

# 多标记众包学习\*

李绍园, 姜远



(计算机软件新技术国家重点实验室(南京大学), 江苏 南京 210023)

通讯作者: 姜远, E-mail: jiangy@lamda.nju.edu.cn

**摘要:** 传统的多标记学习任务要求训练数据拥有完整的或者至少部分的真实标记, 而真实标记耗费昂贵并且难以获取. 不同于由昂贵受限的专家标注真实标记, 众包环境下, 多标记任务被分配给多个容易获取的非专家标注, 学习目标是先从错误的非专家标注中估计样本的真实标记. 这一问题的关键在于如何融合非专家标注. 以往的众包学习主要集中在单标记任务上, 忽视了多标记任务的标记相关性; 而多标记任务上的众包工作集中在局部标记相关性的利用如标记共同出现的概率, 标记间条件相关性, 其估计很敏感地受到标记数量和质的影响. 考虑到多标记任务上多个标注者的标注结果整体上存在低秩结构关系, 提出一种基于低秩张量矫正的方法. 首先, 将标注结果组织成三维的张量(样本, 标记, 标注者), 用低秩张量补全的方法对收集到的标注做预处理, 以同时达到两个目的: (1) 优化已有标注; (2) 补全标注者在其未标注的标记上的标注结果. 然后, 对所有标注融合, 测试了 3 种融合方法, 分别从不同的方面考虑标注的置信度. 真实数据上的实验结果验证了所提方法的有效性.

**关键词:** 多标记学习; 众包; 低秩; 张量近似; 融合

**中图分类号:** TP181

中文引用格式: 李绍园, 姜远. 多标记众包学习. 软件学报, 2020, 31(5): 1497–1510. <http://www.jos.org.cn/1000-9825/5673.htm>

英文引用格式: Li SY, Jiang Y. Multi-label crowdsourcing learning. Ruan Jian Xue Bao/Journal of Software, 2020, 31(5): 1497–1510 (in Chinese). <http://www.jos.org.cn/1000-9825/5673.htm>

## Multi-label Crowdsourcing Learning

LI Shao-Yuan, JIANG Yuan

(State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210023, China)

**Abstract:** Previous multi-label learning requires that all or at least a subset of ground truth labels is given for the training example. This study investigates how to utilize the wisdom of crowds for multi-label tasks, where rather than high cost ground truth labels, imperfect annotations from crowds are collected for learning. The target is to infer the instances' ground truth labels. The key challenge lies in how to aggregate the noisy annotations. Different from previous crowdsourcing works on single-label problems which ignore the correlation between labels, and multi-label works which consider local label correlations whose estimation heavily depends on the annotations' quality and quantity, this study proposes an approach considering the global low rank structure of the crowds' annotations. Regarding the crowds' annotations for multi-label tasks as a three-way tensor (instance, label, worker), the crowds' annotations are firstly preconditioned using low rank tensor completion, such that it is able to simultaneously correct the observed noisy annotations and at the same time predict the unobserved annotations. After that, the preconditioned annotations are aggregated through some heuristic methods. Three aggregation methods taking into account the crowds' annotation confidence from different perspectives are tested. Experimental results on real world multi-label crowdsourcing data sets demonstrate the superiority of the proposed approach.

**Key words:** multi-label learning; crowdsourcing; low rank; tensor approximation; aggregation

\* 基金项目: 国家自然科学基金(61673201)

Foundation item: National Natural Science Foundation of China (61673201)

收稿时间: 2017-12-29; 修改时间: 2018-05-16; 采用时间: 2018-09-17

传统的单标记学习假设每个样本只与单个类别相关,而在真实世界的任务中,一个样本可以同时与多个概念类别相关.例如,在图片分类中,一张图片中可以有多个对象,如城市、道路<sup>[1]</sup>;在生物信息学中,每个基因序列有多种功能,如新陈代谢、转录、蛋白质合成<sup>[2]</sup>.为了处理这样的任务,多标记学习成为机器学习领域的重要研究方向,并被广泛应用于各个领域<sup>[3,4]</sup>.

经典的多标记学习方法要求训练样本拥有完整或者至少部分的真实标记,对多标记任务而言,真实标记是代价昂贵且不容易获取的资源.一方面,标注真实标记需要领域专家知识,而专家资源是昂贵且有限的;另一方面,标注多标记样本需要逐个检查候选标记是否与当前样本相关,当标记数量比较多时,标记正确的类别非常耗时耗力.随着众包平台比如 AMT(Amazon mechanical turk),Crowdsflower 的出现,众包提供了一种新的更容易获取标记的方式.不同于要求领域专家标注,众包将标注任务分配给容易访问的非专家,收集大量的监督信息<sup>[5,6]</sup>.由于非专家的标注可能犯错,为了获得高质量的标记,众包通常将任务分配给多个标注者,然后融合多个标注结果以估计样本的真实标记.

本文考虑众包环境下的多标记学习,不同于由代价昂贵的专家标注真实标记,任务被分配给多个容易获取的非专家标注,学习目标是从可能犯错的非专家标注中估计样本的真实标记.这一问题的关键在于如何融合非专家标注.以往的众包相关工作主要集中在单标记问题上<sup>[7-15]</sup>,将多标记任务分解成多个单标记任务,单标记众包学习方法可以直接应用.但是这种方法忽略了标记的相关性,而合理地利用标记关系有助于提高学习效果.多标记学习中众包的研究工作还很少,文献[16-18]考虑通过众包方式收集标记及标记间关系,估计多个标记间层次结构关系的问题;文献[19,20]则分别考虑从标注中估计标记共同出现的概率及标记间条件相关性,以恢复样本真实标记,其问题在于这两种局部标记相关性的估计结果很敏感地受到标注数量和质量的影响.

本文则从张量的概念出发,考虑标注的全局结构关系,使用张量低秩近似的工具来解决多标记众包学习问题.考虑到以下两点:1) 对于多标记任务来说,真实标记矩阵可能存在低秩结构;2) 在标注者好意标注的前提下,多个标注者标注同样的任务,其标注结果会互相接近.因此,多个非专家的标注结果整体上存在低秩的结构关系.同时考虑真实标记间和多个标注者标注结果间的低秩结构,我们提出一种基于低秩张量近似的思路.首先,我们将标注结果组织成三维的张量(样本,标记,标注者),用低秩张量补全的方法对收集到的标注做预处理,以同时达到两个目的:1) 优化已有的标注;2) 补全标注者在其未标注的样本上的标注结果.之后,我们对优化补全之后的标注融合,测试了 3 种不同的融合方法.在真实多标记众包数据集上,在不同的标注数目情形下,与多个对比方法的实验结果表明,通过考虑全局的低秩结构信息,本文的方法在多种评价指标上都显著优于对比方法.

本文第 1 节介绍多标记众包学习的相关工作.第 2 节提出本文的方法.第 3 节给出实验结果.最后总结全文.

## 1 相关工作

本节中,我们简要介绍多标记众包学习的相关工作,主要包括多标记学习、众包学习、低秩张量近似.

### • 多标记学习

过去的 10 多年,很多方法被提出来,以处理多标记学习问题.最直接的策略是将多标记任务分解为多个独立的二分类问题,然后应用单标记的学习方法<sup>[21]</sup>.这个策略忽略了显式或隐含的标记间关系,而标记间关系被广泛认为有助于提升学习效果.在文本分类任务中,标签通常按层次结构组织,这启发了多标记学习中利用标记间外部结构的方法,如文献[22]考虑标记间树形结构,文献[23]考虑标记间树形和有向无环图结构.除了显式的标记结构之外,也有探索利用隐含标记关系的工作,如标记排序的思路<sup>[2,24]</sup>,将多标记任务转化为标记排序问题,要求各个样本的正标记排在负标记之前;样本表示增强的思路<sup>[25,26]</sup>,使用标记信息构造样本的元特征,以使样本的表示融合隐式的标记间关系.经典的多标记学习方法要求训练样本具有真实标记,在实际应用中真实标记是昂贵不易获取的资源.

### • 众包学习

随着众包平台如 AMT(Amazon mechanical turk),Crowdsflower 等的出现,众包成为一种获取监督信息的便捷收集方式.由于众包标注来自可能犯错的非专家,因此从标注中估计真实标记成为众包学习的重要关注点.以

往的工作主要集中在单标记任务上,通过对标注者的标注能力建模,在用户标注和真实标记间建立关联,通过概率图模型的方法同时估计标注者的标注能力和真实标记<sup>[8-12,14,15]</sup>.多标记任务方面,文献[13,18]收集样本所有可能的标记,利用标记共同出现的概率以估计标记层次结构关系;文献[16,17]通过询问标记间父子关系重构标记层次结构.至于估计样本真实标记,相关工作主要通过扩展单标记众包学习方法,加入标记间关系的利用.文献[18]提出一种考虑标记共同出现的概率关系估计真实标记的方法,文献[19]探索了3种标记关系,通过将标记分组,或者学习标记间的条件相关性将多标记问题转化为多分类问题,以估计样本真实标记.类似于文献[18],文献[20]也提出了一种利用标记共同出现概率关系的方法.以上工作的标记关系主要集中在标记间局部关系,其估计质量很敏感地受到标注的数量和质量的影响.本文则从标注整体上存在低秩结构关系来考虑标记相关性.

