

# 利用标注者相关性的深度生成式众包学习<sup>\*</sup>

李绍园, 韦梦龙, 黄圣君



(南京航空航天大学 计算机科学与技术学院/人工智能学院, 江苏 南京 211106)

通信作者: 李绍园, E-mail: lisy@nuaa.nju.edu.cn

**摘要:** 传统监督学习需要训练样本的真实标记信息, 而在很多情况下, 真实标记并不容易收集. 与之对比, 众包学习从多个可能犯错的非专家收集标注, 通过某种融合方式估计样本的真实标记. 注意到现有深度众包学习工作对标注者相关性建模不足, 而非深度众包学习方面的工作表明, 标注者相关性建模利用有助于改善学习效果. 提出一种深度生成式众包学习方法, 以结合深度神经网络优势及利用标注者相关性. 该模型由深度神经网络分类器先验和标注生成过程组成, 其中, 标注生成过程通过引入各类别内标注者能力的混合模型以建模标注者相关性. 为自适应地匹配数据及模型复杂度, 实现了完全贝叶斯推断. 基于结构变分自编码器的自然梯度随机变分推断技术, 将共轭参数变分消息传递与神经网络参数随机梯度下降结合到统一框架, 实现端到端的高效优化. 在 22 个真实众包数据集上的实验结果验证了该方法的有效性.

**关键词:** 众包学习; 深度生成式模型; 标注者相关性; 贝叶斯; 自然梯度随机变分推断

**中图法分类号:** TP18

中文引用格式: 李绍园, 韦梦龙, 黄圣君. 利用标注者相关性的深度生成式众包学习. 软件学报, 2022, 33(4): 1274–1286. <http://www.jos.org.cn/1000-9825/6479.htm>

英文引用格式: Li SY, Wei ML, Huang SJ. Deep Generative Crowdsourcing Learning with Worker Correlation Utilization. Ruan Jian Xue Bao/Journal of Software, 2022, 33(4): 1274–1286 (in Chinese). <http://www.jos.org.cn/1000-9825/6479.htm>

## Deep Generative Crowdsourcing Learning with Worker Correlation Utilization

LI Shao-Yuan, WEI Meng-Long, HUANG Sheng-Jun

(College of Computer Science and Technology/College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

**Abstract:** Traditional supervised learning requires the ground truth labels for the training data, which can be difficult to collect in many cases. In contrast, crowdsourcing learning collects noisy annotations from multiple non-expert workers and infers the latent true labels through some aggregation approach. This study notices that existing deep crowdsourcing work do not sufficiently model worker correlations, which however is shown to be helpful for learning by previous non-deep learning approaches. A deep generative crowdsourcing learning model is proposed to combine the strength of deep neural networks (DNN) and at the same time exploit the worker correlations. The model comprises a DNN classifier as a priori for the true labels, and one annotation generation process in which a mixture model of workers' reliabilities within each class is introduced for inter-worker correlation. To automatically trade-off between the model complexity and data fitting, fully Bayesian inference is developed. Based on the natural-gradient stochastic variational inference techniques developed for structured variational autoencoder (SVAE), variational message passing is combined for conjugate parameters and stochastic gradient descent for DNN under a unified framework to conduct efficient end-to-end optimization. Experimental results on 22 real world crowdsourcing data sets demonstrate the effectiveness of the proposed approach.

**Key words:** crowdsourcing learning; deep generative model; worker correlations; Bayesian; natural-gradient stochastic variational inference

\* 基金项目: 国家自然科学基金(61906089); 江苏省基础 Research 计划(BK20190408); 中国博士后基金(2019TQ0152)

本文由“面向开放场景的鲁棒机器学习”专刊特约编辑陈恩红教授、李宇峰副教授、邹权教授推荐.

收稿时间: 2021-05-31; 修改时间: 2021-07-16; 采用时间: 2021-08-27; jos 在线出版时间: 2021-10-26

作为机器学习中研究最多、应用最广的一种学习范式, 监督学习需要训练样本及其标记信息. 然而在很多现实任务中, 收集真实标记并不容易: 一方面, 真实标记需要领域专家标注, 而专家资源通常很昂贵并且是有限的; 另一方面, 当样本较多时, 标注非常耗时耗力. 随着众包平台比如 AMT (Amazon mechanical turk), Crowdsflower 的出现, 众包提供了一种更容易获取标记的方式. 通过将标注任务分配给网络上容易访问的非专家<sup>[1]</sup>, 众包可以快速地收集大量的监督信息, 在自然语言处理<sup>[2]</sup>、医学诊断<sup>[3]</sup>、图像识别<sup>[4]</sup>和命名实体识别<sup>[5]</sup>等领域得到广泛的应用.

