智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

# 面向众包数据的特征扩维标签质量提高方法

李易南，王士同

引用本文:
李易南, 王士同. 面向众包数据的特征扩维标签质量提高方法[J]. 智能系统学报, 2020, 15(2): 227–234.
LI Yinan, WANG Shitong. A feature augmentation method for enhancing the labeling quality of crowdsourcing data[J]. *CAAI Transactions on Intelligent Systems*, 2020, 15(2): 227–234.

在线阅读 View online: https://dx.doi.org/10.11992/tis.201810014

**DOI**: 10.11992/tis.201810014
**网络出版地址**: http://kns.cnki.net/kcms/detail/23.1538.TP.20190829.0906.002.html

# 面向众包数据的特征扩维标签质量提高方法

李易南, 王士同
(江南大学 数字媒体学院, 江苏 无锡 214122)

**摘 要**: 众包是一个新兴的收集数据集标签的方法。虽然它经济实惠, 但面临着数据标签质量无法保证的问题。尤其是当客观原因存在使得众包工作者工作质量较差时, 所得的标签会更加不可靠。因此提出一个名为基于特征扩维提高众包质量的方法 (FA-method), 其基本思想是, 首先由专家标注少部分标签, 再利用众包者标注的数据集训练模型, 对专家集进行预测, 所得结果作为专家数据集新的特征, 并利用扩维后的专家集训练模型进行预测, 计算每个实例为噪声的可能性以及噪声数量上限来过滤出潜在含噪声标签的数据集, 类似地, 对过滤后的高质量集再次使用扩维的方法进一步校正噪声。在 8 个 UCI 数据集上进行验证的结果表明, 和现有的结合噪声识别和校正的众包标签方法相比, 所提方法能够在重复标签数量较少或标注质量较低时均取得很好的效果。

**关键词**: 众包; 标签质量; 扩维; 专家标注; 噪声识别; 噪声校正; 噪声可能性; 噪声数量上限
**中图分类号**: TP181 **文献标志码**: A **文章编号**: 1673−4785(2020)02−0227−08

**中文引用格式**: 李易南, 王士同. 面向众包数据的特征扩维标签质量提高方法 [J]. 智能系统学报, 2020, 15(2): 227–234.
**英文引用格式**: LI Yinan, WANG Shitong. A feature augmentation method for enhancing the labeling quality of crowdsourcing data[J]. CAAI transactions on intelligent systems, 2020, 15(2): 227–234.

# A feature augmentation method for enhancing the labeling quality of crowdsourcing data

LI Yinan, WANG Shitong
(School of Digital Media, Jiangnan University, Wuxi 214122, China)

**Abstract**: Crowdsourcing is a new method of collecting the labels of data. Although it is economical, crowdsourcing faces an unavoidable problem, i.e., the quality of the labels cannot be guaranteed. In particular, when the quality of labeling work is low because of the existence of objective causes, the result of crowdsourcing will be unreliable. In this study, a feature augmentation method for enhancing the labeling quality of crowdsourcing data is proposed. In the proposed method, first, a small amount of expert data is labeled by several people with professional knowledge. Then, the crowdsourcing data are used to create the classifiers and predict the expert data. The resultant predicted labels are used to augment the expert data. Then, the augmented expert data are used to create the classifiers, predict the original data, and calculate the probability of noise for each instance and the upper limit of noise number to filter out the high-quality dataset from potentially noisy labels. Similarly, the filtered high-quality dataset is utilized to further correct the noisy labels using the proposed feature augmentation method. The experiments conducted on eight UCI datasets show that the proposed feature augmentation method has achieved encouraging results when the number of repeated labels is comparatively small or the quality of labeling is comparatively low.

**Keywords**: crowdsourcing; labeling quality; feature augmentation; expert labeling; noise identification; noise correction; noise probability; upper limit of noise number

**收稿日期**: 2018−10−15. **网络出版日期**: 2019−08−29.
**基金项目**: 国家自然科学基金项目 (61272210).
**通信作者**: 李易南. E-mail: 1920898036@qq.com.

众包, 一般认为是由美国作家杰夫·豪 (Jeff Howe) 在 2006 年 6 月于《众包: 大众力量缘何推

动商业未来》(Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business) 一书中首次提出并做了详细的阐释。书中关于众包的定义为:“一个公司或机构把过去由员工执行的工作任务,以自由自愿的形式外包给非特定的 (而且通常是大型的) 大众网络的做法。众包的任务通常是由个人来承担,但如果涉及到需要多人协作完成的任务,也有可能以依靠开源的个体生产的形式出现”。

众包是不准确监督的一种典型场景[1],包括机器学习、数据挖掘等在内的技术,通常需要大量带有标记的数据对算法以及模型进行训练、预测以及验证。数据数量越大,准确率越高,算法及模型的效果就越好。结合众包系统获取标签成本低、单个质量不可靠的特点,通常采取对同一实例由多个不同标记者进行标记,并用真值推理算法将所得的多重标签进行整合。基础的真值推理算法有多数投票 (MV),此方法要求保证工人的平均正确率大于 50%,则工人数量越多,标签质量越高。其他还有一些以投票法为基础的改进形式[2],但实验表明,仅仅通过真值推理的方式效果差别并不明显[3]。在上述真值推理算法的基础上,结合噪声识别和校正是一种行之有效的方案[4],但需要较高的标记质量 (如增加众包人数或者寻求更好的标记者),否则结果偏差较大。

