 0 & 0 & 0 & 0 & 0.75 \end{bmatrix}\end{array}$$

图 9 对角矩阵 $\Lambda'$ 示意

$$\begin{array}{cc} & \begin{array}{ccccc}\text{标签1} & \text{标签2} & \cdots & & \text{标签5}\end{array} \\ \begin{array}{c}\text{标签1}\\ \text{标签2}\\ \\ \vdots \\ \text{标签5}\end{array} & \begin{bmatrix} 1.5 & -0.25 & -0.5 & -0.5 & -0.25 \\ -0.25 & 0.75 & 0 & 0 & -0.5 \\ -0.5 & 0 & 1.5 & -1 & 0 \\ -0.5 & 0 & -1 & 1.5 & 0 \\ -0.25 & -0.5 & 0 & 0 & 0.75 \end{bmatrix}\end{array}$$

图 10 规范化 Laplacian 矩阵 $M_L$ 示意

观察对角矩阵 $\Lambda'$ 和规范化 Laplacian 矩阵 $M_L$ 可知，对角矩阵 $\Lambda'$ 中元素的值是标签相关矩阵 $L$ 对应行所有元素的总和；规范化 Laplacian 矩阵 $M_L$ 是对角矩阵 $\Lambda'$ 与标签相关矩阵 $L$ 的差值。

步骤六：根据实例特征矩阵 $X$、预估标签矩阵 $\tilde{Y}$、对角矩阵 $V$、Laplacian 矩阵 $M_S$ 和 $M_L$、式（13）和式（14）解出预测矩阵 $W$ 和标签偏差 $b$；

$$\begin{array}{cc} & \begin{array}{ccccc}\text{实例1} & \text{实例2} & \cdots & & \text{实例5}\end{array} \\ \begin{array}{c}\text{实例1}\\ \text{实例2}\\ \\ \vdots \\ \text{实例5}\end{array} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\end{array}$$

图 11 对角矩阵 $V$ 示意

$$\begin{array}{cc} & \begin{array}{ccccc}\text{标签1} & \text{标签2} & \cdots & & \text{标签5}\end{array} \\ \begin{array}{c}\text{特征1}\\ \text{特征2}\\ \\ \vdots \\ \\ \text{特征6}\end{array} & \begin{bmatrix} 10.44 & -2.36 & -3.37 & -1.05 & -2.19 \\ -5.55 & -4.81 & 2.85 & 2.12 & -0.88 \\ 6.13 & 3.48 & -3.27 & -1.54 & -0.23 \\ -12.66 & -3.14 & 5.44 & 1.13 & 1.69 \\ -0.71 & 0.69 & 0.61 & 0.18 & 1.29 \\ -0.39 & 2.82 & -1.15 & -0.72 & 1.98 \end{bmatrix}\end{array}$$

图 12 预测矩阵 $W$ 示意

$$\begin{array}{cc} & \begin{array}{ccccc}\text{标签1} & \text{标签2} & \cdots & & \text{标签5}\end{array} \\ \text{实例} & \begin{bmatrix} 0.92 & 1.57 & -0.10 & 0.31 & -0.12 \end{bmatrix}^T\end{array}$$

图 13 标签偏差 $b$ 示意

观察如图 12~13 所示的预测矩阵 $W$ 和标签偏差 $b$ 可知，预测矩阵 $W$ 显示了特征和标签间的对应关系，标签偏差 $b$ 则是不同标签对应的计算偏差。（注：对角矩阵 $V$ 如图 11 所示，它的前 3 个对角元素是 1，后 2 个对角元素为 0，原因是前 3 个实例为弱标签实例，后 2 个实例为无标签实例。）

步骤七：根据预测矩阵 $W$、标签偏差 $b$ 和式（2）返回 $f(x)=W^T x+b$，即得到预测模型。

### 4.3 标签预测

本文选取如图 14 所示的无标签实例 6 作为待预测实例，将实例 6 的特征向量输入到预测模型中，计算得到的预测标签向量如图 15 所示，将标签阈值设置为 0.5，得到如图 16 所示的最终预测标签向量。（算例部分仅为对算法 SWCMR 的流程进行推演，预测效果分析详见实验结果部分。）

$$\begin{array}{cc} & \begin{array}{cccccc}\text{特征1} & \text{特征2} & \cdots & & & \text{特征6}\end{array} \\ \text{实例6} & \begin{bmatrix} 0.4 & 0.7 & 0.5 & 0.3 & 0.8 & 0.1 \end{bmatrix}^T\end{array}$$

图 14 实例 6 特征向量示意

$$\begin{array}{cc} & \begin{array}{ccccc}\text{标签1} & \text{标签2} & \cdots & & \text{标签5}\end{array} \\ \text{实例6} & \begin{bmatrix} -0.13 & -1.10 & 0.91 & 1.01 & 0.25 \end{bmatrix}^T\end{array}$$

图 15 实例 6 初始预测标签向量示意

$$\begin{array}{cc} & \begin{array}{ccccc}\text{标签1} & \text{标签2} & \cdots & & \text{标签5}\end{array} \\ \text{实例6} & \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \end{bmatrix}^T\end{array}$$

图 16 实例 6 最终预测标签向量示意

# 5 实验结果

## 5.1 实验数据

本文在 Cal500、Yeast 和 Delicious 三个多标签数据集上进行实验。其中，Cal500 是一个音乐多标签分类数据集；Yeast 是基因功能分类数据集；Delicious 是大型文本分类数据集，数据规模相对较大，存在大量噪声和冗余特征，为减少维数影响和过滤罕见标签，本文进行了数据预处理，仅保留前 20%常用词和前 30%频繁标签。这些数据集可从开源项目 mulan[23] 的主页（http://mlkd.csd.auth.gr-/multilabel.html）下载得到，详细信息如表 1 所示。

