

基于实例结构的不完备多标签学习

陈天柱^{1,2,3}, 李凤华^{1,2}, 郭云川^{1,2}, 李子孚¹

(1. 中国科学院信息工程研究所, 北京 100093; 2. 中国科学院大学网络空间安全学院, 北京 100049;
3. 中国电子科技集团公司信息科学研究院, 北京 100086)

摘 要: 针对现有标签缺失下多标签学习方案未能有效解决标签缺失的问题, 提出了基于实例结构的不完备多标签学习方案, 考虑实例特征和标签结构特点, 利用数据标签向量几何相似度来补全缺失标签, 利用加权排序来降低正关系学为负关系所带来的模型偏差, 并利用低秩结构来俘获模型低秩结构。具体地, 通过确保数据预测标签几何相似度与数据标签几何相似度的一致性来俘获数据流型结构; 通过度量完备标签下和不完备标签下的排序损失来区分标签与实例的相关程度。实验结果表明, 所提方案优于典型的标签缺失下的多标签学习方案, 甚至在一些评估标准下其精度比最好对比方案提升了 10% 以上。

关键词: 多标签学习; 低秩结构; 流型正则; 标签排序

中图分类号: TP302

文献标识码: A

DOI: 10.11959/j.issn.1000-436x.2021186

Instance structure based multi-label learning with missing labels

CHEN Tianzhu^{1,2,3}, LI Fenghua^{1,2}, GUO Yunchuan^{1,2}, LI Zifu¹

1. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

2. School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

3. Information Science Academy of China Electronics Technology Group Corporation, Beijing 100086, China

Abstract: To address the problem that the existing methods in multi-label learning did not efficiently deal with the problems, the instance structure based multi-label learning scheme with missing labels was proposed. By considering the feature and label structure of instance, the similarity of label vectors were exploit to fill the missing labels and the weight rank loss was exploit to reduce the model bias. Meanwhile, the weight rank loss was also exploit to reduce the model bias. More specially, the manifold structure was capture by forcing the consistency of the geometry similarity of labels and one of the predicted labels. By measuring ranking loss for complete labels and incomplete labels, the relevance of label was distinguish to instance. Experiment results show that the superior performances of the proposed approach compared with the state-of-the-art methods and the accuracy is improved by more than 10% compared with the best comparison scheme under some evaluation criteria.

Keywords: multi-label learning, low rank structure, manifold structure, label ranking

1 引言

物联网中实体、资源和服务常用唯一标识来识别, 每个标识可能包含多个具有语义信息的标签。

由于物联网中资源的海量性, 采用人工方式标注资源的语义面临巨大挑战。为了自动标注语义, 学术界和工业界提出了多标签学习技术。然而, 在一些应用场景中, 训练集数据由网络志愿者标记而获

收稿日期: 2021-07-19; 修回日期: 2021-10-09

通信作者: 李子孚, lizifu@iie.ac.cn

基金项目: 国家重点研发计划基金资助项目 (No.2018YFB2100402); 国家自然科学基金资助项目 (No.U1836203); 山东省重点研发计划基金资助项目 (No.2019JZY020127)

Foundation Items: The National Key Research and Development Program of China (No.2018YFB2100402), The National Natural Science Foundation of China (No.U1836203), ShanDong Provincial Key Research and Development Program (No.2019JZY020127)

得, 其通常呈现标签缺失的现象^[1]。

现有标签缺失下的多标签学习方法主要利用标签关联来填充缺失标签。这些方法大体可分为显式标签关联法和隐式标签关联法。其中, 显式标签关联法的核心思想是利用标签向量直接计算标签关联, 并约束预测标签也具有该关联。例如, Wu 等^[2]利用标签向量计算了标签相似度, 并约束语义相似标签的预测向量也相似。Li 等^[3]从维基百科等辅助源中抽取标签知识创建标签关联矩阵。Wu 等^[4]通过实例特征向量和标签向量直接构建无向标签关联图, 在此基础上融合标签语义层次结构(例如, 动物标签是马标签的父节点)创建了标签关联混合图。然而, 上述方法仅利用标签间的直接关联, 忽略了非直接关联, 这使这些方法的预测表现通常较差。

隐式标签关联法基于低秩假设, 利用随机投影、典型相关分析、线性变化等方法压缩高维标签为低维空间。随后在低维空间中预测标签并将预测标签翻转到原标签空间而获得预测标签。其中, 隐式标签关联嵌入标签压缩中。例如, Zhang 等^[5]利用典型相关分析技术挖掘压缩标签与预测标签间的线性关联, 从而将标签关联蕴含于多标签学习中。Yu 等^[6]在特征矩阵低秩假设下, 基于矩阵低秩分解技术俘获了拟合矩阵低秩结构。然而, 上述方法并未有效解决标签缺失下相似度准确度量问题, 也未结合显式关联与隐式关联来解决标签缺失下的多标签标记问题。此外, 在标签完备情况下, 标签排序^[7-8]也是提升模型预测精度的重要方法。然而, 现有工作很少从标签排序的角度来解决标签缺失下的多标签标记问题。

针对上述不足, 本文从实例的特征和标签结构出发, 利用标签矩阵低秩结构和数据标签流型结构来补全缺失标签, 利用加权排序损失体现标签对数据的相关程度并降低正关系学为负关系所带来的模型偏差程度。具体地, 通过确保数据预测标签几何相似度与数据标签几何相似度的一致性来俘获数据流型结构; 通过度量完备标签下和不完备标签下的排序损失来区分正负标签对实例的相关程度。实验结果表明本文方案优于典型的标签缺失下的多标签学习方案。本文的贡献如下。

1) 本文设计了新的流型结构, 该结构降低了对标签缺失率的依赖。具体地, 本文假设实例正标签是准确标记的, 利用 k-means 聚类方法对每个标签

所有正实例的特征向量进行聚类, 并利用类中心的最小主角计算标签相似度。同时, 本文通过度量完备标签下和非完备标签下的排序损失, 提出了新的排序正则化方法。

2) 为高效优化目标函数, 本文松弛目标函数精确解为非精确解, 提出了线性 ADMM (LADMM, linear alternating direction method of multipliers)。通过线性近似迭代获得子问题的近似解, 该近似解可确保线性 ADMM 的收敛性。

3) 在 6 个数据集上的实验结果表明, 本文方案优于代表性的标签缺失下的多标签学习方案。在一些评估标准下, 本文方案的精度比最好的对比方案提升了 10% 以上。

2 相关工作

标签关联法是处理标签缺失情况下多标签标记问题的一般方法, 其可分为显式方法和隐式方法。在显式方法中, Li 等^[3]提出包含特征输入层、标签输出层和潜标签输出层的三层玻尔兹曼机, 并利用潜输出层学习高阶标签关联, 同时利用标签先验知识来俘获标签关联, 利用其正则化模型。Wu 等^[4]以标签为节点, 以标签一致性和实例相似度构建节点无向边, 以标签语义层次构建节点有向边, 创建标签关联混合图。其中, 标签语义层次是指某一标签的语义信息包含了另一标签的语义信息, 如植物标签和玫瑰标签。Zhu 等^[9]利用标签在数据集上的分布计算标签相似度, 并确保相似度高的标签的标签预测分布距离小, 相似度低的标签的标签预测分布距离大。具体地, 将训练集聚类成子集, 在子集和数据集中分别俘获局部和全局标签结构。Huang 等^[10]从缺失训练数据中训练高阶标签关联, 对每个标签学习与其最相关的特征而获得数据的低维表示。该方案同时学习标签关联与低维表示。然而, 上述方法将缺失标签看作负标签来计算标签关联, 造成标签直接关联度量错误, 该错误随着标签缺失率的变大而越来越不准确。同时, 这些方法也忽略了标签非直接关联。

针对以上不足, 研究者提出隐式方法来提升模型预测精度, 其大体分为基于标签矩阵方法和基于拟合矩阵方法。基于标签矩阵方法的核心思想是利用典型相关分析和线性变化等技术压缩高维标签空间为低维空间, 随后在低维空间中预测标签并将其翻转到原标签空间。例如, Tai 等^[11]对标签矩阵