#### • 低秩张量近似

在数据科学中,矩阵是最常见的数据组织形式,作为矩阵的重要概念,矩阵的秩刻画了数据的一种全局相似性结构信息.当数据集合可以由其子集线性组合重构时,矩阵被称为具有低秩结构.文献[27]从理论上证明,通过将矩阵的数据恢复问题形式化为最小化秩的优化目标,只需要很少的观测值,低秩矩阵就可以被高概率地恢复.这个工作启发了一系列利用矩阵低秩结构做信息重构/数据近似的工作,包括各种低秩矩阵补全的算法类工作<sup>[28,29]</sup>及推荐系统、图像补全等方面的应用工作<sup>[30,31]</sup>.但在很多实际应用中,由于数据收集方式的丰富、数据维度的增长,使得张量,也就是多维的矩阵,成为更自然的数据表示.相比于矩阵,张量能够保持数据多个维度上更丰富的自然结构信息.尽管矩阵低秩近似的工作很多,但利用张量的整体结构信息来帮助数据近似并不那么直接,主要原因在于张量的低秩结构难以直观刻画.为了解决这一问题,文献[32-34]提出使用张量各个维度展开的矩阵的低秩来近似表达张量的低秩结构,通过最小化矩阵的秩做张量的低秩近似/补全.此外,基于张量分解的思路,比如 Tucker 张量分解<sup>[35]</sup>、Parafac 张量分解<sup>[36]</sup>、展开矩阵分解<sup>[37]</sup>的方式也被尝试用来做张量的近似/补全.到目前为止,关于张量近似/补全的方法的理论保证仍然是开放式问题,文献[34]中的实验表明,其所提的 LRTC 方法表现更为稳定.我们注意到,文献[38]将张量补全的方法用于单标记分类任务,通过在张量上增加一层真实标记层,首先通过先验知识估计真实标记的初值,然后利用张量补全的方法补全真实标记层.这个问题在于,真实标记的质量依赖于估计初值,而估计初值需要从额外的信息中获取.

本文将多标记任务的所有标注者的标注结果组织成一个三维张量(样本,标记,标注者),考虑到样本的真实标记存在低秩结构,且多个标注者对同一任务的标注结果相近,采用文献[34]的想法对张量各维度的展开矩阵做低秩约束,利用低秩张量近似/补全的方法补全标注张量,然后使用3种标注融合方法估计样本的真实标记.

## 2 基于低秩张量矫正的多标记众包方法

### 2.1 符号表示

我们首先说明本文出现的符号表示含义.我们用大写字母表示矩阵,如  $X$ ;用小写字母表示矩阵中的元素,如  $x_{ij}$ .矩阵  $X$  的 Frobenius 范数被定义为  $\|X\|_F := (\sum_{ij} x_{ij}^2)^{\frac{1}{2}}$ , 迹范数被定义为矩阵的所有奇异值的和,即  $\|X\|_* = \sum_i \sigma_i(X)$ , 其中,  $\sigma_i(X)$  表示矩阵  $X$  的第  $i$  个最大的奇异值.张量用手写体字母表示,如  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_n}$  表示  $n$  维的张量,其元素表示为  $x_{i_1, i_2, \dots, i_n}$ ,  $1 \leq i_k \leq I_k$ . 可以通过展开操作(unfold)将张量转换成矩阵,比如,张量  $\mathcal{X}$  在第  $k$  维的展开被定义为  $unfold_k(\mathcal{X}) := \mathcal{X}_{(k)} \in \mathbb{R}^{I_k \times (I_1 \dots I_{k-1} I_{k+1} \dots I_n)}$ , 其反操作 fold 被定义为  $fold(\mathcal{X}_{(k)}) := \mathcal{X}$ . 张量的 Frobenius 范数被定义为  $\|\mathcal{X}\|_F := (\sum_{i_1, i_2, \dots, i_n} |x_{i_1, i_2, \dots, i_n}|^2)^{\frac{1}{2}}$ .

### 2.2 问题形式化

在多标记众包问题中,给定  $N$  个样本、 $L$  个候选标记、 $M$  个标注者,每个标注者的标注结果可以表示成  $N \times L$  的矩阵  $X \in \{+1, -1, 0\}^{N \times L}$ , 其中,  $x_{ij} = +1(-1)$  表示第  $i$  个样本的第  $j$  个标记被标注为正(负)标记,  $x_{ij} = 0$  表示第  $i$  个样本的第  $j$  个标记没有被标注,即标注缺失.在众包标记收集方式下,标注结果通常都不是完整的,即不是所有的标注者都标注了所有样本的所有标记.我们的学习目标是给定收集的标注结果,估计样本的真实标记矩阵  $Z$ , 其中,

$z_{ij}=0$  表示第  $i$  个样本在  $j$  个标记上的结果.

使用张量的数据组织形式,将  $M$  个标注者的标注结果叠起来,我们得到  $N \times N \times L$  的三维标注张量  $\chi \in \{+1, -1, 0\}^{M \times N \times L}$ , 其中,  $\chi_{tij}$  表示第  $t$  个标注者在第  $i$  个样本第  $j$  个标记上的标注. 由于标注结果来自非专家,与真实标记相比,  $\chi$  中一方面有缺失的标注,另一方面含有部分的错误标记. 若要估计样本的真实标记,则需要对标注结果做矫正. 考虑到如下的因素: 1) 相似样本的标记也应该相似,因此样本集的真实标记矩阵  $Z$  可能存在低秩结构; 2) 由于多个标注者标注的是同一任务,在标注者质量可靠的情况下,多个标注者的标注结果也是相似的,因此,标注张量  $\chi$  整体上应该存在低秩的结构关系. 我们提出了一种基于低秩张量补全的思路. 首先,用低秩张量补全的方法对收集到的标注做预处理,以同时达到两个目的: 1) 矫正已有的可能会犯错的标注; 2) 补全各标注者缺失的标注结果. 之后,我们对矫正补全之后的标注做融合,尝试了 3 种标注融合方法,最终估计样本的真实标注.

### 2.3 低秩张量矫正

我们首先介绍矩阵的低秩补全,然后基于此给出张量低秩近似的实现. 在低秩矩阵补全问题中,最常用的方法是最小化矩阵秩. 由于矩阵的秩最小化是非凸问题,文献[27]证明了矩阵的迹范数是矩阵秩的最紧的凸闭包,放松到凸问题,矩阵补全被形式化为最小化矩阵的迹范数:

$$\min_X \|X\|_* \text{ s.t. } X_\Omega = A_\Omega \quad (1)$$

其中,  $\|X\|_*$  表示目标矩阵  $X$  的迹范数,  $\Omega$  表示观测到的元素的下标集合,  $A$  表示观测矩阵.

作为矩阵的高维扩展,张量的低秩结构不像矩阵可以写出直观的优化目标,通常的实现都是通过启发式的优化目标设计,实现对张量低秩的约束. 到目前为止,关于张量补全的各种算法还缺乏类似矩阵低秩恢复的理论保证,即以某种优化目标,在何种观测样本复杂度下能够以较高的概率恢复真实的张量. 考虑到文献[34]中所提方法 SiLRTC (simple low rank tensor completion) 简单易实现,并且其文中的多种实验对比也表明 SiLRTC 方法表现更为稳定,本文中我们采用 SiLRTC 方法作为多标记众包标记标注矫正/补全的方法.

以下简单介绍一下 SiLRTC 方法. 文献[34]定义张量  $\chi$  的迹范数为

$$\|\chi\|_* = \sum_{l=1, \dots, n} \alpha_l \|\chi_{(l)}\|_* \text{ s.t. } \sum_{l=1, \dots, n} \alpha_l = 1, \alpha_l \geq 0, l = 1, \dots, n \quad (2)$$

这里,  $\alpha_l$  为预先设定的参数. 类似矩阵补全的优化目标,张量补全的目标可以被形式化为

$$\min_{\chi} \sum_{l=1, \dots, n} \alpha_l \|\chi_{(l)}\|_* \text{ s.t. } \chi_\Omega = \mathcal{T}_\Omega \quad (3)$$

由于各个维度展开的矩阵  $\chi_{(l)}$  并不互相独立,为了能够分别独立地求解  $\chi_{(l)}$ , 同样数目的矩阵  $M_l$  被引入到优化问题中,上述的优化问题被转化为如下的形式.

$$\min_{\chi, M_l} \sum_{l=1, \dots, n} \alpha_l \|M_l\|_* + \frac{\beta_l}{2} \|\chi_{(l)} - M_l\|_F^2 \text{ s.t. } \chi_\Omega = \mathcal{T}_\Omega \quad (4)$$

将优化变量分解成  $\chi, M_1, M_2, \dots, M_n$  共  $n+1$  个组,可以采用分块坐标下降法求解上述优化问题,优化其中一组变量时固定其他变量.