**表 1 实验数据集(Avg.对应每个实例相关标签的平均个数)**

| 数据集 | 实例数 | 特征维度 D | 标签数 C | Avg. |
|---|---|---|---|---|
| Cal500 | 502 | 68 | 174 | 26.044 |
| Yeast | 2417 | 103 | 14 | 4.237 |
| Delicious | 16105 | 295 | 100 | 15.323 |

### 5.2 对比算法及评价准则

为对比分析方法 SWCMR 的性能，本文与 ML-LOC（监督多标签学习方法）[8]、MLR-GL（弱标签学习方法）[10]、Tram（半监督多标签学习方法）[12]、SSWL（半监督弱标签学习方法）[22]和 S4VM（半监督多标签学习方法）[3]五种具有代表性的相关方法对比，并在引言中对这五种相关方法做了介绍。Tram、SSWL 和 S4VM 是直推式分类方法，仅能预测训练集中实例的标签，无法预测训练集以外未知新实例的标签。为进行对比实验，本文设置各直推式分类方法中新实例的标签为训练集中与其距离最近实例的标签，训练集中无标签实例的标签通过直推分类获得。本文按照各对比方法原文建议的参数范围选取最优参数进行实验，并通过在训练集上进行 5 折交叉验证优化 SWCMR 中的参数 $\alpha$ 和 $\beta$ 的取值，$\alpha$ 和 $\beta$ 的取值范围是 0.01~1，步长为 0.01；最终设置 $\alpha = 0.8$，$\beta = 0.12$。

多标签分类问题有多种评价指标，不同的评价指标度量不同方面的性能。半监督弱标签学习是多标签学习的分支，故本文直接选用多标签分类问题的评价指标。本文选用汉明损失（HammLoss）、覆盖率（Coverage）、排序损失（RankLoss）、平均精度（AvgPrec）和 AUC（ROC 曲线下面积）来综合评价 SWCMR 和上述对比方法的性能。其中，文献[3]定义了前四种评价标准，文献[8]定义了 AUC（ROC 曲线下面积），具体公式参见原文。汉明损失（HammLoss）考察预测标签与真实标签的不一致程度；覆盖率（Coverage）衡量了为覆盖 $x$ 的所有相关标签，平均需要从 $f(x)$ 的标签列表跨越标签的个数；排序损失（RankLoss）表明预测结果中不相关标签置信度高于真实相关标签置信度的次数；平均精度（AvgPrec）反映预测标签与真实标签的一致程度；ROC 曲线下面积（AUC）将每个测试实例的所有标签按照预测似然值降序排列，使预测标签的数量从 1 变化到 C，并通过计算真阳性率和假阳性率绘制接收算子曲线，然后计算 ROC 曲线下面积，利用该面积的大小来评价多标签分类的性能。为与其他评价指标保持一致，实验中汇报 1-HammLoss 和 1-RankLoss 的值。类似其它评价指标（Coverage 除外），1-HammLoss 和 1-RankLoss 的值越大，表明性能越好。然而，要使一种方法的性能在所有评价指标上皆优于其它方法是非常困难的。

### 5.3 实验结果

对于每个数据集，本文随机选取 70%的实例作为训练样本，30%的实例作为测试样本。其中，本文考虑到训练样本中有标签实例的两种比例：10%和 30%（10%代表训练样本中 10%为有标签实例，90%为无标签实例；30%同理）。为进行实验，本文假设每个实验数据集的已有标签信息是完整的，考虑到不完整标签率（I. L. Ratio），通过随机隐藏有标签训练样本{20%,40%,60%}的标签来模拟标签不完整的场景，对每个数据集采用相同的数据设置来比较所有方法。为降低随机性影响，本文针对每个数据集上的每个方法重复进行 10 次独立随机实验，记录每个方法在不同不完整标签率下的 10 次平均结果。各方法在 Cal500、Yeast 和 Delicious 数据集上的实验结果(均值±方差)如表 2~表 7 所示。其中，↑(↓)表示该评价指标越大(越小)对应方法性能越好，表中粗体表明其在 95%置信度水平下的配对检验中取得最好结果。

观察表中数据可知，SWCMR 在大多数情况下都获得了相对更好的实验结果，且效果稳定。对于文本分类数据集 Delicious，SWCMR 在各个评价指标上的性能皆好于其它相关对比方法，原因是本文方法最大限度地利用了文本数据间的标签相关性，

