基于自适应图正则化与联合低秩矩阵分解的数字文化遗产多标签众包答案聚合方法 *

王春雪 [1, 2†]，徐琳琳 [3]，俞天秀 [1, 2]

(1. 甘肃省敦煌文物保护研究中心，甘肃 敦煌 736200; 2. 敦煌研究院 文物数字化研究所，甘肃 敦煌 736200; 3. 内蒙古财经大学 计算机信息管理学院，呼和浩特 010070)

摘 要：多标签答案聚合问题是通过融合众包收集的大量非专家标注来估计样本的真实标签。由于数字文化遗产数据具有标注成本高、样本类别多、分布不均衡等特点，给数据集多标签答案聚合问题带来极大挑战。以往的方法主要集中在单标签任务，忽视了多标签任务的标签关联性；大部分多标签聚合方法虽然一定程度上考虑了标签相关性，但是很敏感地受噪声和离群值的影响。为解决这些问题，提出一种基于自适应图正则化与联合低秩矩阵分解的多标签答案聚合方法 AGR-JMF。首先，将标注矩阵分解成纯净标注和噪声标注两部分；然后，对纯净标注采用自适应图正则化方法构建标签间的关联矩阵；最后，利用标注质量、标签关联性、标注人员行为属性相似性等信息指导低秩矩阵分解，以实现多标签答案的聚合。真实数据集和莫高窟壁画数据集上的实验表明，AGR-JMF 相较于现有算法在聚合准确率、识别欺诈者等方面具有明显优势。

关键词：多标签众包答案聚合；纯净标注数据；自适应图正则化；低秩矩阵分解

中图分类号：TP399 doi: 10.19734/j.issn.1001-3695.2022.09.0442

Multi-label crowd answer aggregation of digital cultural heritage based on adaptive graph regularization and joint low-rank matrix factorization

Wang Chunxue[1, 2†], Xu Linlin[3], Yu Tianxiu[1, 2]

(1. *Gansu Provincial Research Center for Conservation of Dunhuang Cultural Heritage*, *Dunhuang Gansu* 736200, *China*; 2. *Cultural Heritage Digitization Institute*, *Dunhuang Academy*, *Dunhuang Gansu* 736200, *China*; 3. *School of Computer Information Management*, *Inner Mongolia University of Finance & Economics*, *Hohhot* 010070, *China*)

Abstract: Multi-label answer aggregation problem aims to estimate the ground truth labels of samples by aggregating a large number of non-expert annotations collected by crowdsourcing. Due to the high annotation cost, multiple sample categories and uneven distribution of digital cultural heritage data, it brings great challenges to multi-label answer aggregation of datasets. Previous methods mainly focused on single-label problems, ignoring the label relevance of multi-label tasks; To some extent, most multi-label aggregation methods consider label correlations but are sensitive to noises and outliers. To solve these problems, this paper proposed a multi-label answer aggregation method based on adaptive graph regularization and joint low rank matrix factorization AGR-JMF. Firstly, it divided the input annotation matrix into two parts: pure annotations and noise annotations; Then, it constructed the association matrix between labels by adaptive graph regularization method for pure annotations; Finally, to realize the multi-label answer aggregations, it used labeling quality, label relevance, and the behavior attributes similarity between annotators to guide the low rank matrix factorization. Experiments on real-world datasets and MGF dataset show that AGR-JMF has obvious advantages over existing algorithms in terms of aggregating accuracy and identifying unreliable annotators.

Key words: multi-label crowd answer aggregation; pure annotations; adaptive graph regularization; low rank matrix factorization

0 引言

自上世纪末，国内外已开展以“数字敦煌”、“数字故宫”、“美国记忆”为代表的文化遗产数字化建设。经过数十年的发展，我国已积累了大规模多种类的珍贵数字文化遗产资源，极大地推动了文物保护、管理、研究和传承。在自然图像、三维模型的分类、分割、识别等方面，近年来以深度学习为代表的人工智能技术取得了长足的进步。但是，数字文化遗产数据标注成本高、样本类别多、分布不均衡等特点极大地制约了智能算法的应用。自 2020 年以来，在国家重点研发计划项目的支持下，敦煌研究院构建了面向敦煌壁画元素分割、分类和识别的数据集，涉及壁画元素类型达 78 种，样本实例 1.6 万余张。由于敦煌石窟壁画内容元素众多、分布广且具有不同程度的病害，高质量的数据集需要专家花费

收稿日期：2022-09-18；**修回日期：**2022-10-27 **基金项目：**甘肃省敦煌文物保护研究中心开放课题(GDW2021YB05)；陇原青年创新创业人才(个人)项目(2022LQGR40)；国家重点研发计划资助项目(2020YFC1522701，2020YFC1522705)

作者简介：王春雪(1988-)，女(通信作者)，山东德州人，副研究馆员，博士，主要研究方向为计算机图形学与图像处理、数值优化等(chunxuewang2019@163.com)；徐琳琳(1988-)，女，山东济宁人，讲师，博士，主要研究方向为稀疏优化、几何建模与处理等；俞天秀(1981-)，男，甘肃武威人，研究馆员，文物数字化研究所所长，在读博士，主要研究方向为人工智能、文化遗产数字化.

大量的时间和精力完成，而专家资源是昂贵且有限的。为了降低标注成本，通过借鉴众包标注思想[1,2]，本文聘用高校的学生经过专业培训后标注。同时，考虑到标注任务的成本要远远高于检查任务的成本，且非专家的标注可能犯错，标注任务只由一个标注人员完成，审核任务可由具有五年以上壁画图像拼接经验的若干名专业技术人员给出诸如“正确”“标签名称错误”“标签类型错误”等多标签答案。因此，如何获得高质量的审核结果是一个典型的多标签答案聚合问题[3,4]，直接决定着数字文化遗产智能检索、智能分析与理解等技术的应用效果。

由于众包标注往往存在标注空间巨大、标签稀疏且含有不同程度的噪声，高质量的多标签答案聚合面临较大的挑战。以往的答案聚合相关工作主要集中在单标签问题上[5~9]，通过将多标签任务转换成多个单标签任务求解，但是忽略了标签以及标注人员标注行为的相关性。为了克服单标签任务的不足，文献[10]考虑通过众包方式收集样本标签及标签间关系，估计多个标签间层次结构关系；文献[11,12]分别考虑从标注中估计标签共同出现的概率及标签间条件相关性来恢复样本真实标签。这两种方法仅仅考虑了局部标签相关性，很容易受标注质量和数量的影响。Tu 等[13]从标注整体存在低秩结构关系入手，对不同标注者的样本-标签关联矩阵进行矩阵分解，同时考虑标签的关联性以及不同标注者的标注相似性来推断真值标签；类似的想法，李等[14]则采用低秩张量矫正模型和标注融合策略两步优化估计样本的真实标签。以上方法均直接对不同标注者的样本-标签关联矩阵进行建模，很容易受到噪声和离群值的影响而产生较大的误差。

基于对上述研究工作的观察和总结分析，本文提出了一种鲁棒的多标签答案聚合方法 AGR-JMF。通过自适应去除原始标签数据中的噪声，同时基于该去噪数据在标签关联性、不同标注者的标注质量及行为属性相似性的指导下进行低秩矩阵的分解优化。本文工作的主要创新点如下：

a) 针对低秩矩阵分解易受到噪声干扰的问题，本文提出一个联合的多标签答案聚合框架，考虑标注人员的标注质量、标签关联性、标注人员行为属性相似性等因素，将去噪、低秩矩阵分解、自适应图正则化等集成到统一的目标函数中进行优化；

b) 针对低秩矩阵分解高度依赖标签关联性的问题，采用自适应图正则化方法获取不同标签之间的关联矩阵；

c) 针对噪声具有随机性和稀疏性等特点，采用 L_1 正则项优化去除标注数据中的噪声；

d) 本文分别在 6 个真实数据集和敦煌壁画数据集上进行了实验，并与当前具有代表性的方法进行了比较。实验结果证明了 AGR-JMF 不仅具有较高的准确率，同时较为鲁棒地识别低质量的标注者甚至欺诈者。

1　相关工作

受标注人员的知识、背景以及图像内容复杂性、质量等影响，不同标注人员反馈的标注结果可能差异较大。为了在众包中快速获得高质量的聚合结果，诸多方法被研究者们相继提出。本文简要介绍众包答案聚合的相关工作，主要包括单标签众包答案聚合方法和多标签众包答案聚合方法。

1.1　单标签答案聚合方法

作为最简单且最有效的众包答案聚合方法，多数投票法(Majority Voting, MV)[15]将所有标注者中大多数的标注作为真实标签的估计。MV 方法一般基于以下两种假设：1)在单标签任务中，标注者的整体准确率大于 50%；2)每个标注者的误差均匀分布在所有标签上。然而，这些基本假设并不适用于复杂的实际应用，尤其在专业性极强的文化遗产领域。此外，由于 MV 并没有考虑标注人员的表现，当存在大量恶意标注者时，MV 的效果会受到很大的误导和干扰。

除了 MV 方法之外，研究者们提出了通过建立众包过程的概率模型，并使用基于期望最大化(Expectation Maximization, EM)或其他推理算法来聚合答案[16]。例如，Dawid 和 Skene(DS)[17]利用 EM 建模每个标注者的混淆矩阵，并迭代估计最有可能是真值的标签。Raykar 等[5]假设每个标注人员的表现独立于特定任务，并使用 two-coin 模型衡量每个标注人员对未知真值的敏感性和特异性，然后利用 EM 算法迭代估计敏感性和特异性。Whitehill 等[18]在条件独立假设下，对标注质量和标注难度建立概率模型(Generative model of Labels Abilities and Difficulties, GLAD)，并应用 EM 算法推导出每个样本最可能的标签。Welinder 等[6]引入标注能力偏差，并进一步将 GLAD 中的概率模型推广为关于任务难度、标注质量和标注能力偏差的高维变量。以上四种方法在标注稀疏的情况下常常出现聚合答案不准确问题。为解决该问题，Demartini 等[19]只通过一个参数建模标注人员的可靠性，以避免在稀疏数据集上变量估计偏差大的问题。此后，研究者们通过考虑更多附加特性(如标签的偏见、置信度、意图等)提出了更复杂的模型和推理算法[8,20]。Liu 等[21]将众包问题转换为图模型中的变分推理问题，并利用包括置信传播和均值场在内的变分推理工具对标签进行推理。但是，他们方法的性能在很大程度上依赖于标注人员可靠性先验知识的选择。与此同时，基于最小最大熵的概率模型也陆续被应用。Zhou 等[22]假设标签由标注人员和任务的概率分布生成，该概率分布的熵最大化可导致任务难度和标注质量的提高，而概率分布的熵最小化可推断出真值标签，但该方法往往需要每个标注人员提供大量标签来构建混淆编码矩阵。Ma 等[23]提出联合建模生成任务内容和标注者答案的概率模型(FaitCrowd)，可同时评估标注人员的专业性及标注正确性，大大提高了答案聚合的可靠性和准确性。不过，这些额外信息也引入了更多的噪声和不确定性。Zhang 等[8]提出了正标签频率阈值法(Positive LAbel frequency Threshold, PLAT)，在计算每个样本的正标签数后自动搜索阈值并将样本分类为两类，在解决有偏标注问题和不平衡类问题方面具有明显效果，但对多标签问题偏差却难以有效建模。

除了以上基于概率模型的方法外，研究者们引入其他相关技术技巧来提高答案聚合算法的性能，包括改进优化现有方法[9,24]、聚类[25,26]和深度学习[27,28]等。例如，Zhang 等[9]提出自适应加权多数投票算法(Adaptive Weighted Majority Voting, AWMV)，利用每个样本的多个有噪声标签中正例的频率估计偏好率，并基于偏好率分配权重给正例和负例。Zhang 等[26]提出了双层聚类方法(Bi-Layer Clustering, BLC)，首先提取概念级特征对样本进行聚类，然后使用物理级特征再次执行聚类，同时物理层上的估计标签校正可能在概念层上错误的聚合标签。Atarashi 等[28]提出生成式深度学习模型，通过引入潜在特征以有效利用未标注数据解决了潜变量后验概率难以处理的问题。以上方法大大提高了单标签答案聚合，但由于没有考虑多标签的全局关联性，在多标签任务上仍表现欠佳。

1.2　多标签答案聚合方法

与单标签聚合方法相比，多标签答案聚合问题的研究相对较少[13,14]。最初的多标签答案聚合方法相关工作主要通过

一些先验知识来扩展单标签众包学习方法。Nowak 等[29]发现使用多数投票策略从多个标注集生成一个标注集可在一定程度上剔除非专家的噪声标注。Duan 等[30]提出了一种概率级联方法(Cascaded estimation with Dawid-Skene, C-DS)，利用源分类中的标签集与目标分类中的标签集之间的语义距离建立二者之间的映射。然而，这两种方法均忽视了标签之间的关联性。Yoshimura 等[31]通过合并 GLAD[18]到(RAndom k-labEL sets, RAkEL)[32]中提出了 RAkEL-GLAD 方法来平衡多标签答案聚合的估计精度和计算复杂度。Hung 等[33]提出贝叶斯非参数一致性方法，通过建模标签之间的共现依赖关系，将答案相似的标注者分为一组来实现对标注者之间的部分聚合答案。以上多标签答案聚合方法均忽略了对标注人员的建模。为解决该问题，Zhang 等[34]提出一个更通用的多分类多标签依赖模型(Multi-Class Multi-Label Dependency, MCMLD)，首先通过对每个标注者建立一个多标签混淆矩阵，然后采用 EM 算法来推理每个样本的真值。Tu 等[13]提出了一种多标签众包聚合方法(Multi-Label Crowd Consensus, MLCC)，利用低秩矩阵分解方法对标签的关联性、不同标注者的相似性、标注质量进行建模。李等[14]则采用低秩张量矫正模型和标注融合策略两步优化估计样本的真实标签。以上方法虽能在一定程度上识别欺诈人员，但对于噪声较大的标注数据仍存在聚合准确率低的问题。

