

基于多视角低秩表征的短视频多标签学习模型

吕卫¹, 李德盛^{1*}, 谭浪², 井佩光¹, 苏育挺¹

¹天津大学电气自动化与信息工程学院, 天津 300072;

²北京智芯微电子科技有限公司, 北京 102200

摘要 提出一种基于多视角低秩表征的短视频多标签分类模型。该模型将低秩表征和多标签学习结合到同一框架中, 利用不同类型特征的一致性学习本征稳定的低秩表示。同时为了获得标签相关性的潜在表示, 构建了标签相关性学习项来自适应地捕获标签的相关性矩阵。此外, 模型利用监督信息进一步提高了其表征能力。大量的实验结果证实了所提方法的优越性。

关键词 图像处理; 低秩表征; 多标签学习; 多视角学习; 短视频

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP57.221012

Microvideo Multilabel Learning Model Based on Multiview Low-Rank Representation

Lü Wei¹, Li Desheng^{1*}, Tan Lang², Jing Peiguang¹, Su Yuting¹

¹School of Electronics and Information Engineering, Tianjin University, Tianjin 300072, China;

²Beijing Smartchip Microelectronics Technology Co., Ltd., Beijing 102200, China

Abstract We propose a microvideo multilabel learning model based on a multiview low-rank representation, which combines the low-rank representation and multilabel learning into a unified framework and uses the consistency in different features to learn an intrinsically robust low-rank representation. Meanwhile, to represent the potential label correlations, our proposed model constructs a label correlation learning term to adaptively capture the labels' correlation matrix. Furthermore, the supervised information is exploited to further improve the representation ability of our model. Extensive experiments on a large-scale public dataset show the effectiveness of the proposed scheme.

Key words image processing; low-rank representation; multi-label learning; multi view learning; micro-video

OCIS codes 100.5010; 100.3008; 110.2960

1 引言

现如今, 短视频作为一种新型的媒体形式出现在大众面前, 因各大网络、社交平台允许用户录制、加工、上传自己的短视频, 故大量短视频原始数据会周期性出现, 且这些数据作为多媒体内容常常携带有多个标签, 如“美妆”和“女生自拍”经常出现在同一短视频中。在大量数据源背景下, 对短视频数据进行有效的多标签学习逐渐体现出其必要性。

多标签学习一直是机器学习领域的热门话题, 其基本任务就是将多个标签分配到每一个未知的

数据样本中^[1]。二元关联算法^[2]是最早解决多标签学习的方法之一, 但这种方法忽略了标签与标签之间的相关性。建模时如果没有利用标签之间的依赖性和独立性, 则无法实现最佳的预测性能。受此思想影响, 文献[3]提出了一种分类器链, 即将多标签问题转化为一串二元分类问题来对标签之间的相关性进行建模; 文献[4]提出了一种采用特征诱导的标签信息增强的多标签学习方法, 其将多标签样本的类别信息通过特征空间的结构化信息加以增强。然而以上方法由于未将特征表征和预测模型纳入同一框架下, 学习过程整体性欠缺, 容易

收稿日期: 2020-03-10; 修回日期: 2020-04-06; 录用日期: 2020-04-10

基金项目: 国家自然科学基金(61802277)

*E-mail: lidesheng1996@tju.edu.cn

因为特征表征和模型的不匹配而产生次优的学习结果。

同时,由于多媒体环境的特点,在处理短视频的相关问题时,通常会提取多个不同视角的特征来描述其内容^[5]。现有的研究成果针对多视角问题的处理往往是将各个视角串联起来,这样会造成视角之间影响的不确定性、串联后特征的冗余性以及特征维度过大等问题,因此在处理短视频多视角特征时需要考虑一种更有效的特征表征方法。

为解决上述问题,本文从短视频本身的特点出发,将多视角低秩表征和多标签学习整合到同一个框架下,利用多视角特征的一致性和低秩表征对于多媒体内容理解的优势训练得到本征稳定的低秩表示;同时通过矩阵正态分布原理构造标签相关性学习项,得到标签相关性矩阵,以此进一步适应多标签学习的需求。此外,模型通过低秩表示到标签空间的映射构造损失函数项,将低秩表示与模型本身加以匹配。本文模型保证了学习过程的整体性,提高了最终结果的准确性和对数据噪声的鲁棒性。

2 低秩表征

低秩表征由压缩感知理论^[6]衍生而来,是指将原有的特征矩阵变化为字典集和一个新的特征表示的乘积,同时要求该特征表示的秩尽量小,从而避免了特征的冗余性;由于其对于多媒体内容的理解能力和在数据噪声存在情况下的稳健能力,低秩表征被广泛应用于图像处理^[7-8]、机器学习^[9]、计算机视觉^[10-12]等诸多领域。