做奇异值分解，利用奇异值向量矩阵（低维正交矩阵）做投影矩阵，利用其转置矩阵做解码矩阵（正交矩阵的逆矩阵是其转置矩阵）。Zhang 等^[5]利用 2 个线性变换分别压缩标签和预测标签，通过典型相关分析寻找压缩标签空间最优预测方向来俘获标签关联。随后，在文献[5]方案的框架下，Zhang 等^[12]保持标签距离结构与嵌入标签距离结构一致性，并最大化数据预测标签与其他数据预测标签的边界距离，提出了嵌入标签可区分性和可预测性的多标签学习方案。然而，随着标签规模的变大，确保整体数据距离结构的方案^[12]需要极大的计算开销。针对该不足，Jin 等^[13]仅约束实例预测标签与其嵌入标签距离小于实例预测标签与其近邻数据嵌入标签距离。

在基于拟合矩阵的方法中，Yu^[6]等分解拟合矩阵为 2 个低维矩阵，利用一个矩阵压缩特征空间为低维空间，另一个矩阵拟合压缩后的特征从而获得标签预测。此外，Jing 等^[14]为俘获有效标签个数，约束拟合矩阵核范数从而实现拟合矩阵低秩结构，同时引入流型结构来确保特征相似度高的数据的预测标签相似度也高。Li 等^[1]利用与标签个数等维的标签关联方阵来映射实例标签向量，从而恢复缺失标签，并利用一个具有低秩结构的去噪声编码器来学习标签关联方阵。Zhu 等^[9]分解标签矩阵为 2 个低维矩阵的和，利用一个矩阵表示标签的潜标签空间，另一个矩阵表示原始标签投影到潜标签空间的投影矩阵。然而，上述方法未结合标签显式关联和隐式关联来解决标签缺失下的标签标记问题。Li 等^[15]对标签矩阵做低秩分解来补全缺失标签，并利用希尔伯特-施密特独立性准则来俘获标签空间和特征空间的关联，从而提升模型预测精度。此外，Gupta 等^[16]将训练集聚成多个子类，基于单词嵌入方法的思想，分别将每个子类的数据实例和其最近邻实例映射为主题模型中的单词和上下文，利用 Skip Gram Negative Sampling 将每个实例的特征向量和标签向量嵌入低维特征空间和标签空间上。进一步地，Liu 等^[17]利用平滑剪裁的绝对偏差和最小最大凹惩罚两类非凸函数来消除子类数据预测时的模型偏差。

实例正标签比其负标签更能体现实例语义信息，约束模型正标签输出值大于负标签输出值可正则化模型。基于此观察，研究者针对标签完备情况，利用线性模型、支持向量机、神经网络提出了各种

标签排序方案。Usunier 等^[18]基于有序加权平均函数和铰链函数提出了有序加权排序损失函数。Weston 等^[7]利用铰链函数来近似 0-1 排序损失，提出了函数性质为光滑的加权损失函数。然而，上述方案利用传统的数据特征作为模型输入，导致预测精度较低。Gong 等^[8]训练了包含几个卷积层和 dense 连接层的神经网络结构，以近似处理后的加权损失函数为基础构造 top-k 排序损失，提出了基于深度卷积排序的图像多标签标记方法。Durand 等^[19]以 4 096 维的深度图像特征为模型输入，利用 2 个潜变量分别预测标签的正向证据和负向证据。标签排序已广泛用于多媒体数据标记、语音标记中，并获得了较大成功，然而标签排序很少用于解决标签缺失下的多标签学习问题。

3 模型描述

3.1 模型动机

模型流型正则和标签排序正则步骤如图 1 所示。本文研究动机基于如下 3 个事实。1) 在较小规模和中等规模标签的数据集中，标签矩阵和特征矩阵通常呈现一定的低秩结构。通过学习标签矩阵和特征矩阵的低秩结构来俘获标签内在关联，并利用该标签关联补充缺失标签。2) 对于 2 个在特征向量上相似度越高的实例，其标签向量也越相似；对于 2 个在特征向量上相似度不高的实例，其标签向量也不相似。利用实例准确的特征信息来度量实例相似度，并利用该相似度约束预测标签的相似度以实现缺失标签补全。3) 实例的正标签比负标签更能描述实例的语义信息，因此在模型训练过程中确保正标签的输出值大于负标签的输出值具有较大意义。综上所述，本文从低秩结构出发俘获了标签间的隐式关联；利用数据流型结构确保数据特征间距离与标签间距离的一致性，进而俘获了标签显式关联；利用标签加权排序区分了正负标签对数据的描述能力，俘获了每个实例的标签显式关系。

3.2 标签标记流程

本节简略介绍标签标记过程，其可粗略分为 4 个子过程：训练集处理、模型建模、模型训练和标签标记，如图 2 所示。训练集处理过程中，首先，专业人员根据现实需求和数据语义信息挑选标签并制定标签字典。然后，标签标记者从标签字典中挑选合适的标签来标记收集的数据，从而形成训练集。由于标记者工作疏忽，训练集中的数据可能呈

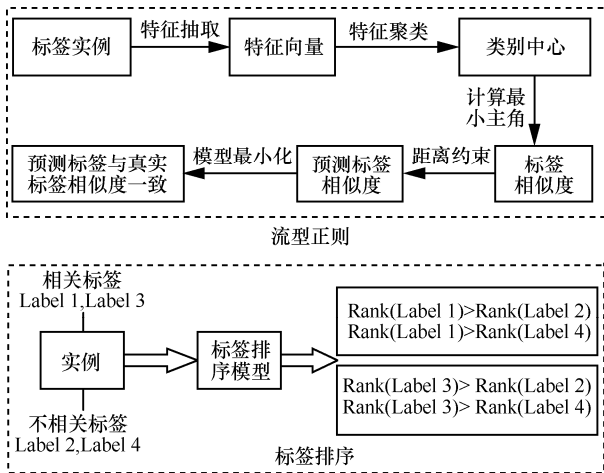


图 1 模型流型正则和标签排序正则步骤

现标签缺失、标签冗余和标签噪声现象。最后，专业人员对训练集数据做特征处理，将数据转化为可度量形式。例如，依据图像数据的纹理、色彩和像素信息将数据转化为特征向量表示。模型建模过程根据训练集特点，选取线性模型、随机森林模型和深度学习模型来预测数据标签，并根据标签关联关系、流型结构和 L2 范数等对模型进行正则化。模型训练过程根据目标函数的特点，挑选恰当的优化方法来最小化目标函数。典型的优化方法包括梯度下降法、线性搜索法、牛顿法和翻转牛顿法等。标签标记过程将检

测数据的向量表示传递给预测模型获得数据标签预测。本文标签标记过程不涉及训练集处理步骤。

3.3 拟合矩阵的低秩结构

设 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ 为训练集， $x_i = (x_{i1}, x_{i2}, \dots, x_{id}) \in R^d$ 是 d 维特征向量， $y_i = (y_{i1}, y_{i2}, \dots, y_{il}) \in \{-1, 0, 1\}^l$ 是 l 维标签向量。其中， $y_{ij} = 1$ 表示第 i 个实例 x_i 标记了第 j 个标签（正标签）， $y_{ij} = -1$ 表示第 i 个实例与第 j 个标签无关（负标签）， $y_{ij} = 0$ 表示第 i 个实例的第 j 个标签缺失。本节的核心目标是寻找恰当矩阵 $M \in R^{l \times d}$ 来准确预测新实例 x 的标签 $f(x) = Mx$ 。具体地，如果 $M_i x \geq 0$ ，实例 x 标记第 i 个标签；否则，实例 x 未标记第 i 个标签。本节称 Y 为标签矩阵，称 Mx 为预测标签矩阵。

本文利用二次损失来度量预测标签与真实标签的差异，即 $\|Y - MX\|_2^2$ （也称为损失函数），其含义是寻找各点到拟合直线距离和最小的直线，也就是平方和最小。通过约束拟合矩阵 M 的秩获得如下目标函数

$$\begin{aligned} \min_M L(Y, MX) &= \|Y - MX\|_2^2 \\ \text{s.t. } \text{rank}(M) &< k_0 \end{aligned}$$