2　基于自适应图正则化与联合低秩矩阵分解的众包标注答案聚合

2.1　符号表示

首先说明本文出现的符号表示含义。本文用大写黑体字母表示矩阵，如 $\boldsymbol{X}$；用小写黑体字母表示矩阵的行或列(向量)，如 x_i；用小写字母表示矩阵中的元素，如 x_{ij}。矩阵 $\boldsymbol{X}$ 的 Frobenius 范数被定义为 $\|\boldsymbol{X}\|_F := (\sum_{ij} x_{ij}^2)^{\frac{1}{2}}$。矩阵 $\boldsymbol{X}$ 的 L_1 范数被定义为 $\|\boldsymbol{X}\|_1 := \sum_{ij} |x_{ij}|$。矩阵 $\boldsymbol{X}$ 的迹被定义为 $tr(\boldsymbol{X}) := \sum_i x_{ii}$。$\boldsymbol{1}$ 是一个元素全部为 1 的向量。$\boldsymbol{X} \geq 0$ 表示矩阵 $\boldsymbol{X}$ 的所有元素均为非负。向量 x 满足 $0 \leq \mathrm{x} \leq 1$ 表示 x 的所有元素都属于[0,1]。

2.2　问题定义

表 1 列举了由 6 名审核人员($W_1 - W_6$)对图 1 中 4 张已标注图像($i_1 - i_4$)提供的审核意见。为描述方便，本文用数字 1~6 分别表示候选标签{正确、标注范围不准确、漏标、多标、标注类型错误、标签名称错误}，符号−表示标注人员认为当前图像不具有该标签。作为一种简单且广泛采用的答案聚合方法，MV[15]倾向于选择票数最多的候选标签作为估计标签。对比表 1 中的真值，发现 MV 的结果要么出现部分不正确要么出现部分不完整。

为求解以上多标签答案聚合问题，首先给出问题定义及相关符号说明。给定候选标签集 $\mathcal{L} \triangleq \{1, \cdots, l\}$，假设 m 个标注者为 n 个样本提供标注，每个样本可被多个标注人员提供 $\mathcal{L}$ 中的一个或多个标签。那么，每个标注者都关联着一个 $n \times l$ 的样本-标签标注矩阵，如下所示。

$$\boldsymbol{A}_w \triangleq \begin{pmatrix} a_{11}^w & \cdots & a_{1l}^w \\ \vdots & \ddots & \vdots \\ a_{n1}^w & \cdots & a_{nl}^w \end{pmatrix}$$

其中，$a_{ij}^w = \{-1, 0, 1\}, 1 \leq i \leq n, 1 \leq j \leq l, 1 \leq w \leq m$。$a_{ij}^w = 1(-1)$ 表示第 w 个标注者给第 i 个样本标注(不标注)第 j 个标签(即正(负)标签)，$a_{ij}^w = 0$ 表示第 w 个标注者不给第 i 个样本提供任何标签(即标注缺失)。

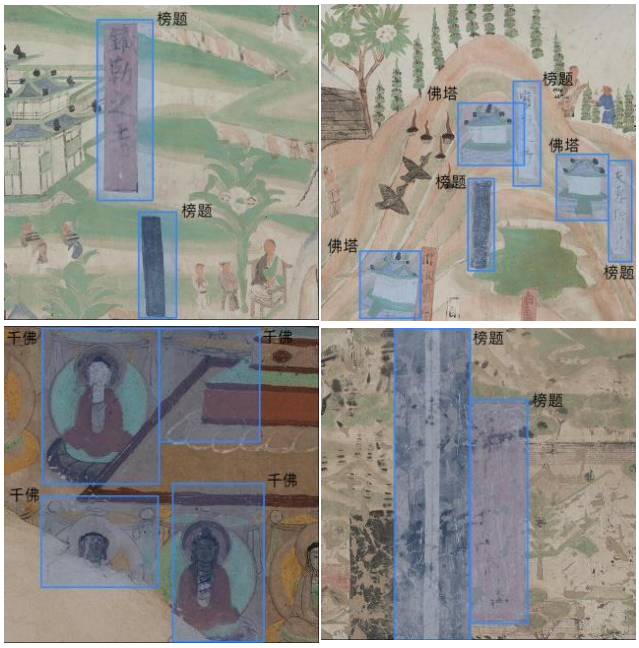

图 1　四张已标注壁画图像

Fig. 1　Four labeled mural images

表 1　6 名审核人员在 4 张已标注壁画图像上给出的审核结果

Tab. 1　Checking results given by 6 auditors on 4 labeled mural images

	W_1	W_2	W_3	W_4	W_5	W_6	MV	真值
i_1	2,3	2,3	1	3	2	1	2,3	2
i_2	3,6	3,6	2	3	3	2	3	2,3,6
i_3	3	1	4	4,5	4	2	4	2,3,4
i_4	-2,3	6	6	1	6	1	6	3,6

注：1：正确;2：标注范围不准确;3：漏标;4：多标;5：标注类型错误;6：标签名称错误

在众包标注模式下，标注结果往往是不完整且带有噪声的，即标注人员仅标注了部分样本的部分标签，甚至给出偏离真实标签的值。本文的目标是从标注结果 $\mathcal{A} = \{\boldsymbol{A}_w\}_{w=1}^m \in \mathbb{R}^{m \times n \times l}$ 估计样本的真实标签矩阵 $\boldsymbol{A}^* \in \mathbb{R}^{n \times l}$。

2.3　AGR-JMF 优化模型提出

为了从收集的样本-标签标注矩阵 $\mathcal{A}$ 中准确估计所有样本的真实标签 $\boldsymbol{A}^* \in \mathbb{R}^{n \times l}$，考虑从以下几方面对标注结果进行建模优化：a) 标注数据往往存在大量稀疏的噪声和离群值，直接用来建模会存在不可靠、不稳定等问题；b) 考虑到多个标注者标注同一任务，在标注者质量可靠的情况下，多个标注者的标注结果是一致的，因此纯净的标注数据整体上应该存在低秩结构并可以从矩阵分解的角度考虑多标签聚类问题；c) 标签的关联矩阵可以较好地体现出标注数据的共现信息，且依赖于标注样本之间的距离。也就是说，如果标注样本的距离计算不准确，将得到错误的标签关联性，进而影响答案聚合效果。为此，本文将输入的样本-标签标注矩阵分为纯净标注和噪声标注两部分，对纯净标注进行矩阵分解、标注质量、标签关联性、标注行为属性相似性等联合学习，并通过交替迭代的方法进行优化。

1) 稀疏噪声去除

基于文献[35,36]的研究，原始数据一般可以分为低秩结构数据和稀疏的震荡/噪声部分，并广泛应用在推荐系统、图像处理和计算机视觉等应用中。一般来说，噪声去除会根据噪声的先验分布来确定采用不同范数的正则化。例如，当噪

声服从独立同分布的高斯分布时，F 范数可以较为容易地恢复噪声，并有快速稳定的求解算法[35]；但当噪声部分具有稀疏、较大梯度等特性时，从数学上来说使用 L_0 正则化约束来选择少量特征以满足噪声稀疏性的要求。但是 L_0 范数的优化是 NP 难的，为了容易优化求解一般将该部分松弛为 L_1 凸优化[36]。考虑到收集到的样本-标签数据中噪声分布稀疏且因壁画内容复杂程度、标注人员背景等导致的标注质量差异较大，选择 L_1 过滤噪声标注数据。因此，本文通过极小化以下目标函数来恢复纯净标注数据 $\{\boldsymbol{D}_w\}_{w=1}^m$ 和噪声标注数据 $\{\boldsymbol{N}_w\}_{w=1}^m$：

$$\min_{\boldsymbol{D}_w,\boldsymbol{N}_w}\frac{1}{2}\|\boldsymbol{A}_w-\boldsymbol{D}_w-\boldsymbol{N}_w\|_F^2+\xi\|\boldsymbol{N}_w\|_1, \tag{1}$$

其中，ξ 是非负正则化参数。很显然，优化问题(1)是关于 $\{\boldsymbol{D}_w\}_{w=1}^m$ 和 $\{\boldsymbol{N}_w\}_{w=1}^m$ 的凸优化问题，使用梯度下降法可以保证收敛到最优解。

2) 纯净标注数据的低秩矩阵分解

作为一种有效降维方法，低秩矩阵分解通过探索和利用行列之间潜在的语义(或结构)关系，在大数据分析中起着越来越重要的作用。给定纯净标注数据 $\{\boldsymbol{D}_w\}_{w=1}^m$，本文希望借助低秩矩阵分解对收集的标注数据挖掘标签间存在的潜在关联性、标注人员行为特征等，采用以下方式联合分解 $\{\boldsymbol{D}_w\}_{w=1}^m$：

$$\begin{aligned}&\min_{\boldsymbol{\mu},\boldsymbol{U},\boldsymbol{V}\ge 0}\sum_{w=1}^m\boldsymbol{\mu}_w\|\boldsymbol{D}_w-\boldsymbol{U}_w\boldsymbol{S}\boldsymbol{V}^T\|_F^2+\lambda\|\boldsymbol{\mu}\|_2^2,\\&\text{s.t. }\boldsymbol{\mu}\boldsymbol{1}=\boldsymbol{1},0\le\boldsymbol{\mu}\le 1.\end{aligned} \tag{2}$$

其中，$\boldsymbol{U}_w\in\mathbb{R}^{l\times k}$ 和 $\boldsymbol{V}\in\mathbb{R}^{l\times k}$ 分别表示第 w 个标注人员的个体矩阵及共享的低秩标签矩阵，$k<(n,l)$ 是两个低秩矩阵的秩。$\boldsymbol{S}\in\mathbb{R}^{k\times k}$ 是 k 个奇异值构成的对角矩阵，用以保证 $\boldsymbol{U}_w\in\mathbb{R}^{l\times k}$ 和 $\boldsymbol{V}\in\mathbb{R}^{l\times k}$ 的非负性。$\boldsymbol{\mu}_w$ 是根据矩阵分解的近似程度给每个标注人员分配的非负权重，λ 是平衡低秩矩阵分解和 $\boldsymbol{\mu}$ 正则化的非负参数。具体来说，如果移除第二个正则化项($\lambda=0$)，本文将得到平凡解 $\boldsymbol{\mu}_w=1$，即只为 $\|\boldsymbol{D}_w-\boldsymbol{U}_w\boldsymbol{S}\boldsymbol{V}^T\|_F^2$ 最小的标注者分配权重1，其他权重为0；相反，如果 $\lambda\to\infty$，所有权重将分配给相同的值，即不考虑不同标注人员的标注质量。通过优化能量函数(2)，本文希望为与 $\boldsymbol{D}_w$ 近似误差较小的标注人员分配较大的权重。换句话说，如果 $\boldsymbol{D}_w$ 不能被很好地近似，也就是该标注人员更大地可能提供了与其他标注人员不一致的答案，通常认为该标注人员不可靠而被赋予更小的权重。因此，该部分通过有选择地降低低质量标注者的影响以提高答案聚合的准确率。

3) 基于自适应图正则化的标签关联性构建

在多标签答案聚合应用中，标签关联性构建至关重要。例如，“正确”标签与其他任何错误标签不可能同时出现，而在缺损严重的壁画中“标注范围不准确”与“漏标”往往同时出现，本文希望能将这些内在的关联性较为准确地嵌入到关联性构建过程中。现有的大部分方法都是直接基于一个预定义的模型(如高斯核函数[37]、余弦相似度[13]等)对未处理的标注数据构建标签关联性，但受限于预定义模型的表达能力而达不到最优结果。同时观察到，大多数方法对关联性的计算都是基于带噪声的标注数据样本计算距离，受数据噪声和离群值的影响得到不准确的距离会导致质量很差的关联性矩阵，进而影响低秩矩阵分解的质量，最终产生不理想的答案聚合结果。鉴于此，考虑从每步迭代得到的纯净数据出发，通过优化平均样本标签的图正则化来自适应构建标签关联矩阵。

记 $\overline{\boldsymbol{D}}=\sum_{w=1}^m\boldsymbol{D}_w/m$ 为平均样本标签，标签 i 和 j 的相关性为 c_{ij}，可以通过以下优化求解 c_{ij}：

$$\begin{aligned}&\min_{\boldsymbol{c}_i}\sum_{i,j=1}^l(\|\overline{\boldsymbol{D}}_i-\overline{\boldsymbol{D}}_j\|_2^2c_{ij},\\&\text{s.t. }\boldsymbol{c}_i\boldsymbol{1}=\boldsymbol{1},0\le\boldsymbol{c}_i\le 1.\end{aligned} \tag{3}$$

其中，向量 $\boldsymbol{c}_i\in\mathbb{R}^l$ 的第 jth 个元素是 c_{ij}。然而，与式(2)中的 λ 类似，问题(3)也会得到 $\boldsymbol{c}_i$ 的一个平凡解。为了解决这个问题，根据局部保留投影算法(Locality Preserving Projections, LPP)[38]，添加 $\boldsymbol{c}_i$ 的正则项 $\eta\sum_{i,j=1}^l c_{ij}^2$，建立以下优化问题：

$$\begin{aligned}&\min_{\boldsymbol{c}_i}\sum_{i,j=1}^l(\|\overline{\boldsymbol{D}}_i-\overline{\boldsymbol{D}}_j\|_2^2c_{ij}+\eta c_{ij}^2),\\&\text{s.t. }\boldsymbol{c}_i\boldsymbol{1}=\boldsymbol{1},0\le\boldsymbol{c}_i\le 1.\end{aligned} \tag{4}$$