从成本和质量两个角度考虑,我们可以由专家首先标记少部分实例,并适当减少众包工人人数和水平来达到成本和质量的最优解。针对这一场景,本文提出了一个适应性较强的基于特征扩维提高众包质量的框架 (FA-method),结合少量的专家标签,运用特征扩维的方法对真值推理算法所得的集成标签集进行处理,通过计算每个实例为噪声的可能程度以及集成标签的噪声数量的上限区分噪声,并进一步提高标签质量和可靠性。实验证明,本方法在不同标记质量下均可以取得较好的效果,在众包标记质量较低或众包工人较少的情况下,依然可以得到较高质量的标签。并且以校正后的数据集训练所得模型的泛化能力有进一步的提高。和现有的结合噪声识别和校正的框架相比也具有优势。

## 1　相关工作

通常,在人工智能领域,采用众包的方法获取标签有两个目的:其一是为了获得大量的带有标签的数据,专注于标签本身的质量,以所得标签与真值相比的准确率为度量;其二是获取更多数据以训练模型,专注于所得模型的质量,以模型的泛化能力为度量。

对于前者,常见思路是通过众包标签集推断真实标签,如上文所述,目前最常用且简单有效的方法为多数投票 (MV),即针对每个实例的多重标签集合,取多数为最终标签[5]。该方法的前提是工作人员的准确率需高于 50%,在此基础上,随着工作人员数量的增加,其正确率会不断提高。其他方法还有 GLAD 方法[6]、贝叶斯方法 RY[7]、ZenCrowd 算法[8] 等,一些对不同数据集进行实验证明,这些真值推理方法效果差异并不明显[3, 9]。因此,仅仅通过设计新颖的推理算法来提高标签质量比较困难。

对于后者,一种思路是在上面真值推理方法所得的集成标签基础上进一步识别出疑似噪声样本并在数据集中将其去掉。一项实验证明,简单去除噪声可以提高所训练模型的质量[10],因此一些噪声识别方法可以应用于此。例如,基于阈值的方法[11-12],用一些特殊的标准 (例如熵) 来对每个实例进行打分。如果分数超过阈值,则该实例将被视为噪声;基于模型预测的过滤算法,通过模型对实例进行分类并识别噪声,有如表决过滤 (VF)[13],对数据集进行多次交叉验证,若超过一半不相同则认为该实例为噪声样本。

目前常用的方法,或者标签质量不够理想,或者需要去除部分实例以保证质量。为达到既能保证质量,又能完整保留数据集全部实例的目的,Zhang 等首先提出了将噪声识别和校正相结合的方法 AVNC[4]。在集成标签的基础上,首先由多次交叉验证识别噪声,计算噪声可能程度和噪声数量范围以识别噪声,然后采用集成学习的方式进行校正。此方法在标记质量较高时有不错的效果,当标记质量较低时,仅仅通过交叉验证识别会产生较大的偏差,从而导致后续校正无法继续。

为解决上述问题,本文在 AVNC 所采用的噪声识别和校正相结合的基础框架上做了改进,引入了少量专家标记,并采用特征扩维的方法进行噪声识别和校正,保证了当标记质量较低时所得结果质量的稳定。

## 2　基于特征扩维的众包质量提高方法

### 2.1　基本框架

图 1 所示为本方法的基本框架。首先,由每

个众包工人标记的结果训练一个模型, 并对专家集进行预测, 所得结果作为专家集新的特征。之后用扩维后的专家集训练模型并对原始数据进行预测, 与集成标签相比较, 计算每个实例的集成标签是噪声的可能程度以及噪声数量。将集成标签划分为质量可靠的保留集和需要校正的含噪声集, 同时形成用于辅助校正的 $M$ 个辅助集。之后用辅助集对保留集扩维, 对含噪声集合进行校正, 最终将校正后的数据和保留集合并为最终结果。整个框架的关键在于噪声识别和噪声校正两个部分, 将在后文分别介绍。

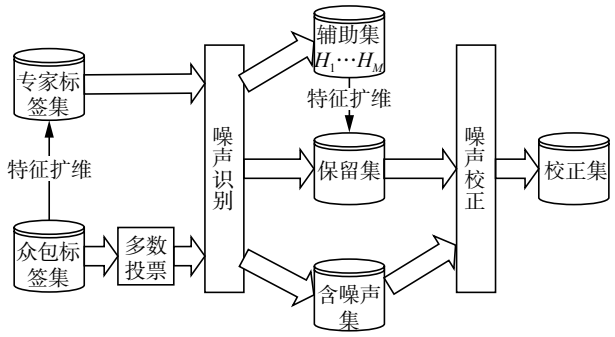



**图 1　基于特征扩维的众包质量提高方法的基本框架**

**Fig. 1　The framework of the feature-augmentation method of enhancing the labeling quality of crowdsourcing data**

## 2.2　特征扩维方法

本文所用的特征扩维方法类似于文献 [14], 是一种两层学习器结合的方法, 以噪声识别部分为例: 首先由众包数据训练出一组初级学习器, 然后输出所得的类概率作为专家数据集的新的特征, 专家数据集的标签仍作为新数据集的标签。这些增加的特征实质上可以打开原始输入空间的流形结构, 从而可以增强的专家数据集的分类性能。与由专家集直接训练模型相比, 其在泛化性能上能够有进一步的提高。