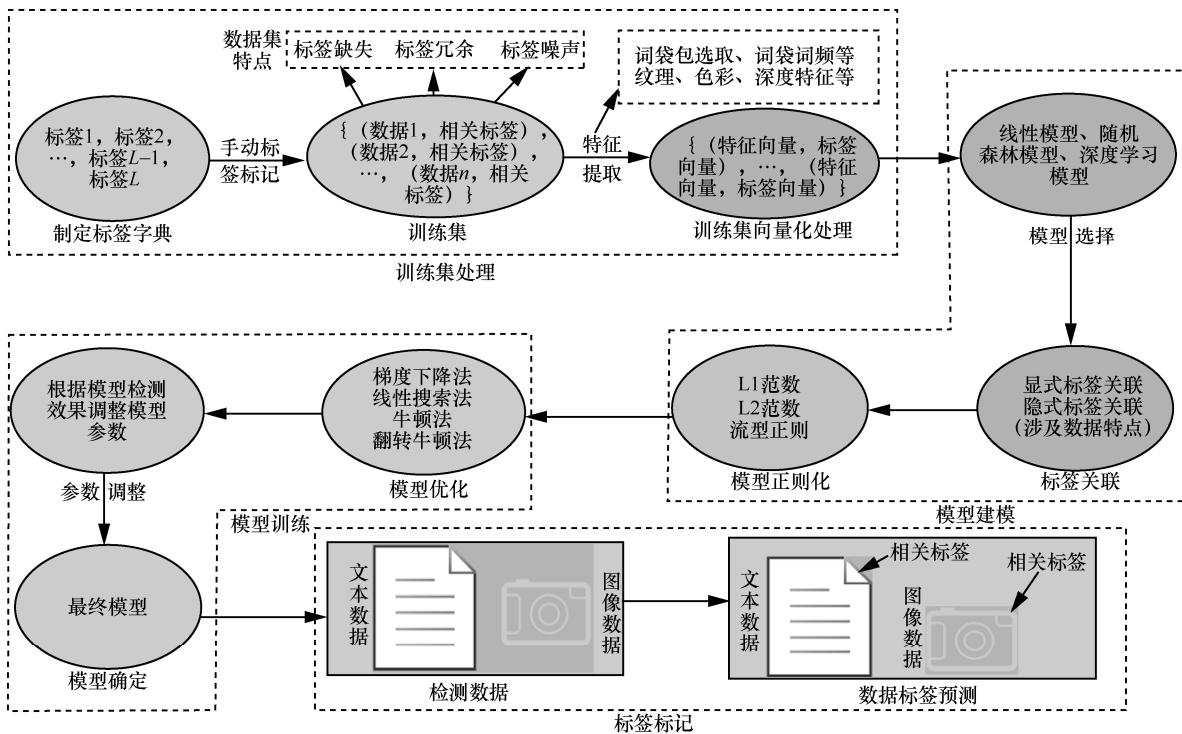


图 2 标签标记流程

其中, k_0 用于约束拟合矩阵的秩, 约束条件 $\text{rank}(\mathbf{M}) < k_0$ 是非凸优化条件^[20]。为降低优化困难, 本文将该约束条件转化为拟合矩阵核范数约束, 获得等价形式为

$$\begin{aligned} \min_{\mathbf{M}} L(\mathbf{Y}, \mathbf{M}\mathbf{X}) &= \|\mathbf{Y} - \mathbf{M}\mathbf{X}\|_2^2 \\ \text{s.t. } \|\mathbf{M}\|_* &< k_1 \end{aligned} \quad (1)$$

其中, 参数 k_1 与 k_0 一一对应。

3.4 流型正则

机器学习的核心目标是设计一个在训练集和检测集上都具有较好效果的算法。研究者们设计了一些正则化方法^[21]来减少测试误差(可能以增大训练误差为代价)。正则化方法大体分为两类: 控制模型复杂度、控制实例的几何结构^[22]。本节利用实例分布设计了一个流型正则用于控制实例的几何结构, 其核心挑战是寻找一个相似度度量函数来满足如下性质: 对于任何 2 个实例, 如果两者的特征向量越相似, 两者的预测标签距离就越小, 反之亦然。

3.4.1 几何相似度

本节利用 k-means 聚类方法对每个标签所有正实例的特征向量进行聚类, 其中特征向量间的距离采用欧氏距离来度量。具体地, 第 i 个标签对应正实例聚成 p 个类 L_i^p , 簇中心集为 S_i , 每个元素 $\mathbf{s} \in S$ 是一个 d 维特征向量。对于任何一个簇中心 \mathbf{s} 和簇中心集 S , 本节用最小主角来度量 \mathbf{s} 到 S 的距离 $\text{dist}(\mathbf{s}, S)$ ^[23], 即

$$\text{dist}(\mathbf{s}, S) = \max_{\mathbf{t} \in S} \frac{\mathbf{s}^T \mathbf{t}}{\|\mathbf{s}\|_2 \|\mathbf{t}\|_2}$$

其中, \mathbf{s}^T 是 \mathbf{s} 的转置向量。基于该距离定义, 本节定义标签 i 与标签 j 的几何相似度为

$$\text{gs}(i, j) = \frac{1}{2}(\text{gs}(i, j) + \text{gs}(j, i))$$

其中, $\text{gs}(i, j) = \frac{1}{|S_i|} \sum_{\mathbf{s} \in S_i} \text{dist}(\mathbf{s}, S_j)$, $|S_i|$ 是 S_i 的基数。几何相似度具有对称性, 即 $\text{gs}(i, j) = \text{gs}(j, i)$ 。

3.4.2 标签预测距离

对于任意标签 i 和标签 j , 本文定义 2 个标签对应正实例 $L_i^p \cup L_j^p$ 的预测距离为 $\|\mathbf{M}_i \mathbf{X}^{(i,j)} - \mathbf{M}_j \mathbf{X}^{(i,j)}\|_2^2$, 其中, \mathbf{M}_i 是矩阵 \mathbf{M} 的第 i 行, $\mathbf{X}^{(i,j)}$ 是集合 $L_i^p \cup L_j^p$ 的矩阵形式, 即 $\mathbf{X}_{\{i,j\}} = [\mathbf{X}_1, \dots, \mathbf{X}_m] \in R^{m \times n}$ 。 \mathbf{x}_k 是 $L_i^p \cup L_j^p$ 的第 k 元素的特征向量, m 是集合 $L_i^p \cup L_j^p$

的基数, $\mathbf{M}_i \mathbf{X}^{(i,j)}$ 是第 i 个标签在正实例集 $\mathbf{X}^{(i,j)}$ 上的标签预测分布。

融合几何相似度与标签预测距离, 本文获得如下形式的流型正则

$$\sum_{i=1}^{l-1} \sum_{j=1}^{l-i} \text{gs}(i, j) \|\mathbf{M}_i \mathbf{X}^{(i,j)} - \mathbf{M}_j \mathbf{X}^{(i,j)}\|_2^2$$

该流型正则最终作为目标函数的一部分。在对目标函数最小化的过程中, $\text{gs}(i, j)$ 值越大, 标签预测分布距离 $\|\mathbf{M}_i \mathbf{X}^{(i,j)} - \mathbf{M}_j \mathbf{X}^{(i,j)}\|_2^2$ 越小; $\text{gs}(i, j)$ 值越小, 标签预测分布距离 $\|\mathbf{M}_i \mathbf{X}^{(i,j)} - \mathbf{M}_j \mathbf{X}^{(i,j)}\|_2^2$ 越大。更具体地, 如果 $\text{gs}(i, j) > \text{gs}(i, k)$, 通常有 $\|\mathbf{M}_i \mathbf{X}^{(i,j)} - \mathbf{M}_j \mathbf{X}^{(i,j)}\|_2^2 < \|\mathbf{M}_i \mathbf{X}^{(i,k)} - \mathbf{M}_k \mathbf{X}^{(i,k)}\|_2^2$ 成立。

3.5 排序正则

排序正则的核心目标是确保每个实例的正标签排在其负标签的前面, 可定义为

$$\mathbf{M}_i \mathbf{X} > \mathbf{M}_j \mathbf{X}, \forall \mathbf{X} \in (i, j)^{p,n}, 1 \leq i, j \leq l$$

其中, $(i, j)^{p,n} = \{\mathbf{x} | \text{实例 } \mathbf{x} \text{ 的第 } i \text{ 个标签是正标签, 第 } j \text{ 个标签是负标签}\}$ 。为实现该目标, 本文定义标签完备下任意标签对的排序损失。