其中，η 是非负图正则化参数。很明显，$\|\overline{\boldsymbol{D}}_i-\overline{\boldsymbol{D}}_j\|_2^2$ 越小，标签 i 和标签 j 之间的关联性 c_{ij} 越大。通过优化能量函数(4)，可以自适应地从纯净数据中得到标签间的关联性。

考虑到低秩矩阵 $\boldsymbol{V}$ 是在 k 维空间中 l 个标签依赖关系的编码，本文希望利用标签间的关联性 c_{ij} 去进一步指导低秩矩阵 $\boldsymbol{V}$ 的优化[13]，如下所示。

$$\min_{\boldsymbol{v}\ge 0}\frac{1}{2}\sum_{i,j=1}^l c_{ij}\|\boldsymbol{v}_i-\boldsymbol{v}_j\|_2^2=tr(\boldsymbol{V}^T(\boldsymbol{K}-\boldsymbol{C})\boldsymbol{V})=tr(\boldsymbol{V}^T\boldsymbol{L}_C\boldsymbol{V}), \tag{5}$$

其中，矩阵 $\boldsymbol{C}\in\mathbb{R}^{l\times l}$ 是式(4)中 c_{ij} 构成的标签关联性矩阵，$\boldsymbol{v}_i$ 表示非负矩阵 $\boldsymbol{V}$ 的第 i 行。$\boldsymbol{K}$ 是一个对角矩阵，且 $\boldsymbol{K}_{ii}=\sum_{j=1}^l\boldsymbol{C}_{ij}$，$\boldsymbol{L}_C=\boldsymbol{K}-\boldsymbol{C}$。通过优化能量函数式(5)，可以估计出标注人员共享的低秩标签矩阵(k 维空间)。

4) 标注人员行为属性的相似性

与标签关联性构建类似，本文希望用标注人员行为属性的相似性指导标注者个体低秩矩阵 $\boldsymbol{U}_w$ 的求解。换句话说，如果两个标注人员的行为属性(知识背景、经验等)是相似的，那么他们很有可能在同一样本上给出相似的标签。受到文献[13]的启发，首先计算标注人员的相似度矩阵 $\boldsymbol{R}_{wp}$，然后建立基于标注人员行为相似性的正则项，如下所示。

$$\begin{aligned}&\min_{\boldsymbol{U}_w\ge 0}\frac{1}{2}\sum_{w\ne p}\boldsymbol{R}_{wp}\|\boldsymbol{U}_w-\boldsymbol{U}_p\|_F^2=\\&\min_{\boldsymbol{U}_w\ge 0}\frac{1}{2}\sum_{w\ne p}\boldsymbol{R}_{wp}tr((\boldsymbol{U}_w-\boldsymbol{U}_p)^T(\boldsymbol{U}_w-\boldsymbol{U}_p))\end{aligned} \tag{6}$$

其中，$\boldsymbol{R}_{wp}$ 表示标注人员 w 和 p 之间的相似度，可由修正的 RV 系数、Pearson 相关系数、余弦相似度等求解。$\boldsymbol{R}_{wp}\in[0,1]$，其值越大说明二者的相似度越大。其中，余弦相似度通过计算两个向量的夹角的余弦值来度量它们之间的相似性，但没有考虑数据的去中心化(减去平均值)；Pearson 相关系数可以看做是数据去中心化后的余弦相似度，能较好地衡量两个变量之间线性相关性，但对非线性的相关关系往往效果不好；修正的 RV 系数则是从范数角度巧妙地度量高维数据矩阵的公共信息，简单而全面地探求生物数据、数据集(或数据矩阵)对之间的相似性。鉴于此，本文采用式(7)中修正的 RV 系数估计标注者相似度：

$$\boldsymbol{R}_{wp}=\frac{tr(\boldsymbol{A}_w'\boldsymbol{A}_p')}{\sqrt{tr(\boldsymbol{A}_w'\boldsymbol{A}_w')tr(\boldsymbol{A}_p'\boldsymbol{A}_p')}}, \tag{7}$$

其中，$\boldsymbol{A}_w'=\boldsymbol{A}_w\boldsymbol{A}_w^T-\mathrm{diag}(\boldsymbol{A}_w\boldsymbol{A}_w^T)$。同时，本文也通过实验发现 $\boldsymbol{R}_{wp}$ 采用修正的 RV 系数估计标注者相似度要优于采用 Pearson 相关系数或余弦相似度。

基于以上分析，本文将去噪、低秩矩阵分解、标签关联性、标注人员行为属性相似性等优化集成到统一的目标函数 $\Phi(\boldsymbol{D},\boldsymbol{N},\boldsymbol{U},\boldsymbol{S},\boldsymbol{V},\boldsymbol{\mu},\boldsymbol{C})$ 中，如下所示。

$$\min_{\boldsymbol{U},\boldsymbol{V},\boldsymbol{S},\boldsymbol{D},\boldsymbol{N},\boldsymbol{\mu},\boldsymbol{C}} \rho\sum_{w=1}^{m}\|\boldsymbol{A}_w-\boldsymbol{D}_w-\boldsymbol{N}_w\|_F^2+\xi\sum_{w=1}^{m}\|\boldsymbol{N}_w\|_1+\sum_{w=1}^{m}\boldsymbol{\mu}_w\|\boldsymbol{D}_w-\boldsymbol{U}_w\boldsymbol{S}\boldsymbol{V}^T\|_F^2+\lambda\|\boldsymbol{\mu}\|_2^2+\alpha tr(\boldsymbol{V}^T\boldsymbol{L}_C\boldsymbol{V})+\beta\sum_{w\neq p}\boldsymbol{R}_{wp}tr\big((\boldsymbol{U}_w-\boldsymbol{U}_p)^T(\boldsymbol{U}_w-\boldsymbol{U}_p)\big)+\gamma\Big(\sum_{i,j=1}^{l}(\|\overline{\boldsymbol{D}}_i-\overline{\boldsymbol{D}}_j\|_2^2c_{ij}+\eta c_{ij}^2)\Big),$$
$$\text{s.t. }\boldsymbol{\mu}\boldsymbol{1}=\boldsymbol{1},0\le\boldsymbol{\mu}\le 1,\boldsymbol{U}\ge 0,\boldsymbol{V}\ge 0,\boldsymbol{c}_i\boldsymbol{1}=\boldsymbol{1},0\le\boldsymbol{c}_i\le 1. \tag{8}$$

其中，ρ，ξ，λ，α，β，γ 以及 η 是控制各部分能量的非负权重。

当优化完 $\boldsymbol{D},\boldsymbol{N},\boldsymbol{U},\boldsymbol{S},\boldsymbol{V},\boldsymbol{\mu},\boldsymbol{C}$ 之后，自适应地整合 m 个标注人员的标注矩阵得到最终的聚合答案：

$$\boldsymbol{A}^*=\sum_{w=1}^{m}\boldsymbol{\mu}_w\boldsymbol{U}_w\boldsymbol{S}\boldsymbol{V}. \tag{9}$$

最后的推断结果 $\boldsymbol{A}^*$ 不仅通过数据分离、低秩矩阵近似来去除 $\boldsymbol{A}_w$ 中的噪声或离群值标注，还通过给标注者分配不同的权重值来降低太过嘈杂的标注矩阵的影响。

2.4　算法求解

问题(8)是关于 $\boldsymbol{D},\boldsymbol{N},\boldsymbol{U},\boldsymbol{S},\boldsymbol{V},\boldsymbol{\mu},\boldsymbol{C}$ 的带约束优化问题，并不能保证同时对所有变量是凸的。因此，期望找到该优化问题的全局最优解是不现实的。为此，本文采用交替迭代优化的更新策略分别优化 $\boldsymbol{D},\boldsymbol{N},\boldsymbol{U},\boldsymbol{S},\boldsymbol{V},\boldsymbol{\mu},\boldsymbol{C}$，即在固定其他变量为常量的同时，只对一个变量进行优化。原问题被转换为 7 个可单独求解的子问题，重复这个过程直到收敛为止。

a)　$\boldsymbol{D}$ 子问题

给定 $(\boldsymbol{N}^k,\boldsymbol{U}^k,\boldsymbol{S}^k,\boldsymbol{V}^k,\boldsymbol{\mu}^k,\boldsymbol{C}^k)$，考虑到 $\sum_{i,j=1}^{l}\|\overline{\boldsymbol{D}}_i-\overline{\boldsymbol{D}}_j\|_2^2c_{ij}$ 该项主要作用是指导标签关联性求解，且关于 $\boldsymbol{D}_w$ 很难直接求导，在该子问题中本文忽略这一项的优化。因此，对任意 $1\le w\le m$，得到关于 $\boldsymbol{D}_w$ 的能量函数：

$$\min\Phi_1(\boldsymbol{D}_w)=\rho\|\boldsymbol{A}_w-\boldsymbol{D}_w-\boldsymbol{N}_w^k\|_F^2+\boldsymbol{\mu}_w^k\|\boldsymbol{D}_w-\boldsymbol{U}_w^k\boldsymbol{S}^k(\boldsymbol{V}^k)^T\|_F^2,$$

令 $\frac{\partial\Phi_1(\boldsymbol{D}_w)}{\partial\boldsymbol{D}_w}=0$，得到以下更新公式：

$$\boldsymbol{D}_w^{k+1}=\frac{\boldsymbol{\mu}_w^k\boldsymbol{U}_w^k\boldsymbol{S}^k(\boldsymbol{V}^k)^T+\rho(\boldsymbol{A}_w-\boldsymbol{N}_w^k)}{\rho+\boldsymbol{\mu}_w^k}. \tag{10}$$

b)　$\boldsymbol{N}$ 子问题

给定 $(\boldsymbol{D}^{k+1},\boldsymbol{U}^k,\boldsymbol{S}^k,\boldsymbol{V}^k,\boldsymbol{\mu}^k,\boldsymbol{C}^k)$，对任意 $1\le w\le m$，得到关于 $\boldsymbol{N}_w$ 的能量函数：

$$\min\Phi_2(\boldsymbol{N}_w)=\rho\|\boldsymbol{A}_w-\boldsymbol{D}_w^{k+1}-\boldsymbol{N}_w\|_F^2+\xi\|\boldsymbol{N}_w\|_1,$$

该优化问题有封闭解：

$$\boldsymbol{N}_w^{k+1}=\text{shrink}(\boldsymbol{A}_w-\boldsymbol{D}_w^{k+1},\frac{\xi}{2\rho}). \tag{11}$$

其中，$\text{shrink}(s,t)=\frac{s}{|s|_1}\max(|s|_1-t,0)$。

c)　$\boldsymbol{C}$ 子问题

给定 $(\boldsymbol{D}^{k+1},\boldsymbol{N}^{k+1},\boldsymbol{U}^k,\boldsymbol{S}^k,\boldsymbol{V}^k,\boldsymbol{\mu}^k)$，得到关于 $\boldsymbol{C}$ 的目标方程：

$$\min\Phi_3(\boldsymbol{C})=\alpha tr((\boldsymbol{V}^k)^T\boldsymbol{L}_C(\boldsymbol{V}^k))+\gamma\Big(\sum_{i,j=1}^{l}(\|\overline{\boldsymbol{D}}_i^k-\overline{\boldsymbol{D}}_j^k\|_2^2c_{ij}+\eta c_{ij}^2)\Big),$$
$$\text{s.t. }\boldsymbol{c}_i\boldsymbol{1}=\boldsymbol{1},0\le\boldsymbol{c}_i\le 1. \tag{12}$$

令 $b_{ij}^p=\|\overline{\boldsymbol{D}}_i^k-\overline{\boldsymbol{D}}_j^k\|_2^2$，$b_{ij}^v=\|\boldsymbol{v}_i^k-\boldsymbol{v}_j^k\|_2^2$，式(12)转换为

$$\min\Phi_3(\boldsymbol{C})=\gamma\Big(\sum_{i,j=1}^{l}(b_{ij}^p+\frac{\alpha}{\gamma}b_{ij}^v)c_{ij}+\eta c_{ij}^2\Big),$$
$$\text{s.t. }\boldsymbol{c}_i\boldsymbol{1}=\boldsymbol{1},0\le\boldsymbol{c}_i\le 1. \tag{13}$$

进一步，对任意 $1\le i\le l$，定义行向量 $\boldsymbol{b}_i=\boldsymbol{b}_i^p+\frac{\alpha}{\gamma}\boldsymbol{b}_i^V\in\mathbb{R}^{1\times l}$，式(13)可以重新写为

$$\min\Phi_3(\boldsymbol{c}_i)=\boldsymbol{b}_i\cdot\boldsymbol{c}_i+\eta\|\boldsymbol{c}_i\|_2^2,$$
$$\text{s.t. }\boldsymbol{c}_i\boldsymbol{1}=\boldsymbol{1},0\le\boldsymbol{c}_i\le 1. \tag{14}$$

为了求解带约束的优化问题(14)，本文引入以下定理。

定理 1　假设常向量 $\mathrm{k}=(k_1,\cdots,k_l)\in\mathbb{R}^{1\times l}$ 和非负常数 λ，关于 $\mathrm{x}=(x_1,\cdots,x_l)\in\mathbb{R}^{1\times l}$ 的优化问题

$$\min_{\mathrm{x}\in\mathbb{R}^{1\times l}}\boldsymbol{k}\cdot\mathrm{x}+\lambda\|\mathrm{x}\|_2^2\quad\text{s.t. }\mathrm{x}\cdot\boldsymbol{1}=\boldsymbol{1},0\le\mathrm{x}\le 1. \tag{15}$$