设 $J$ 个众包人员标记的众包标签集为 $C=\{(X_1,y_1),(X_2,y_2),\cdots,(X_J,y_J)\}$, 专家集 $P(X_P,y_P)$, 基础算法 $\mathcal{L}$, 原始数据集 $D(X_D)$。该方法的伪代码如下:

for $j=1,2,\cdots,J$ do:

$h_j=\mathcal{L}(X_j,y_j)$;

$y_{Pj}=h_j(X_P)$

$y_{Dj}=h_j(X_D)$

end for

$h'=\mathcal{L}((X_P,y_{P1},\cdots,y_{PJ}),y_P)$

**输出**: $z=h'(X_D,y_{D1},\cdots,y_{DJ})$

## 2.3　噪声识别部分

定义一个众包系统, 以下均针对二分类情况。设众包数据共有 $I$ 个实例, 每个实例均经过众包工人标记了 $J$ 个标签。对于每个实例 $i$, 其特征记为 $X_i$, $J$ 个标签集合记为 $\{l_i^1,l_i^2,\cdots,l_i^J\}$, 其中 $l_i^1\in\{-1,1\}$(分别代表负例和正例), 未知的真实标签记为 $y_i$, 经过标签集成算法处理后的集成标签记为 $\hat{y}_i$。这里需要明确的是, 每个工人的正确率需要大于 50%, 否则众包系统无法成立。在实际中, 众包工作者通常会经过初步的筛选, 以防止恶意标注者 (标注正确率低于 50%) 的出现。

对于每个标记者, 可以简单按式 (1) 估计出其错误率:

$$q_j=\sum_{i=1}^{I}I\left(l_i^j\neq\hat{y}_i\right)/I \tag{1}$$

对于每个实例 $i$, 由基础分类算法 (如决策树) 进行 $M$ 轮预测后会得到 $M$ 个分类标签 $\{l_i^1,l_i^2,\cdots,l_i^M\}$, 将其作为实例新的特征, 并和初始标签 $\hat{y}_i$ 进行对比, 计算不相同次数 $c_i$。

$$c_i=\sum_{m=1}^{M}I(l_i^m\neq\hat{y}_i) \tag{2}$$

式 (1) 和式 (2) 中函数 $I(\cdot)$ 是一个指示器函数, 如果括号内条件满足则返回 1, 不满足则返回 0。

第 $m$ 轮扩维将实例 $i$ 预测为正例的概率为 $c_1^m$, 预测为负例的概率为 $c_{-1}{}^m$。则实例 $i$ 的标签不确定程度可由式 (3) 来度量, 即:

$$e_i=-\sum_{m=1}^{M}\left[c_{-1}^{(m)}\log c_{-1}^{(m)}+c_1^{(m)}\log c_1^{(m)}\right] \tag{3}$$

下面定义一个量 $\alpha_i$ 来表示实例 $i$ 的集成标签, $\hat{y}_i$ 是噪声的可能程度:

$$\alpha_i=c_i+\left(e_i/\sum_{i=1}^{I}e_i\right) \tag{4}$$

$\alpha_i$ 是一种双层排序, 以不相同次数 $c_i$ 作为整数部分, 以不确定度做小数部分。不相同次数 $c_i$ 将标签集分为 $M$+1 组, $c_i$ 越大, 意味着有更多的模型将这个标签标记为噪声。在每组内又按照不确定程度进行排序, 不确定程度越大, 意味着越有可能是噪声。按 $\alpha_i$ 大小对所有实例进行排序, 显然 $\alpha_i$ 越大, 就意味着集成标签 $\hat{y}_i$ 越有可能是噪声。

接下来计算噪声可能的数量, 以期将集成标签进行划分。由于我们采用的是投票法对众包标签集进行初步整合, 为保证识别出的高质量集的

质量可靠，本着“宁缺毋滥”的原则，因此我们需要计算出噪声比例的上限。

由式 (1) 可计算贴标者错误率为 $q$，通过投票法整合 $J$ 个标记者，超过半数正确则集成结果正确。在这里，各个标记者错误率可视为相互独立，可由 Hoeffding 不等式推导出集成错误率的上限[15]：

$$Q=\sum_{k=0}^{\lfloor J/2\rfloor}\binom{J}{k}q^{J-k}(1-q)^k\leqslant \exp\left(-\frac{1}{2}J(1-2q)^2\right)=Q_{\max} \tag{5}$$

$$\text{noise\_num}=Q_{\max}\cdot I=\exp\left(-\frac{1}{2}J(1-2q)^2\right)\cdot I \tag{6}$$

按 $\alpha_i$ 大小对所有实例进行降序排序，我们可以将集成标签集分为两部分：前 noise_num 个标签为待进一步处理的含噪声集，其余为可靠的保留集。

### 2.4　噪声校正部分

目前，关于噪声校正的研究数量较少，且一些实验也证明校正噪声是比较困难的。一种直接的思路是，在分离出含噪声的集合后，用高质量集直接训练模型对噪声集进行校正，但效果不理想。

为了提高分离出的高质量集所训练的模型的泛化能力，我们再次使用扩维的方法：噪声识别过程中，每轮扩维预测后，每个实例 $i$ 均获得一个标签 $l_i$，若 $l_i=\hat{y}_i$，则将实例 $i$ 加入辅助集 $H_m$ 中，共获得 $M$ 个辅助集 $\{H_1,H_2,\cdots,H_M\}$。和识别部分类似，用之前得到的辅助集 $\{H_1,H_2,\cdots,H_M\}$ 分别训练得到模型 $\{h_C^1,h_C^2,\cdots,h_C^M\}$，对噪声集和保留集扩维并进行预测，重复 $M$ 轮，用投票法整合所得的 $M$ 个标签集合，作为对噪声集的校正结果。把校正后的结果和保留集合并为最终结果。