- 优化  $\chi$

固定  $M_1, M_2, \dots, M_n$  为已知变量,关于  $\chi$  的优化问题为

$$\min_{\chi} \sum_{l=1, \dots, n} \frac{\beta_l}{2} \|\chi_{(l)} - M_l\|_F^2 \text{ s.t. } \chi_\Omega = \mathcal{T}_\Omega \quad (5)$$

因此,  $\chi$  存在闭式解:

$$\chi_{i_1, i_2, \dots, i_n} = \begin{cases} \left( \frac{\sum_i \beta_i \text{fold}_i(M_i)}{\sum_i \beta_i} \right)_{i_1, i_2, \dots, i_n}, & (i_1, i_2, \dots, i_n) \notin \Omega \\ \mathcal{T}_{i_1, i_2, \dots, i_n}, & (i_1, i_2, \dots, i_n) \in \Omega \end{cases} \quad (6)$$

- 优化  $M_i$

固定  $\chi, M_1, M_2, \dots, M_{i-1}, M_{i+1}, \dots, M_n$  为已知变量,关于  $M_i$  的优化问题为

$$\min_{M_i} \frac{\beta_i}{2} \|M_i - \chi_{(i)}\|_F^2 + \alpha_i \|M_i\|. \quad (7)$$

上述优化问题存在闭式解<sup>[31]</sup>.最优解的形式为  $M_i = D_i(\chi_{(i)})$ ,  $\tau = \alpha_i / \beta_i$ . 这里,  $D_i(X)$  表示对  $X$  的奇异值分解  $X = U \Sigma V^T$  的收缩算子, 其定义为  $D_i(X) = U \Sigma_i V^T$ , 其中,  $\Sigma_i = \text{diag}(\max(\sigma_i - \tau, 0))$ .

由于公式(4)为凸优化问题, 并且非光滑项是可以分别独立优化的, 由文献[39]可知, 使用分块坐标下降法可以保证收敛到最优解.

#### 2.4 3种标注融合方法

通过使用上面的低秩张量补全方法对标注张量做预处理, 利用标注整体上存在的低秩结构关系作为约束, 我们得到了处理之后的标注结果: 1) 在每个样本的每个标记上, 我们获得了所有标注者的可能标注值预测; 2) 相比原本收集的离散取值为  $\{+1, -1, 0\}$  (分别表示正标记、负标记、无标记) 的标注, 处理之后的标注为有符号的实值, 其范围在连续的实数范围内. 在本节中, 我们对处理之后的标注结果做融合, 尝试了 3 种标注融合方法来估计样本的真实标注.

将处理之后的标注张量表示为  $\tilde{\mathcal{X}}$ , 则有  $\tilde{\mathcal{X}} \in \mathbb{R}^{M \times N \times L}$ ,  $\tilde{\mathcal{X}}_{ij}$  表示第  $t$  个标注者在第  $i$  个样本第  $j$  个标记上的标注, 我们估计各个样本上各个标记的预测值  $\tilde{z}_{ij}$ .

##### 1) Signed Voting

不考虑标注的值信息, 只利用标注结果的符号信息. 在每个样本的每个标记上, 对其标注结果的符号做投票, 即

$$\tilde{z}_{ij} = \sum_t \text{sign}(\tilde{\mathcal{X}}_{ij}) / M \quad (8)$$

##### 2) Valued Voting

考虑到标注的值可以反映标注者的置信度信息, 比如对比两个标注者  $t, t'$ , 在同一个样本的同一标记上, 如果  $\tilde{\mathcal{X}}_{ij} = 0.9$ ,  $\tilde{\mathcal{X}}_{ij'} = -0.1$ , 则表示标记  $j$  更有可能是样本  $i$  的正标记. 因此, Valued Voting 加入对标注值的利用:

$$\tilde{z}_{ij} = \sum_t \tilde{\mathcal{X}}_{ij} / M \quad (9)$$

##### 3) Entropy Voting

借鉴文献[40]中的无监督集成方法, 在每个标记上, 对每个标注者估计一个基于信息熵的权重, 使用加权投票的方法估计  $\tilde{z}_{ij}$ . 根据信息论中的知识, 标注者在样本上的标注的概率分布的 entropy(熵)越小(大), 表示其标注内容的信息量越大(小), 标注者也就越可信(不可信). 因此, 使用如下的方法计算标注者权重.

首先, 对每个标记任务  $j$ , 对标注者  $t$ , 将其标注结果转化为  $[0, 1]$  范围的标注概率:

$$\tilde{P}_{ij} = (\tilde{\mathcal{X}}_{ij} - \min \tilde{\mathcal{X}}_{i, \dots}) / (\max \min \tilde{\mathcal{X}}_{i, \dots} - \min \tilde{\mathcal{X}}_{i, \dots}) \quad (10)$$

然后在每个标记  $j$  上计算标注者  $t$  的标注熵:

$$H_j^t = -(1/N) \sum_i \tilde{P}_{ij} \log(\tilde{P}_{ij}) + (1 - \tilde{P}_{ij}) \log(1 - \tilde{P}_{ij}) \quad (11)$$

则标记  $j$  上, 标注者  $t$  的系数为

$$c_j^t = \frac{1 - \kappa e^{H_j^t}}{\sum_i 1 - \kappa e^{H_j^i}} \quad (12)$$

最终标记的预测结果为

$$\tilde{z}_{ij} = \sum_t (c_j^t \tilde{\mathcal{X}}_{ij}) / M \quad (13)$$

结合张量补全, 分别使用这 3 种标记融合方法, 我们得到 3 种方法: TencSign、TencVal 和 TencEnt.

注意到本文算法的输出  $\tilde{z}_{ij}$  为实值的标记预测, 根据标记的值大小排序可以得到标记的排序结果. 以往的众包学习方法的输出也是实值的结果, 除了 MV 以外, 其他基于概率图模型的单标记或者多标记众包方法的输出均是各个类别为正标记的概率. 为了获得分类评价指标, 需要对预测值做阈值划分. 以往的概率输出的众包工作简单地以 0.5 作为阈值, 而 MV 以 0 作为阈值. 但在多标记学习中, 分类结果受划分阈值的影响很大. 在本文的实

验部分,为了获得公平的分类结果对比,在计算分类指标时,我们对各个方法在各个样本上的标记预测值做排序,将预测值最大的  $k$  个标记预测为正标记,其他预测为负标记,其中,  $k$  为该样本的真实正标记的个数.在这里,为了计算分类指标,我们用到了各个样本的真实标记数目信息,而这在实际应用中是不可行的.实际应用中,我们的方法和以往的众包学习方法可以用于标记的排序预测.而要估计样本确切的分类结果,可以借鉴多标记工作中的阈值学习方法,对小部分的样本收集其真实标记,使用文献[4]提到的阈值校准方法学习关于样本的正负标记阈值函数,从而可以得到分类预测结果.

本文算法的流程图如下所示.

**算法.** TencSign, TencVal, TencEnt.

输入:  $M$  个标注者在  $N$  个样本  $L$  个标记上的标注张量  $\chi \in \{+1, -1, 0\}^{M \times N \times L}$ , 参数  $\alpha = [1/3, 1/3, 1/3]$ ,  $\beta = [0.1, 0.1, 0.1]$ .

输出: 真实标记预测矩阵  $\tilde{Z} \in \mathbb{R}^{N \times L}$ , 其中  $\tilde{z}_{ij}$  表示第  $i$  个样本上第  $j$  个标记的预测值.