假设给定一组数据集 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbf{R}^{D \times N}$, 其中 \mathbf{R} 为实数集, D 为特征维数, N 为样本个数;低秩表示的中心思想就是找到一个合适的字典集 $\mathbf{U} \in \mathbf{R}^{D \times d}$, 使得原始数据矩阵 \mathbf{X} 通过该字典集可以映射到一个最低秩的线性组合解;假设用 $\mathbf{Z} \in \mathbf{R}^{d \times N}$ 来表示该最低秩表征的系数矩阵,则有 $\mathbf{X} = \mathbf{UZ}$, 其中 d 表示最低秩表征的维数。由于在实际求解过程中矩阵的秩不方便优化,常常用矩阵核范数代替矩阵的秩进行求解;另外,为了提高最终结果的鲁棒性,低秩模型常常引入稀疏噪声,由此得到低秩表示的基本模型为

$$\min_{\mathbf{Z}, \mathbf{E}} \|\mathbf{U}\|_* + \lambda_1 \|\mathbf{E}\|_1, \text{ s. t. } \mathbf{X} = \mathbf{UZ} + \mathbf{E}, \quad (1)$$

其中 $\mathbf{E} \in \mathbf{R}^{D \times N}$ 为稀疏噪声项; $\|\cdot\|_*$ 为矩阵的核范数,代表了矩阵奇异值的和; λ_1 为平衡参数; $\|\cdot\|_1$ 为矩阵的 l_1 范数,对稀疏噪声数据进行规范化。

3 基于多视角低秩表征的短视频多标签学习模型

在处理诸如短视频等多媒体内容的相关问题时,往往需要提取多种特征来描述同一内容。但在提取过程中无法保证不同特征维度之间的不相关性,例如一个关于“美妆”的短视频中不仅包含与人脸相关的属性特征,还可能包含人脸中的其他信息。由此就会产生大量的特征冗余的情况,造成模型复杂度增加和预测结果准确性降低。由于低秩表征本身存在消除特征冗余的特性,本文利用其内在规律,同时为了适应多标签学习的需求,提出了一种基于多视角低秩表征的短视频多标签学习模型。该模型使得多视角低秩表征和多标签学习在同一个框架下进行,包含多视角低秩表征、损失函数和标签相关性学习三个部分,进而保证了学习过程的整体性。本文模型如图 1 所示,图中实线箭头代表训练过程,虚线箭头代表测试过程。

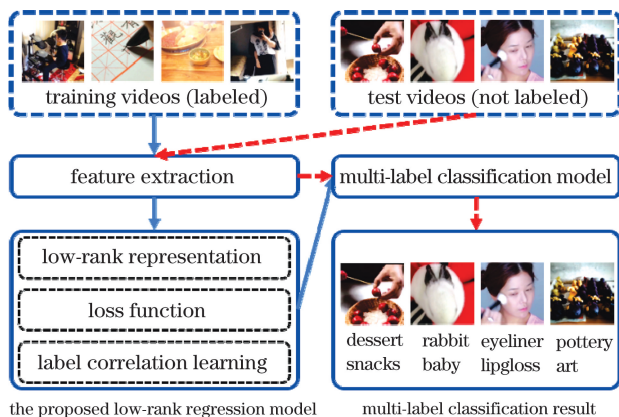


图 1 本文模型示意图

Fig. 1 Illustration of proposed model

3.1 算法模型

假设一组用多视角表示的短视频数据集: $\{\mathbf{X}_i \mid \mathbf{X}_i \in \mathbf{R}^{D_i \times N}\}_{i=1}^M$, 其中 M 为视角个数, D_i 为第 i 视角的特征维度, N 为短视频样本个数;同时假设用 \mathbf{Y} 来表示标签矩阵: $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]^T$, 其中 \mathbf{y}_p 是一个 $C \times 1$ 的矩阵,元素取值为 -1 或 1 ,用于表征第 p 个视频的标签分布情况, C 为标签总数;如果第 p 个短视频属于第 q 个标签,则 $y_{p,q} = 1$, 否则 $y_{p,q} = -1$ 。

1) 多视角低秩表征: 将其原本不同视角的特征矩阵 \mathbf{X}_i 通过各自的字典集 $\mathbf{U}_i \in \mathbf{R}^{D_i \times d}$ 映射到公共的低秩表示 $\mathbf{Z} \in \mathbf{R}^{d \times N}$ 上,同时要求 \mathbf{Z} 的秩最小,即

$$\min_{\mathbf{Z}, \mathbf{E}=\{\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_M\}} \sum_{i=1}^M \|\mathbf{U}_i\|_* + \lambda_1 \sum_{i=1}^M \|\mathbf{E}_i\|_1, \text{ s. t. } \mathbf{X}_i = \mathbf{U}_i \mathbf{Z} + \mathbf{E}_i, i=1, 2, \dots, M, \quad (2)$$