其最优解可表示为

$$x_i=\begin{cases}\dfrac{\theta-k_i}{2\lambda}, & i\le q,\\ 0, & i>q.\end{cases} \tag{16}$$

其中，$\theta=\dfrac{2\lambda+\sum_{i=1}^{q}k_i}{q}$，$\mathrm{k}=(k_1,\cdots,k_l)$ 为 k 按升序重排列向量。

证明　为了求解问题(15)，采用拉格朗日乘子法将其转换为

$$L(\mathrm{x},\boldsymbol{\zeta},\theta)=\mathrm{k}\cdot\mathrm{x}+\lambda\|\mathrm{x}\|_2^2-\boldsymbol{\zeta}\cdot\mathrm{x}-\theta(\mathrm{x}\cdot\boldsymbol{1}-1),$$

其中，向量 $\boldsymbol{\zeta}=(\zeta_1,\cdots,\zeta_l)\ge 0$ 和 $\theta\ge 0$ 为拉格朗日乘子分别用来约束 $0\le\boldsymbol{x}\le 1$ 和 $\boldsymbol{x}\cdot\boldsymbol{1}=1$。$L(\boldsymbol{x},\boldsymbol{\zeta},\theta)$ 关于 $\boldsymbol{x}$ 的偏导数为

$$\frac{\partial L(\mathrm{x},\boldsymbol{\zeta},\theta)}{\partial\mathrm{x}}=\mathrm{k}+2\lambda\mathrm{x}-\boldsymbol{\zeta}-\theta\cdot\boldsymbol{1},$$

由 Karush-Kuhn-Tucker(KKT)条件可知，x 的最优解应满足以下四个条件：

松弛互补性条件：对任意 $1\le i\le l$，$\zeta_i x_i=0$；

稳定性条件：对任意 $1\le i\le l$，$k_i+2\lambda x_i-\zeta_i-\theta=0$；

可行性条件：对任意 $1\le i\le l$，$x_i\ge 0,\sum_i x_i=1$；

对偶可行性条件：对任意 $1\le i\le l$，$\zeta_i\ge 0$。

由条件 ii)可知，$x_i=\frac{\zeta_i+\theta-k_i}{2\lambda}$，通过分情况讨论如下：

如果 $\theta>k_i$，根据条件 i)和 iv)可得到 $\zeta_i=0$，$x_i=\frac{\theta-k_i}{2\lambda}$；

如果 $\theta=k_i$，根据条件 i)和 iv)可得到 $\zeta_i=x_i=0$；

如果 $\theta<k_i$，根据条件 i)、ii)及 iv)可得到 $\zeta_i>0$，$x_i=0$。

假设 $\mathrm{k}=(k_1,\cdots,k_l)$ 是 k 以升序重新排列的向量，对于给定的 λ，存在某个 $q\in\{1,\cdots,l\}$ 使得 $k_q<\theta$，$k_{q+1}\ge\theta$，且满足 $\mathrm{x}\cdot\boldsymbol{1}=\sum_{i=1}^{q}\frac{\theta-k_i}{2\lambda}=1$。因此，可得到以下解：

$$x_i=\begin{cases}\dfrac{\theta-k_i}{2\lambda}, & i\le q,\\ 0, & i>q.\end{cases}$$

其中，$\theta=\dfrac{2\lambda+\sum_{i=1}^{q}k_i}{q}$。证明毕。

基于定理1，本文通过逐行求解 $\boldsymbol{c}_i$ 来构建标签关联性矩阵 $\boldsymbol{C}$。

d)　$\boldsymbol{V}$ 子问题

给定 $(\boldsymbol{D}^{k+1},\boldsymbol{N}^{k+1},\boldsymbol{U}^k,\boldsymbol{S}^k,\boldsymbol{\mu}^k,\boldsymbol{C}^{k+1})$，得到关于 $\boldsymbol{V}$ 的能量函数：

$$\min\Phi_4(\boldsymbol{V})=\sum_{w=1}^{m}\boldsymbol{\mu}_w^k\|\boldsymbol{D}_w^{k+1}-\boldsymbol{U}_w^k\boldsymbol{S}^k\boldsymbol{V}^T\|_F^2+\alpha tr(\boldsymbol{V}^T\boldsymbol{L}_C\boldsymbol{V}),$$
$$\text{s.t.}\boldsymbol{V}\ge 0.$$

$\Phi_4(\boldsymbol{V})$ 关于 $\boldsymbol{V}$ 的偏导数为

$$\frac{\partial\Phi_4(\boldsymbol{V})}{\partial\boldsymbol{V}}=\sum_{w=1}^{m}\boldsymbol{\mu}_w^k(2\boldsymbol{V}(\boldsymbol{S}^k)^T(\boldsymbol{U}_w^k)^T\boldsymbol{U}_w^k\boldsymbol{S}^k-2(\boldsymbol{D}_w^k)^T\boldsymbol{U}_w^k\boldsymbol{S}^k)+2\alpha\boldsymbol{L}_C\boldsymbol{V},$$

利用 KKT 条件来限制 $\boldsymbol{V}$ 的非负性，得到 $\boldsymbol{V}$ 的更新式：

$$\boldsymbol{V}^{k+1}=\boldsymbol{V}^{k}\sqrt{\frac{\sum_{w=1}^{m}\boldsymbol{\mu}_w^k\left((\boldsymbol{D}_w^{k+1})^T\boldsymbol{U}_w^k\boldsymbol{S}^k\right)^{+}+\sum_{w=1}^{m}\boldsymbol{\mu}_w^k\boldsymbol{V}\left((\boldsymbol{S}^k)^T(\boldsymbol{U}_w^k)^T\boldsymbol{U}_w^k\boldsymbol{S}^k\right)^{-}+\alpha\boldsymbol{L}^{-}\boldsymbol{V}}{\sum_{w=1}^{m}\boldsymbol{\mu}_w^k\left((\boldsymbol{D}_w^{k+1})^T\boldsymbol{U}_w^k\boldsymbol{S}^k\right)^{-}+\sum_{w=1}^{m}\boldsymbol{\mu}_w^k\boldsymbol{V}\left((\boldsymbol{S}^k)^T(\boldsymbol{U}_w^k)^T\boldsymbol{U}_w^k\boldsymbol{S}^k\right)^{+}+\alpha\boldsymbol{L}^{+}\boldsymbol{V}}},\tag{17}$$

其中，$\boldsymbol{O}^{+}=\frac{|\boldsymbol{O}|+\boldsymbol{O}}{2}$，$\boldsymbol{O}^{-}=\frac{|\boldsymbol{O}|-\boldsymbol{O}}{2}$。

e) $\boldsymbol{U}$ **子问题**

给定 $(\boldsymbol{D}^{k+1},\boldsymbol{N}^{k+1},\boldsymbol{V}^{k+1},\boldsymbol{S}^{k},\boldsymbol{\mu}^{k},\boldsymbol{C}^{k+1})$，本文通过优化以下目标函数逐一更新 $\boldsymbol{U}_w$：

$$\begin{aligned}\min\Phi_5(\boldsymbol{U}_w)=\boldsymbol{\mu}_w^k\|\boldsymbol{D}_w^{k+1}-\boldsymbol{U}_w\boldsymbol{S}^k(\boldsymbol{V}^k)^T\|_F^2+\\ \beta\sum_{w\neq p}\boldsymbol{R}_w tr((\boldsymbol{U}_w-\boldsymbol{U}_p^k)(\boldsymbol{U}_w-\boldsymbol{U}_p^k)^T),\\ \text{s.t.}\,\boldsymbol{U}_w\geq 0.\end{aligned}$$

$\Phi_5(\boldsymbol{U}_w)$ 关于 $\boldsymbol{U}_w$ 的偏导数为

$$\begin{aligned}\frac{\partial\Phi_5(\boldsymbol{U}_w)}{\partial\boldsymbol{U}_w}=\boldsymbol{\mu}^k(2\boldsymbol{U}_w(\boldsymbol{S}^k)(\boldsymbol{V}^{k+1})^T\boldsymbol{V}^{k+1}(\boldsymbol{S}^k)^T-\\ 2\boldsymbol{D}_w^{k+1}\boldsymbol{V}^{k+1}(\boldsymbol{S}^k)^T)+2\beta\sum_{w\neq p}\boldsymbol{R}_w(\boldsymbol{U}_w-\boldsymbol{U}_p^k),\end{aligned}$$

与 $\boldsymbol{V}$ 的求解类似，利用 KKT 条件来限制 $\boldsymbol{U}_w$ 的非负性，得到 $\boldsymbol{U}_w$ 的更新式：

$$\boldsymbol{U}_w^{k+1}=\boldsymbol{U}_w^{k}\sqrt{\frac{\boldsymbol{\mu}_w^k(\boldsymbol{D}_w^{k+1}\boldsymbol{V}^{k+1}(\boldsymbol{S}^k)^T)^{+}+\boldsymbol{\mu}_w^k(\boldsymbol{S}^k(\boldsymbol{V}^k)^T\boldsymbol{V}^k(\boldsymbol{S}^k)^T)^{-}+2\beta\sum_{w\neq p}\boldsymbol{R}_{wp}\boldsymbol{U}_p^k}{\boldsymbol{\mu}_w^k(\boldsymbol{D}_w^{k+1}\boldsymbol{V}^{k+1}(\boldsymbol{S}^k)^T)^{-}+\boldsymbol{\mu}_w^k(\boldsymbol{S}^k(\boldsymbol{V}^k)^T\boldsymbol{V}^k(\boldsymbol{S}^k)^T)^{+}+2\beta\sum_{w\neq p}\boldsymbol{R}_{wp}\boldsymbol{U}_w^k}}.\tag{18}$$

f) $\boldsymbol{S}$ 子问题

给定 $(\boldsymbol{D}^{k+1},\boldsymbol{N}^{k+1},\boldsymbol{V}^{k+1},\boldsymbol{U}^{k+1},\boldsymbol{\mu}^{k},\boldsymbol{C}^{k+1})$，关于 $\boldsymbol{S}$ 的目标函数是经典的二次凸优化问题：

$$\min\Phi_6(\boldsymbol{S})=\sum_{w=1}^{m}\boldsymbol{\mu}_w^k\|\boldsymbol{D}_w^{k+1}-\boldsymbol{U}_w^k\boldsymbol{S}(\boldsymbol{V}^{k+1})^T\|_F^2,$$

令 $\frac{\partial\Phi_6(\boldsymbol{S})}{\partial\boldsymbol{S}}=0$，得到 $\boldsymbol{S}$ 的更新公式：

$$\begin{aligned}\boldsymbol{S}^{k+1}=\left(\sum_{w=1}^{m}\boldsymbol{\mu}_w^k(\boldsymbol{U}_w^{k+1})^T\boldsymbol{U}_w^{k+1}\right)^{-1}\\ \left(\sum_{w=1}^{m}\boldsymbol{\mu}_w^k(\boldsymbol{U}_w^{k+1})^T\boldsymbol{D}_w^{k+1}\boldsymbol{V}^{k+1}\right)\left((\boldsymbol{V}^{k+1})^T\boldsymbol{V}^{k+1}\right)^{-1}.\end{aligned}\tag{19}$$

g) $\boldsymbol{\mu}$ 子问题

给定 $(\boldsymbol{D}^{k+1},\boldsymbol{N}^{k+1},\boldsymbol{V}^{k+1},\boldsymbol{U}^{k+1},\boldsymbol{S}^{k+1},\boldsymbol{C}^{k+1})$，得到关于 $\boldsymbol{\mu}$ 的目标方程如下：

$$\begin{aligned}\min\Phi_7(\boldsymbol{\mu})=\sum_{w=1}^{m}\boldsymbol{\mu}_w\|\boldsymbol{D}_w^{k+1}-\boldsymbol{U}_w^{k+1}\boldsymbol{S}^{k+1}(\boldsymbol{V}^{k+1})^T\|_F^2+\lambda\|\boldsymbol{\mu}\|_2^2,\\ \text{s.t.}\,\boldsymbol{\mu}\boldsymbol{1}=\boldsymbol{1},0\leq\boldsymbol{\mu}\leq 1.\end{aligned}\tag{20}$$

令 $\mathrm{k}_w=\|\boldsymbol{D}_w^{k+1}-\boldsymbol{U}_w^{k+1}\boldsymbol{S}^{k+1}(\boldsymbol{V}^{k+1})^T\|_F^2$，则式(20)转换为

$$\min\Phi_7(\boldsymbol{\mu})=\mathrm{k}\cdot\boldsymbol{\mu}+\lambda\|\boldsymbol{\mu}\|_2^2,\ \text{s.t.}\ \boldsymbol{\mu}\boldsymbol{1}=\boldsymbol{1},0\leq\boldsymbol{\mu}\leq 1.\tag{21}$$

式(21)可运用定理 1 求解。

综上，AGR-JMF 通过以上 7 个子问题的迭代求解得到，算法详细流程如算法 1 所示。

算法 1　本文提出的 AGR-JMF 多标签答案聚合方法

输入：m 个标注人员的标注矩阵 $\mathcal{A}\in\mathbb{R}^{m\times n\times l}$；$\alpha$，$\beta$，$\gamma$，$\eta$，$\rho$，$\xi$，$\lambda$ 正则项权重；低秩矩阵的秩 k；迭代终止阈值 ϵ 及最大迭代步数 maxIter