//基于低秩张量补全的标注矫正

(1)  $\varepsilon = 1e^{-5}$ ;

(2) While  $\|\tilde{\mathcal{X}} - \mathcal{X}\|_F^2 \geq \varepsilon$

(3)  $M_i = D_r(\chi_{(i)})$  for  $i=1, 2, \dots, n$

(4) 
$$\mathcal{X}_{i_1, i_2, \dots, i_n} = \begin{cases} \left( \frac{\sum_i \beta_i \text{fold}_i(M_i)}{\sum_i \beta_i} \right)_{i_1, i_2, \dots, i_n}, & (i_1, i_2, \dots, i_n) \notin \Omega \\ \mathcal{T}_{i_1, i_2, \dots, i_n}, & (i_1, i_2, \dots, i_n) \in \Omega \end{cases}$$

(5) End

//标注融合

(6) TencSign:  $\tilde{z}_{ij} = \sum_i \text{sign}(\tilde{\mathcal{X}}_{ij}) / M$  (公式(8))

(7) TencVal:  $\tilde{z}_{ij} = \sum_i \tilde{\mathcal{X}}_{ij} / M$  (公式(9))

(8) TencEnt:  $\tilde{z}_{ij} = \sum_i (c'_i \tilde{\mathcal{X}}_{ij}) / M$  (公式(13)), 其中,  $c'_i$  由公式(12)计算得到

### 3 实验

#### 3.1 实验数据

为了获得真实世界的多标记众包数据,我们在 AMT 平台上发布了样本和标记数目不同的两个图像标注任务,随机分配图片给标注者,要求标注者从候选标记集中给图片标记类别.这两项任务的详细信息如下.

- **Image1:** Image1 数据集包含 700 张图片,6 个类别{沙漠、沙滩、海、山、树、日出/日落}.平均每张图片有 1.24 个标记.共有 18 名标注者,每人标记的图片数目不少于 70 张.平均每个人标注了 267 张图片,每张图片被标注了 7 次.共收集到 28 878 个标注.
- **Image2:** Image2 数据集包含 1 695 张图片,16 个类别{沙漠、沙滩、海、山、树、日出/日落、船、花、瀑布、建筑、城市、花园、车、人、室内、天空}.平均每张图片有 1.80 个标记.共有 15 名标注者,每人标记的图片数目不少于 100 张.平均每个人标注了 397 张图片,每张图片被标注了 10 次.共收集到 96 640 个标注.

对于每个数据集,其真实标记由本文作者标注.为了对收集到的众包标记质量有所了解,我们做了一些初步的标记质量分析.对每个标注者的标注结果,我们对其标注的样本的标注和真实标记做对比,用 Macro F1 衡量其标注准确度. Macro F1 是一种常用的多标记分类性能评价指标,计算方式为每个类别上 F1 score 的平均,其取值范围为 [0,1]. Macro F1 越大,说明标注结果越接近真实标记.两个数据集上各个标注者的 Macro F1 结果统计如下: Image1 上,18 名标注者的 Macro F1 值分别为 {0.616, 0.6880, 0.758, 0.765, 0.781, 0.785, 0.796, 0.797, 0.801, 0.821, 0.826, 0.829, 0.829, 0.833, 0.835, 0.847, 0.851, 0.911}; 在 Image2 上,15 名标注者的 Macro F1 值分别是 {0.597, 0.705, 0.7530, 0.7550, 0.7570, 0.760, 0.761, 0.765, 0.777, 0.777, 0.780, 0.794, 0.779, 0.827, 0.847}. 由此我们可以看到,各个标注

者的 Macro  $F1$  结果各不相同,但主要分布在[0.700,0.800]之间,说明大部分人的标注水平比较接近,标注质量也比较好.

### 3.2 对比方法及评价准则

我们对比了 4 种考虑标记间关系的多标记众包方法 D-DS、P-DS、ND-DS<sup>[19]</sup>、MLNB<sup>[18]</sup>.其中,D-DS、P-DS 使用标记子集(label powerset)、标记对(label pair)的分组方式将多个标记的组合当做单个类别,将多标记问题转化为多分类问题,然后使用单标记众包方法 DS 来学习;ND-DS 从收集到的标记集合中学习标记间的条件相关性以估计样本真实标记;MLNB 在单标记众包方法 DS 中加入标记共同出现的概率信息作为标记关系.此外,我们还与 4 种最常用效果最稳定的单标签众包方法 MV、KOS<sup>[11]</sup>、DS<sup>[14]</sup>、MaxEn<sup>[12]</sup>做了比较.我们将多标记任务看作多个独立的单标记任务,在每个标记上分别使用单标记众包方法.对于各个方法的参数,对比方法的参数设置依照其对应文献的推荐参数设置.对于本文所提方法的参数 $\alpha_i$ 设置为 1/3, $\beta_i$ 设置为 0.1.之后,我们会详细讨论本文所提方法的参数影响.

多标记学习的性能可以从多个不同的角度评价.为了综合评价各个方法的效果,本文选用 Hamming loss、Macro  $F1$ 、Average precision 和 Ranking loss 这 4 种常用的多标记评价指标来衡量学习性能,其中,Hamming loss、Macro  $F1$  为分类指标,Average precision、Ranking loss 为排序指标.Macro  $F1$ 、Average precision(Hamming loss、Ranking loss)的值越大(小),表明预测结果越接近真实标记.这些指标的具体定义参见文献[4].由于各个众包方法的预测结果是每个类别为正标记的概率,为了获得分类评价指标,需对概率预测做阈值划分.以往的众包工作简单地以 0.5 作为阈值,而在多标记学习中,分类结果受划分阈值的影响很大.本文中,为了获得相对公平的分类结果对比,在计算各方法的分类指标时,我们将每个样本的各个标记的预测值做排序,将预测值最大的  $k$  个标记预测为正标记,其他的预测为负标记,其中, $k$  为该样本的真实正标记的个数.这里,为了计算分类指标,我们用到了真实标记信息.而这在实际应用中是不可行的,在以后的工作中,我们将考虑把阈值确定也放到学习过程中.

### 3.3 实验结果

类似于之前的众包工作<sup>[9,28]</sup>,我们考察了标注量对众包学习方法效果的影响.我们从收集到的标注集合中随机采样一定比例的标注  $p$ ,在这部分标注上学习,估计数据集的真实标记.以 0.1 为间隔,我们测试了  $p$  在[0.1,1]的实验.为了缓解采样对学习算法的随机影响,我们在每种情形下进行了 10 次重复实验,记录下实验均值和标准差.两个数据集的实验结果分别如图 1~图 4 所示.

由实验结果可以看出,对比 TencSign、TencVal、TencEnt 这 3 种使用不同融合策略的方法,在两个数据集上,TencEnt 和 TencVal 都明显优于 TencSign;而在 Image1 数据集上,TencVal 和 TencEnt 的差别并不明显;在 Image2 数据集上,TencEnt 略微优于 TencVal.由此可以看出,通过张量补全方法对标注预处理之后,实值的标注结果包含了每个标注者的置信度信息,简单地使用标注符号而不考虑标注者置信度,忽视了各个标注者标注质量的差异.

由图 1~图 4 可知,随着标注数量由少到多增加,各个方法整体上都呈现出性能提升的趋势;在标注数量增长到一定程度之后,学习效果趋于稳定.与对比方法相比,TencEnt 和 TencVal 在各个指标上的结果都显著最优,并且更快地收敛.在标注数量比较少时,如  $p < 40\%$  时,对比方法与 TencEnt、TencVal 之间存在明显的差距.这说明为了达到同样好的学习效果,TencEnt 和 TencVal 所需的标注更少.

对比 4 种多标记众包方法 D-DS、P-DS、ND-DS、MLNB,由图 1 和图 3 可以看出,MLNB 的效果最差.其可能的原因有两个.一个原因是,MLNB 提出的目的是学习标记深层结构关系,主要通过加入标记共同出现的概率来考虑标记关系.对于标记集合巨大并且存在深层标记结构关系的数据,这种标记关系常见并且容易准确地计算.而对于缺乏这样深层标记结构关系的数据,这种类型的标记关系并不明显.另一个原因是,MLNB 算法的实现中,对所有的标注者使用了同样的参数,从而忽略了标注者标注质量的不同.对比 D-DS、P-DS、ND-DS,考虑所有标记间关系的 P-DS、ND-DS 的效果更优.

对比 4 种单标记众包方法 MV、DS、KOS、MaxEn,由图 2 和图 4 可知,简单的 MV 效果最差,而考虑各个标注者在各个样本上的标注能力,并且采用最优化最坏熵的 MaxEn 效果最好.对比图 1 和图 2、图 3 和图 4 中

的多标记众包方法和单标记众包方法,我们可以看出,考虑标记间关系的 4 种多众包标记方法并不一定优于将多个标记独立学习的单标记众包方法.这是因为这些方法考虑标记间关系时主要集中在标记间的局部关系,其估计依赖于收集到的标注,很敏感地受到标注质量和数量的影响.而本文提出的基于低秩张量矫正的方法则从标注整体存在结构关系来同时考虑标记相关性和标注者的相似性,从而更可靠地估计样本真实标记.