其中 d 为公共低秩表示的特征维度, $\mathbf{E}_i \in \mathbf{R}^{D_i \times N}$ 为第 i 个视角的稀疏噪声项, λ_1 为调节参数。由此整体构成多视角低秩表征项。由第 2 节引申可知,多视角低秩表征将多个视角的特征通过字典集转化为维度更低的公共低秩表示 \mathbf{Z} , 其通过最小化字典集的核范数, 达到消除冗余性的作用, 并对稀疏噪声进行规则化, 使得最终结果的抗干扰能力更强。

2) 损失函数: 为了建立低秩表示和标签空间之间的联系, 同时使实验结果和真实的标签矩阵更为接近, 模型利用已标注短视频数据的监督信息, 引入了损失函数项来确保最终结果的鲁棒性, 损失函数为

$$\min_{\mathbf{Z}, \mathbf{W}} \|\mathbf{Z}^T \mathbf{W} - \mathbf{Y}\|_F^2, \quad (3)$$

其中 $\|\cdot\|_F$ 为矩阵的 F 范数, $\mathbf{W} \in \mathbf{R}^{d \times C}$ 为权重矩阵, 同时也是低秩表示到标签空间的映射, 由此构成损失函数项。其将低秩表示 \mathbf{Z} 通过权重矩阵 \mathbf{W} 映射到标签空间中, 是建立从输入特征到输出结果的重要组成部分, 其通过最小化 $\mathbf{Z}^T \mathbf{W}$ 与 \mathbf{Y} 的差, 使特征表示更能贴合输出需求, 进而提高输出结果的鲁棒性。

3) 标签相关性学习: 为了研究标签之间的相关性, 使最终结果更符合多标签分类的要求, 模型引入了标签相关性矩阵 $\mathbf{S} \in \mathbf{R}^{C \times C}$ 来表示标签与标签之间的关系。在权重矩阵 \mathbf{W} 固定的前提下, \mathbf{S} 可由矩阵正态分布的原理进行优化, 权重矩阵 \mathbf{W} 的概率密

度函数为

$$p(\mathbf{W}|\mathbf{S}) = \frac{\exp\left[-\frac{1}{2}\text{tr}(\mathbf{W}^T \mathbf{W} \mathbf{S}^{-1})\right]}{(2\pi)^{C/2} |\mathbf{S}^{C/2}|}. \quad (4)$$

将(4)式取对数后转化为

$$\log(p(\mathbf{W}|\mathbf{S})) = -\frac{1}{2}\text{tr}(\mathbf{W}^T \mathbf{W} \mathbf{S}^{-1}) - \frac{dC}{2}\log(2\pi) - \frac{C}{2}\log|\mathbf{S}^{C/2}|, \quad (5)$$

其中 $\text{tr}(\cdot)$ 为矩阵的迹, $(\cdot)^{-1}$ 为矩阵的逆, $\log(\cdot)$ 为对数函数。与单标签学习问题不同, 多标签学习模型需要将单个特征映射到多个标签, 即达到一对多的输出目的。标签相关性学习通过最大化 \mathbf{W} 的概率密度函数, 使得其列向量对应的类别信息更加明显, 进而实现对于潜在标签相关性矩阵 \mathbf{S} 的求解, 并在迭代过程中利用 \mathbf{S} 使得 \mathbf{W} 能够更好地实现从低秩表示到多个标签的映射, 从而使模型更加符合多标签学习问题的需求。同时为了保证获取足够的标签相关性, 需要最小化其逆矩阵的 l_1 范数, 由此得到标签相关性学习项, 即

$$\min_{\mathbf{S}, \mathbf{W}} \lambda_3 \|\mathbf{S}^{-1}\|_1 + \lambda_4 \text{tr}(\mathbf{W}^T \mathbf{W} \mathbf{S}^{-1}) + \lambda_5 \log|\mathbf{S}|, \quad (6)$$