输出：通过式(9)返回聚合答案 $\boldsymbol{A}^{*}$

初始化：$\boldsymbol{U}$ 和 $\boldsymbol{V}$ 为随机 $n\times k$ 矩阵，$\boldsymbol{S}=\boldsymbol{I}_{k\times k}$，$\boldsymbol{D}=\boldsymbol{0}$，$\boldsymbol{N}=\boldsymbol{A}$，$\boldsymbol{C}=\boldsymbol{I}_{l\times l}$，$\boldsymbol{\mu}=\frac{1}{m}\cdot\boldsymbol{1}$，$t=0$，$\Phi^0=0$

```
while |Φ^{t+1} − Φ^t| > ϵ & t < maxIter
  t = t + 1;
  通过式(10)更新 D^t ;
  通过式(11)更新 N^t ;
  通过定理 1 更新 C^t 的每一行;
  通过式(17)更新 V^t ;
  通过式(18)更新 U^t ;
  通过式(19)更新 S^t ;
  通过定理 1 更新 μ^t ;
  通过式(8)计算能量函数值 Φ^t ;
end while
```

3　实验与结果

3.1　数据集描述

为了验证和比较本文提出算法的性能，本节首先在 6 个真实数据集上进行实验，这些数据集的详细信息如表 2 所示。其中，Movie 数据集是一个关于电影类别分类的数据集[33]，Affective 是一个包含 100 个标题样本和 6 种情绪类别，由来自 Amazon Mechanic Turk 平台的标注人员被要求为每个情绪提供 0 到 100 之间的分数[39]。本文将每个标签分为两类：negative(分值=0)和 positive (分值>0)。其余 4 个数据集来自于“Apple”和“Love”两部小说中的人物情感标注，被 Duan 等[11]用于情感分析中。

表 2　用于实验的 6 个真实数据集统计

Tab. 2　Statistics of six real-world datasets used in the experiment

Datasets	Workers	Instances	Labels	Workers/Instance	Label/Instance
Movie	89	100	19	35	1.95
AppleEkman	68	78	6	30	1.27
AppleNakamura	57	78	10	30	1.18
LoveEkman	54	63	6	30	1.05
LoveNakamura	41	63	10	31	1.53
Affective	38	100	6	10	6

此外，为了解决数字文化遗产众包标注的人工审核效率低的问题，本文收集整理了包含了榜题、供养人、草庐等 9 类壁画元素的 700 张图像的审核标注结果。该数据集主要以目标检测和实例语义分割为主，标注信息由天津大学的 6 名研究生经敦煌研究院的敦煌学专家培训后提供。考虑到标注成本远远高于审核成本，本文将该数据集的审核任务分配给从事文物壁画处理工作 5 年以上的 18 名专业技术人员，对每张图像给出了 6 个候选审核标签{正确、漏标、多标、标注范围不准确、标签类型错误、标签名称错误}中的一个或多个，每人审核的样本数量不少于 89 个。平均每个人审核了 288 张图像，平均每张图像被标注了 7.4 次，共收集到 31098 个审核标注。为了描述方便，本文记莫高窟壁画数据集(Mogao Grottoes Frescoes)为 MGF。

MGF 真实标签是由本文作者标注。为了粗略了解审核人员的质量，对收集到的审核标签做了初步分析。对每个审核人员提供的标签和真实标签做对比，并计算了每个人的审核准确度。准确度越高，说明审核结果越接近真实标签。18 名审核人员的准确率分别是[0.621, 0.574, 0.482, 0.773, 0.708, 0.640, 0.484, 0.482, 0.501, 0.858, 0.521, 0.778, 0.769, 0.489, 0.503, 0.788, 0.554, 0.493]。由此可以看到，各位审核人员的

准确率各不相同，主要集中在[0.48, 0.86]之间。

3.2　对比方法与评价指标

分别与经典 MV[7]、2 个代表性单标签答案聚合方法 PLAT[8] 和 AWMV[9] 以及 C-DS[30]、RAkEL-GLAD[33]、MCMLD[34]和 MLCC[13]的 4 个先进的多标签答案聚合方法进行比较。

为了便于与 MV[7]、PLAT[8] 和 AWMV[9]比较，将多标签任务看做多个独立的单标签任务，在每个标签上分别使用单标签众包方法。各对比方法的代码均来自于文章源代码，参数设置依照其对应论文或代码中的推荐参数设置。AGR-JMF 的参数缺省设置为：$\alpha=1$，$\beta=0.01$，$\gamma=1$，$\eta=10$，$\rho=10^5$，$\xi=10^4$，$\lambda=10^3$，$k=\lceil l/2\rceil+1$，$\epsilon=10^{-5}$ 及 maxIter= 1000。本文的对比实验均在 3.2GHz、8G 内存的八核台式机上分别采用 Python 3.6 和 MATLAB R2018b 进行。

为量化分析上述各方法的性能，本文采用 Average precision、Ranking loss、Hamming loss 和 Macro F1 这 4 种常用的性能评价度量。其中，Average precision 和 Ranking loss 是由聚合结果 $\boldsymbol{A}^*$ 直接计算的排序度量，Hamming loss 和 Macro F1 是由聚合结果 $\boldsymbol{A}^*$ 转换成二进制结果计算的分类度量，并且 Average precision、Macro F1 (Ranking loss、Hamming loss)的值越大(小)，表明聚合结果越接近真实标签。这些度量的具体定义参见文献[3]。本文中，根据每个样本的真实标签的数量选择预测值最大的 p 个标签作为该样本的二进制聚合标签。其中，p 为该样本的真实正标签的个数。这里，为了获得与以往聚合方法的公平对比，分类度量的计算用到了真实标签信息。在实际数字化应用过程中，只需要记录预测值最大的标签为非“正确”标签的样本，并反馈给数据集审核管理员进行对应样本的标注编辑与修改。

3.3　实验结果

为了降低初始化对算法的随机影响，记录了 10 次重复实验的均值和标准差。表 3 报告了不同答案聚合方法在 6 个真实数据集上的结果。

表 3　6 个真实数据集上 4 种常用评价度量的对比结果(mean ± std.)

Tab. 3　Comparison of four commonly used evaluation metrics (mean ± std.) On six real-world datasets.

Metrics	MV	PLAT	AWMV	C-DS	RAkEL-GLAD	MCMLD	MLCC	AGR-JMF
					Movie			
Average precision	.811±.000	.828±.000	.846±.000	.877±.001	.881±.012	.894±.010	.918±.003	**.939±.009**
Ranking loss	.329±.000	.292±.000	.269±.000	.217±.008	.211±.009	.193±.003	.161±.007	**.142±.003**
Hamming loss	.311±.000	.278±.000	.255±.000	.214±.004	.199±.011	.184±.008	.152±.011	**.137±.008**
Macro F1	.683±.000	.719±.000	.741±.000	.783±.012	.802±.014	.817±.008	.858±.010	**.884±.009**
					AppleEkman			
Average precision	.821±.000	.847±.000	.863±.000	.894±.001	.909±.021	.912±.003	.933±.011	**.951±.007**
Ranking loss	.317±.000	.289±.000	.242±.000	.196±.007	.177±.01	.169±.007	.147±.008	**.122±.009**
Hamming loss	.308±.000	.274±.000	.228±.000	.185±.006	.164±.009	.154±.008	.141±.009	**.114±.008**
Macro F1	.703±.000	.721±.000	.767±.000	.809±.010	.833±.013	.849±.007	.866±.010	**.891±.011**
					AppleNakamura			
Average precision	.833±.000	.853±.000	.872±.000	.902±.001	.912±.012	.906±.003	.941±.011	**.958±.009**
Ranking loss	.311±.000	.271±.000	.232±.000	.177±.008	.153±.009	.169±.003	.129±.008	**.111±.012**
Hamming loss	.288±.000	.253±.000	.216±.000	.169±.004	.147±.011	.163±.008	.118±.010	**.092±.008**
Macro F1	.711±.000	.739±.000	.778±.000	.833±.012	.851±.013	.839±.009	.889±.007	**.903±.011**
					LoveEkman			
Average precision	.838±.000	.851±.000	.873±.000	.899±.001	.911±.014	.919±.002	.946±.012	**.962±.005**
Ranking loss	.291±.000	.269±.000	.223±.000	.182±.009	.166±.010	.151±.008	.117±.007	**.094±.009**
Hamming loss	.283±.000	.258±.000	.211±.000	.164 ±.006	.147±.008	.139±.007	.096±.008	**.088±.007**
Macro F1	.715±.000	.741±.000	.783±.000	.822±.011	.851±.012	.862±.009	.889±.009	**.896±.012**
					LoveNakamura			
Average precision	.825±.000	.841±.000	.868±.000	.904±.014	.903±.022	.911±.003	.927±.021	**.946±.013**
Ranking loss	.338±.000	.319±.000	.272±.000	.211±.010	.205±.011	.193±.009	.137±.009	**.109±.007**
Hamming loss	.311±.000	.299±.000	.253±.000	.197±.006	.188±.009	.171±.007	.129±.009	**.092±.009**
Macro F1	.692±.000	.709±.000	.746±.000	.807±.021	.819±.040	.828±.027	.874±.011	**.912±.011**
					Affective			
Average precision	.749±.000	.756±.000	.773±.000	.826±.001	.828±.019	.836±.002	.896±.020	**.929±.009**
Ranking loss	.491±.000	.474±.000	.451±.000	.347±.011	.336±.009	.326±.011	.201±.030	**.169±.012**
Hamming loss	.319±.000	.310±.000	.298±.000	.250±.010	.246±.014	.239±.016	.180±.020	**.148±.009**
Macro F1	.657±.000	.669±.000	.709±.000	.764±.014	.766±.017	.774±.013	.811±.021	**.857±.011**

表 3 中加粗的结果表明，在不同的数据集上，AGR-JMF 在配对检验(95%置信度)中显著优于其他方法。MV、PLAT、AWMV 将多标签任务转换为多个单标签任务而忽略了标签间的关联性，它们的结果差于 RAkEL-GLAD、MCMLD 以及 MLCC(利用了标签相关性)。这一现象表明，标签相关性在多标签答案聚合中很重要。虽然 AWMV 和 PLAT 是单标签方法，但因 AWMV 为不同类型的标签分配了不同的权重而效果优于 PLAT。多标签答案聚合方法中，除 C-DS 方法之外，

其余的方法均考虑了标签相关性。因此，RAkEL-GLAD、MCMLD 以及 MLCC 的结果都优于 C-DS，但劣于 AGR-JMF。这是因为，虽然 MLCC 考虑了标注者的质量并通过矩阵分解减少了少量噪声标注的影响，但当存在较多标注质量较差的标注人员时，标签间关联性矩阵不够准确而影响了低秩矩阵分解的质量。AGR-JMF 能够自适应地去除原始标注数据中的噪声，同时基于该去噪数据在标签关联性以及不同标注者行为属性相似性的指导下进行低秩矩阵的分解优化，大大提高了答案聚合的准确率。

此外，本文还对比了在莫高窟壁画数据集上的结果，如表 4 所示。由于敦煌壁画元素涵盖内容广、绘画风格迥异、专业性较强，加之有不同程度的病害，标注和审核人员提供的答案参差不齐(收集的数据可能具有更多的噪声和离群值)。从表 4 数据可以清晰地看出，AGR-JMF 明显优于其他方法。

表 4　MGF 数据集上 4 种常用评价度量的对比结果(mean ± std.)

Tab. 4　Comparison of four commonly used evaluation metrics (mean ± std.) On the MGF datasets

Metrics	MV	PLAT	AWMV	C-DS	RAkEL-GLAD	MCMLD	MLCC	AGR-JMF
Average precision	.757±.000	.772±.000	.789±.000	.818±.001	.822±.012	.832±.010	.873±.051	**.924±.003**
Ranking loss	.183±.000	.168±.000	.157±.000	.145±.008	.137±.009	.126±.003	.090±.040	**.039±.002**
Hamming loss	.148±.000	.136±.000	.130±.000	.122±.004	.115±.011	.110±.008	.085±.031	**.058±.002**
Macro F1	.628±.000	.632±.000	.667±.000	.711±.012	.715±.014	.726±.008	.787±.082	**.866±.004**

综上所述，这些实验结果不仅证明了去噪对低秩矩阵分解、自适应获取标签相似性的重要性，也证实了在聚合众包答案时需要考虑标签关联性、标注者质量及标注行为属性相似性等先验知识。

3.4　参数讨论与分析

在上面的实验中，AGR-JMF 使用了固定参数。由于 ρ 和 ξ 分别是约束数据逼近和噪声正则化的参数，在优化中为了避免纯净标注数据严重偏离输入的标注数据，分别缺省设置为较大的正实数 10^5 和 10^4，本文不做过多讨论。因此，本节中本文依次讨论 α，β，γ，η，λ 及矩阵 $\boldsymbol{S}$ 的秩 k 这 6 个参数对 AGR-JMF 的影响。

图 2 展示了 AGR-JMF 在 Affective 和 MGF 数据集上不同 α 和 β 设置组合下的结果。从图中可以观察到，当 α 固定时，$\beta \in [10^{-4},1]$ 取得的结果明显优于其他取值。这是因为太小的 β 忽略了标注人员的行为属性相似性，太大的 β 则夸大了标注人员的行为属性相似性。事实上，为了节省成本，众包标注收集到的样本-标签标注矩阵在某些样本上是稀疏的，进而导致标注者个体矩阵 $\boldsymbol{U}_w$ 也是稀疏的，因此标注人员之间共享较低的行为相似度。基于以上分析，AGR-JMF 倾向于选择不能过大的 β。当固定 β 时，$\alpha \geq 10^{-2}$ 比 $\alpha < 10^{-2}$ 取得了更稳定的结果。这是因为太小的 α 忽略了标签间的关联性这一内在规律导致推理答案没有一致性。综上，适当地考虑标注人员的行为属性相似性和标签的关联性有助于提高多标签答案聚合的准确率。