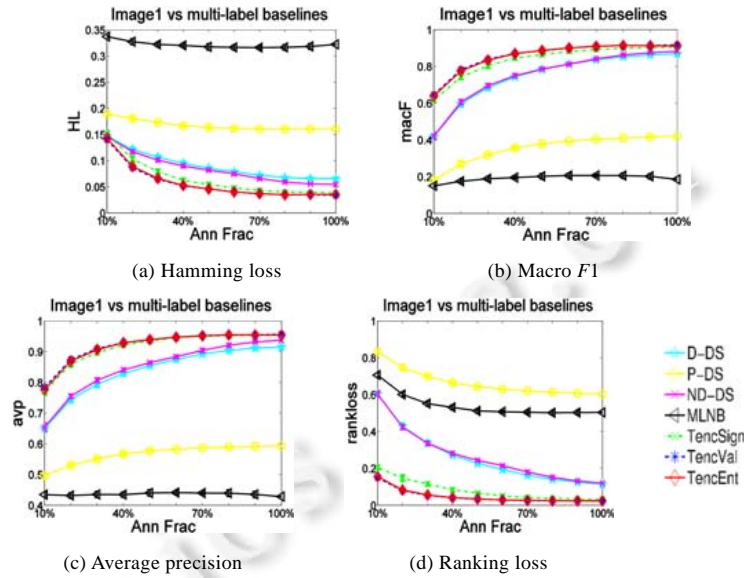


Fig.1 Results of four commonly used multi-label performance measure with multi-label crowdsourcing baselines on Image1

图 1 Image1 数据集上 4 种常用的多标记性能评价指标结果,对比多标记众包学习方法

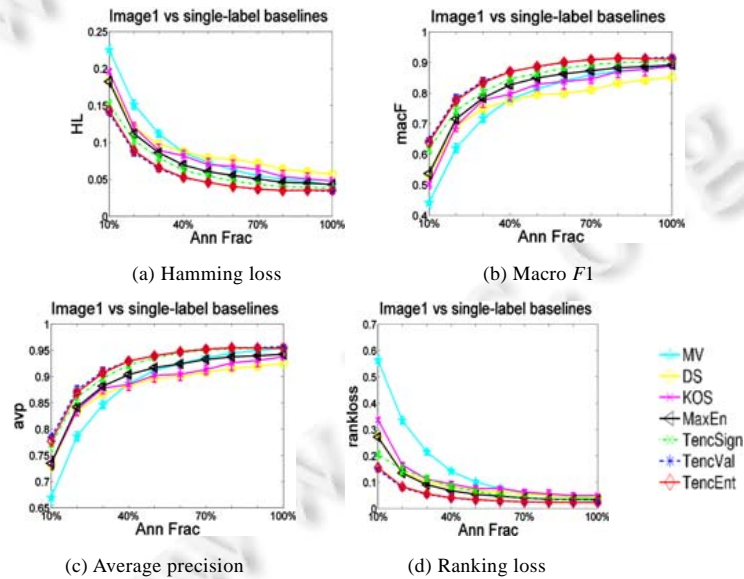


Fig.2 Results of four commonly used multi-label performance measure with single-label crowdsourcing baselines on Image1

图 2 Image1 数据集上 4 种常用的多标记性能评价指标结果,对比单标记众包学习方法



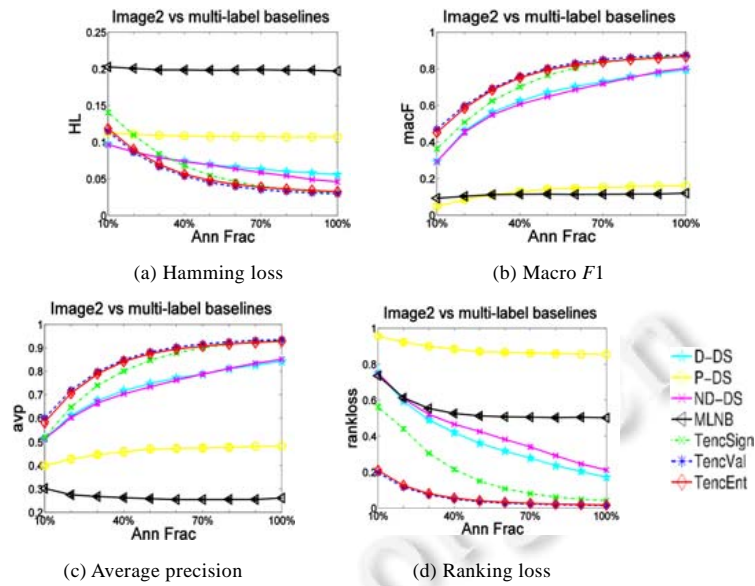


Fig.3 Results of four commonly used multi-label performance measure with multi-label crowdsourcing baselines on Image2

图 3 Image2 数据集上 4 种常用的多标记性能评价指标结果,对比多标记众包学习方法

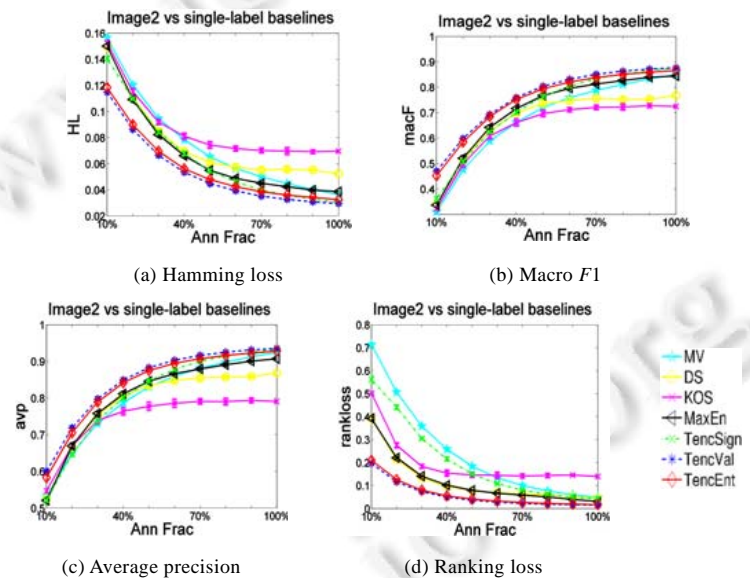


Fig.4 Results of four commonly used multi-label performance measure with single-label crowdsourcing baselines on Image2

图 4 Image2 数据集上 4 种常用的多标记性能评价指标结果,对比单标记众包学习方法

此外,我们还考察了另外一种情形的标注数量对各种方法学习效果的影响.我们依次增加各个数据集上的标注者数量,加入其对应的标注结果,测试各个方法的学习效果.类似前面,我们记录了 10 次重复实验的均值和标准差.两个数据集上的实验结果分别见表 1 和表 2.加粗的结果表明,其在配对检验(95%置信度)中显著优于其他结果,或与最优结果之间无显著性差异.