**表 2 SWCMR 与对比方法在 CAL500 数据集（10%有标签实例）上的实验结果**

| I. L. Ratio | Methods | 1-HammLoss (↑) | Coverage (↓) | 1-RankLoss (↑) | AvgPrec (↑) | AUC (↑) |
|---|---|---|---|---|---|---|
| 20% | SWCMR | **0.861 ± 0.002** | **90.693 ± 0.391** | 0.722 ± 0.002 | 0.505 ± 0.003 | 0.718 ± 0.002 |
| | ML-LOC | 0.839 ± 0.003 | 93.169 ± 0.273 | 0.696 ± 0.002 | 0.484 ± 0.002 | 0.697 ± 0.003 |
| | MLR-GL | 0.834 ± 0.001 | 91.312 ± 0.353 | 0.724 ± 0.002 | 0.502 ± 0.002 | **0.721 ± 0.002** |
| | Tram | 0.845 ± 0.003 | 90.824 ± 0.217 | 0.645 ± 0.003 | 0.411 ± 0.003 | 0.687 ± 0.003 |
| | SSWL | 0.857 ± 0.002 | 90.854 ± 0.328 | **0.730 ± 0.002** | **0.507 ± 0.003** | 0.707 ± 0.003 |
| | S4VM | 0.792 ± 0.003 | 94.173 ± 0.311 | 0.687 ± 0.002 | 0.392 ± 0.001 | 0.633 ± 0.002 |
| 40% | SWCMR | **0.857 ± 0.002** | 90.837 ± 0.295 | **0.726 ± 0.001** | **0.506 ± 0.003** | **0.723 ± 0.003** |
| | ML-LOC | 0.831 ± 0.003 | 94.251 ± 0.163 | 0.676 ± 0.003 | 0.454 ± 0.003 | 0.675 ± 0.001 |
| | MLR-GL | 0.839 ± 0.002 | 91.075 ± 0.406 | 0.722 ± 0.001 | 0.503 ± 0.002 | 0.718 ± 0.002 |
| | Tram | 0.843 ± 0.001 | 90.470 ± 0.361 | 0.649 ± 0.003 | 0.409 ± 0.004 | 0.686 ± 0.003 |
| | SSWL | 0.856 ± 0.002 | **89.452 ± 0.244** | 0.711 ± 0.001 | 0.505 ± 0.003 | 0.708 ± 0.002 |
| | S4VM | 0.787 ± 0.001 | 95.256 ± 0.287 | 0.678 ± 0.003 | 0.372 ± 0.003 | 0.628 ± 0.003 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 60% | SWCMR | **0.853 ± 0.002** | **90.783 ± 0.248** | **0.724 ± 0.003** | 0.504 ± 0.003 | **0.718 ± 0.003** |
| | ML-LOC | 0.830 ± 0.002 | 94.469 ± 0.146 | 0.674 ± 0.004 | 0.463 ± 0.004 | 0.673 ± 0.003 |
| | MLR-GL | 0.841 ± 0.001 | 91.108 ± 0.348 | 0.709 ± 0.002 | 0.493 ± 0.002 | 0.714 ± 0.002 |
| | Tram | 0.838 ± 0.003 | 91.865 ± 0.276 | 0.646 ± 0.002 | 0.411 ± 0.003 | 0.681 ± 0.001 |
| | SSWL | 0.851 ± 0.001 | 90.891 ± 0.241 | 0.692 ± 0.001 | **0.505 ± 0.001** | 0.705 ± 0.002 |
| | S4VM | 0.783 ± 0.002 | 95.462 ± 0.137 | 0.659 ± 0.002 | 0.351 ± 0.003 | 0.611 ± 0.002 |

而音乐和基因数据间标签相关强度相对较弱，故在Cal500 和 Yeast 上的表现有所下降。在三个数据集上的 90 种（3 个数据集×2 种有标签实例比例×5 种评价指标×3 种不完整标签率）对比实验中，SWCMR 的结果好于 ML-LOC、MLR-GL、Tram、SSWL 和 S4VM 的比率分别为 98.89%、95.56%、96.64%、78.89%、98.89%。

ML-LOC 和 SWCMR 是归纳式学习方法，且利用了标签相关性，但 SWCMR 性能稍好于 ML-LOC，原因是 SWCMR 考虑到标签信息的不完整性和在训练过程中利用了无标签实例；MLR-GL 也是归纳式学习方法，利用弱标签实例预测无标签实例的标签，但实验效果稍逊色于 SWCMR，原因是 SWCMR 在训练过程中利用了无标签实例；Tram 利用了无标签实例，但实验效果相对 SWCMR 较差，这与SWCMR 考虑了标签的不完整性且无需估计先验概率有关；S4VM 由于忽略了标签之间的相关性，学习性能受限。以上结果表明考虑标签信息缺失和标签相关性以及利用无标签实例可以提升多标签分类的性能。

SSWL 是最近提出的半监督弱标签学习方法，假设相似实例有相似标签概念组合，分别为弱标签实例和无标签实例构建模型，然后通过协同正则化框架进行集成，相较其它方法预测性能较好。但是

表 3　SWCMR 与对比方法在 CAL500 数据集（30%有标签实例）上的实验结果

| I. L. Ratio | Methods | 1-HammLoss (↑) | Coverage (↓) | 1-RankLoss (↑) | AvgPrec (↑) | AUC (↑) |
|---|---|---|---|---|---|---|
| 20% | SWCMR | **0.863 ± 0.002** | **89.921 ± 0.286** | **0.738 ± 0.003** | 0.521 ± 0.002 | **0.733 ± 0.002** |
| | ML-LOC | 0.848 ± 0.003 | 92.585 ± 0.295 | 0.710 ± 0.002 | 0.496 ± 0.003 | 0.704 ± 0.003 |
| | MLR-GL | 0.856 ± 0.001 | 90.344 ± 0.353 | 0.729 ± 0.002 | 0.492 ± 0.001 | 0.723 ± 0.002 |
| | Tram | 0.859 ± 0.003 | 90.564 ± 0.217 | 0.522 ± 0.003 | 0.342 ± 0.003 | 0.656 ± 0.003 |
| | SSWL | 0.861 ± 0.002 | 90.054 ± 0.246 | 0.736 ± 0.002 | **0.525 ± 0.003** | 0.717 ± 0.003 |
| | S4VM | 0.801 ± 0.001 | 93.873 ± 0.348 | 0.693 ± 0.002 | 0.407 ± 0.003 | 0.683 ± 0.002 |
| 40% | SWCMR | 0.857 ± 0.003 | 90.417 ± 0.252 | **0.737 ± 0.001** | 0.518 ± 0.001 | **0.731 ± 0.003** |
| | ML-LOC | 0.841 ± 0.001 | 93.456 ± 0.154 | 0.696 ± 0.003 | 0.482 ± 0.003 | 0.694 ± 0.001 |
| | MLR-GL | 0.854 ± 0.002 | 90.485 ± 0.406 | 0.726 ± 0.001 | 0.491 ± 0.003 | 0.726 ± 0.002 |
| | Tram | 0.850 ± 0.003 | 90.070 ± 0.378 | 0.514 ± 0.003 | 0.334 ± 0.004 | 0.654 ± 0.003 |
| | SSWL | **0.858 ± 0.003** | **89.217 ± 0.235** | 0.731 ± 0.001 | **0.521 ± 0.002** | 0.712 ± 0.002 |
| | S4VM | 0.798 ± 0.001 | 94.784 ± 0.247 | 0.686 ± 0.003 | 0.398 ± 0.003 | 0.648 ± 0.003 |
| 60% | SWCMR | 0.855 ± 0.002 | **89.743 ± 0.248** | **0.736 ± 0.002** | 0.516 ± 0.003 | **0.731 ± 0.001** |
| | ML-LOC | 0.839 ± 0.002 | 94.269 ± 0.151 | 0.698 ± 0.003 | 0.485 ± 0.004 | 0.693 ± 0.003 |
| | MLR-GL | **0.856 ± 0.001** | 90.448 ± 0.329 | 0.727 ± 0.002 | 0.493 ± 0.002 | 0.726 ± 0.002 |
| | Tram | 0.845 ± 0.003 | 90.357 ± 0.256 | 0.536 ± 0.002 | 0.345 ± 0.003 | 0.660 ± 0.003 |
| | SSWL | 0.852 ± 0.001 | 90.451 ± 0.241 | 0.726 ± 0.001 | **0.519 ± 0.003** | 0.711 ± 0.002 |
| | S4VM | 0.797 ± 0.003 | 95.036 ± 0.157 | 0.664 ± 0.003 | 0.376 ± 0.003 | 0.642 ± 0.003 |