其中 $\lambda_3, \lambda_4, \lambda_5$ 为调节参数。

综合上述三个函数可得整体目标函数为

$$\min_{\mathbf{U}, \mathbf{W}} \sum_{i=1}^M \|\mathbf{U}_i\|_* + \lambda_1 \sum_{i=1}^M \|\mathbf{E}_i\|_1 + \lambda_2 \|\mathbf{Z}^T \mathbf{W} - \mathbf{Y}\|_F^2 + \lambda_3 \|\mathbf{S}^{-1}\|_1 + \lambda_4 \text{tr}(\mathbf{W}^T \mathbf{W} \mathbf{S}^{-1}) + \lambda_5 \log|\mathbf{S}|, \quad (7)$$

s. t. $\mathbf{X}_i = \mathbf{U}_i \mathbf{Z} + \mathbf{E}_i, i=1, 2, \dots, M.$

由此将多视角低秩表征和多标签学习整合到同一个框架下, 二者在求解迭代过程中互相约束, 从而保证了学习过程的整体性。

3.2 模型求解

根据模型本身的数学特点, 目标函数采用增广拉格朗日乘子法的线性交替方向法进行求解, 在目标函数中引入 $\{\mathbf{Q}_i | \mathbf{Q}_i \in \mathbf{R}^{D_i \times N}\}_{i=1}^M$ 作为拉格朗日乘子矩阵, 得到增广拉格朗日乘子式, 即

$$L(\mathbf{U}, \mathbf{E}, \mathbf{Z}, \mathbf{W}, \mathbf{S}) = \sum_{i=1}^M \|\mathbf{U}_i\|_* + \lambda_1 \sum_{i=1}^M \|\mathbf{E}_i\|_1 + \lambda_2 \|\mathbf{Z}^T \mathbf{W} - \mathbf{Y}\|_F^2 + \lambda_3 \|\mathbf{S}^{-1}\|_1 + \lambda_4 \text{tr}(\mathbf{W}^T \mathbf{W} \mathbf{S}^{-1}) + \lambda_5 \log|\mathbf{S}| + \sum_{i=1}^M \langle \mathbf{Q}_i, \mathbf{X}_i - \mathbf{U}_i \mathbf{Z} - \mathbf{E}_i \rangle + \frac{\mu}{2} \sum_{i=1}^M \|\mathbf{X}_i - \mathbf{U}_i \mathbf{Z} - \mathbf{E}_i\|_F^2, \quad (8)$$

其中, μ 为惩罚参数, $\langle \cdot \rangle$ 为矩阵内积算子。由此可以通过固定其他矩阵变量的迭代过程求解上述增广拉格朗日乘子式, 主要矩阵变量的求解式为

$$\mathbf{U}_i^{(g+1)} = \underset{\mathbf{U}_i}{\text{argmin}} \left\| \mathbf{U}_i \right\|_* + \frac{\mu}{2} \left\| \mathbf{X}_i - \mathbf{U}_i \mathbf{Z}^{(g)} - \mathbf{E}_i^{(g)} + \frac{\mathbf{Q}_i^{(g)}}{\mu} \right\|_F^2, \quad (9)$$

$$\mathbf{W}^{(g+1)} = \operatorname{argmin}_{\mathbf{W}} \lambda_2 \left\| (\mathbf{Z}^{(g+1)})^T \mathbf{W} - \mathbf{Y} \right\|_{\text{F}}^2 + \lambda_4 \operatorname{tr}(\mathbf{W}^T \mathbf{W} (\mathbf{S}^{(g)})^{-1}), \quad (10)$$

$$\mathbf{S}^{(g)} = \operatorname{argmin}_{\mathbf{S}} \lambda_3 \left\| \mathbf{S}^{-1} \right\|_1 + \lambda_4 \operatorname{tr}((\mathbf{W}^{(g)})^T \mathbf{W}^{(g)} \mathbf{S}^{-1}) + \lambda_5 \log |\mathbf{S}|, \quad (11)$$

其中 g 为迭代次数, $(\cdot)^{(g)}$ 为 g 次迭代时的变量。至此求得原始特征到低秩表示的字典集 \mathbf{U}_i 以及低秩表示到标签空间的映射 \mathbf{W} , 测试集的输出结果 $\hat{\mathbf{Y}}$ 可表示为

$$\hat{\mathbf{Y}} = \mathbf{X}_i^T \mathbf{U}_i (\mathbf{U}_i^T \mathbf{U}_i)^{-1} \mathbf{W}, \text{ s. t. } i = 1, 2, \dots, M.$$

4 实验和结果分析

4.1 实验数据及设置

本文使用的数据集为美拍 MLSV-2018 短视频数据集, 该数据集从官方后台数据库中筛选出 12 万个短视频用作训练样本并进行人工多标签标注, 主要标注原则为短视频中出现的主体、场景和动作三个部分, 每个短视频标注 1~3 个标签, 数据集样例如图 2 所示, 其中第 1 行的 4 个短视频样例都是根据视频主体和场景进行标注, 第 2 行的 4 个短视频样例部分标签是根据视频的动作来标注的; 鉴于数