3.5.1 标签完备下的排序损失

本文假设训练集 D 中所有实例都被完全标记, 即任意实例 \mathbf{x} 的任意标签 i 和标签 j 为正标签或负标签, 不存在任何缺失。针对该情况, 对于任意标签 i 和标签 j 的排序损失定义为

$$\text{rlc}(i, j) = \sum_{\mathbf{x} \in (i, j)^{p,n}} h(j, i, \mathbf{x}) + \sum_{\mathbf{x} \in (i, j)^{n,p}} h(i, j, \mathbf{x})$$

其中 $(i, j)^{n,p} = \{\mathbf{x} | \text{实例 } \mathbf{x} \text{ 是第 } i \text{ 个负标签, 实例 } \mathbf{x} \text{ 的第 } j \text{ 个标签是正标签}\}$, 铰链损失 $h(j, i, \mathbf{x})$ 为 $\max\{0, 1 + \mathbf{M}_j \mathbf{x} - \mathbf{M}_i \mathbf{x}\}$ 。因此, 所有标签对的排序损失和为实例排序损失 $\text{wrlc}(\mathbf{M}) = \sum_{1 \leq i, j \leq l} \text{rlc}(i, j)$ 。

由于无法区分一个实例不同正标签对实例描述程度的差异, 本文未对实例正标签进行排序; 同理, 本文也未对实例负标签进行排序。

3.5.2 非完备标签下的排序损失

针对任意标签对 (i, j) , 本文假设其未被完全标记, 即训练集 D 中必存在一实例, 其标签 i 和标签 j 存在一个标签缺失。针对该情况, 训练集 D 中的所有实例可分为 5 个类别: $(i, j)^{p,m}$ 、 $(i, j)^{n,m}$ 、 $(i, j)^{m,p}$ 、

$(i, j)^{m,n}$ 和 $(i, j)^{m,m}$ 。其中, $(i, j)^{p,m} = \{\mathbf{x} | \text{实例 } \mathbf{x} \text{ 的第 } i \text{ 个标签是正标签, 第 } j \text{ 个标签是缺失标签}\}$, 剩余 4 种类别的定义类似该定义。上述 5 种情况的排序损失函数 $\text{rli}(i, j)^{p,m}$ 、 $\text{rli}(i, j)^{n,m}$ 、 $\text{rli}(i, j)^{m,p}$ 、 $\text{rli}(i, j)^{m,n}$ 和 $\text{rli}(i, j)^{m,m}$ 分别定义为

$$\begin{aligned} \text{rli}(i, j)^{p,m} &= (1 - \theta_j) \sum_{\mathbf{x} \in (i, j)^{p,m}} h(j, i, \mathbf{x}) \\ \text{rli}(i, j)^{n,m} &= \theta_j \sum_{\mathbf{x} \in (i, j)^{n,m}} h(i, j, \mathbf{x}) \\ \text{rli}(i, j)^{m,p} &= (1 - \theta_i) \sum_{\mathbf{x} \in (i, j)^{m,p}} h(i, j, \mathbf{x}) \\ \text{rli}(i, j)^{m,n} &= \theta_i \sum_{\mathbf{x} \in (i, j)^{m,n}} h(j, i, \mathbf{x}) \\ \text{rli}(i, j)^{m,m} &= \theta_i (1 - \theta_j) \sum_{\mathbf{x} \in (i, j)^{m,m}} h(j, i, \mathbf{x}) + \\ &\theta_j (1 - \theta_i) \sum_{\mathbf{x} \in (i, j)^{m,m}} h(i, j, \mathbf{x}) \end{aligned}$$

其中, θ_i 和 θ_j 分别是标签 i 和标签 j 的正实例个数与整体实例个数的比值。完整的排序损失函数为

$$\begin{aligned} \text{wrli}(\mathbf{M}) &= \sum_{i,j} \text{rli}(i, j)^{p,m} + \text{rli}(i, j)^{n,m} + \\ &\text{rli}(i, j)^{m,p} + \text{rli}(i, j)^{m,n} + \text{rli}(i, j)^{m,m} \end{aligned}$$

则标签排序损失 $\text{RL}(\mathbf{M})$ 为标签完备下的排序损失与标签非完备情况下的排序损失和, 即

$$\text{RL}(\mathbf{M}) = \text{wrlc}(\mathbf{M}) + \text{wrli}(\mathbf{M})$$

3.6 目标函数

通过融合流型正则、排序正则与式(1), 并通过惩罚法转化有约束优化问题为无约束优化问题, 获得如下形式的目标函数

$$\begin{aligned} \min_{\mathbf{M}} O(\mathbf{M}) &= \|\mathbf{Y} - \mathbf{MX}\|_2^2 + \lambda_3 \|\mathbf{M}\|_* + \\ &\lambda_1 \sum_{i,j} \text{gs}(i, j) \|\mathbf{M}_i \mathbf{X}^{(i,j)} - \mathbf{M}_j \mathbf{X}^{(i,j)}\|_2^2 + \lambda_2 \text{RL}(\mathbf{M}) \end{aligned} \quad (2)$$

其中, λ_1 和 λ_2 用于控制相关部分的权值。

4 模型优化

在目标函数式(2)中, 部分项 $g(\mathbf{M}) = \lambda_1 \sum_{i,j} \text{gs}(i, j) \|\mathbf{M}_i \mathbf{X}^{(i,j)} - \mathbf{M}_j \mathbf{X}^{(i,j)}\|_2^2$ 和 $\text{RL}(\mathbf{M})$ 是非连续不规则函数, 其导函数是间断函数, 这使目标函

数不具备精确解。同时, 目标函数中的核范数 $\|\mathbf{M}\|_*$ 与二次项的和可用高效算法来优化^[24]。针对上述特点, 本节借鉴 ADMM 的可分性, 将目标函数中核范数 $\|\mathbf{M}\|_*$ 与二次项组合, 并与剩余项分割开, 利用文献[24]算法来优化核范数部分。对于剩余项, 本节利用线性近似法求解其近似解。上述方法融合了近似法与传统 ADMM^[1], 本文称之为线性 LADMM, 其详细过程如下。

首先, 引入辅助变量 $\mathbf{M} = \mathbf{Z}$, 将变量 \mathbf{M} 的优化问题转化为变量 \mathbf{M} 和 \mathbf{Z} 的优化问题, 其目标函数为

$$\begin{aligned} L(\mathbf{M}, \mathbf{Z}) &= \|\mathbf{Y} - \mathbf{MX}\|_2^2 + \lambda_3 \|\mathbf{Z}\|_* + \\ &\lambda_1 g(\mathbf{M}) + \lambda_2 \text{RL}(\mathbf{M}) \\ \text{s.t. } &\mathbf{M} = \mathbf{Z} \end{aligned} \quad (3)$$

其次, 转化式(3)问题为如下增广拉格朗日形式

$$\begin{aligned} L(\mathbf{M}, \mathbf{Z}, \boldsymbol{\gamma}) &= \|\mathbf{Y} - \mathbf{MX}\|_2^2 + \lambda_3 \|\mathbf{Z}\|_* + \lambda_1 g(\mathbf{M}) + \\ &\lambda_2 \text{RL}(\mathbf{M}) + \frac{\beta}{2} \|\mathbf{M} - \mathbf{Z}\|_2^2 + \text{tr}(\boldsymbol{\gamma}^T (\mathbf{M} - \mathbf{Z})) \end{aligned} \quad (4)$$

其中, β 是正的乘子参数, $\boldsymbol{\gamma} \in R^{d \times l}$ 是正的拉格朗日常数。本节算法包含三部分: 更新 \mathbf{M} 、更新 \mathbf{Z} 、更新 $\boldsymbol{\gamma}$ 。

4.1 更新 \mathbf{M}

固定式(4)问题中的变量 \mathbf{Z} 和 $\boldsymbol{\gamma}$, 式(4)问题变为求解变 \mathbf{M} 的最小化问题

$$\begin{aligned} \min_{\mathbf{M}} &\|\mathbf{Y} - \mathbf{MX}\|_2^2 + \lambda_1 g(\mathbf{M}) + \lambda_2 \text{RL}(\mathbf{M}) + \\ &\frac{\beta}{2} \|\mathbf{M} - \mathbf{Z}\|_2^2 + \text{tr}(\boldsymbol{\gamma}^T (\mathbf{M} - \mathbf{Z})) \end{aligned} \quad (5)$$