表 4　SWCMR 与对比方法在 Yeast 数据集（10%有标签实例）上的实验结果

| I. L. Ratio | Methods | 1-HammLoss (↑) | Coverage (↓) | 1-RankLoss (↑) | AvgPrec (↑) | AUC (↑) |
|---|---|---|---|---|---|---|
| 20% | SWCMR | 0.770 ± 0.002 | 3.685 ± 0.149 | **0.723 ± 0.002** | **0.325 ± 0.003** | 0.793 ± 0.002 |
| | ML-LOC | 0.768 ± 0.003 | 4.560 ± 0.230 | 0.715 ± 0.002 | 0.225 ± 0.002 | 0.614 ± 0.003 |
| | MLR-GL | 0.756 ± 0.003 | 4.892 ± 0.096 | 0.713 ± 0.001 | 0.217 ± 0.002 | 0.721 ± 0.002 |
| | Tram | 0.735 ± 0.001 | 4.742 ± 0.090 | 0.708 ± 0.003 | 0.269 ± 0.001 | 0.772 ± 0.003 |
| | SSWL | **0.772 ± 0.002** | **3.618 ± 0.054** | 0.721 ± 0.002 | 0.290 ± 0.003 | **0.796 ± 0.003** |
| | S4VM | 0.733 ± 0.001 | 5.276 ± 0.136 | 0.697 ± 0.003 | 0.208 ± 0.001 | 0.563 ± 0.002 |
| 40% | SWCMR | **0.768 ± 0.002** | **3.394 ± 0.146** | 0.719 ± 0.001 | **0.323 ± 0.002** | **0.794 ± 0.002** |
| | ML-LOC | 0.693 ± 0.003 | 4.922 ± 0.269 | 0.711 ± 0.003 | 0.229 ± 0.003 | 0.610 ± 0.001 |
| | MLR-GL | 0.739 ± 0.002 | 5.392 ± 0.192 | 0.696 ± 0.001 | 0.236 ± 0.002 | 0.723 ± 0.002 |
| | Tram | 0.702 ± 0.001 | 5.390 ± 0.149 | 0.705 ± 0.003 | 0.282 ± 0.004 | 0.698 ± 0.003 |
| | SSWL | 0.761 ± 0.002 | 3.728 ± 0.054 | **0.721 ± 0.001** | 0.295 ± 0.003 | 0.792 ± 0.002 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | S4VM | 0.696 ± 0.003 | 5.522 ± 0.014 | 0.693 ± 0.003 | 0.212 ± 0.003 | 0.558 ± 0.003 |
| 60% | SWCMR | **0.767 ± 0.002** | **3.368 ± 0.177** | **0.719 ± 0.003** | **0.324 ± 0.003** | 0.793 ± 0.003 |
| | ML-LOC | 0.654 ± 0.003 | 5.203 ± 0.293 | 0.707 ± 0.004 | 0.229 ± 0.004 | 0.608 ± 0.003 |
| | MLR-GL | 0.719 ± 0.001 | 5.587 ± 0.163 | 0.688 ± 0.002 | 0.233 ± 0.002 | 0.722 ± 0.002 |
| | Tram | 0.692 ± 0.003 | 5.667 ± 0.175 | 0.704 ± 0.003 | 0.296 ± 0.003 | 0.706 ± 0.001 |
| | SSWL | 0.759 ± 0.003 | 3.722 ± 0.054 | 0.716 ± 0.001 | 0.298 ± 0.002 | **0.794 ± 0.003** |
| | S4VM | 0.632 ± 0.002 | 5.953 ± 0.068 | 0.688 ± 0.002 | 0.216 ± 0.003 | 0.551 ± 0.002 |

SWCMR 的性能在大多数情况下都略好或近似于 SSWL，尤其是在有标签实例比例较低和缺失标签比例较高时。原因是 SWCMR 针对缺失标签进行了初步预估，并基于平滑性假设构造了两个正则化项，且 SWCMR 为归纳式分类方法，可直接对新实例的标签进行预测。上述实验结果体现了本文方法 SWCMR 的整体性能。

### 5.4 有效性分析

为分析 SWCMR 预估缺失标签、利用实例相似性和标签相关性的有效性，本文引入 SWCMR 的三个变体：SWCMR-NC，SWCMR-NS 和 SWCMR-NL。SWCMR-NC 不考虑标签信息存在缺失，即公式（10）中的$\tilde{Y}=Y$；SWCMR-NS 不考虑实例相似性，即公式（10）中的$\alpha=0$；SWCMR-NL 不考虑标签相关性，即公式（10）中的$\beta=0$。与 5.3 中的实验设置类似，本部分实验在有标签实例比例为 30%下进行，这 4 种方法在 Delicious 数据集上的实验结果如图 17 所示。

经实验表明，SWCMR 在绝大多数情况下都获得了相对更好的分类效果。具体来说，在 Delicious 数据集上的 15 种（1 个数据集×1 种有标签实例比例×5 种评价指标×3 种不完整标签率）对比实验中，经 Friedman 检验，这四种方法的性能排序为