据集标注的特点, 本文通过视觉几何组^[11] (VGG) 和轨迹合并的深度卷积描述符^[13] (TDD) 提取出短视频的视觉特征和轨迹特征两个视角的特征, 2 个特征的维度均为 2048 维。实验采用五倍交叉验证的方法, 将数据集随机分为 5 份, 其中 4 份作为训练集, 1 份作为测试集进行实验。根据经验设置参数为 $\lambda_1 = 0.01, \lambda_2 = 0.0001, \lambda_3 = 10^{-10}, \lambda_4 = 10^{-10}, \lambda_5 = 10^{-10}$, 求解过程中的惩罚参数和迭代次数分别为: 1 和 25。参考文献[1], 本研究使用的多标签学习评价指标有以下 5 种: 平均精度 (Average Precision), 汉明损失 (Hamming Loss), 排序损失 (Ranking Loss), 覆盖率 (Coverage) 和 1 - 错误率 (One-error)。其中平均精度越高代表预测效果越好, 其余 4 项越低代表预测效果越好。本文综合考虑以上 5 种评价指标, 从正样本及负样本全体角度出发, 综合评价模型的多标签学习性能。

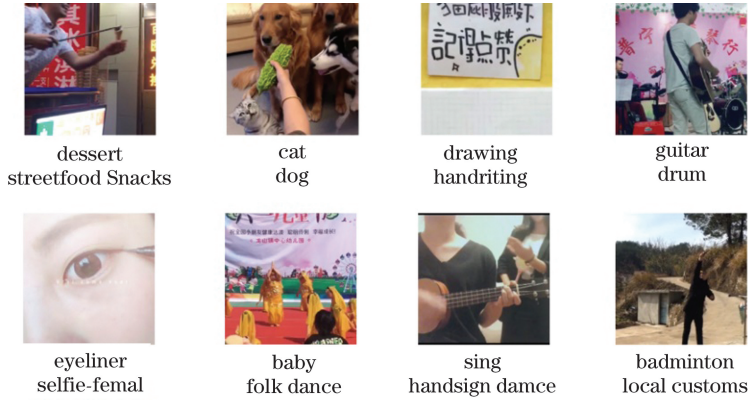


图 2 数据集中具有不同标签的视频示例

Fig. 2 Sample video with different labels selected from dataset

4.2 结果分析

4.2.1 算法收敛性

为证明本文算法的可行性和求解方式的合理性, 收敛性验证以低秩表示迭代前后的绝对差值为第一指标, 公式为: $Z_{\text{diff}} = \left\| \mathbf{Z}^{(g)} - \mathbf{Z}^{(g-1)} \right\|_1$, 并以迭代过程中平均精度的变化为第二指标。收敛性验证结果如图 3 所示, 从图中可以看出, 低秩表示的差值随着迭代次数快速递减, 平均精度的值快速上升。从数据上看, 大约 20 次左右达到收敛 (迭代前

后差值小于 10^{-4})。这证实了模型的可收敛性和算法的可行性。在后续实验中, 设置迭代次数为 25 次。

4.2.2 参数对比实验

为了研究模型最佳参数, 本研究在迭代过程中改变其中一个参数同时固定其他参数, 直到模型最终收敛并观察结果, 找到对应最佳参数后再将此参数固定改变其他参数, 从而找到全部最佳参数。

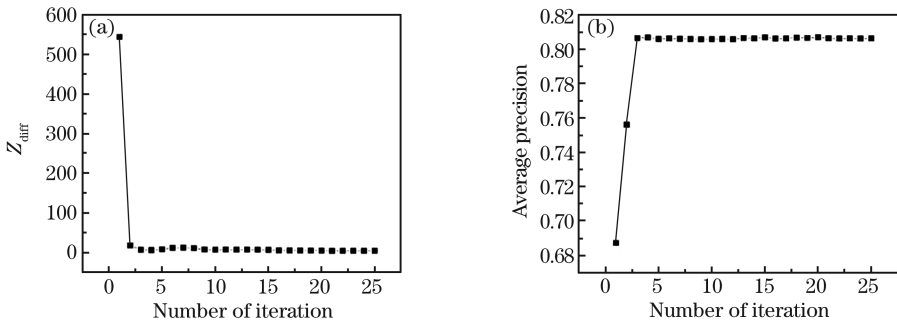


图3 收敛性验证图。(a) Z_{diff} 随模型迭代的变化;(b)平均精度随模型迭代的变化

Fig. 3 Convergence verification graphs. (a) Variation of Z_{diff} with model iteration; (b) variation of average precision with model iteration