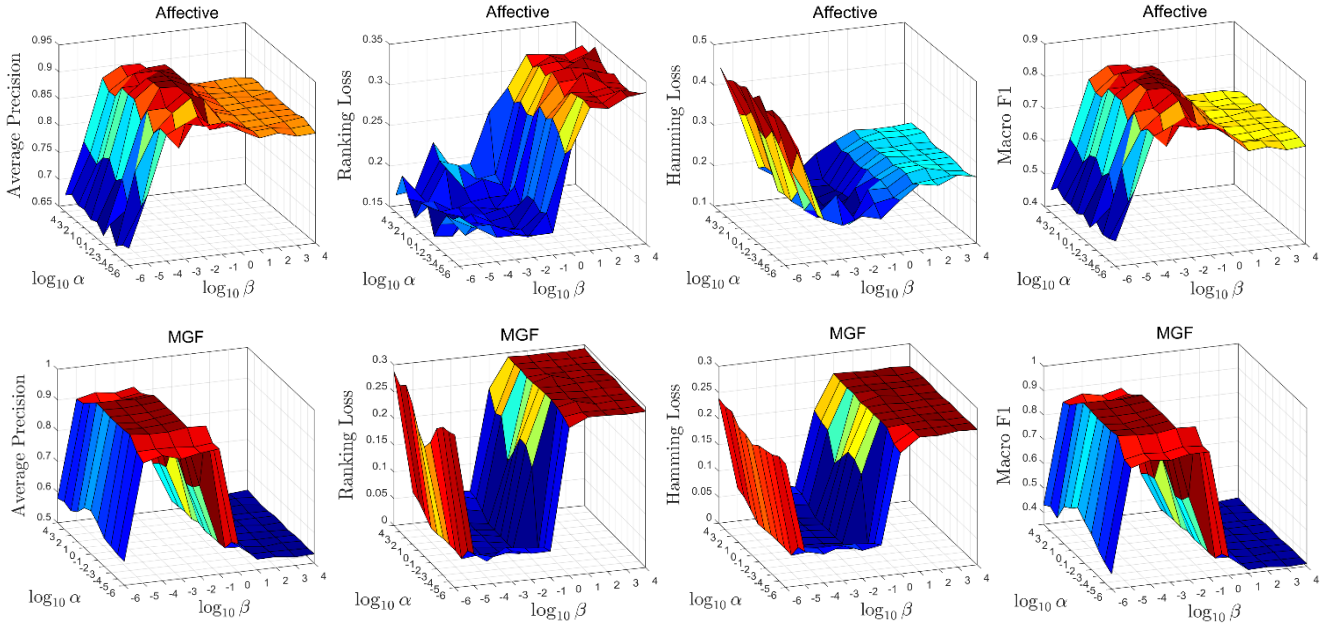

图 2　Affective 和 MGF 数据集上不同 α 和 β 组合下的 4 种常用多标记性能评价指标结果

Fig. 2　Four commonly used evaluation metrics on Affective and MGF under different combinations of α and β

图 3 展示了 AGR-JMF 在 Affective 和 MGF 数据集上不同 γ 和 η 设置组合下的结果。从图中结果可以发现，当 $\gamma > 10^{-1}$ 且 $\eta > 10^{-1}$ 时，AGR-JMF 取得了较稳定的结果。这是因为太小的 γ 和 η 降低了纯净数据对标签间相似矩阵的自适应优化与正则性约束。事实上，纯净数据中蕴涵着丰富的标签内在相似性，较大的正则项系数能保证标签间的相似性是基于更准确的样本距离来计算的，进而更准确地指导低秩矩阵分解。

图 4 展示了 AGR-JMF 在 6 个真实数据集及 MGF 上取不同 λ 的结果。从折线图发现，当 $\lambda \approx 10^3$ 时，AGR-JMF 取得最好结果；当 $\lambda < 1$ 时，AGR-JMF 结果越来越不稳定。这是因为，由第 2.4 部分中 $\boldsymbol{\mu}$ 的计算可知，太小的 λ 导致个人矩阵的权重分配上没有足够的正则化影响，会导致 $\boldsymbol{\mu}$ 取值是稀疏的，即只选择少数标注人员作为可信标注人员，从而损失了大量重要的标注数据；太大的 λ 会因为强大的正则化效应导致所有标注人员获得几乎一样的权重，无法较好地区分哪些标注人员更可靠。

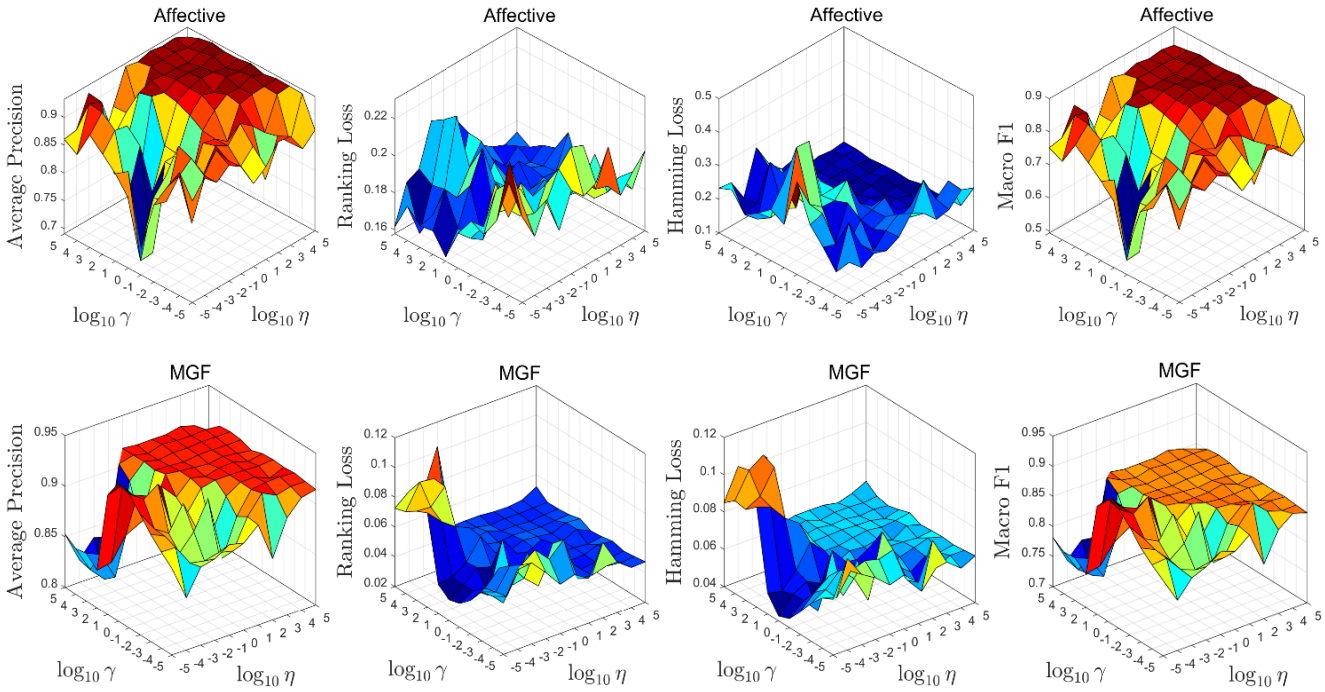

图 3　Affective 和 MGF 数据集上不同 γ 和 η 组合下的 4 种常用多标记性能评价指标结果

Fig. 3　Four commonly used evaluation metrics on Affective and MGF under different combinations of γ and η

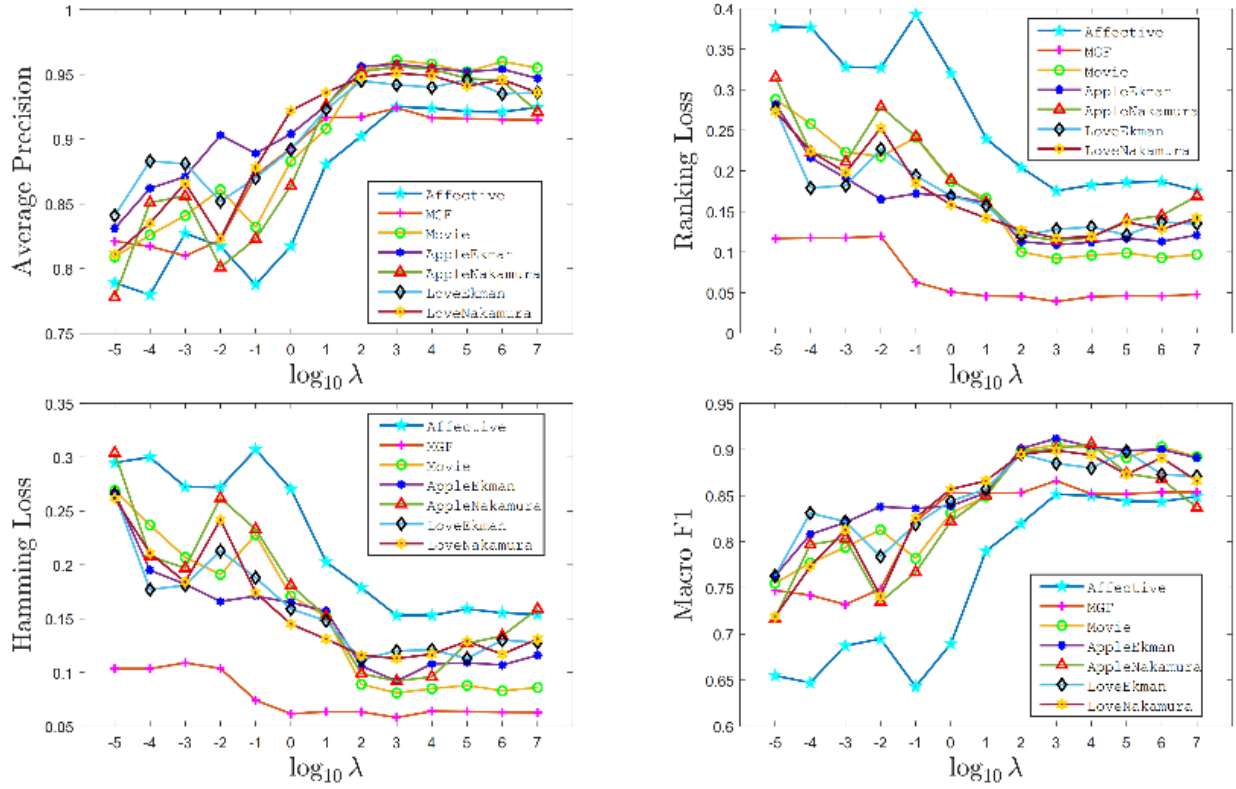

图 4　7 个数据集上不同 λ 取值下的 4 种常用多标记性能评价指标结果

Fig. 4　Four commonly used evaluation metrics on 7 datasets under different input values of λ

为了进一步分析这些结果，选择 Affective 和 MGF 两个数据集观察不同 λ 对标注人员权重的影响，如图 5 所示。从图 5 中有如下发现：1) 当 $\lambda=1$ 时，只有很少一部分标注人员的标注矩阵被选择；当 $\lambda=10^5$ 时，所有标注人员的标注矩阵都被选择，并且被选择的权重几乎一样，这一现象符合上面的分析；2) 加权处理实现了不同标注人员的标注矩阵信息互补，因而 $\lambda=10^3$，$\lambda=10^5$ 取得的结果要比 $\lambda=1$ 更理想；3) 标注人员分配的权重大小与其标注质量是正相关的，即标注质量越好，权重也越大。正如 MGF 数据集，无论是 $\lambda=10^3$ 还是 $\lambda=1$，权重较小甚至为 0 的标注人员恰好是标注准确率相对较低的标注人员。综上所述，以上实验结果证明 AGR-JMF 能够较好地识别标注质量低的标注人员，并通过低秩矩阵分解选择性地整合不同标注质量的标注矩阵。

此外，本文还讨论分析了矩阵 $\boldsymbol{S}$ 的秩 k 对 AGR-JMF 性能的影响，如图 6 所示。通过观察发现，随着 k 的增加，AGR-JMF 性能一开始增加直到 $k>\lceil l/2\rceil+1$ 趋于稳定或者减小。在大部分数据集上，$k=\lceil l/2\rceil+1$ 几乎取得了最好的结果。这一现象证实了通过低秩矩阵近似来估计标注数据的全局结构信息是可行的。