**Table 1** Results (mean±std.) on Image1, *nw* denotes the number of workers

**表 1** Image1 数据集上多种方法的实验结果(mean±std.),*nw* 表示标注者数目

	算法	<i>nw</i> =1	<i>nw</i> =3	<i>nw</i> =5	<i>nw</i> =7	<i>nw</i> =9	<i>nw</i> =11	<i>nw</i> =13	<i>nw</i> =15
Hamming loss	D-DS	<b>.131±.056</b>	.096±.007	.089±.009	.084±.013	.077±.011	.072±.005	.069±.001	.066±.001
	P-DS	.174±.024	.172±.009	.172±.005	.168±.007	.165±.006	.164±.003	.163±.001	.161±.000
	ND-DS	<b>.131±.056</b>	<b>.088±.009</b>	.080±.007	.074±.010	.068±.008	.063±.004	.059±.002	.056±.000
	MLNB	.343±.019	.320±.004	.321±.004	.324±.001	.315±.008	.317±.006	.317±.004	.317±.002
	MV	.214±.114	.115±.023	.091±.022	.079±.018	.067±.013	.055±.006	.051±.002	.047±.002
	DS	.199±.101	.110±.024	.102±.013	.096±.018	.085±.012	.076±.011	.070±.003	.060±.016
	KOS	.369±.000	.116±.020	.094±.022	.081±.022	.072±.011	.066±.013	.067±.002	.057±.010
	MaxEn	.197±.103	.113±.018	.087±.019	.071±.019	.061±.015	.054±.007	.049±.003	.047±.003
	TencSign	.203±.100	.113±.024	.086±.016	.073±.013	.061±.009	.051±.006	.047±.001	.042±.002
	TencVal	.168±.077	<b>.089±.017</b>	<b>.070±.017</b>	<b>.060±.015</b>	<b>.051±.010</b>	<b>.044±.005</b>	<b>.041±.003</b>	<b>.037±.002</b>
	TencEnt	.168±.077	<b>.089±.016</b>	<b>.070±.017</b>	<b>.060±.015</b>	<b>.050±.011</b>	<b>.844±.007</b>	<b>.041±.004</b>	<b>.039±.003</b>
MacF1	D-DS	.466±.341	.723±.065	.768±.053	.797±.050	.826±.037	.846±.016	.855±.003	.860±.004
	P-DS	.243±.183	.350±.064	.365±.045	.381±.040	.397±.028	.408±.013	.414±.004	.417±.003
	ND-DS	.467±.341	.734±.070	.780±.053	.812±.049	.836±.037	.857±.017	.868±.008	.875±.004
	MLNB	.141±.041	.192±.014	.189±.011	.186±.006	.207±.016	.204±.017	.204±.011	.206±.008
	MV	.465±.280	.703±.062	.766±.057	.798±.047	.829±.035	.859±.014	.870±.008	.881±.006
	DS	.496±.259	.718±.064	.737±.036	.748±.044	.777±.031	.801±.027	.815±.008	.844±.016
	KOS	.082±.023	.707±.054	.765±.053	.806±.056	.828±.029	.845±.030	.840±.007	.864±.024
	MaxEn	.505±.259	.709±.048	.778±.050	.820±.049	.846±.038	.865±.017	.876±.010	.880±.008
	TencSign	.499±.241	.712±.064	.779±.042	.815±.032	.846±.023	.871±.014	.882±.004	.894±.006
	TencVal	<b>.577±.199</b>	<b>.770±.047</b>	<b>.821±.047</b>	<b>.848±.040</b>	<b>.873±.028</b>	<b>.890±.015</b>	<b>.899±.008</b>	<b>.907±.006</b>
	TencEnt	<b>.577±.199</b>	<b>.769±.042</b>	<b>.820±.046</b>	<b>.846±.040</b>	<b>.874±.030</b>	<b>.890±.017</b>	<b>.898±.012</b>	<b>.904±.009</b>
Average precision	D-DS	.691±.189	.825±.029	.850±.033	.868±.035	.888±.026	.901±.011	.907±.003	.911±.006
	P-DS	.552±.087	.575±.043	.568±.023	.575±.019	.582±.014	.588±.006	.590±.003	.591±.002
	ND-DS	.690±.189	.839±.038	.865±.029	.887±.027	.907±.023	.921±.011	.928±.005	.932±.003
	MLNB	.437±.003	.446±.008	.436±.003	.432±.010	.442±.012	.439±.014	.441±.009	.437±.004
	MV	.690±.189	.851±.043	.886±.035	.908±.030	.927±.021	.939±.010	.945±.003	.950±.002
	DS	.687±.173	.835±.041	.849±.020	.859±.028	.878±.023	.895±.018	.905±.005	.918±.008
	KOS	.434±.000	.825±.033	.863±.032	.883±.030	.900±.017	.907±.014	.909±.006	.920±.014
	MaxEn	.692±.167	.829±.028	.868±.033	.896±.029	.914±.024	.926±.009	.931±.004	.936±.004
	TencSign	.705±.169	.853±.042	.889±.031	<b>.910±.025</b>	.927±.018	.939±.009	.944±.002	<b>.950±.003</b>
	TencVal	<b>.746±.148</b>	<b>.865±.040</b>	<b>.894±.034</b>	<b>.911±.028</b>	<b>.928±.019</b>	<b>.940±.009</b>	<b>.946±.005</b>	<b>.951±.003</b>
	TencEnt	<b>.746±.148</b>	.861±.035	<b>.892±.034</b>	.905±.030	<b>.929±.020</b>	<b>.940±.010</b>	<b>.945±.007</b>	.949±.005
Ranking loss	D-DS	.529±.343	.268±.071	.222±.064	.192±.061	.162±.040	.141±.019	.131±.007	.126±.008
	P-DS	.770±.173	.671±.064	.657±.044	.642±.039	.626±.027	.615±.012	.610±.004	.607±.003
	ND-DS	.529±.343	.267±.077	.219±.060	.187±.052	.165±.038	.145±.019	.133±.009	.128±.006
	MLNB	.727±.192	.553±.036	.542±.032	.531±.021	.512±.011	.507±.020	.498±.011	.507±.004
	MV	.529±.343	.229±.084	.165±.069	.123±.059	.087±.039	.063±.017	.052±.006	.045±.006
	DS	.396±.257	.160±.061	.126±.042	.104±.038	.082±.027	.065±.015	.057±.005	.048±.006
	KOS	1.00±.150	.205±.075	.145±.066	.120±.049	.096±.034	.083±.013	.071±.008	.066±.012
	MaxEn	.386±.257	.163±.054	.118±.049	.088±.042	.065±.028	.051±.011	.044±.003	.038±.003
	TencSign	.482±.290	.224±.084	.151±.054	.109±.042	.078±.027	.058±.016	.047±.004	.040±.005
	TencVal	<b>.221±.077</b>	<b>.096±.039</b>	<b>.068±.027</b>	<b>.055±.023</b>	<b>.042±.016</b>	<b>.033±.005</b>	<b>.029±.003</b>	<b>.026±.002</b>
	TencEnt	<b>.221±.077</b>	<b>.096±.041</b>	<b>.072±.033</b>	<b>.058±.030</b>	<b>.042±.017</b>	<b>.033±.006</b>	<b>.029±.004</b>	<b>.027±.002</b>

**Table 2** Results (mean±std.) on Image2, *nw* denotes the number of workers

**表 2** Image2 数据集上多种方法的实验结果(mean±std.),*nw* 表示标注者数目

	算法	<i>nw</i> =1	<i>nw</i> =3	<i>nw</i> =5	<i>nw</i> =7	<i>nw</i> =9	<i>nw</i> =11	<i>nw</i> =13	<i>nw</i> =15
Hamming loss	D-DS	<b>.109±.001</b>	.071±.016	.069±.011	.067±.008	.059±.002	.057±.002	.056±.000	.055±.000
	P-DS	.114±.000	.109±.001	.108±.001	.107±.001	.107±.001	.107±.000	.106±.000	.107±.000
	ND-DS	.126±.000	<b>.063±.001</b>	<b>.062±.001</b>	.059±.001	.050±.001	.048±.000	.048±.000	.091±.000
	MLNB	.204±.000	.200±.002	.199±.002	.199±.001	.199±.001	.198±.001	.199±.001	.198±.001
	MV	.192±.002	.104±.037	.089±.029	.075±.022	.054±.007	.044±.002	.040±.001	.036±.000
	DS	.191±.003	.100±.037	.087±.028	.073±.020	.057±.006	.054±.001	.054±.001	.052±.003
	KOS	.205±.000	.102±.036	.096±.022	.086±.016	.070±.003	.067±.004	.066±.003	.069±.000
	MaxEn	.191±.002	.100±.039	.086±.031	.072±.023	.051±.007	.043±.001	.040±.001	.038±.000
	TencSign	.187±.012	.098±.030	.083±.025	.067±.020	.047±.006	.038±.002	.035±.001	<b>.030±.000</b>
	TencVal	.152±.005	.076±.031	<b>.065±.026</b>	<b>.054±.018</b>	<b>.039±.006</b>	<b>.032±.000</b>	<b>.031±.000</b>	<b>.029±.000</b>
	TencEnt	.152±.005	.080±.030	.068±.026	.057±.017	.042±.005	.035±.001	.034±.000	.032±.000

**Table 2** Results (mean±std.) on Image2,  $nw$  denotes the number of workers (Continued)**表 2** Image2 数据集上多种方法的实验结果(mean±std.), $nw$  表示标注者数目(续)