### 2.5　完整框架

该算法主要时间消耗在于对专家集扩维、对保留集扩维两个部分，且与所选择的基础算法 $\mathcal{L}$ 有关。设基础学习算法对实例数为 $n$ 的众包集训练及预测的时间复杂度为 $T(n)$，则本方法的时间复杂度为 $M[J\cdot T(n)+T(nr)]+M[M\cdot T(n')+T(n'')]$，其中，$M$ 为预设重复轮数，$J$ 为众包者数量，$r$ 为专家集比例，$n'$为辅助集实例数，$n''$为噪声集实例数，以上均为常数，且 $M\ll n$，$J\ll n$，$0<r<1$，$n'<n$，$n''<n$。由此可见本方法的时间复杂度取决于所选择的基础算法时间复杂度。完整流程图如图 2 所示。

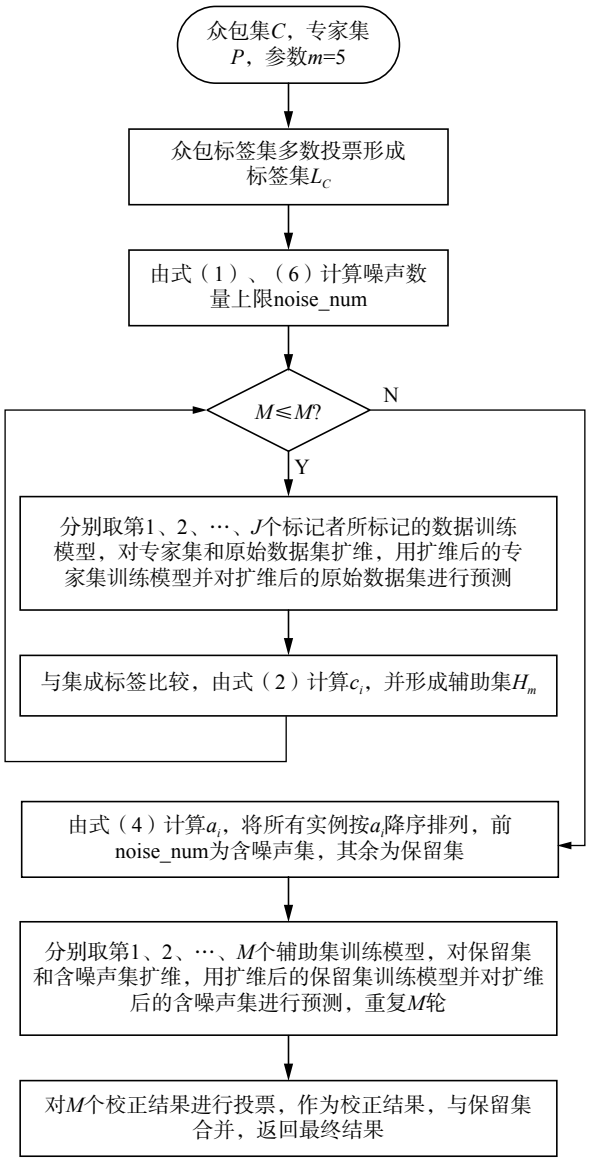



**图 2　算法流程图**

**Fig. 2　Flowchart of the proposed method**

## 3　实验结果分析

由上文所述，众包通常关注于两个结果：1) 标签本身质量；2) 训练所得模型的质量。因此在这里分别进行实验。基础的学习模型均采用决策树，由 python 的 sklearn 库实现，参数均取默认值，取 $M$=5。

### 3.1　实验数据

实验数据来自 UCI 机器学习库的 8 个数据集，它们具有不同的数量的实例，不同的类分布，不同数量的特征及其类型，以便验证本方法在不同情况下的适用性。其中 4 个为较小规模数据集，4 个为较大规模数据集。在模拟实验中不对数据集本身做任何特征处理。数据集具体情况如表 1。

**表 1　8 个 UCI 数据集的基本情况**

**Table 1　Basic conditions of 8 UCI datasets**

| 名称 | 数量 | 正例 | 负例 | 特征数量 | 特征类型 |
|---|---|---|---|---|---|
| mushroom | 8 124 | 3 916 | 4 280 | 23 | 类别 |
| kr-vs-kp | 3 196 | 1 527 | 1 669 | 37 | 类别 |
| spambase | 4 601 | 1 813 | 2 788 | 58 | 数值 |
| sick | 3 772 | 231 | 3 541 | 30 | 混合 |
| biodeg | 1 055 | 356 | 699 | 42 | 数值 |
| tic-tac-toe | 958 | 332 | 626 | 10 | 类别 |
| vote | 435 | 168 | 267 | 17 | 类别 |
| ionosphere | 351 | 126 | 225 | 35 | 数值 |