式(5)问题包含非连续的不规则项 $g(\mathbf{M})$ 和 $\text{RL}(\mathbf{M})$, 其不具备精确解。本文利用二次近似法优化该问题从而获得任意程度的近似解。具体地, 选取初始点 $\mathbf{M}_{(0)}$, 对 2 个项 $g(\mathbf{M})$ 和 $\text{RL}(\mathbf{M})$ 在初始点展开为二次近似形式, 展开后的函数是良定形式, 求其精确解。以精确解为新的展开点重复上述过程直到获得任意近似程度的解。函数 $g(\mathbf{M})$ 和 $\text{RL}(\mathbf{M})$ 在近似点 $\mathbf{M}_{(k)}$ 的近似展开形式为

$$\begin{aligned} g(\mathbf{M}) &= \\ g(\mathbf{M}_{(k)}) &+ \frac{\partial g(\mathbf{M}_{(k)})}{\partial \mathbf{M}} (\mathbf{M} - \mathbf{M}_{(k)}) + \frac{1}{2\eta_{(k)}} \|\mathbf{M} - \mathbf{M}_{(k)}\|_2^2 \end{aligned}$$

$$\text{RL}(\mathbf{M}) = \text{RL}(\mathbf{M}_{(k)}) + \frac{\partial \text{RL}(\mathbf{M}_{(k)})}{\partial \mathbf{M}} (\mathbf{M} - \mathbf{M}_{(k)}) + \frac{1}{2\eta_{(k)}} \|\mathbf{M} - \mathbf{M}_{(k)}\|_2^2$$

其中, $\eta_{(k)}$ 是 k 循环的步长, (k) 是第 k 步循环的索引, $\frac{\partial g(\mathbf{M}_{(k)})}{\partial \mathbf{M}}$ 和 $\frac{\partial \text{RL}(\mathbf{M}_{(k)})}{\partial \mathbf{M}}$ 是 $g(\mathbf{M})$ 和 $\text{RL}(\mathbf{M})$ 在点 $\mathbf{M}_{(k)}$ 的导数, 约束项 $\frac{1}{2\eta_{(k)}} \|\mathbf{M} - \mathbf{M}_{(k)}\|_2^2$ 用于确保第 $k+1$ 次循环点 $\mathbf{M}_{(k+1)}$ 能在第 k 次循环点 $\mathbf{M}_{(k)}$ 附近, 避免出现一次近似法的解振荡问题。

用 $g(\mathbf{M})$ 和 $\text{RL}(\mathbf{M})$ 的近似形式代替式(5)问题中 $g(\mathbf{M})$ 和 $\text{RL}(\mathbf{M})$, 获得如下形式的目标函数

$$\min_{\mathbf{M}} \|\mathbf{Y} - \mathbf{M}\mathbf{X}\|_2^2 + \lambda_1 \frac{\partial g(\mathbf{M}_{(k)})}{\partial \mathbf{M}} \mathbf{M} + \lambda_2 \frac{\partial \text{RL}(\mathbf{M}_{(k)})}{\partial \mathbf{M}} \mathbf{M} + \frac{1}{\eta_{(k)}} \|\mathbf{M} - \mathbf{M}_{(k)}\|_2^2 + \frac{\beta}{2} \|\mathbf{M} - \mathbf{Z}\|_2^2 + \text{tr}(\boldsymbol{\gamma}_{(k)}^T (\mathbf{M} - \mathbf{Z}))$$

求函数关于变量 \mathbf{M} 的导函数, 并设置导函数为 0, 可得

$$2\mathbf{M}\mathbf{X}\mathbf{X}^T + \mathbf{Q}_1 \mathbf{M} = \mathbf{Q}_2 \quad (6)$$

$$\text{其中, } \mathbf{Q}_2 = 2\mathbf{Y}\mathbf{X}^T + \beta\mathbf{Z} - \boldsymbol{\gamma} - \lambda_1 \frac{\partial g(\mathbf{M}_{(k)})}{\partial \mathbf{M}} - \lambda_2 \frac{\partial \text{RL}(\mathbf{M}_{(k)})}{\partial \mathbf{M}} + \frac{1}{\eta_{(k)}} \mathbf{M}_{(k)}, \quad \mathbf{Q}_1 = 2\lambda_3 (\mathbf{C} - \mathbf{I})^T (\mathbf{C} - \mathbf{I}) + \frac{\eta_{(k)}\beta + 2}{\eta_{(k)}} \mathbf{I}.$$

通过分析发现, \mathbf{Q}_1 和 $2\mathbf{X}\mathbf{X}^T$ 是 2 个对称矩阵, 其对称分解为 $\mathbf{Q}_1 = \mathbf{U}\mathbf{A}\mathbf{U}^T$ 和 $2\mathbf{X}\mathbf{X}^T = \mathbf{V}\mathbf{B}\mathbf{V}^T$. 其中 \mathbf{U} 和 \mathbf{V} 是特征向量, \mathbf{A} 和 \mathbf{B} 是对角矩阵, 对角线元素是 \mathbf{Q}_1 和 $2\mathbf{X}\mathbf{X}^T$ 的特征值. 利用对称分解替换式(6)中的 \mathbf{Q}_1 和 $2\mathbf{X}\mathbf{X}^T$, 可得

$$\mathbf{M}\mathbf{V}\mathbf{B}\mathbf{V}^T + \mathbf{U}\mathbf{A}\mathbf{U}^T \mathbf{M} = \mathbf{Q}_2 \quad (7)$$

对式(7)左右两边分别乘以 \mathbf{U}^T 和 \mathbf{V} , 可将其化简为

$$\mathbf{V}^T \mathbf{M}\mathbf{U}\mathbf{A} + \mathbf{B}\mathbf{V}^T \mathbf{M}\mathbf{U} = \mathbf{V}^T \mathbf{Q}_2 \mathbf{U} \quad (8)$$

设 $\mathbf{V}^T \mathbf{M}\mathbf{U} = \vec{\mathbf{M}}$, 式(8)化简为

$$\mathbf{A}\vec{\mathbf{M}}\mathbf{B} + \vec{\mathbf{M}} = \mathbf{V}^T \mathbf{Q}_2 \mathbf{U}$$

在该等式中, 任意片段 $\vec{\mathbf{M}}_{i,j}$ 有 $\mathbf{A}_{ii} \vec{\mathbf{M}}_{i,j} \mathbf{B}_{jj} +$

$$\vec{\mathbf{M}}_{i,j} = (\mathbf{V}^T \mathbf{Q}_2 \mathbf{U})_{i,j} \text{ 成立, 即 } \vec{\mathbf{M}}_{i,j} = \frac{(\mathbf{V}^T \mathbf{Q}_2 \mathbf{U})_{i,j}}{\mathbf{A}_{ii} \mathbf{B}_{jj} + 1}. \text{ 因此,}$$

其解为 $\mathbf{M} = \mathbf{U}\vec{\mathbf{M}}\mathbf{V}^T$.

4.2 更新 \mathbf{Z}

固定式(4)问题中变量 \mathbf{M} 和 $\boldsymbol{\gamma}$, 该问题变为求解变量 \mathbf{Z} 的最小化问题, 即

$$\min_{\mathbf{Z}} \frac{\beta}{2} \|\mathbf{M} - \mathbf{Z}\|_2^2 + \text{tr}(\boldsymbol{\gamma}^T (\mathbf{M}^T - \mathbf{Z})) + \lambda_3 \|\mathbf{Z}\|_* \quad (9)$$

式(9)等价于

$$\min_{\mathbf{Z}} \frac{1}{2} \|\mathbf{M} + \frac{\boldsymbol{\gamma}}{\beta} - \mathbf{Z}\|_2^2 + \frac{\lambda_3}{\beta} \|\mathbf{Z}\|_* \quad (10)$$

式(10)是最小化核范数问题, 可用文献[24]算法来高效优化, 其解为

$$\mathbf{Z} = D_{\frac{\lambda_3}{\beta}} \left(\mathbf{M} + \frac{\boldsymbol{\gamma}}{\beta} \right)$$

其中, $D_{\frac{\lambda_3}{\beta}}$ 是删除输入矩阵小于 $\frac{\lambda_3}{\beta}$ 的奇异值以及对应奇异值向量后形成的矩阵。