根据实验结果,可知 λ_2 和 λ_4 在模型迭代过程中敏感性较高,其中 λ_2 为损失函数项参数, λ_4 为标签相关性学习项参数。图4为这两个参数对于实验结果的影响,实验中 λ_2 取值范围为 $10^{-1} \sim 10^{-7}$, λ_4 取值范围为 $10^{-7} \sim 10^{-13}$,从图中可以看出 λ_2 取 0.0001, λ_4 取 10^{-10} 时模型效果最优。实验结果证明:二者设定过高,相当于削弱了低秩表征的约束,导致最终结果具有次优性;而设定过低又会导致低秩表示映射到标签空间的能力被削弱。由此证明该设定参数可以平衡低秩表征与多标签学习的比重。

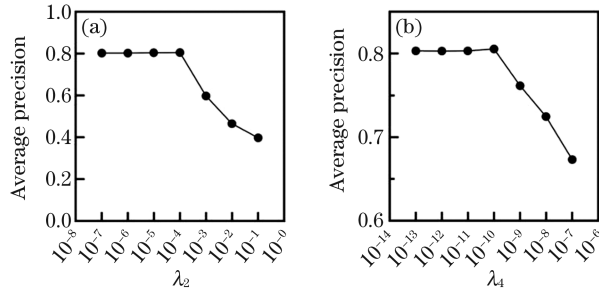


图4 不同参数对于平均精度的影响。(a) λ_2 对于平均精度的影响;(b) λ_4 对于平均精度的影响

Fig. 4 Effect of different parameters on average precision. (a) Effect of λ_2 on average precision; (b) effect of λ_4 on average precision

4.2.3 消融实验

为验证多视角低秩表征、损失函数和标签相关性学习的作用,在实验过程中控制其他变量不变,将三者相对应的参数分别置 0,即无多视角低秩表征(No LR)、无损失函数(No LF)和无标签相关性学习(No LC),并列 3 种不同情况下的结果与原整体模型的差距,可得到如表 1 所示的实验结果。

表 1 消融实验结果

Table 1 Ablation experiment results

Evaluation metrics	No LR	No LF	No LC
Average precision difference	-0.0508	-0.2423	-0.0172
Hamming loss difference	0.0013	0.0029	0.0005
Ranking loss difference	0.0079	0.0521	0.0049
Coverage difference	0.3492	4.6914	0.2851
One-error difference	0.0175	0.1482	0.0230

通过消融实验可以发现:多视角低秩表征与标签相关性学习二者通过消除特征冗余性和提取标签相关性的特性达到了优化最终实验结果的目的;损失函数项的缺失对于最终的结果有较大影响,导致低秩表征的结果不能更好地映射到其原有标签。由此证明将其融合进整体模型中能够使得低秩表征更大程度地贴合标签空间的分布,进而提高了最终结果的鲁棒性。

4.2.4 不同算法的对比实验

为证明本文所提模型的有效性和优越性,同时论证低秩表征和多标签学习整体性的重要性,本研究利用相同数据集数据和特征将本文方法与其他文献中的方法进行比较,对比结果如表 2 所示,表中包含离散非负矩阵分解^[14] (DNMF)、低秩表示^[15] (LRR)、具有全局和局部相关性的多标签学习^[16] (GLOCAL)、多标签 K 最近邻算法^[17] (MLKNN)、谷歌网络^[18] (Googlenet)、三维卷积网络^[19] (C3D) 和典型相关自动编码器^[20] (C2AE)。其中 DNMF、LRR 和 RMSL 包含了特征表征部分,GLOCAL 和 MLKNN 包含了多标签学习部分,Googlenet、C3D 和 C2AE 为深度网络模型。C2AE 的学习过程是将原始特征和标签转化为潜在的空间表示后,结合判

别器输出最终的学习结果。从对比结果可得出以下结论:1)相较于没有将特征表征融入整体模型的算法(GLOCAL, MLKNN),本文模型的学习结果更加优越,由此证明特征表征可以有效地处理短视频多标签学习问题。2)没有将多标签学习融入整体模型的算法(DNMF, LRR),大多没有将转化后的特征形式与预测模型匹配,其最终学习结果具有次优性。3)相较于对原始特征进行处理且与预测

模型处在同一框架的算法(C2AE),本文模型的整体效果更好,由此证明由低秩表征产生的特征表示,更具有鲁棒性,且去除了特征中不必要的冗余信息,故而实现了更好的学习结果。