首先, 在较大规模数据集取 $r$=0.05 的数据, 较小规模数据集取 $r$=0.1 的数据作为专家标签集, 以他们的真实标签作为专家标记的标签。接下来模拟众包的过程: 创建一个模拟标记者, 为数据集中的每个实例标记一次。然后, 第二个模拟标记者执行相同的任务, 直到第 $J$ 个模拟标记者完成同样的任务。最终, 使每个实例都包含 $J$ 个标签。模拟标记者的标记质量均匀分布, 且所有的模拟标记者都有不同的标记质量。所有实例以及它们对应的多个噪声标签集形成一个模拟众包标签数据集。默认 $J$ 取 10, 平均质量较高的情况下, 取模拟标记者质量范围为 [0.6,0.8], 平均为 0.7; 平均质量较低的情况下, 取模拟标记者质量范围为 [0.5,0.7], 平均为 0.6, 每个实例的集成标签由多数投票产生。

### 3.2　准确率对比

众包的一个目的在于获得大量数据的标签, 因此需要考察众包处理后的标签和真实标签的准确率。 由于 AVNC 是首次被提出将噪声识别和校正组合来提升众包质量的方法, 且实验证明其可以进一步提高标签集成算法的质量。因此本次实验选择多数投票为基础标签集成算法, 在其基础上比较 AVNC 和 FA-method 对于众包质量的提高程度。实验方法如下:

1) 直接由多数投票形成的集成标签 (MV);

2) 多轮交叉验证, 计算噪声数和不相同次数划分噪声集, 采用集成学习方法进行校正 (AVNC);