4.3 更新 $\boldsymbol{\gamma}$

固定式(4)问题中变量 \mathbf{M} 和 \mathbf{Z} , 变量 $\boldsymbol{\gamma}$ 可通过如下形式来更新

$$\boldsymbol{\gamma} = \boldsymbol{\gamma} + \beta(\mathbf{M} - \mathbf{Z})$$

4.4 复杂度分析

从式(6)可知, 更新变量 \mathbf{M} 需要计算变量 \mathbf{Q}_1 和 \mathbf{Q}_2 以及 $2\mathbf{X}\mathbf{X}^T$ 和 \mathbf{Q}_1 的对称分解, $2\mathbf{X}\mathbf{X}^T$ 和 \mathbf{Q}_1 为常数矩阵, 其计算和对称分解由预处理步骤处理. 在计算 \mathbf{Q}_2 过程中, 流型正则部分的导数为 $\sum_{i=1}^{l-1} \sum_{j=1}^{l-1} 2\text{gs}(i, j) \mathbf{M}_i \mathbf{X}^{(i,j)} (\mathbf{X}^{(i,j)})^T$, 其时间复杂度为 $O(ld^2)$. 本节假设每个数据以均等概率标注标签和不标注标签, 则排序正则部分的时间复杂度为 $O(l^2 n^2 d)$. 因此, 更新变量 \mathbf{M} 的时间复杂度为 $O(k(ld^2 + l^2 n^2 d + ld))$, 其中 k 是迭代次数. 更新变量 \mathbf{Z} 的时间复杂度为 $O(\min\{ld^2, l^2 d\})$. 更新变量 $\boldsymbol{\gamma}$ 的时间复杂度为 $O(ld)$.

本文方案 (简称为 SLCR) 单次循环时间复杂度为 $O(l^2 n^2 d + k(ld^2 + l^2 n^2 d + ld) + \min\{ld^2, l^2 d\} + ld n_k)$, 其中, n_k 表示近似算法的迭代次数. BR 的时间复杂度为 $O(ld^3 + ldn^2 + d^2 n^2 + d^3)$. 文献[7]方案 (简

称为 WSABIE) 采用批量梯度下降法优化目标函数, 其时间复杂度为 $O(n_s(ld + l^2))$, 其中, n_s 表示单次选取的梯度下降数据数。文献[6]方案(简称为 LEML) 使用交替下降法优化目标函数, 其单次循环时间复杂度为 $O(nnz(\mathbf{Y}) + nnz(\mathbf{X})k + (n + l)k^2)$, 其中, $nnz(\mathbf{Y}) \ll nL$, k 表示拟合矩阵的秩。文献[14]方案(简称为 SLRM) 采用交替下降法优化目标函数, 其单次循环时间复杂度为 $O(d^2n_l + dn^2 + ldn_l + \min\{dl^2, d^2l\} + dl)$, 其中, n_l 表示训练集中标记标签数据个数。文献[25]方案(简称为 ICVL) 采用交替下降法优化目标函数, 其单次循环时间复杂度为 $O(l^2n + d^2n + d^3 + l^3 + ldn)$, 其中, n 表示实例个数, l 表示标签个数, d 表示特征维数。如上所述, 本文方案时间复杂度高于对比方案时间复杂度。

5 实验

本节在 6 个经典的多标签数据集上比较本文方案和代表性标签缺失下的多标签学习方案的预测效果。

5.1 数据集

本文实验的数据集为 Emotions、Scene、Birds、Mediamill、Delicious、NUS-WIDE-B。为有效检测标签缺失下模型的预测效果, 本文剔除了未标记标签的数据, 其详细统计信息如表 1 所示。

5.2 对比方案

为证明本文方案的预测效果, 本节实验与如下代表性的标签缺失下的多标签学习方案进行比较: BR、WSABIE、LEML、SLRM、ICVL。本节使用 Hamming Loss、Recall、F1-Measure、F1 macro、F1 micro 等 5 个标准来评估各方案的预测效果。详细评估标准信息如文献[26]所示。

5.3 实验结果

本文方案与对比方案在 6 个数据集上的预测

效果如表 2~表 6 所示。在 5 个实验评估标准中, Hamming Loss 取值越小其方案效果越好, 其余评估标准取值越大方案效果越好。为准确比较方案的实验效果, 本文从大量实验结果中挑选了具有相似 Hamming Loss 值的实验结果。从表 2~表 6 可以看出, 本文方案获得了较好的预测效果。例如, 在数据集 NUS-WIDE-B 上, 本文方案在 Recall、F1-Measure、F1 macro、F1 micro 评估标准上比最好对比方案分别高 0.073、0.075、0.057 和 0.042。该结果证明了本文方案在处理缺失标签情况下更具稳定性。详细分析如下。

1) LEML、SLRM 和 ICVL 直接或间接考虑了低秩结构来提升预测精度, WSABIE 使用标签排序来约束目标函数。然而, 本文方案比 LEML、SLRM 和 ICVL 有更好的实验效果, 其原因是低秩结构俘获了标签整体结构上的关联, 标签排序和流型正则俘获了个体标签间的关联, 2 种标签关联相互补充, 从而获得了较好的预测效果。

2) 在 Recall、F1-Measure、F1 macro 和 F1 micro 等评估标准下, BR 实验效果与 WSABIE、LEML 和 SLRM 的实验效果接近。该结果证明了低秩结构在某些数据中并不总是有效。同时, 本文方案在 Scene 和 Delicious 上取得了总体优于对比方案的实验效果, 间接证明了流型正则化和排序正则化俘获了更有效的标签关联。

不同标签缺失率下本文方案与对比方案的预测效果如图 3~图 7 所示。该实验在 5 个数据集 Emotions、Scene、Birds、Mediamill 和 NUS-WIDE-B 上进行。从图 3~图 7 可以看出, 尽管本文方案的实验效果随着标签缺失率下降而下降, 但是本文方案相比于对比方案仍获得较好的实验效果。其原因是本文流型结构从标签正实例出发来度量标签关联, 避免了使用标签向量度量标签关联带来的偏差问题。同时, 排序

表 1 实验数据统计信息

数据集	实例个数	特征维数	标签个数	数据类型	实例平均标签个数	领域
Emotions	593	72	6	numeric	1.869	music
Scene	2407	294	6	numeric	1.074	image
Birds	645	258	19	numeric	1.014	audio
Mediamill	43 907	120	101	numeric	4.376	video
Delicious	16 105	500	983	nominal	9.020	text
NUS-WIDE-B	269 648	128	81	numeric	1.869	images

表 2 6 个数据集上的 Hamming Loss 比较

数据集	BR	WSABIE	LEML	SLRM	ICVL	SLCR
Emotions	0.2186	0.2648	0.2054	0.2096	0.2145	0.1125
Scene	0.1169	0.1322	0.1264	0.1158	0.1319	0.2079
Birds	0.0859	0.2858	0.1055	0.0875	0.0994	0.1257
Mediamill	0.0314	0.0331	0.0314	0.0315	0.0314	0.0311
Delicious	0.0181	0.0331	0.0181	0.0203	0.0667	0.0345
NUS-WIDE-B	0.0262	0.0250	0.0251	0.0250	0.0311	0.0260

表 3 6 个数据集上的 Recall 比较

数据集	BR	WSABIE	LEML	SLRM	ICVL	SLCR
Emotions	0.5338	0.5859	0.5850	0.5503	0.5486	0.6411
Scene	0.5247	0.5447	0.5438	0.5113	0.5192	0.5665
Birds	0.3046	0.5429	0.3511	0.3206	0.3928	0.5248
Mediamill	0.4761	0.5142	0.5084	0.5024	0.4621	0.5633
Delicious	0.1193	0.1778	0.1097	0.1210	0.2571	0.2922
NUS-WIDE-B	0.1875	0.2350	0.2210	0.2228	0.1677	0.3078

表 4 6 个数据集上的 F1-Measure 比较

数据集	BR	WSABIE	LEML	SLRM	ICVL	SLCR
Emotions	0.5731	0.6064	0.6206	0.5951	0.5607	0.6587
Scene	0.5258	0.5311	0.5366	0.5131	0.4969	0.5629
Birds	0.3189	0.2595	0.3357	0.3353	0.3744	0.4430
Mediamill	0.5929	0.6014	0.6102	0.6053	0.5456	0.6390
Delicious	0.1860	0.2436	0.1735	0.1863	0.2484	0.2942
NUS-WIDE-B	0.2328	0.2807	0.2657	0.2668	0.1830	0.3556