综合来看,本文所提模型能够结合短视频本身的特点,充分利用低秩表征的优势,将多视角低秩表征与多标签学习整合在同一框架下,实现较好的学习结果。

表2 不同算法的性能对比

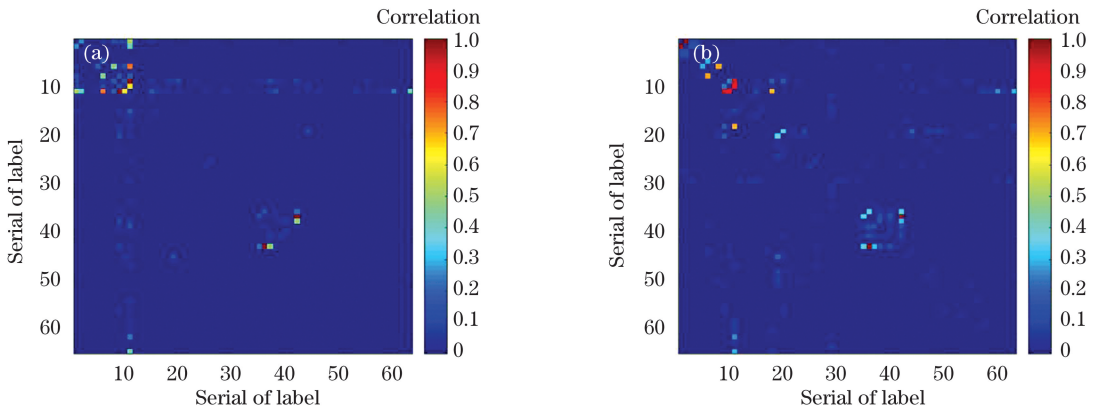
Table 2 Performance comparison of different algorithms

Method	Average precision	Hamming loss	Ranking loss	Coverage	One-error
DNMF	0.4673±0.0063	0.0154±0.0001	0.1077±0.0082	8.3853±0.1621	0.6487±0.0082
LRR	0.5489±0.0057	0.0154±0.0001	0.0991±0.0051	8.4056±0.1803	0.3039±0.0057
GLOCAL	0.7527±0.0064	0.0133±0.0020	0.0515±0.0015	3.9943±0.1056	0.2457±0.0032
MLKNN	0.7843±0.0053	0.0134±0.0001	0.0476±0.0058	4.0204±0.1874	0.3087±0.0058
Googlenet	0.6676±0.0044	0.0176±0.0002	0.4349±0.0066	4.5680±0.0600	0.4349±0.0066
C3D	0.7149±0.0089	0.0146±0.0003	0.3694±0.0028	3.9041±0.2033	0.3694±0.0088
C2AE	0.8013±0.0022	0.0128±0.0001	0.0481±0.0041	3.6942±0.1471	0.2381±0.0026
Proposed	0.8055±0.0028	0.0128±0.0001	0.0432±0.0023	3.6732±0.1274	0.2561±0.0065

4.2.5 标签相关性对比实验

通过对原始数据多标签数量的统计,可以得到数据集中的真实标签相关性矩阵,图5(a)为归一化真实标签相关性矩阵的热力图,可以观察到数据集标签相关性分布有两个主要特点:一是部分标签(如第11个标签)与其他标签普遍相关度较高;二是部分标签(如第36个标签)和特定的标签相关度较高,与其他标签相关度较低。同时由模型的迭代学

习可以得到标签相关性矩阵 \mathbf{S} ,令 \mathbf{S} 的最大值为 s_m ,由此可将 \mathbf{S} 归一化为 $\hat{\mathbf{S}}=\mathbf{S}/s_m-\mathbf{I}$,其中 \mathbf{I} 为单位矩阵,如图5(b)所示。将归一化的矩阵 $\hat{\mathbf{S}}$ 与真实标签相关性比较发现:模型学习到的标签相关性信息能够反映真实标签相关性信息的分布情况,且符合真实标签分布的特征。由此证明从低秩表示到标签的空间映射中学到的潜在标签相关性信息是有效的。

图5 标签相关性矩阵对比。(a)归一化真实标签相关性矩阵;(b)归一化相关性矩阵 $\hat{\mathbf{S}}$ Fig. 5 Label correlation matrix comparison. (a) Normalized correlation matrix for true label; (b) normalized correlation matrix $\hat{\mathbf{S}}$ after the iteration

5 结 论

针对目前短视频多标签学习中遇到的主要问题,提出了一种基于多视角低秩分解的短视频多标签学习模型,并将多视角低秩表征和多标签学习整合到了同一个框架下。该框架利用低秩表征挖掘特征维度之间的相关性,学习更稳健的低秩特征表示,结合损失函数和标签相关性学习增强了模型的整体学习能力。采用基于增广拉格朗日乘子法的线性交替方向法对整体模型进行了求解。在公开数据集上的大量实验表明,模型可以在多视角特征融合的情况下通过去除特征冗余性和学习标签相关性来提高多标签学习的最终效果。