	算法	$nw=1$	$nw=3$	$nw=5$	$nw=7$	$nw=9$	$nw=11$	$nw=13$	$nw=15$
MacF1	D-DS	.076±.016	.569±.220	.621±.161	.670±.100	.752±.028	.777±.008	.784±.006	.793±.000
	P-DS	.014±.015	.131±.210	.142±.152	.148±.110	.157±.030	.157±.009	.162±.007	.161±.000
	ND-DS	.021±.013	<b>.657±.190</b>	.678±.147	.705±.107	.767±.028	.779±.008	.786±.008	.601±.000
	MLNB	.077±.003	.104±.011	.106±.012	.108±.009	.109±.008	.115±.004	.114±.009	.113±.009
	MV	.134±.014	.555±.188	.612±.153	.674±.114	.773±.039	.815±.011	.835±.007	.850±.000
	DS	.141±.015	.578±.184	.629±.143	.686±.101	.759±.032	.765±.005	.760±.005	.770±.003
	KOS	.073±.013	.567±.185	.593±.127	.639±.090	.711±.022	.733±.014	.738±.010	.726±.000
	MaxEn	.141±.013	.575±.195	.631±.161	.690±.119	.786±.038	.826±.007	.836±.004	.844±.000
	TencSign	.156±.055	.583±.159	.644±.134	.708±.105	.804±.033	.845±.007	.858±.005	<b>.876±.000</b>
	TencVal	<b>.294±.019</b>	<b>.661±.160</b>	<b>.706±.136</b>	<b>.757±.099</b>	<b>.831±.033</b>	<b>.865±.003</b>	<b>.871±.002</b>	<b>.877±.000</b>
TencEnt	<b>.294±.019</b>	.647±.153	.694±.136	.747±.091	.822±.027	.852±.004	.859±.001	.867±.000	
Average precision	D-DS	.410±.010	.705±.133	.734±.099	.760±.062	.812±.017	.829±.004	.835±.006	.842±.000
	P-DS	.376±.010	.459±.100	.468±.100	.473±.051	.480±.000	.479±.000	.484±.000	.482±.000
	ND-DS	.363±.011	<b>.753±.100</b>	.768±.101	.785±.053	.825±.000	.830±.000	.840±.000	.686±.000
	MLNB	.353±.003	.281±.034	.276±.018	.266±.018	.254±.006	.256±.005	.254±.008	.252±.009
	MV	.410±.010	.719±.141	.760±.115	.803±.083	.871±.003	.901±.008	.913±.007	.924±.000
	DS	.354±.009	.679±.137	.727±.105	.777±.075	.841±.025	.860±.007	.862±.003	.868±.003
	KOS	.365±.010	.688±.118	.708±.077	.738±.052	.787±.013	.800±.012	.803±.013	.792±.000
	MaxEn	.355±.013	.678±.144	.729±.115	.779±.086	.854±.028	.886±.007	.897±.003	.907±.000
	TencSign	.410±.010	.720±.141	.762±.115	.810±.085	.879±.027	.910±.006	.922±.004	<b>.933±.000</b>
	TencVal	<b>.441±.020</b>	<b>.756±.140</b>	<b>.795±.116</b>	<b>.839±.081</b>	<b>.899±.026</b>	<b>.925±.004</b>	<b>.930±.003</b>	<b>.935±.000</b>
TencEnt	<b>.441±.020</b>	.745±.137	.787±.115	.832±.076	.893±.024	.917±.004	.924±.001	.928±.000	
Ranking loss	D-DS	.919±.019	.423±.220	.357±.161	.297±.093	.216±.027	.189±.011	.179±.001	.171±.000
	P-DS	.986±.015	.881±.210	.871±.152	.864±.100	.857±.030	.857±.009	.853±.007	.854±.000
	ND-DS	.976±.013	.346±.190	.339±.147	.310±.107	.250±.028	.232±.008	.237±.008	.422±.000
	MLNB	.950±.010	.632±.152	.594±.011	.562±.063	.520±.021	.511±.004	.503±.004	.505±.000
	MV	.919±.019	.403±.223	.319±.174	.238±.120	.131±.042	.088±.011	.066±.007	.048±.000
	DS	.693±.013	.292±.169	.230±.125	.169±.081	.096±.028	.071±.015	.058±.006	.043±.000
	KOS	1.00±.010	.352±.246	.297±.175	.243±.088	.160±.025	.135±.008	.142±.014	.139±.000
	MaxEn	.689±.013	.295±.167	.231±.126	.171±.085	.093±.028	.064±.009	.047±.007	.031±.000
	TencSign	.887±.008	.373±.177	.289±.138	.208±.099	.116±.036	.076±.013	.057±.006	.041±.000
	TencVal	<b>.324±.017</b>	<b>.117±.086</b>	<b>.090±.068</b>	<b>.064±.044</b>	<b>.032±.014</b>	<b>.021±.002</b>	<b>.017±.001</b>	<b>.015±.000</b>
TencEnt	<b>.324±.017</b>	<b>.122±.085</b>	<b>.092±.067</b>	<b>.067±.043</b>	<b>.034±.014</b>	<b>.023±.002</b>	<b>.019±.001</b>	<b>.017±.000</b>	

与前面的实验结果类似,从表 1、表 2 中可以看到,TencVal,TencEnt 的结果都显著优于对比方法.而在 Image2 数据集上,表 2 中 TencVal 与 TencEnt 的对比与前面图 3(图 4)的实验对比略有不同.在图 3(图 4)中,TencEnt 略优于 TencVal,而在表 2 的结果中,TencVal 略优于 TencEnt.在实际应用中,推荐使用 TencVal、TencEnt 方法.如果考虑计算简便性,可以采用 TencVal 的融合方法.

### 3.4 参数讨论

在上面的实验中,本文所提的方法使用了固定参数.本节中,我们讨论参数对所提方法的影响.由上面的实验结果可知,TencVal、TencEnt 明显优于 TencSign,并且 TencVal、TencEnt 效果差别不大.考虑到 TencVal 计算简便,本节中给出 TencVal 的结果.限于篇幅,只给出随机采样部分样本情形的结果,标注者数目变化的结果类似.

首先固定  $\beta_i=0.1$ ,我们考察参数  $\alpha=[\alpha_1, \alpha_2, \alpha_3]/(\alpha_1+\alpha_2+\alpha_3)$  不同取值的结果.由于  $\alpha_1, \alpha_2, \alpha_3$  分别对应标注张量在 {样本,标记,标注者} 这 3 个维度的相对重要程度,我们测试了 7 种极端取值,  $[\alpha_1, \alpha_2, \alpha_3]=\{[1,0,0],[0,1,0],[0,0,1],[1,1,0],[1,0,1],[0,1,1],[1,1,1]\}$ ,分别对应 3 个维度的不同重要程度.在 Image1、Image2 两个数据集上的 Macro F1 结果如图 5 所示.

由图 5 可以看出,在标注比例  $p=10\%$  的情况下,  $[\alpha_1, \alpha_2, \alpha_3]=[0,0,1]$  时,即只考虑标注者维度展开的情形,学习效果比其他参数设定差很多.而同时考虑多个维度,包括只考虑 2 个维度和同时考虑 3 个维度,学习效果更鲁棒.因此对于参数  $\alpha$  的取值,可以设为 3 个维度同等重要,即  $\alpha=[1/3,1/3,1/3]$ .

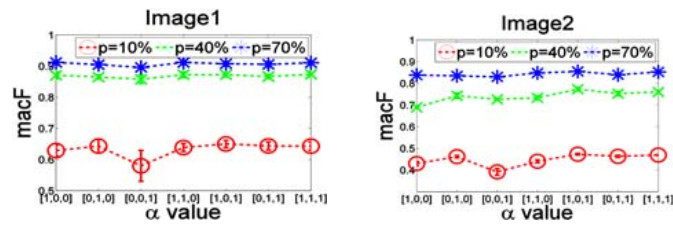


Fig.5 Influence of parameter  $\alpha$  on the results of Image1 and Image2 with three different annotation fraction numbers

图5 在3种标注采样比例情形下,参数 $\alpha$ 不同取值对数据集 Image1、Image2 的学习结果影响

固定 $\alpha=[1/3,1/3,1/3]$ ,我们考察参数 $\beta_i$ 不同取值的结果.由于 $\beta_i$ 对应优化问题(4)中低秩目标和逼近观测值目标的权衡折中,取 $\beta_i=\alpha_i/\gamma$ ,我们测试了7种 $\gamma$ 的取值, $\gamma=[1e^{-3},1e^{-2},\dots,1e^3]$ .在Image1、Image2两个数据集上的 Macro F1 结果如图6所示.

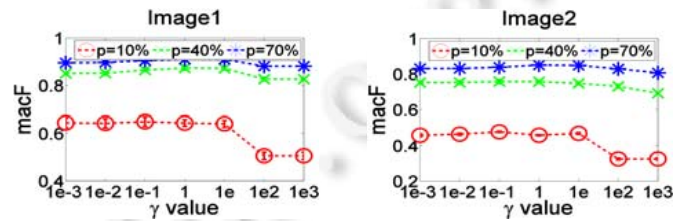


Fig.6 Influence of parameter  $\beta$  on the results of Image1 and Image2 with three different annotation fraction numbers

图6 在3种标注采样比例情形下,参数 $\beta$ 不同取值对数据集 Image1、Image2 的学习结果影响

由图6可以看出,当 $\gamma>10$ 时,Image1和Image2上的学习性能明显下降;当 $\gamma\leq 10$ 时,学习效果比较稳定.也就是说,参数 $\beta_i$ 的值不能太小,否则恢复出来的标注张量会偏离原本的标注太多,从而影响学习效果.在实际使用中,参数 $\beta_i$ 可以设置为0.1~10范围的常数.