表 5 6 个数据集上的 F1 macro 比较

数据集	BR	WSABIE	LEML	SLRM	ICVL	SLCR
Emotions	0.6153	0.5132	0.6527	0.6329	0.6275	0.6702
Scene	0.6139	0.5953	0.6051	0.6110	0.5826	0.6409
Birds	0.4158	0.2793	0.3915	0.4257	0.4268	0.4483
Mediamill	0.5486	0.5533	0.5663	0.5642	0.5419	0.5907
Delicious	0.1911	0.2213	0.1790	0.1755	0.1262	0.2654
NUS-WIDE-B	0.3043	0.3480	0.3310	0.3370	0.2592	0.3895

表 6 6 个数据集上的 F1 micro 比较

数据集	BR	WSABIE	LEML	SLRM	ICVL	SLCR
Emotions	0.5313	0.5960	0.5865	0.5489	0.6072	0.6416
Scene	0.5135	0.5373	0.5350	0.5055	0.5867	0.5543
Birds	0.2625	0.1459	0.2502	0.2681	0.3138	0.3286
Mediamill	0.4194	0.4501	0.4504	0.4475	0.0443	0.4994
Delicious	0.1112	0.1694	0.1026	0.1123	0.0587	0.2856
NUS-WIDE-B	0.1923	0.2328	0.2169	0.2216	0.0377	0.2899