3) 由特征扩维识别并校正噪声 (FA-method)。

每种方法重复实验 10 次, 每次随机取奇数个模拟标记人员所标记的标签 (避免多数投票出现随机值), 对比三者所得到的标签和真实值相比的准确率以及标准差。

图 3 和图 4 分别是平均标记准确率为 0.7 和 0.6 时 4 种方法的准确率。

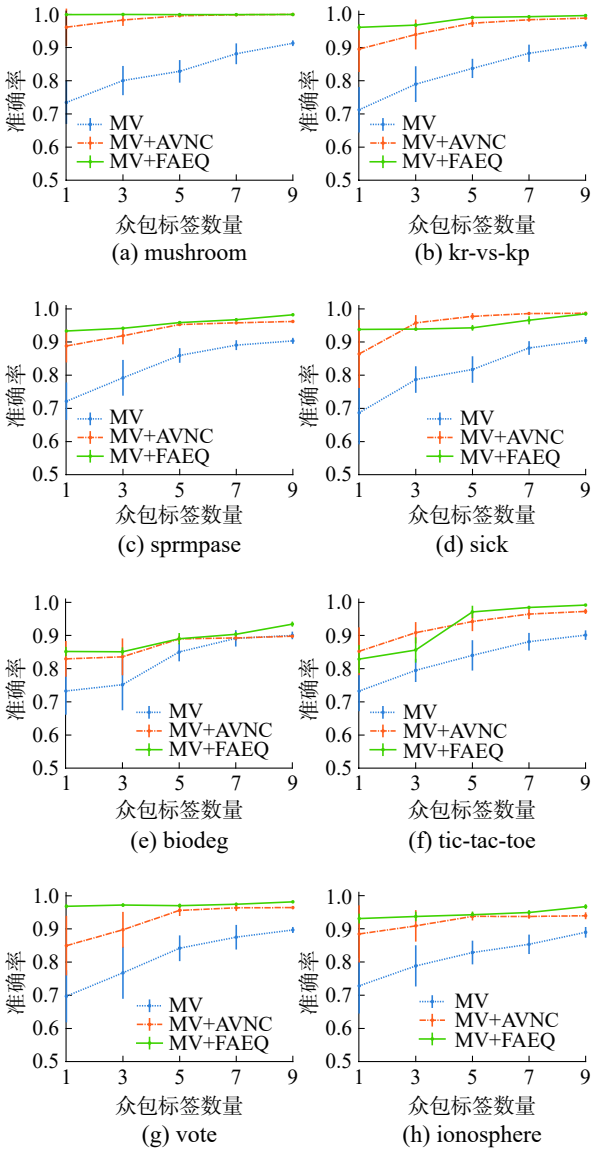



**图 3　高质量标记时众包准确率**

**Fig. 3　Accuracy of crowdsourcing on high quality labeling**

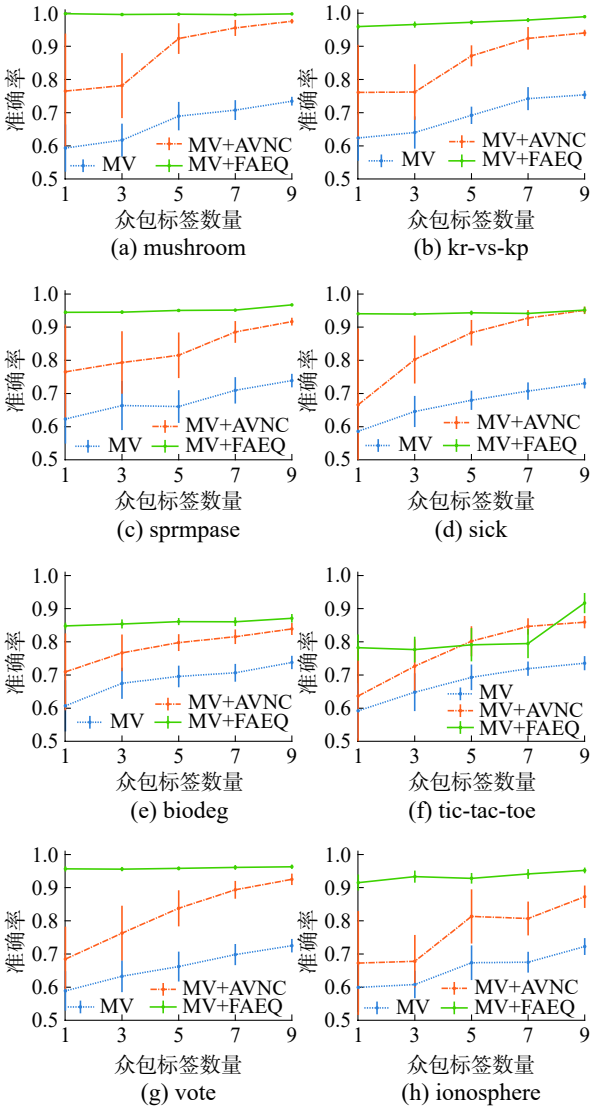



**图 4　低质量标记时众包准确率**

**Fig. 4　Accuracy of crowdsourcing on low quality labeling**

由图 3 和图 4, 可以得出以下结论:

1) 和多数投票所得集成标签相比, 本方法能够在其基础上较大幅度提高标签的质量;

2) 和现有方法相比, 除 sick 和 tic-tac-toe 数据集外, 本方法无论在标记质量较高或较低、标记人数多或少的情况下所得结果均较高;

3) 本方法在标记人数 3 人以下时和其他方法相比有明显优势, 3 人以上时略优于其他方法, 因此可以在适当减少标记人数的情况下保证标记质量;

4) 对比图 3 和图 4 可见, 当标记质量较低时, 对比方法质量明显下降, 而本方法下降幅度与之相比则下降不很明显;

5) sick 和 tic-tac-toe 数据集效果不佳, 其原因在于数据集分布不平衡, 随机选取的专家集可能有严重的偏向, 从而影响识别和校正的性能。

### 3.3　训练模型对比

众包的另一个目的, 是为机器学习的大量数据集进行标注, 以提高模型的泛化能力。因此还需要考察的是修正噪声后的数据集和简单去除噪声的数据集相比, 生成的模型的识别能力是否有所提高。这里选择 VF 进行噪声识别。实验方法如下: 对每一个数据集, 划分 70% 为训练集, 30% 为验证集, 其中训练集按前文方法进行校正, 比较下列 3 种方法:

1) 将众包标签集直接采用投票方法融合, 训练模型对验证集预测;

2) 采用 VF 算法, 多轮交叉验证, 大于一半不相同视为噪声, 直接去除掉噪声, 剩余数据训练模型对验证集预测;