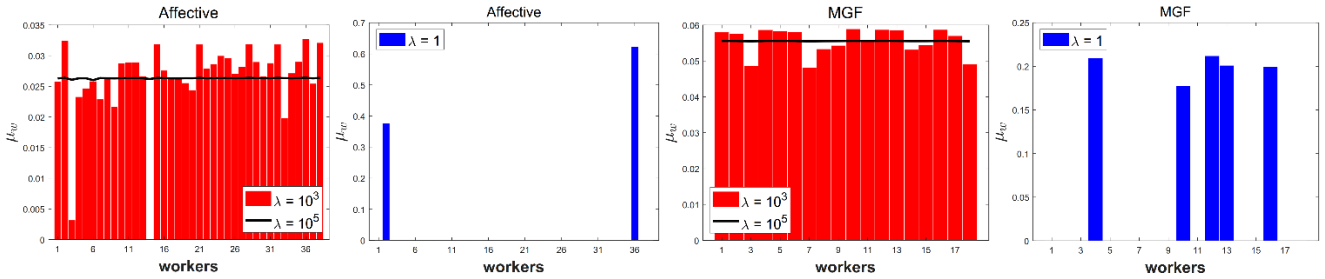

图 5　Affective 和 MGF 数据集上分配给每位标注人员的权重值 $\boldsymbol{\mu}_w$

Fig. 5　Weights assigned to each annotator on Affective and MGF

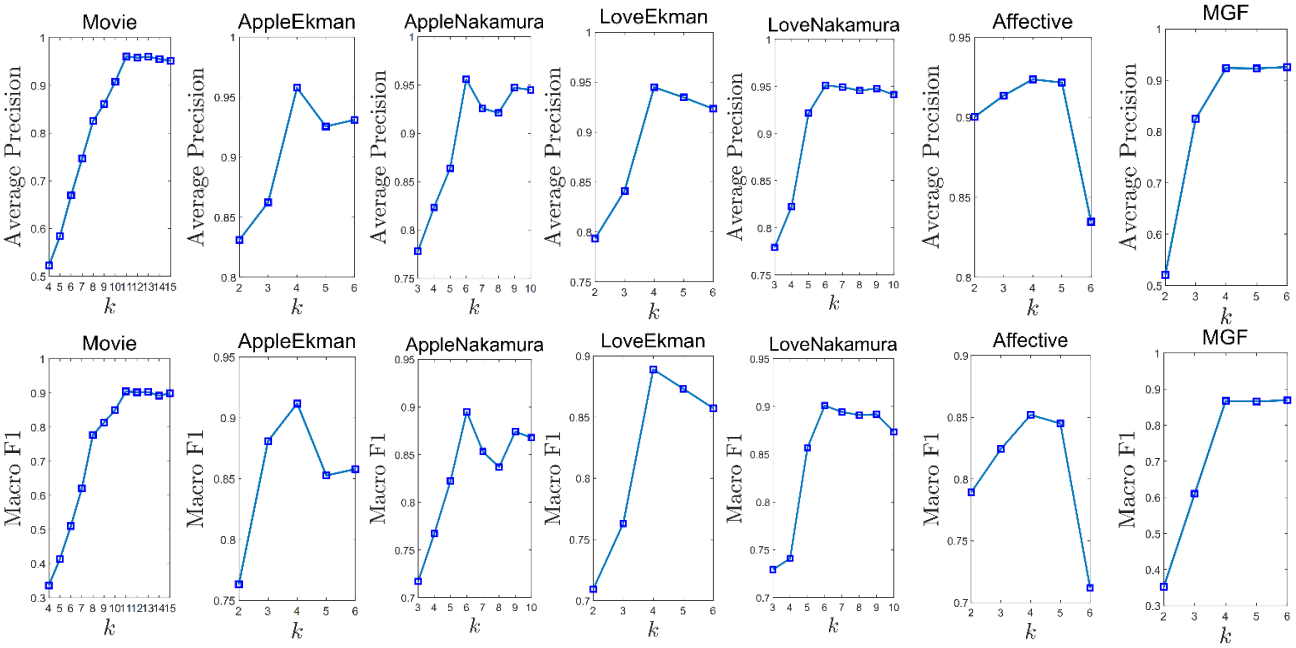

图 6　7 个数据集上不同 k 取值下的 2 种常用多标记性能评价指标结果

Fig. 6　Two commonly used evaluation metrics on 7 datasets under different low-rank sizes k

3.5　欺诈者的鲁棒性分析

由于众包标注并不能实时对标注人员进行监督与约束，收集到的标注结果并不能保证其质量，如何过滤掉提供不可靠答案的欺诈者尤其重要。先前的研究[40]表明欺诈者的比例甚至达到 40%，给高质量的答案聚合带来了极大挑战。鉴于此，本文主要采用 2 种欺诈者添加方式，分别将{10%, 20%, 30%, 40%}的欺诈者添加到原始标注者中，并报告了不同欺诈者比例下的各种对比方法的结果。其中，第一种添加方式是每个欺诈者为每个样本随机分配一个标签，第二种则是每个欺诈者为所有样本随机分配一个标签，对比结果分别见图 7 和 8 所示。

从图中不难发现，无论是哪种添加方式，随着欺诈者比例的增加，所有的聚合方法精度都降低了。原因是更多的欺诈者意味着更多的噪声标注，这甚至可能超过正确的标注，从而给答案聚合带来了极大的困难。MV 对欺诈者最敏感，因为它假设所有的标注者(包括欺诈者)提供的标注质量相同，而忽略了标签关联性。虽然 RAkEL-GLAD、MLCC 等多标签答案聚合方法均考虑了标签关联性，但 AGR-JMF 仍然明显优于它们，尤其当添加的欺诈者比例达到 40%时，AGR-JMF 仍然保持 85%以上的准确率。主要原因可归结为三点：1) 与现有的聚合不同，AGR-JMF 首先对原始标注数据进行了去噪，排除了大量噪声和离群值，保证了标签间相似矩阵、低秩矩阵分解的质量；2) 以往聚合方法都是基于原始标注数据来计算标签的关联矩阵，本文则是对纯净矩阵采用自适应图正则化方法来获取；3) 本文联合低秩矩阵分解有选择地整合标注者的纯净标注矩阵，并通过给欺诈者分配较低(或为零)的权重来显式地减少欺诈者的影响。

综上所述，AGR-JMF 对众包标注结果潜在的欺诈者识别具有鲁棒性。

3.6　实用性分析

为了分析 AGR-JMF 在敦煌壁画数据集构建中的实用性，本文考察了标注量对多答案聚合方法的影响。对收集到的莫高窟壁画数据集的标注数据进行比例为[0.1, 1.0]的随机采样，采样间隔为 0.1，并与已有的多标签答案方法进行性能比较。为了缓解随机采样对各算法的性能影响，在每种情形下均进行 10 次重复实验，并记录其均值和标准差，结果如图 9 所示。图中结果显示，随着标注比例的增大，各聚合方法精度均呈上升趋势。特别地，当标注比例为 80%时，AGR-JMF 已经达到了 85%以上的准确率，各性能指标明显优于其他方法。同时，当标注比例小于 30%时，即标注数量较少时，对比方法与 AGR-JMF 之间的差距明显增大，这说明为了达到同样的聚合效果，AGR-JMF 需要更少数量的标注，一定程度上说明了 AGR-JMF 对稀疏标注更为鲁棒。究其原因，AGR-JMF 在纯净数据上更准确地估计了标签间的相似性，同时考虑了标注人员的标注质量和相似性，保证了低秩矩阵分解从整体上逼近标注数据的整体结构信息。

此外，以上实验对敦煌壁画数据集构建具有指导性意义。正如第3.1 部分所提到的，敦煌壁画数据集中每张图像平均被标注了 7.4 次。为了达到 85%以上的准确率，根据此实验的分析，审核任务分配时至少要保证每张图像被 6 人次标注，这也给未来任务分配策略和主动众包标注研究提供了有力的数据支撑。

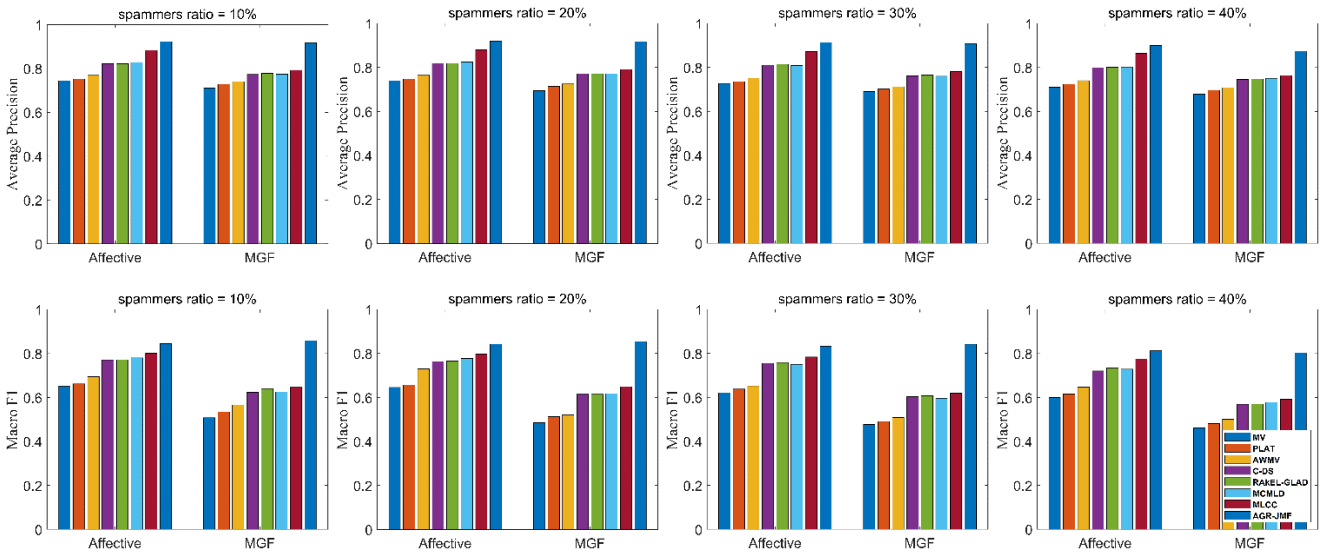

图 7　Affective 和 MGF 数据集上第一种欺诈者添加方式下不同欺诈者比例的 Average precision 和 Macro F1

Fig. 7　Average precision and Macro F1 under different ratios of the first type of spammers on Affective and MGF

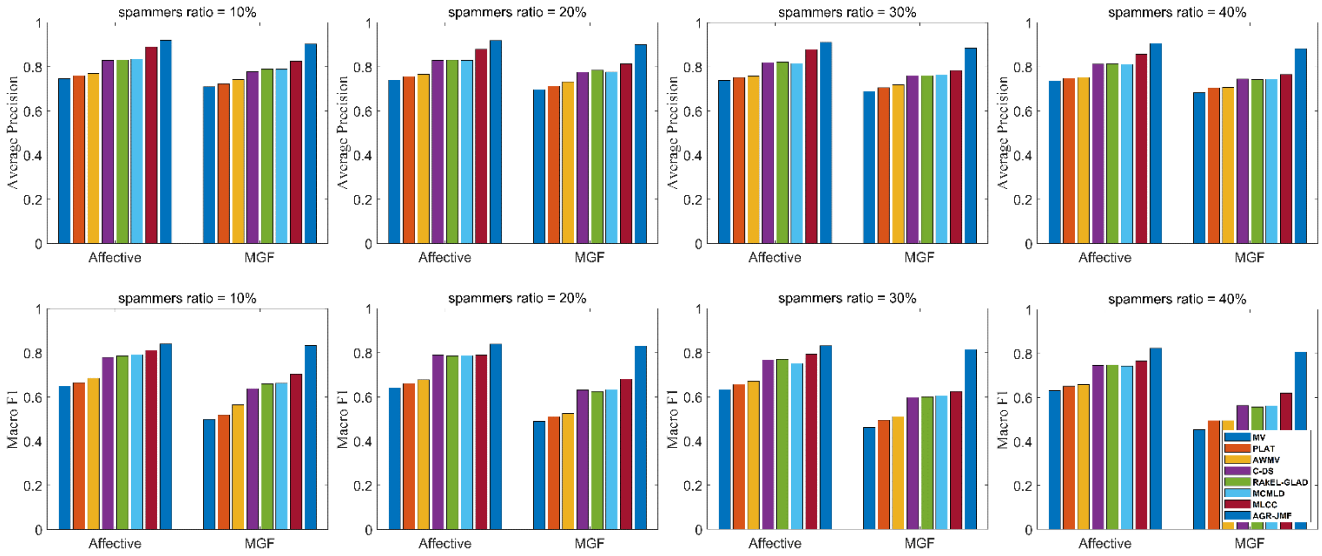

图 8　Affective 和 MGF 数据集上第二种欺诈者添加方式下不同欺诈者比例的 Average precision 和 Macro F1

Fig. 8　Average precision and Macro F1 under different ratios of the second type of spammers on Affective and MGF

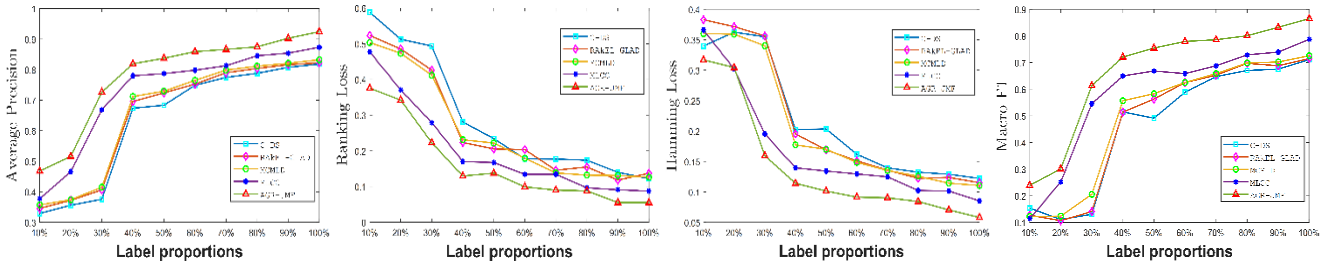

图 9　MGF 数据集上不同比例标签下 4 种常用多标签性能评价指标结果

Fig. 9　Four commonly used evaluation metrics on MGF dataset under different proportions of labels