表 5　SWCMR 与对比方法在 Yeast 数据集（30%有标签实例）上的实验结果

| I. L. Ratio | Methods | 1-HammLoss (↑) | Coverage (↓) | 1-RankLoss (↑) | AvgPrec (↑) | AUC (↑) |
|---|---|---|---|---|---|---|
| 20% | SWCMR | 0.775 ± 0.002 | 3.367 ± 0.060 | **0.734 ± 0.002** | **0.327 ± 0.003** | **0.795 ± 0.002** |
| | ML-LOC | 0.775 ± 0.003 | 4.276 ± 0.106 | 0.720 ± 0.002 | 0.242 ± 0.002 | 0.618 ± 0.003 |
| | MLR-GL | 0.789 ± 0.003 | 4.518 ± 0.137 | 0.704 ± 0.003 | 0.235 ± 0.002 | 0.739 ± 0.002 |
| | Tram | 0.776 ± 0.003 | 4.406 ± 0.103 | 0.715 ± 0.003 | 0.286 ± 0.002 | 0.751 ± 0.003 |
| | SSWL | **0.790 ± 0.002** | **3.449 ± 0.115** | 0.731 ± 0.002 | 0.301 ± 0.003 | 0.788 ± 0.003 |
| | S4VM | 0.776 ± 0.003 | 4.917 ± 0.136 | 0.706 ± 0.001 | 0.226 ± 0.001 | 0.584 ± 0.003 |
| 40% | SWCMR | 0.771 ± 0.002 | 3.368 ± 0.101 | **0.731 ± 0.002** | **0.328 ± 0.002** | **0.794 ± 0.003** |
| | ML-LOC | 0.706 ± 0.003 | 4.587 ± 0.166 | 0.718 ± 0.003 | 0.248 ± 0.003 | 0.613 ± 0.001 |
| | MLR-GL | 0.749 ± 0.002 | 5.117 ± 0.111 | 0.703 ± 0.001 | 0.253 ± 0.002 | 0.734 ± 0.002 |
| | Tram | 0.714 ± 0.003 | 5.008 ± 0.110 | 0.713 ± 0.003 | 0.305 ± 0.004 | 0.717 ± 0.003 |
| | SSWL | **0.775 ± 0.002** | **3.553 ± 0.231** | 0.729 ± 0.002 | 0.308 ± 0.001 | 0.781 ± 0.002 |
| | S4VM | 0.721 ± 0.001 | 5.416 ± 0.109 | 0.705 ± 0.003 | 0.213 ± 0.003 | 0.579 ± 0.004 |
| 60% | SWCMR | 0.769 ± 0.002 | **3.368 ± 0.060** | **0.731 ± 0.002** | **0.328 ± 0.003** | **0.796 ± 0.003** |
| | ML-LOC | 0.695 ± 0.002 | 4.592 ± 0.106 | 0.715 ± 0.004 | 0.249 ± 0.002 | 0.604 ± 0.003 |
| | MLR-GL | 0.725 ± 0.001 | 5.391 ± 0.137 | 0.694 ± 0.002 | 0.252 ± 0.002 | 0.726 ± 0.002 |
| | Tram | 0.701 ± 0.003 | 5.557 ± 0.103 | 0.712 ± 0.003 | 0.306 ± 0.001 | 0.717 ± 0.002 |
| | SSWL | **0.769 ± 0.001** | 3.572 ± 0.004 | 0.728 ± 0.003 | 0.299 ± 0.002 | 0.780 ± 0.003 |
| | S4VM | 0.698 ± 0.002 | 5.791 ± 0.078 | 0.701 ± 0.002 | 0.211 ± 0.003 | 0.572 ± 0.001 |

表 6　SWCMR 与对比方法在 Delicious 数据集（10%有标签实例）上的实验结果

| I. L. Ratio | Methods | 1-HammLoss (↑) | Coverage (↓) | 1-RankLoss (↑) | AvgPrec (↑) | AUC (↑) |
|---|---|---|---|---|---|---|
| 20% | SWCMR | **0.887 ± 0.002** | **20.889 ± 0.351** | **0.798 ± 0.002** | **0.329 ± 0.003** | **0.809 ± 0.000** |
| | ML-LOC | 0.854 ± 0.001 | 22.873 ± 0.133 | 0.731 ± 0.002 | 0.260 ± 0.002 | 0.737 ± 0.003 |
| | MLR-GL | 0.867 ± 0.001 | 26.385 ± 1.512 | 0.671 ± 0.003 | 0.272 ± 0.001 | 0.692 ± 0.002 |
| | Tram | 0.875 ± 0.003 | 22.591 ± 0.074 | 0.568 ± 0.001 | 0.221 ± 0.002 | 0.789 ± 0.003 |
| | SSWL | 0.874 ± 0.002 | 20.933 ± 0.115 | 0.788 ± 0.002 | 0.258 ± 0.003 | 0.764 ± 0.002 |
| | S4VM | 0.804 ± 0.001 | 27.442 ± 0.337 | 0.614 ± 0.003 | 0.236 ± 0.001 | 0.661 ± 0.003 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 40% | SWCMR | **0.885±0.002** | **20.487±0.513** | **0.798±0.002** | **0.326±0.002** | **0.809±0.001** |
| | ML-LOC | 0.853±0.003 | 22.569±0.474 | 0.723±0.003 | 0.256±0.003 | 0.730±0.002 |
| | MLR-GL | 0.865±0.002 | 26.327±1.262 | 0.669±0.003 | 0.273±0.002 | 0.681±0.002 |
| | Tram | 0.872±0.003 | 23.083±0.729 | 0.569±0.003 | 0.223±0.004 | 0.787±0.003 |
| | SSWL | 0.876±0.002 | 21.563±0.172 | 0.789±0.002 | 0.259±0.003 | 0.760±0.002 |
| | S4VM | 0.808±0.003 | 27.754±0.062 | 0.609±0.001 | 0.233±0.003 | 0.657±0.004 |
| 60% | SWCMR | **0.885±0.002** | **20.487±0.512** | **0.797±0.003** | **0.326±0.003** | **0.808±0.001** |
| | ML-LOC | 0.853±0.002 | 22.473±0.437 | 0.723±0.004 | 0.256±0.002 | 0.730±0.003 |
| | MLR-GL | 0.864±0.002 | 26.322±1.251 | 0.668±0.002 | 0.273±0.002 | 0.681±0.002 |
| | Tram | 0.873±0.003 | 23.093±0.728 | 0.568±0.003 | 0.223±0.001 | 0.787±0.003 |
| | SSWL | 0.874±0.001 | 21.563±0.362 | 0.782±0.003 | 0.250±0.002 | 0.750±0.003 |
| | S4VM | 0.803±0.002 | 27.781±0.034 | 0.606±0.002 | 0.233±0.003 | 0.657±0.001 |