3) 采用 FA 识别并校正噪声, 将校正过的数据重新加入数据集并训练模型对验证集预测。

评价标准为 AUC, 即 ROC 曲线下面积。每种方法重复实验 10 次, 每次随机取奇数个模拟标记人员所标记的标签, 比较其标准差和平均值。

图 5 和图 6 分别是平均标记准确率为 0.7 和 0.6 时 4 种方法的所得标签训练模型的效果对比。

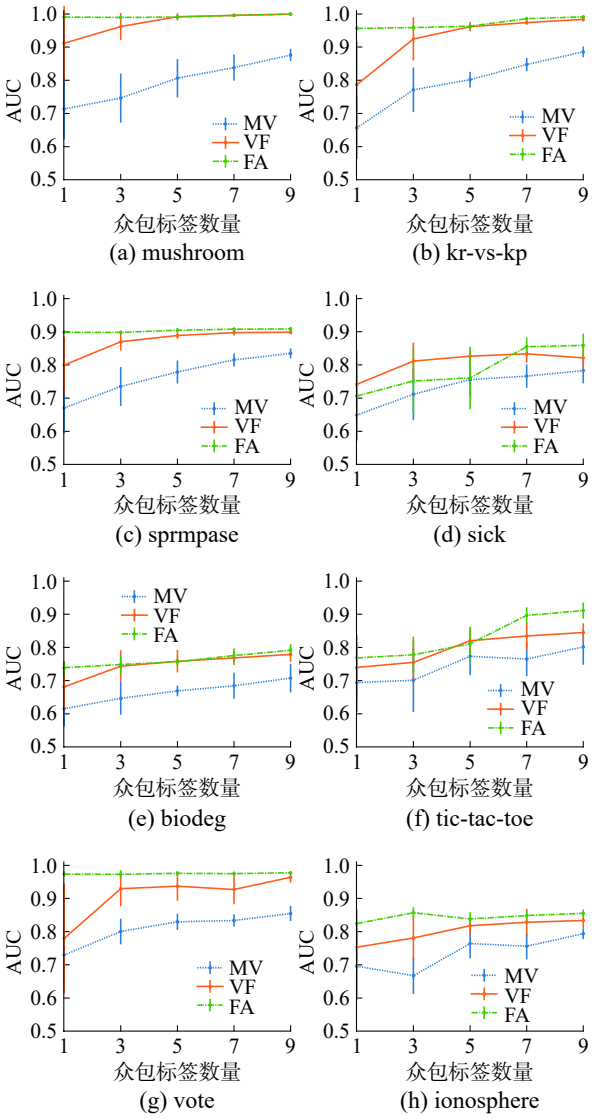



**图 5　高质量标记时众包训练模型 AUC**

**Fig. 5　AUC of model trained by crowdsourcing on high quality labeling**

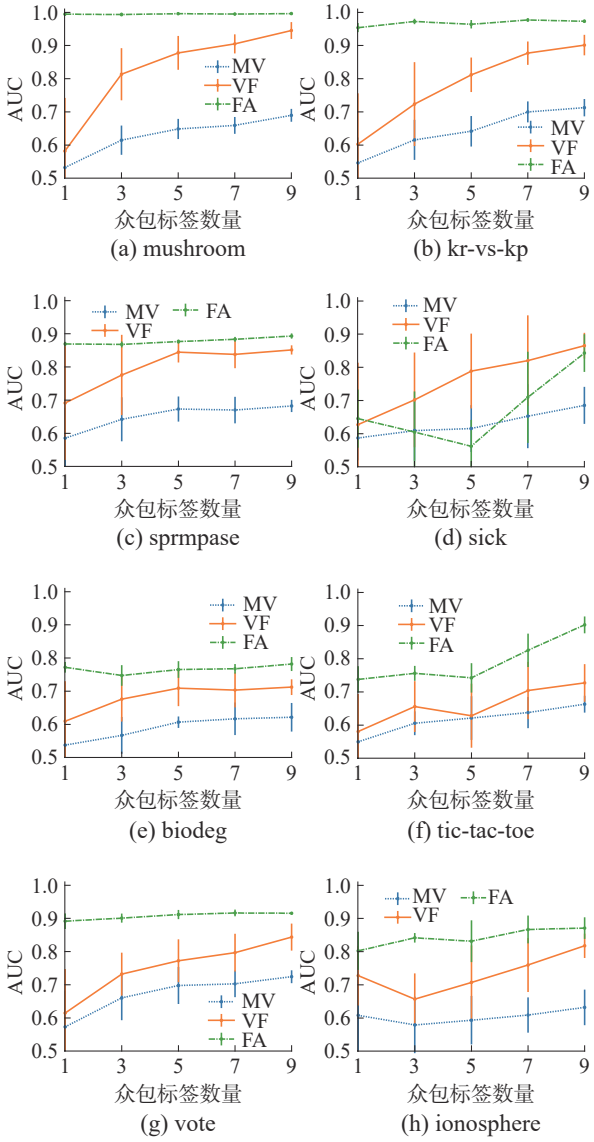



**图 6　低质量标记时众包训练模型 AUC**

**Fig. 6　AUC of model trained by crowdsourcing on high quality labeling**

由图 4 和图 5，有以下观察结果：

1) 简单去除噪声可以提高所训练模型的质量，这验证了文献 [9] 的结论；

2) 和简单去除噪声数据相比，本方法提供了将噪声实例校正并重新加入数据集的机会，从而能够训练出更好的模型。

3) 对比图 5 和图 6，本方法在标记质量下降时，所得众包结果训练模型的 AUC 降幅不大，证明本方法在标记质量较差时能够保持结果的稳定。

4) sick 数据集结果不佳，原因在于其数据分布极不均衡，随机选择的专家集可能会出现严重的偏向，从而影响识别和校正效果。

### 3.4　不平衡数据集

针对 sick 和 tic-tac-toe 等不平衡数据集出现的性能不佳的情况，有如下实验：在选择专家集时，尽量保证正负例数量基本一致，其余条件不变，实验结果如下。