参 考 文 献

- [1] 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.
- [2] Boutell M R, Luo J B, Shen X P, et al. Learning multi-label scene classification [J]. *Pattern Recognition*, 2004, 37(9): 1757-1771.
- [3] Read J, Pfahringer B, Holmes G, et al. Classifier chains for multi-label classification [J]. *Machine Learning*, 2011, 85(3): 333-359.
- [4] Zhang Q W, Zhong Y, Zhang M L. Feature-induced labeling information enrichment for multi-label learning [C] // 32th AAAI Conference on Artificial Intelligence, February 2-7, 2018, New Orleans, Louisiana, USA. Reston, VA: AIAA Press, 2018: 4446-4453.
- [5] Nie L Q, Wang X, Zhang J L, et al. Enhancing micro-video understanding by harnessing external sounds [C] // Proceedings of the 2017 ACM on Multimedia Conference-MM '17, October 19-27, 2017. Mountain View, California, USA. New York: ACM Press, 2017: 1192-1200.
- [6] Lian Q S, Xia C C. Compressed sensing of color images based on local Gaussian model in the dual-tree complex wavelet [J]. *Laser & Optoelectronics Progress*, 2011, 48(10): 101001.
练秋生, 夏长城. 基于双树复数小波局部高斯模型的彩色图像压缩感知[J]. *激光与光电子学进展*, 2011, 48(10): 101001.
- [7] Yang P, Liu D E, Li R X, et al. Damage detection of metal parts by combining information entropy and low-rank tensor analysis [J]. *Laser & Optoelectronics Progress*, 2019, 56(21): 211006.
杨鹏, 刘德儿, 李瑞雪, 等. 结合信息熵与低秩张量分析的金属零件破损检测[J]. *激光与光电子学进展*, 2019, 56(21): 211006.
- [8] Niu Q, Chen X H. Image recognition using joint projection learning algorithm based on latent low-rank representation [J]. *Laser & Optoelectronics Progress*, 2019, 56(14): 141006.
牛强, 陈秀宏. 基于隐式低秩表示的联合投影学习算法及图像识别[J]. *激光与光电子学进展*, 2019, 56(14): 141006.
- [9] Zhang J, Fu J P, Li X H. Low-rank regularized heterogeneous tensor decomposition algorithm for subspace clustering [J]. *Laser & Optoelectronics Progress*, 2018, 55(7): 071003.
张静, 付建鹏, 李新慧. 基于低秩正则化异构张量分解的子空间聚类算法[J]. *激光与光电子学进展*, 2018, 55(7): 071003.
- [10] Yang P, Liu D E, Li R X, et al. Damage detection of metal parts by combining information entropy and low-rank tensor analysis [J]. *Laser & Optoelectronics Progress*, 2019, 56(21): 211006.
杨鹏, 刘德儿, 李瑞雪, 等. 结合信息熵与低秩张量分析的金属零件破损检测[J]. *激光与光电子学进展*, 2019, 56(21): 211006.
- [11] Zhang X H, Hao R F, Li T Y. Hyperspectral abnormal target detection based on low rank and sparse matrix decomposition-sparse representation [J]. *Laser & Optoelectronics Progress*, 2019, 56(4): 042801.
张晓慧, 郝润芳, 李廷鱼. 基于低秩稀疏矩阵分解和稀疏字典表达的高光谱异常目标检测[J]. *激光与光电子学进展*, 2019, 56(4): 042801.
- [12] Hassannejad H, Matrella G, Ciampolini P, et al. Food image recognition using very deep convolutional networks [C] // Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management-MADiMa '16, October 16, 2016, Amsterdam, The Netherlands. New York: ACM Press, 2016: 41-49.
- [13] Wang L M, Qiao Y, Tang X O. Action recognition with trajectory-pooled deep-convolutional descriptors [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 4305-4314.
- [14] Jia Z L, Zhang X, Guan N Y, et al. Gene ranking of RNA-seq data via discriminant non-negative matrix

- factorization [J]. PLoS One, 2015, 10 (9): e0137782.
- [15] Liu G C, Yan S C. Latent Low-Rank Representation for subspace segmentation and feature extraction[C]//2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE Press, 2011: 1615-1622.
- [16] 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.
- [17] Zhang M L, Zhou Z H. ML-KNN: a lazy learning approach to multi-label learning [J]. Pattern Recognition, 2007, 40(7): 2038-2048.
- [18] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 1-9.
- [19] Tran D, Bourdev L, Fergus R, et al. Learning spatiotemporal features with 3D convolutional networks[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 4489-4497.
- [20] Yeh C K, Wu W C, Ko W J, et al. Learning deep latent spaces for multi-label classification [C]. 31th AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California. Reston, VA: AIAA Press, 2017: 2838-2844.