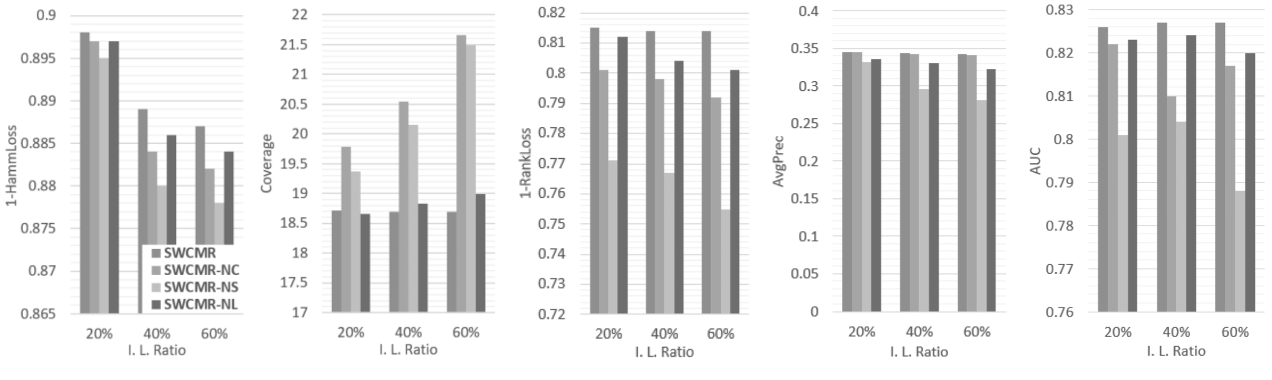
SWCMR>SWCMR-NL>SWCMR-NC>SWCMR-NS。SWCMR 在 5 个评价指标上的结果均显著超过其变体方法，这不仅表明 SWCMR 的预测性能较好，还表明本文模型各部分的有效性。特别地，当$\alpha$=0 时，SWCMR-NS 为弱标签分类方法，即训练模型时未利用无标签实例，故也说明了在标签信息不足时利用大量无标签实例的必要性。

### 5.5 参数敏感性分析

为分析参数$\alpha$和$\beta$对本文方法 SWCMR 分类预测性能的影响，本文对比了参数$\alpha$和$\beta$在{0.005,0.01,0.1,0.2,...,1}这 12 种取值组合情况下的 1-RankLoss 的实验结果。从图 18 中可以观察到 SWCMR 在不同参数取值下的结果都较为稳定，当

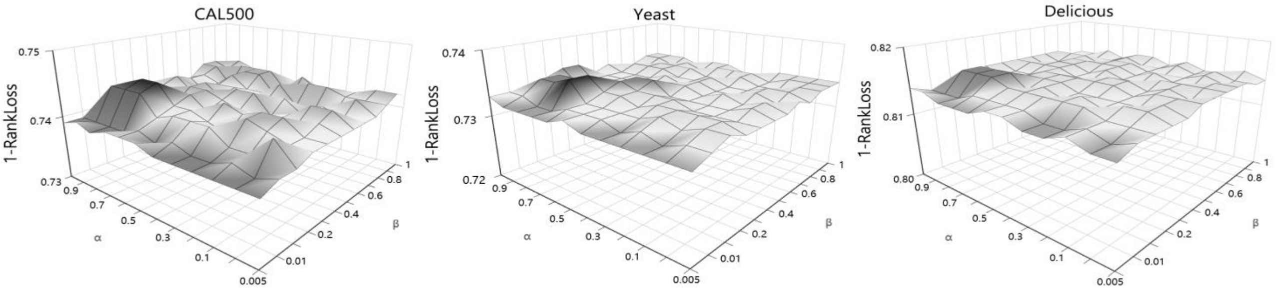
表 7 SWCMR 与对比方法在 Delicious 数据集（30%有标签实例）上的实验结果