首先是准确率对比，图 7 和图 8 为两数据集分别在高质量和低质量标记情况下校正的准确率。

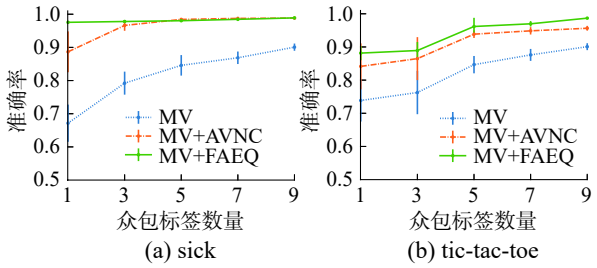



**图 7　高质量标记时众包准确率**

**Fig. 7　Accuracy of Crowdsourcing on high quality labeling**

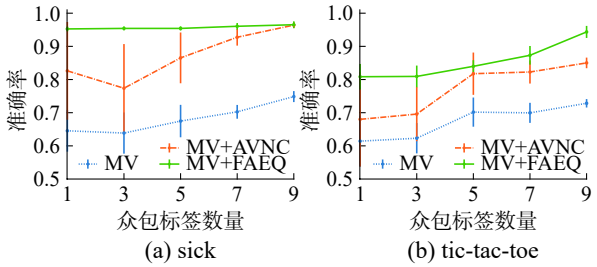



**图 8　低质量标记时众包准确率**

**Fig. 8　Accuracy of Crowdsourcing on low quality labeling**

其次是训练模型对比，图 9 和图 10 为两数据集分别在高质量和低质量标记情况下所得校正结果训练模型的 AUC。

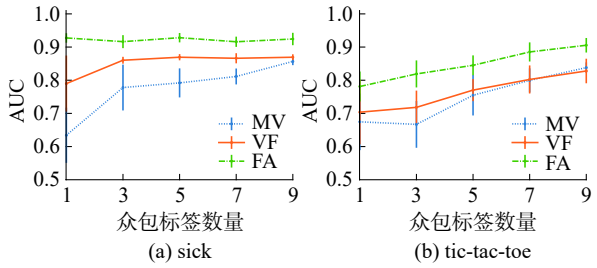



**图 9　高质量标记时众包训练模型 AUC**

**Fig. 9　AUC of model trained by Crowdsourcing on high quality labeling**

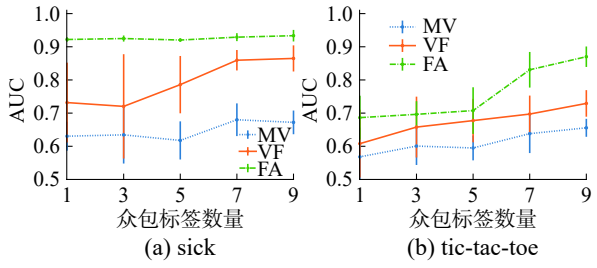



**图 10　低质量标记时众包训练模型 AUC**

**Fig. 10　AUC of model trained by Crowdsourcing on high quality labeling**

由上述实验可知，专家集的选取对本方法有一定影响，在处理不平衡数据集时，随机选择的专家集可能会导致样本偏差较大从而影响到最终的结果，人为选择使得专家集正负样本基本一致会改进这种情况。

## 4　结论

1) 本框架在工作人员数量较低或标记质量较低时均能取得不错的效果，且和增加标记人数或提高标记质量所得结果差异不大。换句话说，可以在适当降低成本的同时获得更高质量的结果；

2) 和现有识别和验证框架比，由于引入专家标签，使得在标记质量较低时也能够取得不错的效果，且标记质量较高时准确率能够进一步提高；

3) 提供了将噪声实例校正并重新加入数据集的机会。

本方法只适用于二分类标签，而扩展到多分类的情况时会变得较为复杂，结果偏差会变大。另外，本方法对于分布极不均衡的数据集效果略差，如何应对也需要做进一步的研究。

## 参考文献：


[1] ZHOU Zhihua. A brief introduction to weakly supervised learning[J]. National science review, 2018, 5(1): 44–53.

[2] HU Huiqi, ZHENG Yudian, BAO Zhifeng, et al. Crowdsourced POI labelling: location-aware result inference and task assignment[C]//Proceedings of 2016 IEEE 32nd International Conference on Data Engineering. Helsinki, Finland, 2016: 61–72.

[3] RODRIGUES F, PEREIRA F C, RIBEIRO B. Gaussian process classification and active learning with multiple annotators[C]//Proceedings of the 31st International Conference on International Conference on Machine Learning. Beijing, China, 2014: II–433–II–441.

[4] ZHANG Jing, SHENG V S, LI Tao, et al. Improving crowdsourced label quality using noise correction[J]. IEEE transactions on neural networks and learning systems, 2018, 29(5): 1675–1688.

[5] IPEIROTIS P G, PROVOST F, SHENG V S, et al. Repeated labeling using multiple noisy labelers[J]. Data mining and knowledge discovery, 2014, 28(2): 402–441.

[6] WHITEHILL J, RUVOLO P, WU Tingfan, et al. Whose vote should count more: optimal integration of labels from labelers of unknown expertise[C]//Proceedings of the 22nd International Conference on Neural Information Processing Systems. Vancouver, British Columbia, Canada, 2009: 2035–2043.

[7] RAYKAR V C, YU Shisheng, ZHAO L H, et al. Learning from crowds[J]. Journal of machine learning research, 2010, 11: 1297–1322.

[8] DEMARTINI G, DIFALLAH D E, CUDRÉ-MAUROUX P. ZenCrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking[C]//Proceedings of the 21st International Conference on World Wide Web. Lyon, France, 2012: 469–478.

[9] MUHAMMADI J, RABIEE H R, HOSSEINI A. A unified statistical framework for crowd labeling[J]. Knowledge and information systems, 2015, 45(2): 271–294.

[10] FRENAY B, VERLEYSEN M. Classification in the presence of label noise: a survey[J]. IEEE transactions on neural networks and learning systems, 2014, 25(5): 845–869.

[11] GAMBERGER D, LAVRAČ N, DŽEROSKI S. Noise elimination in inductive concept learning: a case study in medical diagnosis[C]//Proceedings of the 7th International Workshop on Algorithmic Learning Theory. Sydney, Australia, 1996: 199–212.

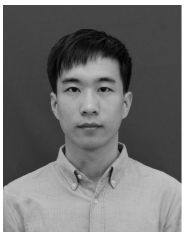
[12] SUN Jiangwen, ZHAO Fengying, WANG Chongjun, et al. Identifying and correcting mislabeled training instances[C]//Proceedings of Future Generation Communication and Networking. Jeju, South Korea, 2007: 244–250.

[13] BRODLEY C E, FRIEDL M A. Identifying mislabeled training data[J]. Journal of artificial intelligence research, 1999, 11(1): 131–167.

[14] ZHOU Ta, ISHIBUCHI H, WANG Shitong. Stacked-structure-based hierarchical Takagi-Sugeno-Kang fuzzy classification through feature augmentation[J]. IEEE transactions on emerging topics in computational intelligence, 2017, 1(6): 421–436.

[15] ZHOU Zhihua. Ensemble methods: foundations and algorithms[M]. Boca Raton: Taylor & Francis, 2012.


## 作者简介：

李易南，硕士研究生，主要研究方向为人工智能与模式识别。

王士同，教授，博士生导师，主要研究方向为人工智能与模式识别。发表学术论文近百篇。