3.7　复杂性分析

由于式(8)的 $\Phi(\boldsymbol{D},\boldsymbol{N},\boldsymbol{U},\boldsymbol{S},\boldsymbol{V},\boldsymbol{\mu},\boldsymbol{C})$ 关于每个优化变量是局部凸的，因此本文提出的交替迭代算法可以保证每个子问题在迭代过程中能量逐渐下降。但由于各未知量耦合在一起，且有等式及不等式约束，因此很难给出算法的全局收敛性证明。鉴于此，本文绘制了 $\Phi(\boldsymbol{D},\boldsymbol{N},\boldsymbol{U},\boldsymbol{S},\boldsymbol{V},\boldsymbol{\mu},\boldsymbol{C})$ 在 Affective 和 MGF 数据集上的能量函数值，如图 10 所示。从图中可以看出，在算法迭代优化初期能量函数很快下降，并随着优化过程逐渐稳定。实验发现，在其他数据集上也有类似的能量下降走势。

由于各对比方法采用不同的语言实现，直接比较其运行时间是没有意义的。本小节中，假设 t 为迭代次数，分别给出 5 种多标签答案聚合方法的理论计算复杂度。C-DS 计算源标签和目标标签的联合分布需要 $O(mnl^2+ml^3)$，计算每个样本中每个标签的概率需要 $O(l^3)$，因此总计算复杂度为 $O(mnl^2t+ml^3t)$。RAkEL-GLAD 创建每个标签的幂集需要 $O(mnl)$，每个标签的平均可能性需要 $O(2^k mnM)$（k 为标签子集中候选标签数量，M 为随机标记子集数），因此总计算复杂

度为 $O(mnlt+2^k mnMt)$。MCMLD 计算协方差矩阵的特征值需要 $O(mnl)$，EM 迭代中的 E-step 和 M-step 分别需要 $O(ml^2)$ 和 $O(2l^2mn+nR)$（R 为聚类数量），因此总计算复杂度为 $O(mnl+ml^2t+2l^2mnt+nRt)$。MLCC 每步迭代更新 $\boldsymbol{V}$，$\boldsymbol{U}_w$ 和 $\boldsymbol{S}$ 分别需要 $O(mnk^2)$，$O(nlk)$ 和 $O(mnlk)$，以及更新 $\boldsymbol{\mu}$ 需要 $O(m)$，因此总计算复杂度为 $O(tmnk^2+tmnlk+tm)$。AGR-JMF 每步迭代更新 $\boldsymbol{D}$，$\boldsymbol{N}$ 及 $\boldsymbol{\mu}$ 分别需要 $O(nk^2)$，$O(mnl)$ 和 $O(m)$，更新 $\boldsymbol{V}$，$\boldsymbol{U}_w$ 和 $\boldsymbol{S}$ 分别需要 $O(mnk^2)$，$O(nlk)$ 和 $O(mnlk)$，以及更新 $\boldsymbol{C}$ 需要 $O(m^2)$，因此总计算复杂度为 $O(tmnk^2+tmnlk+tm^2)$。由于三种单标签答案聚合方法(MV、PLAT 和 AWMV)单独聚合每个标签答案，所以计算复杂度低于多标签方法。由于 $k<\{n,l\}$，AGR-JMF 复杂度低于 RAkEL-GLAD 和 C-DS，但是高于 MLCC 和 MCMLD。

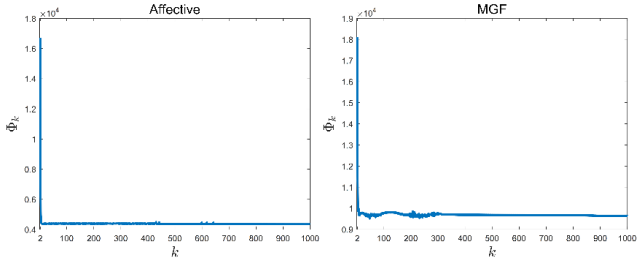

图 10　AGR-JMF 在 Affective 和 MGF 数据集上的收敛曲线

Fig. 10　Convergence curve of AGR-JMF on Affective and MGF datasets

此外，AGR-JMF 将大大提高文化遗产数据集标注审核的效率。例如，给定 1000 个样本，每个样本标注完成后平均需要 7 个审核人员审核，采用本算法在几分钟之内即可完成自动审核。如果采用人工验收审核结果，则至少需要 1000*7*3 mins ≈14.6 days。当数字文化遗产数据集样本数量庞大时，自动审核算法的优势会更加明显。

4　结束语

考虑到数字文化遗产领域采用专家标注昂贵而稀缺，本文将多标签任务分配给多个容易访问的非专家收集标注信息，并从含有大量噪声的标注中估计样本的真实标签。以往的单标签答案聚合方法忽视了多标签任务的标签关联性，而多标签聚合方法直接从标注数据中估计标签关联性，很敏感地受噪声和离群值的影响。针对以上问题，本文提出了一种鲁棒的多标签答案聚合方法 AGR-JMF。通过 L_1 正则项优化去除原始标记数据中的噪声，同时基于该去噪数据自适应估计标签间的关联矩阵，并结合标注人员的标注质量、标注行为属性相似性来指导低秩矩阵分解，进而实现高质量的多标签答案聚合。在 6 个真实数据集和莫高窟壁画数据集上都验证了 AGR-JMF 的合理性与有效性。此外，AGR-JMF 已经在敦煌数字文化遗产数据集构建过程中得到了实际应用，大大提高了数字文化遗产数据集的审核效率。

实验结果表明，AGR-JMF 在准确率、噪声鲁棒性方面显示出明显的优越性和先进性，但仍有局限性。例如，算法依赖于 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的初始化，本文采用的随机初始化不一定是最优的初始化。AGR-JMF 需要输入 α，β 等 7 个正则化参数，这些参数基于实验经验来设置，如何自动确定每个参数的最佳值也值得进一步深入研究。此外，在未来的工作中，可以进一步考虑最优化任务请求者的成本预算、自适应任务分配方案等更多众包标注属性。

参考文献：

[1] Kovashka A, Russakovsky O, Li Feifei, *et al.* Crowdsourcing in computer vision [J]. CoRR, 2016, abs/1611. 02145 (3) .

[2] Meng Rui, Tong Yongxin, Chen Lei, *et al.* CrowdTC: Crowdsourced taxonomy construction [C]// Proc of IEEE International Conference on Data Mining. Piscataway, NJ: IEEE Press, 2015: 913-918.

[3] Zhang Minling, Zhou Zhihua. A review on multi-label learning algorithms [J]. IEEE transactions on knowledge and data engineering, 2013, 26 (8): 1819-1837.

[4] Gibaja E, Ventura S. A tutorial on multilabel learning [J]. ACM Computing Surveys, 2015, 47 (3): 1-38.

[5] Raykar V C, Yu Shiping, Zhao Linda, *et al.* Learning from crowds [J]. Journal of Machine Learning Research, 2010, 11 (4): 1297-1322.

[6] Welinder P, Branson S, Belongie S, *et al.* The multidimensional wisdom of crowds [J]. Advances in Neural Information Processing Systems, 2010, 23: 2424-2432.

[7] Bragg J, Mausam, Weld D S. Crowdsourcing multi-label classification for taxonomy creation [C]// Proc of the 1st AAAI Conference on Human Computation and Crowdsourcing. 2013: 25-33.

[8] Zhang Jing, Wu Xindong, Sheng V S. Imbalanced multiple noisy labeling [J]. IEEE Trans on Knowledge and Data Engineering, 2014, 27 (2): 489-503.

[9] Zhang Jing, Sheng V S, Li Qianmu, *et al.* Consensus algorithms for biased labeling in crowdsourcing [J]. Information Sciences, 2017, 382: 254-273.

[10] Sun Yuyin, Singla A, Fox D, *et al.* Building hierarchies of concepts via crowdsourcing [C]// Proc of the 24th International Joint Conference on Artificial Intelligence. 2015: 844-851.

[11] Duan Lei, Oyama S, Sato H, *et al.* Separate or joint?Estimation of multiple labels from crowdsourced annotations [J]. Expert Systems with Applications, 2014, 41 (13): 5723-5732.

[12] Tam N T, Viet H H, Hung N Q V, *et al.* Multi-label answer aggregation for crowdsourcing [J]. Technique report, 2016: 1-13.

[13] Tu Jinzheng, Yu Guoxian, Domeniconi C, *et al.* Multi-label crowd consensus via joint matrix factorization [J]. Knowledge and Information Systems, 2020, 62 (4): 1341-1369.

[14] 李绍园，姜远. 多标记众包学习 [J]. 软件学报, 2020, 31 (05): 1497-1510. (Li Shaoyuan, Jiang Yuan. Multi-label crowdsourcing learning. Ruan Jian Xue Bao/Journal of Software, 2020, 31 (5): 1497-1510.)

[15] Lee J, Cho H, Park J W, *et al.* Hybrid entity clustering using crowds and data [J]. The VLDB Journal, 2013, 22 (5): 711-726.

[16] Zhang Jing, Wu Xindong, Sheng V S. Learning from crowdsourced labeled data: a survey [J]. Artificial Intelligence Review, 2016, 46 (4): 543-576.

[17] Dawid A P, Skene A M. Maximum likelihood estimation of observer error-rates using the EM algorithm [J]. Journal of the Royal Statistical Society: Series C (Applied Statistics) , 1979, 28 (1): 20-28.

[18] Whitehill J, Wu Tingfa, Bergsma J, *et al.* Whose vote should count more: Optimal integration of labels from labelers of unknown expertise [J]. Advances in Neural Information Processing Systems, 2009, 22: 2035–2043.

[19] Demartini G, Difallah D E, Cudre-Mauroux P. Zencrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking [C]// Proc of the 21st International Conference on World Wide Web. 2012: 469-478.

[20] Kurve A, Miller D J, Kesidis G. Multicategory crowdsourcing accounting for variable task difficulty, worker skill, and worker intention

[J]. IEEE Trans on Knowledge and Data Engineering, 2014, 27 (3): 794-809.

[21] Liu Qiang, Peng Jian, Ihler A T. Variational inference for crowdsourcing [J]. Advances in neural information processing systems, 2012, 25: 692-700.

[22] Zhou Dengyong, Basu S, Mao Yi, *et al.* Learning from the wisdom of crowds by minimax entropy [J]. Advances in Neural Information Processing Systems, 2012, 25: 1-9.

[23] Ma Fenglong, Li Yaliang, Li Qi, *et al.* Faitcrowd: Fine grained truth discovery for crowdsourced data aggregation [C]// Proc of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015: 745-754.

[24] Li Qi, Li Yaliang, Gao Jing, *et al.* A confidence-aware approach for truth discovery on long-tail data [J]. Proceedings of the VLDB Endowment, 2014, 8 (4): 425-436.

[25] Zhang Jing, Sheng V S, Wu Jian, *et al.* Multi-class ground truth inference in crowdsourcing with clustering [J]. IEEE Trans on Knowledge and Data Engineering, 2015, 28 (4): 1080-1085.

[26] Zhang Jing, Sheng V S, Li Tao. Label aggregation for crowdsourcing with bi-layer clustering [C]// Proc of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2017: 921-924.

[27] Rodrigues F, Pereira F C. Deep learning from crowds [C]// Proc of the 32th AAAI Conference on Artificial Intelligence. 2018, 32 (1): 1611-1618.

[28] Atarashi K, Oyama S, Kurihara M. Semi-supervised learning from crowds using deep generative models [C]// Proc of the 32th AAAI Conference on Artificial Intelligence. 2018, 32 (1): 1555–1562.

[29] Nowak S, Ryuger S. How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation [C]// Proc of International Conference on Multimedia Information Retrieval. 2010: 557-566.

[30] Duan Lei, Oyama S, Kurihara M, *et al.* Crowdsourced semantic matching of multi-label annotations [C]// Proc of the 24th International Joint Conference on Artificial Intelligence. 2015: 3483-3489.

[31] Yoshimura K, Baba Y, Kashima H. Quality control for crowdsourced multi-label classification using RAkEL [C]// Proc of International Conference on Neural Information Processing. 2017: 64-73.

[32] Tsoumakas G, Katakis I, Vlahavas I. Random k-labelsets for multilabel classification [J]. IEEE Trans on Knowledge and Data Engineering, 2010, 23 (7): 1079-1089.

[33] Hung N Q V, Viet H H, Tam N T, *et al.* Computing crowd consensus with partial agreement [J]. IEEE Trans on Knowledge and Data Engineering, 2017, 30 (1): 1-14.

[34] Zhang Jing, Wu Xindong. Multi-label inference for crowdsourcing [C]// Proc of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018: 2738-2747.

[35] 史加荣, 郑秀云, 魏宗田, 等. 低秩矩阵恢复算法综述 [J]. 计算机应用研究, 2013, 30 (06): 1601-1605. (Shi Jiarong, Zheng Xiuyun, Wei Zongtian, *et al.* Survey on algorithms of low-rank matrix recovery. Application Research of Computers, 2013, 30 (06): 1601-1605.)

[36] Kang Zhao, Pan Haiqi, Hoi S C H, *et al.* Robust graph learning from noisy data [J]. IEEE transactions on cybernetics, 2019, 50 (5): 1833-1843.

[37] 于进, 钱锋. 基于粒子群优化的高斯核函数聚类算法 [J]. 计算机工程, 2010, 36 (14): 22-28. (Yu Jin and Qian Feng. Gauss kernel function clustering algorithm based on particle swarm optimization. Computer Engineering, 2010, 36 (14): 22-28.)

[38] He Xiaofei, Niyogi P. Locality preserving projections [J]. Advances in Neural Information Processing Systems, 2003, 16: 153–160.

[39] Snow R, O'connor B, Jurafsky D, *et al.* Cheap and fast–but is it good?evaluating non-expert annotations for natural language tasks [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2008: 254-263.

[40] Vuurens J, de Vries A P, Eickhoff C. How much spam can you take?an analysis of crowdsourcing results to increase accuracy [C]// Proc of ACM SIGIR Workshop on Crowdsourcing for Information Retrieval. 2011: 21-26.