| I. L. Ratio | Methods | 1-HammLoss (↑) | Coverage (↓) | 1-RankLoss (↑) | AvgPrec (↑) | AUC (↑) |
|---|---|---|---|---|---|---|
| 20% | SWCMR | **0.898±0.002** | **18.716±0.261** | **0.815±0.002** | **0.345±0.003** | **0.826±0.000** |
| | ML-LOC | 0.868±0.001 | 22.389±0.082 | 0.740±0.002 | 0.292±0.001 | 0.745±0.002 |
| | MLR-GL | 0.871±0.013 | 26.246±1.155 | 0.781±0.001 | 0.293±0.002 | 0.702±0.002 |
| | Tram | 0.876±0.002 | 21.808±0.136 | 0.601±0.003 | 0.226±0.004 | 0.798±0.003 |
| | SSWL | 0.878±0.002 | 19.371±0.588 | 0.793±0.002 | 0.295±0.003 | 0.756±0.003 |
| | S4VM | 0.839±0.003 | 26.776±0.749 | 0.643±0.003 | 0.249±0.003 | 0.661±0.001 |
| 40% | SWCMR | **0.889±0.002** | **18.692±0.241** | **0.814±0.002** | **0.344±0.002** | **0.827±0.000** |
| | ML-LOC | 0.867±0.003 | 21.963±0.012 | 0.739±0.003 | 0.287±0.003 | 0.757±0.001 |
| | MLR-GL | 0.869±0.002 | 26.003±1.715 | 0.776±0.001 | 0.295±0.002 | 0.693±0.002 |
| | Tram | 0.874±0.003 | 21.750±0.919 | 0.609±0.003 | 0.229±0.004 | 0.796±0.002 |
| | SSWL | 0.875±0.002 | 18.993±0.915 | 0.791±0.002 | 0.297±0.001 | 0.763±0.002 |
| | S4VM | 0.837±0.003 | 26.682±0.276 | 0.621±0.004 | 0.238±0.001 | 0.670±0.003 |
| 60% | SWCMR | **0.887±0.003** | **18.693±0.241** | **0.814±0.003** | **0.344±0.003** | **0.827±0.000** |
| | ML-LOC | 0.863±0.002 | 21.978±0.011 | 0.738±0.004 | 0.287±0.003 | 0.757±0.003 |
| | MLR-GL | 0.868±0.001 | 26.003±1.755 | 0.776±0.002 | 0.290±0.002 | 0.693±0.002 |
| | Tram | 0.874±0.003 | 21.759±0.934 | 0.609±0.001 | 0.229±0.001 | 0.763±0.002 |
| | SSWL | 0.876±0.003 | 18.998±0.145 | 0.792±0.003 | 0.297±0.003 | 0.796±0.003 |
| | S4VM | 0.821±0.002 | 26.692±0.371 | 0.609±0.002 | 0.235±0.003 | 0.665±0.001 |



（a）1-HammLoss（b）Coverage（c）1-RankLoss（d）AvgPrec（e）AUC

图 17 SWCMR 与其变体方法在 Delicious 数据集（30%有标签实例）上的实验结果



（a）CAL500 数据集（b）Yeast 数据集（c）Delicious 数据集

图 18 不同参数组合下 SWCMR 在 Cal500、Yeast 和 Delicious 数据集上的实验结果

$\alpha \in [0.7, 0.9]$，$\beta \in [0.08, 0.2]$时，SWCMR 在三个数据集上皆可取得较好结果。以在 Delicious 数据集上的实验结果为例，SWCMR 在 144 种参数组合情况下的实验结果都好于或等于其他对比方法。当$\beta < 0.02$时，SWCMR 的预测性能开始下降，这表明利用标签相关性可以提升预测性能，也进一步证明了利用标签相关性的合理性。

### 5.6 复杂度和运行时间分析

时间复杂度是衡量算法优劣的重要指标，方法 SWCMR 的时间复杂度主要由矩阵的乘法和逆运算决定，而计算标签相关矩阵$L$、预估标签矩阵$\tilde{Y}$、实例相似矩阵$S$和标签偏差$b$的时间复杂度相对较小，故方法 SWCMR 的时间复杂度主要取决于预测矩阵$W$的计算。又由于不同数据集样本数$n$、特征数$d$和标签数$c$的数量规模不同和度量实例相似性时的近邻数$k \ll n$，SWCMR 训练和预测的时间复杂度可分别表示为$O_T = O(dn^2 + dnc + nd^2 + c^2 d^2)$、$O_P = O(dnc + n^2)$。时间复杂度定性的描述了执行算法所需要的计算工作量，而算法执行所耗费的具体时间则无法从理论上得出。

为分析方法 SWCMR 与各对比方法的实际运行效率，本文统计了各方法在相同实验平台（CPU i5-4590,16GBRAM,Win10,Matlab2018a）下不同数据集上的运行时间，并计算出各方法 5 次独立运行的平均时间，如表 8 所示：

表 8 各方法运行时间对比（单位：秒）

| Methods | Cal500 | Yeast | Delicious | Total |
|---|---|---|---|---|
| SWCMR | 6.54 | 23.28 | 848.03 | 877.85 |
| ML-LOC | 7.74 | 48.65 | 8489.03 | 8545.42 |
| MLR-GL | 2.51 | 14.94 | 342.25 | 359.70 |
| Tram | 2.32 | 13.83 | 312.67 | 328.82 |
| SSWL | 12.54 | 127.37 | 2936.89 | 3076.80 |
| S4VM | 4.67 | 89.37 | 3257.51 | 3351.55 |

由表 8 可知，SWCMR 的运行时间总是小于除 Tram 和 MLR-GL 外的其他对比方法。Tram 和 SWCMR 都是构造邻域图，通过求解平滑性正则化问题来预测无标签实例的标签，但 SWCMR 求解过程中涉及矩阵的逆，因此相对耗时较多。MLR-GL 训练过程仅利用了 30%的有标签实例，且通过优化的组稀疏和排序损失求解，耗时较少。ML-LOC 训练过程中也只利用了有标签实例，但耗时较多，主要原因是使用了聚类算法生成 LOC 码。SSWL 和 S4VM 相较其他方法耗时最多，两种方法皆为半监督集成学习方法，模型复杂度较高。S4VM 基于低密度分离假设，使用多个低密度分隔符来逼近决策边界，通过模拟退火算法和采样方式求解全局最优解。SSWL 通过协同正则化框架集成多个模型，将问题表述成双凸优化问题，最后通过块坐标下降方法求解。上述实验结果表明，本文提出的方法 SWCMR 除了获得相对其它方法更好的预测效果外，也获得了相对较高的运行效率（Tram 和 MLR-GL 除外）。

## 6 结束语

SWCMR 是基于正则化的一种半监督弱标签归纳式分类方法，具有同时兼顾实例相似性和标签相关性的特点。实验表明，该方法适用于半监督弱标签分类问题，且获得了相对较高的预测精度和运行效率，尤其是标签信息较少时，分类效果提升更显著。在未来工作中，将考虑如何在稀疏空间利用高阶策略更准确的刻画标签相关性和设计新的半监督弱标签分类方法，进一步提高模型预测性能。

### 参考文献

[1] ZhangML, ZhouZH. A review on multilabel learning algorithms. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819-1837.