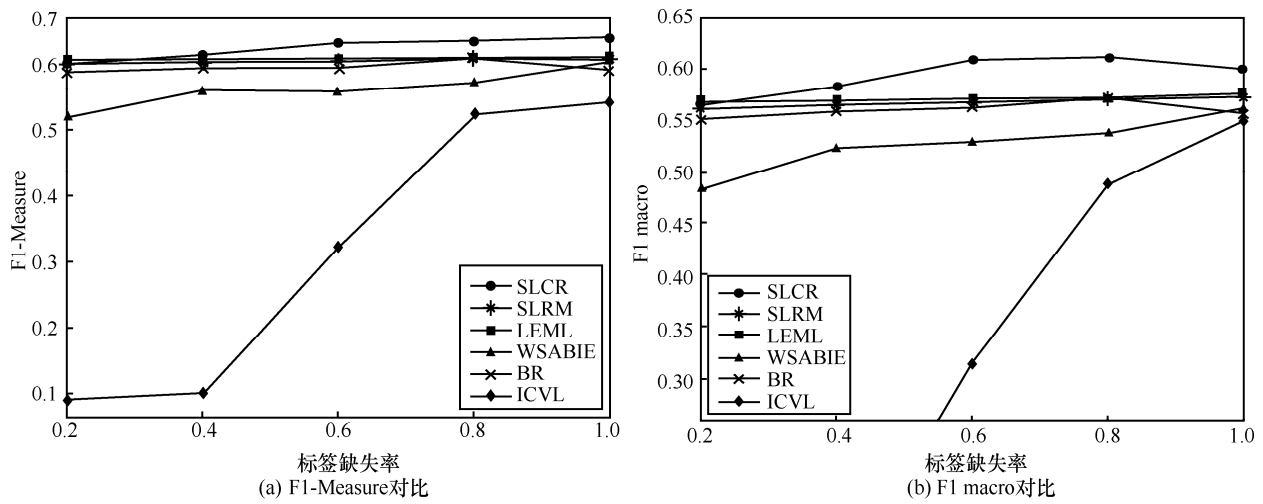


图3 Mediamill数据集上预测效果对比

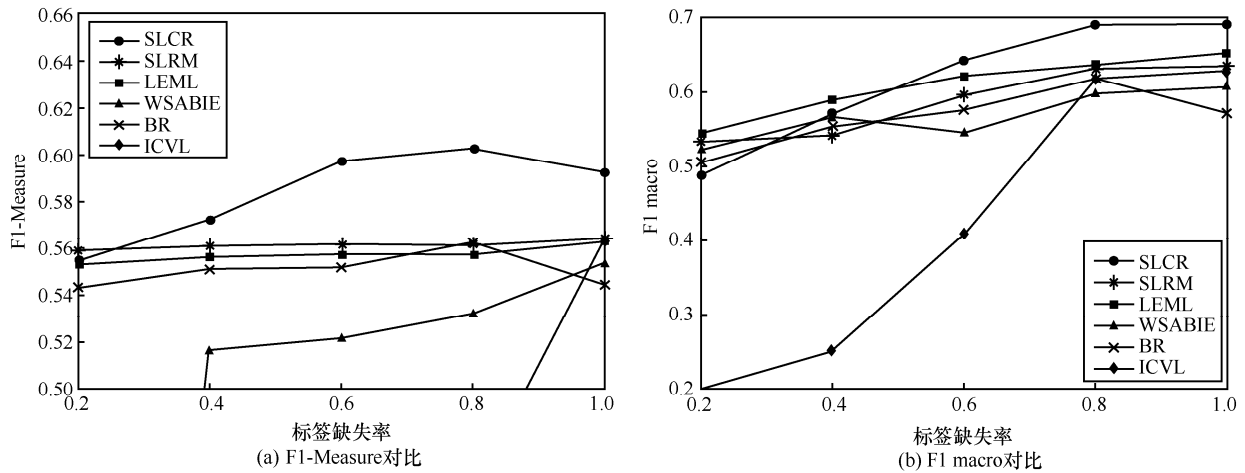


图4 Emotions数据集上预测效果对比

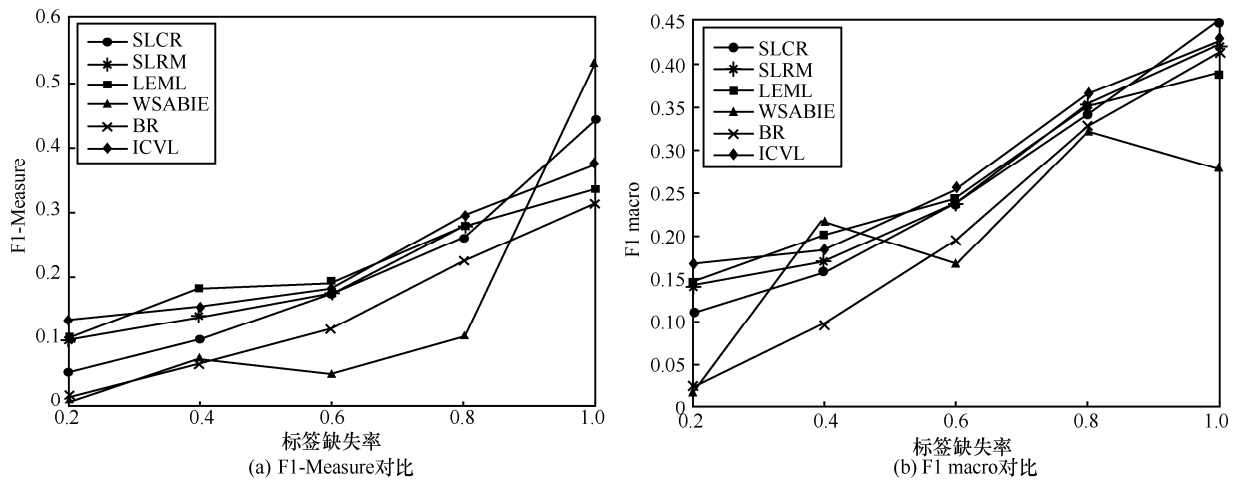


图5 Birds数据集上预测效果对比

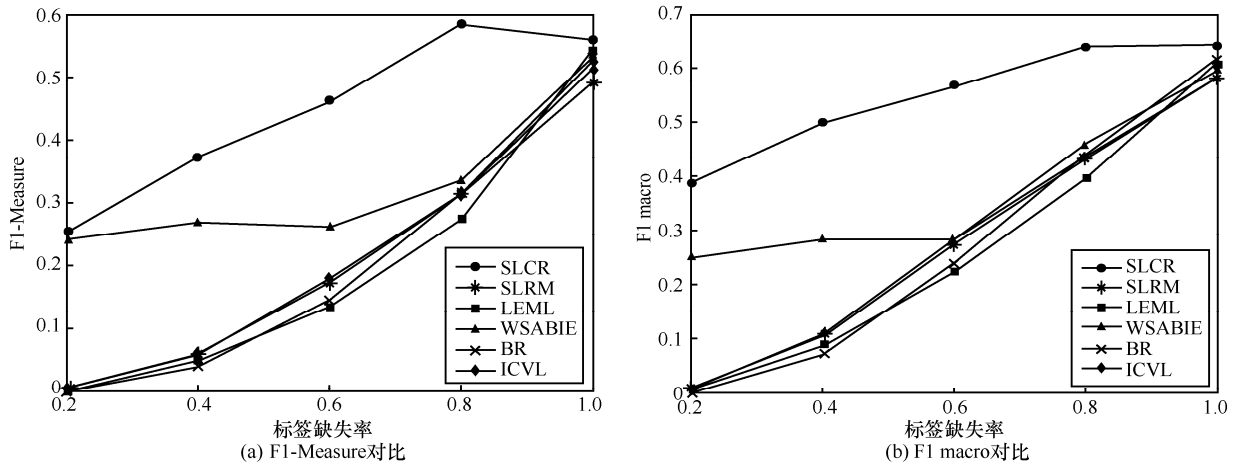


图 6 Scene 数据集上预测效果对比

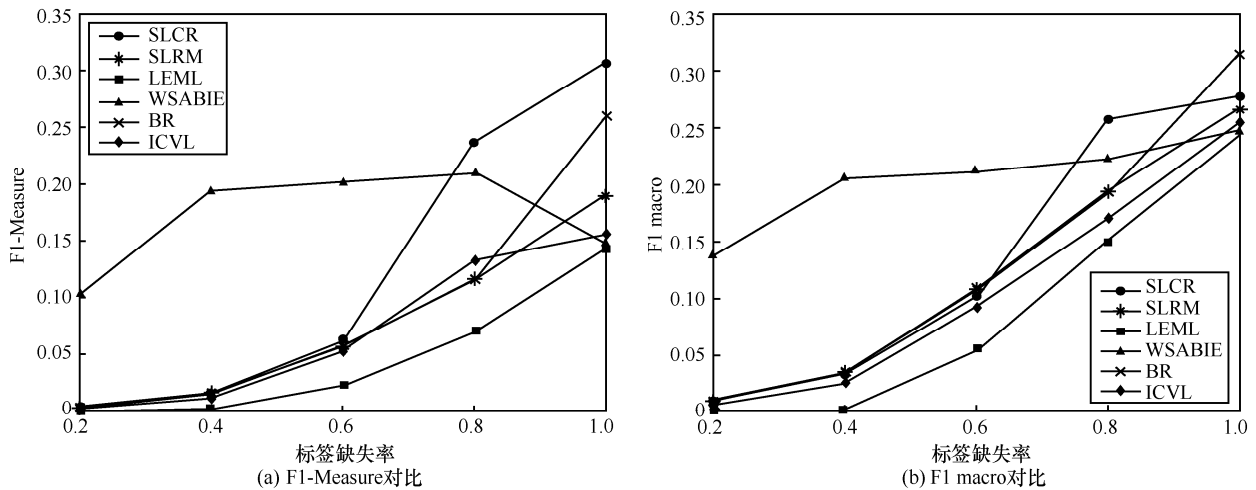


图 7 NUS-WIDE-B 数据集上预测效果对比

正则化在一定程度上缓解了标签错误标记带来的模型偏差。

6 结束语

本文针对标签缺失下的多标签标记问题，从数据流型角度出发俘获了标签关联，利用加权排序降低了模型偏差，并利用低秩结构正则化模型，从而有效填充了缺失标签。具体地，通过确保数据标签几何相似度与标签预测距离一致性俘获数据流型结构；通过度量完备标签下和不完备标签下的排序损失来对模型进行正则。实验结果表明，本文方案优于现有标签缺失下的多标签学习方案。

参考文献：

[1] LI Y M, XU Z L, ZHANG Z F. Learning with incomplete la-

bels[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2018: 3588-3595.

[2] WU B Y, LIU Z L, WANG S F, et al. Multi-label learning with missing labels[C]//Proceedings of 2014 22nd International Conference on Pattern Recognition. Piscataway: IEEE Press, 2014: 1964-1968.

[3] LI X, ZHAO F P, GUO Y H. Conditional restricted Boltzmann machines for multi-label learning with incomplete labels[C]//Proceedings of the International Conference on Artificial Intelligence and Statistics. Massachusetts: MIT Press, 2015: 635-643.

[4] WU B Y, LYU S W, GHANEM B. ML-MG: multi-label learning with missing labels using a mixed graph[C]//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2015: 4157-4165.

[5] ZHANG Y, SCHNEIDER J G. Multi-label output codes using canonical correlation analysis[C]//Proceedings of International Conference on Artificial Intelligence and Statistics. Berlin: Springer, 2011: 873-882.

[6] YU H F, JAIN P, KAR P, et al. Large-scale multi-label learning with missing labels[C]//Proceedings of the International Conference on Machine Learning. Saarland: DBLP, 2014: 593-601.

- [7] WESTON J, BENGIO S, USUNIER N. WSABIE: scaling up to large vocabulary image annotation[C]//Proceedings of the International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2011: 1-7.
- [8] GONG Y, JIA Y, LEUNG T, et al. Deep convolutional ranking for multi-label image annotation[J]. arXiv Preprint, arXiv: 1312.4894, 2013.
- [9] 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.
- [10] HUANG J, QIN F, ZHENG X, et al. Improving multi-label classification with missing labels by learning label-specific features[J]. Information Sciences, 2019, 492: 124-146.
- [11] TAI F, LIN H T. Multi-label classification with principle label space transformation[J]. Neural Computation, 2012, 24(9): 2508-2542.
- [12] ZHANG Y, SCHNEIDER J G. Maximum margin output coding[C]//Proceedings of the International Conference on Machine Learning. New York: ACM Press, 2012: 1-8.
- [13] JIN R, WANG S J, ZHOU Z H. Learning a distance metric from multi-instance multi-label data[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2009: 896-902.
- [14] JING L P, YANG L, YU J, et al. Semi-supervised low-rank mapping learning for multi-label classification[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2015: 1483-1491.
- [15] LI Y C, YANG Y L. Label embedding for multi-label classification via dependence maximization[J]. Neural Processing Letters, 2020, 52(2): 1651-1674.
- [16] GUPTA V, WADBUDE R, NATARAJAN N, et al. Distributional semantics meets multi-label learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 3747-3754.
- [17] LIU W W, SHEN X B. Sparse extreme multi-label learning with oracle property[C]//Proceedings of the International Conference on Machine Learning. New York: ACM Press, 2019: 4032-4041.
- [18] USUNIER N, BUFFONI D, GALLINARI P. Ranking with ordered weighted pairwise classification[C]// Proceedings of the 26th Annual International Conference on Machine Learning. New York: ACM Press, 2009: 1057-1064.
- [19] DURAND T, THOME N, CORD M. MANTRA: minimum maximum latent structural SVM for image classification and ranking[C]//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2015: 2713-2721.
- [20] RECHT B, FAZEL M, PARRILO P A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization[J]. SIAM Review, 2010, 52(3): 471-501.
- [21] BENGIO Y, GOODFELLOW I, COURVILLE A. Deep learning[M]. The MIT Press, 2016.
- [22] BELKIN M, NIYOGI P, SINDHWANI V. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples[J]. Journal of Machine Learning Research, 2006, 7(1):2399-2434.
- [23] WANG J, SHI D M, CHENG D S, et al. LRSR: low-rank-sparse representation for subspace clustering[J]. Neurocomputing, 2016, 214: 1026-1037.
- [24] CAI J F, CANDÈS E J, SHEN Z W. A singular value thresholding algorithm for matrix completion[J]. SIAM Journal on Optimization, 2010, 20(4): 1956-1982.
- [25] FAN R E, CHANG K W, HSIE C J, et al. LIBLINEAR: a library for large linear classification[J]. Journal of Machine Learning Research, 2008: 9:1871-1874.
- [26] 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.

[作者简介]



陈天柱（1987- ），男，河北秦皇岛人，博士，中国电子科技集团公司工程师，主要研究方向为自然语言处理。



李凤华（1966- ），男，湖北浠水人，博士，中国科学院信息工程研究所研究员、博士生导师，主要研究方向为网络与系统安全、信息保护、隐私计算。



郭云川（1977- ），男，四川营山人，博士，中国科学院信息工程研究所正高级工程师、博士生导师，主要研究方向为访问控制、网络安全。



李子孚（1992- ），女，内蒙古赤峰人，博士，中国科学院信息工程研究所工程师，主要研究方向为网络与系统安全、访问控制。