#### 4 结束语

本文考虑多标记众包学习问题,将多标记任务分配给多个容易访问的非专家,收集标注信息,以便从可能犯错的标注中估计样本的真实标记.以往的单标记众包方法忽略了多标记任务中标记之间的相关性,而多标记众包方法主要集中在局部的标记关系学习利用上,其估计依赖于收集到的标注,很敏感地受到标注质量和数量的影响.本文从标注整体存在低秩结构关系来考虑标记间相关性和标注者的相似性,提出基于低秩张量矫正的方法,并提出3种标注融合策略.实验结果表明,本文所提的方法 TencVal、TencEnt 在多种标注数量情形下都显著优于对比方法.在后续工作中,我们将加入对标注收集过程的控制,设计更有针对性的标注收集机制,从而以更少的标记代价收集更可靠、更有利于学习的标注.

#### References:

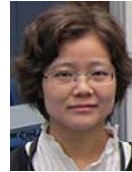
- [1] Boutell MR, Luo J, Shen X, Brown CM. Learning multilabel scene classification. Pattern Recognition, 2004,37(9):1757-1771.
- [2] Elisseeff A, Weston J. A kernel method for multi-labelled classification. In: Dietterich TG, Becker S, Ghahramani Z, eds. Proc. of the Advances in Neural Information Processing Systems 14. 2001. 681-687.
- [3] Tsoumakas G, Katakis I, Vlahavas I. Mining Multi-label Data. Oded M, Lior R, eds. Data Mining and Knowledge Discovery Handbook. Berlin: Springer-Verlag, 2010. 667-685.
- [4] Zhang ML, Zhou ZH. A review on multi-label learning algorithms. IEEE Trans. on Knowledge and Data Engineering, 2014,26(8): 1819-1837.

- [5] Horvitz E. Reflections on challenges and promises of mixed-initiative interaction. *AI Magazine*, 2007,28(2):13–22.
- [6] Weld D, Lin C, Bragg J. Artificial Intelligence and Collective Intelligence. In: Malone T, Bernstein M, eds. *Collective Intelligence Handbook*. 2015.
- [7] Snow R, O'Connor B, Jurafsky D, Ng A. Cheap and fast—but is it good? Evaluating non-expert annotations for natural language tasks. In: *Proc. of the Conf. on Empirical Methods in Natural Language Processing*. Honolulu, 2008. 254–263.
- [8] Whitehill J, Ruvolo P, Wu T, Bergsma J, Movellan J. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In: Bengio Y, Schuurmans D, Lafferty JD, Williams CKI, Culotta A, ed. *Proc. of the Advances in Neural Information Processing Systems 22*. 2009. 2035–2043.
- [9] Raykar V, Yu S, Zhao L, Valadez G, Florin C, Bogoni L, Moy L. Learning from crowds. *Journal of Machine Learning Research*, 2010,11:1297–1322.
- [10] Welinder P, Branson S, Belongie S, Perona P. The multidimensional wisdom of crowds. In: Lafferty J, Williams CI, Shawe-Taylor J, Zemel R, Culotta A, eds. *Proc. of the Advances in Neural Information Processing Systems 23*. 2010. 2024–2432.
- [11] Karger DR, Oh S, Shah D. Iterative learning for reliable crowdsourcing systems. In: Shawe-Taylor J, Zemel RS, Bartlett PL, Pereira FCN, Weinberger KQ, eds. *Proc. of the Advances in Neural Information Processing Systems 24*. 2011. 1953–1961.
- [12] Zhou D, Basu S, Mao Y, Platt JC. Learning from the wisdom of crowds by minimax entropy. In: Bartlett PL, Pereira FCN, Burges CJC, Bottou L, Weinberger KQ, eds. *Proc. of the Advances in Neural Information Processing Systems 25*. 2012. 2195–2203.
- [13] Chilton L, Little G, Edge D, Weld D, Landay J. Cascade: Crowdsourcing taxonomy creation. In: *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*. 2013. 1999–2008.
- [14] Dawid AP, Skene AM. Maximum likelihood estimation of observer error-rates using the em algorithm. *Journal of the Royal Statistical Society*, 1979,28(1):20–28.
- [15] Sheng VS, Provost FJ, Ipeirotis PG. Get another label? Improving data quality and data mining using multiple, noisy labelers. In: *Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. Las Vegas, 2008. 614–622.
- [16] Sun Y, Singla A, Fox D, Krause A. Building hierarchies of concepts via crowdsourcing. In: *Proc. of the 24th Int'l Joint Conf. on Artificial Intelligence*. 2015. 844–853.
- [17] Duan L, Oyama S, Kurihara M, Sato H. Crowdsourced semantic matching of multi-label annotations. In: *Proc. of the 24th Int'l Joint Conf. on Artificial Intelligence*. 2015. 3483–3489.
- [18] Bragg J, Mausam, Weld DS. Crowdsourcing multi-label classification for taxonomy creation. In: *Proc. of the 1st AAAI Conf. on Human Computation and Crowdsourcing*. 2013.
- [19] Duan L, Oyama S, Sato H, Kurihara M. Separate or joint? Estimation of multiple labels from crowdsourced annotations. *Expert Systems with Applications*, 2014,41(13):5723–5732.
- [20] Tam NT, Viet HH, Hung NQV, Weidlich M, Yin H, Zhou X. Multi-label answer aggregation for crowdsourcing. Technical Report, 2016.
- [21] Comite FD, Gilleron R, Tommasi M. Learning multi-label alternating decision tree from texts and data. In: *Proc. of the 3rd Int'l Conf. on Machine Learning and Data Mining in Pattern Recognition*. Leipzig, 2003. 35–49.
- [22] Cesa-Bianchi N, Gentile C, Zaniboni L. Hierarchical classification: Combining bayes with SVM. In: *Proc. of the 23rd Int'l Conf. on Machine Learning*. Pittsburgh, 2006. 177–184.
- [23] Bi W, Kwok J. Multi-label classification on tree- and dagstructured hierarchies. In: *Proc. of the 28th Int'l Conf. on Machine Learning*. Bellevue, 2011. 17–24.
- [24] Schapire RE, Singer Y. BoosTexter: A boosting-based system for text categorization. *Machine Learning*, 2000,39(2-3):135–168.
- [25] Yang Y, Gopal S. Multilabel classification with meta-level features in a learning-to-rank framework. *Machine Learning*, 2012, 88(1-2):47–68.
- [26] Zhang ML. Lift: Multi-label learning with label-specific features. In: *Proc. of the 22nd Int'l Joint Conf. on Artificial Intelligence*. Barcelona, 2011. 1609–1614.
- [27] Candès EJ, Recht B. Exact matrix completion via convex optimization. In: *Proc. of the Foundations of Computational Mathematics*. 2009. 717–772.

- [28] Candès EJ, Tao T. The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. on Information Theory*, 2010, 56(5):2053–2080.
- [29] Recht B. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 2011,12:3413–3430.
- [30] Rennie J, Srebro N. Fast maximum margin matrix factorization for collaborative prediction. In: *Proc. of the 22nd Int'l Conf. on Machine Learning (ICML)*. 2005. 713–719.
- [31] Ma S, Goldfarb D, Chen L. Fixed point and bregman iterative methods for matrix rank minimization. *Mathematical Programming*, 2009,128(1):321–353.
- [32] Gandy S, Recht B, Yamada I. Tensor completion and low-nrank tensor recovery via convex optimization. *Inverse Problem*, 2011, 27(2):Article No.025010.
- [33] Signoretto M, Lathauwer LD, Suykens JAK. Nuclear norms for tensors and their use for convex multilinear estimation. In: *Proc. of the Submitted to Linear Algebra and Its Applications*. 2010.
- [34] Liu J, Musialski P, Wonka P, Ye JP. Tensor completion for estimating missing values in visual data. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2013,35(1):208–220.
- [35] Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 1966,31:279–311.
- [36] Harshman RA. Foundations of the parafac procedure: Models and conditions for an “explanatory” multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 1970,16:1–84.
- [37] Xu Y, Hao R, Yin W, Su Z. Parallel matrix factorization for low-rank tensor completion. In: *Proc. of the CoRR*. 2013.
- [38] Zhou Y, He J. Crowdsourcing via tensor augmentation and completion. In: *Proc. of the 25th Int'l Joint Conf. on Artificial Intelligence*. New York, 2016. 2435–2441.
- [39] Tseng P. Convergence of block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory Application*, 2001,109:475–494.
- [40] Zhou Y, Ying L, He J. MultiC2: An optimization framework for learning from task and worker dual heterogeneity. In: *Proc. of the 2017 SIAM Int'l Conf. on Data Mining*. Houston, 2017. 579–587.



李绍园(1989—),女,河南鹤壁人,博士,CCF专业会员,主要研究领域为机器学习,数据挖掘.



姜远(1976—),女,博士,教授,博士生导师,CCF专业会员,主要研究领域为机器学习,数据挖掘.