[2] Wu X, Hiramatsu K, Kashino K, et al. Label Propagation with Ensemble of Pairwise Geometric Relations: Towards Robust Large-Scale Retrieval of Object Instances. International Journal of Computer Vision, 2018, 126(7): 689-713.

[3] Li YF, Zhou ZH. Towards Making Unlabeled Data Never Hurt. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(1): 175-188.

[4] Zhang ML, Zhang K. Multi-label learning by exploiting label dependency//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington, USA, 2010: 999-1008.

[5] Tsoumakas G, Katakis I, Vlahavas I, Random k-labelsets for multilabel classification,IEEE Transactions on Knowledge & Data Engineering, 2011, 23(7): 1079-1089.

[6] Huang J, Qin F, Zheng X, et al. Improving multi-label classification with missing labels by learning label-specific features. Information Sciences, 2019, 492: 124-146.

[7] Wang S, Chen S, Chen T, et al. Learning with privileged information for multi-Label classification. Pattern Recognition, 2018, 81: 60-70.

[8] Huang SJ, Zhou ZH, Multi-label learning by exploiting label correlations locally//Proceedings of the 26th AAAI Conference on Artificial Intelligence. Toronto, Canada, 2012: 949–955.

[9] Bucak SS, Jin R, Jain A K, et al. Multi-label learning with incomplete class assignments//Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA, 2011: 2801-2808.

[10] Wu BY, Jia F, Liu W, et al. Multi-label Learning with Missing Labels Using Mixed Dependency Graphs. International Journal of Computer Vision, 2018, 126(8): 875-896.

[11] Durand T, Mehrasa N, Mori G, et al. Learning a Deep ConvNet for Multi-Label Classification With Partial Labels//Proceedings of the 32th IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 647-657.

[12] Tan Q, Yu GX, Domeniconi C, et al. Incomplete Multi-View Weak-Label Learning//Proceedings of the International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018: 2703-2709.

[13] Duan J, Li X, Mu D. Learning Multi Labels from Single Label—An Extreme Weak Label Learning Algorithm. Wuhan University Journal of Natural Sciences, 2019, 24(2): 161-168.

[14] Wang Q, Yang L, Li Y, et al. Learning from Weak-Label Data: A Deep Forest Expedition//Proceedings of the National Conference on Artificial Intelligence. New York, USA, 2020:6251-6258.

[15] Liu Y, Jin R, Yang L. Semi-supervised multilabel learning by constrained non-negative matrix factorization//Proceedings of the 21st National Conference on Artificial Intelligence. Boston, USA, 2006: 421-426.

[16] Tang C, Liu X, Wang P, et al. Adaptive Hypergraph Embedded Semi-Supervised Multi-Label Image Annotation. IEEE Transactions on Multimedia, 2019, 21(11): 2837-2849.

[17] Kong X, Ng MK, Zhou ZH, et al. Transductive Multilabel Learning via Label Set Propagation. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(3): 704-719.

[18] Zhan W, Zhang ML. Inductive Semi-supervised Multi-Label Learning with Co-Training// Proceedings of the Acm Sigkdd International Conference. Halifax, Canada, 2017: 1305-1314.

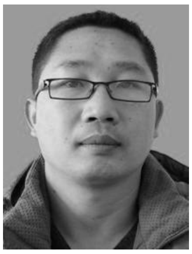
[19] Guo BL, Tao H, Hou CP, et al. Semi-supervised multi-label feature learning via label enlarged discriminant analysis. Knowledge and Information Systems, 2020, 62(6): 2383-2417.

[20] Yu GX, Domeniconi C, Rangwala H, et al. Protein function prediction using dependence maximization//Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer. Prague, Czechia, 2013: 574-589.

[21] Wu BY, Lyu S, Ghanem B, et al. Constrained submodular minimization for missing labels and class imbalance in multi-label learning// Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, USA, 2016: 2229-2236.

[22] Dong HC, Li YF, Zhou ZH, et al. Learning from Semi-Supervised Weak-Label Data// Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 2926-2933.

[23] Tsoumakas G, Vilcek J, Xioufis ES. Mulan: a java library for multi-label learning [Online], available: http://mulan.sourceforge.net/datasets.html, January 1, 2010.



**DING Jia-Man,** M. S., associate professor. His current research interests include data mining, cloud computing and machine learning.

**LIU Nan,** M. S. candidate. Her current research interests include data mining and machine learning.

**ZHOU Shu-Jie,** M. S. candidate. His current research interests include data mining and machine learning.

**JIA Lian-Yin,** Ph. D.,associate professor. His current research interests include database, data mining, information retrieval and parallel computing.

**LI Run-Xin,** Ph. D. His current research interests include data mining and otimization algorithm.

**Background**

Semi-supervised weak-labellearning is a hot topic in data mining and machine learning and to solve the problem of replenishing missed labels of partially labeled instances and classifying new instances in multi-label learning. The concept of semi-supervised weak-labellearning was formally proposed by the group of ZHOU Zhi-Huaofthe 32nd AAAI Conference on Artificial Intelligence and attracted the attention of numerous scholars. Recently, many semi-supervised weak-label classification methods are proposed, but most of them are the transductive method and unable to predict the labels of unknown new instances outside of the training data.

Our research group is committed to data mining and machine learning. this paper proposes a semi-supervised weak-label classification method by regularization(SWCMR), which takes into account both instance similarity and label correlation.SWCMR considers that the existing label information of multi-label instances may be missing. It can incrementally label the missing labels of the instances or predict the labels of unknown new instances based on weak-label instances and unlabeled instances. Experiments on various public multi-label datasets show that SWCMR improves the classification performance, especially when the label information is less, the classification effect is more significant.

This work has been supported by the National Natural Science Foundation of China(61562054).