由于标注来自非专家, 为了降低数据错误, 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真实标记. 根据是否使用深度神经网络分类器模型, 现有众包方法可以分为非深度众包学习和深度众包学习. 多数投票法是一种常用的仅利用标注信息的方法, 其对每个样本采用被标注次数最多的类别作为真实标记. 多数投票法简单快速, 容易实现, 但是忽视了不同标注者的能力差异. 通过将真实标记看作未知的隐变量, 针对不同标注者使用不同参数刻画其标注能力, 基于概率图模型的方法被提出. 作为这方面的早期代表性工作, Dawid 和 Skene<sup>[6]</sup>提出了 DS 模型, 其使用分类准确率刻画标注者能力, 并基于期望最大 (expectation maximization, EM) 优化方法迭代地估计真实标记与标注者准确率, 用于解决从多个非专家(医学生)的医疗诊断中估计更可靠的诊断结论的问题. 在这之后, 很多工作从标注者能力刻画、样本困难程度刻画、优化实现、标注相关性建模等方面对 DS 模型进行改进和扩展, 并取得了良好的效果<sup>[7-12]</sup>.

为了利用样本特征的信息, Raykar 等人<sup>[3]</sup>在 DS 模型中引入逻辑回归分类器作为真实标记的先验, 并基于 EM 迭代式地求解分类器及标注者参数. 这一思路随后被自然地扩展到其他类型的分类器先验, 例如高斯过程分类器<sup>[13]</sup>. 随着深度学习在视觉语音等多个领域取得重大进展<sup>[14]</sup>, 引入神经网络以结合其表示学习优势的深度众包学习成为众包领域的研究趋势<sup>[15-17]</sup>. 由于 EM 式的优化需要在每次迭代中最优化分类器, 当神经网络模型比较复杂时, 其计算复杂度无疑很高<sup>[15]</sup>. 因此, 高效的优化是深度众包学习的一个关注重点. 解决该问题的一种方式是从神经网络多层结构角度, 在神经网络分类器的输出层之后增加一层系数刻画标注者能力, 其最终输出层对应众包标注预测<sup>[16,17]</sup>. 优化则可以采用端到端方式, 即在众包标注的预测损失上对网络中所有参数进行随机梯度下降, 同时估计分类器和标注者能力. 该实现虽然避免了迭代式优化的计算复杂度, 但丢失了概率图模型的可解释结构化表示性质, 同时也无法保证最大化标注似然或者标注似然的下界.

本文基于深度生成式模型进行众包学习, 同时保留深度神经网络的表示学习优势及概率图模型的结构化表示形式. 这方面的工作大多是基于变分自编码器模型及其改进<sup>[18-20]</sup>, 目前还比较少<sup>[21-25]</sup>, 并且忽视了标注者之间的相关性利用. 而已有的非深度众包工作表明<sup>[10-12]</sup>, 标注者相关性建模及利用有助于改善众包学习效果. 本文提出一种利用标注者相关性的深度生成式众包学习方法, 通过扩展文献[12]中的子类混合模型刻画标注者相关性, 使用深度神经网络作为分类器. 该模型在每个子类上使用混淆矩阵描述标注者能力及众包标注生成过程, 子类混合模型、神经网络分类器及标注生成过程共享真实标记隐变量. 为避免手动参数挑选, 自适应地拟合参数, 本文实现了完全贝叶斯模型, 使用概率分布描述模型参数. 为了实现贝叶斯推断, 基于结构变分自编码器优化技术<sup>[20]</sup>, 本文结合共轭参数变分消息传递及自然梯度变分推断, 端到端地对所有参数进行高效更新, 整个优化过程可以保证在最大化众包标注的似然下界. 在 22 个真实众包分类数据集上与多个对比方法的实验结果表明, 深度表示学习及标注者相关性结合可以有效提升众包学习效果. 本文的贡献总结如下: (1) 首次提出了利用标注者相关性的深度生成式众包学习方法, 兼具表示学习优势及可解释性; (2) 实现了高效端到端自然梯度随机变分推断, 其时间复杂度与真实标记下训练深度神经网络分类器相当; (3) 大量真实数据集上的实验结果验证了所提方法的有效性.

本文第 1 节介绍相关工作. 第 2 节给出众包学习问题的形式化描述及相关背景. 第 3 节和第 4 节分别提出本文方法和优化实现. 第 5 节给出实验结果. 最后总结全文.

## 1 相关工作

由于众包标注可能存在错误, 如何估计真实标记是众包学习的研究重点之一. 作为一种直接的思路, 多

数投票法使用被标注次数最多的类别作为真实标记. 由于其假设多数标注正确, 当错误较严重时, 多数投票法效果并不理想. 通过将真实标记看作未知的隐变量, 对标注生成过程建模, 概率图模型提供了另一种研究思路. 作为这方面的早期代表性工作, DS 模型<sup>[6]</sup>提出使用准确率刻画各个标注者能力, 基于最大标注似然的目标, 使用 EM 方法迭代式地估计标注者准确率和真实标记. 后续很多工作对 DS 模型进行了改进和扩展, 例如: 文献[7]从优化角度提出了包括信念传播和均值场模型在内的变分推断优化实现; 文献[8]引入了样本难度参数, 使得标注质量可以同时与标注者及样本相关; 文献[9]为每个样本使用一个混淆矩阵参数刻画标注者的标注质量, 并基于最小最大化熵原则估计样本真实标记及参数. 文献[10,26,27]从贝叶斯角度对 DS 模型进行了扩展, 引入准确率参数的狄利克雷先验, 并分别基于吉布斯采样、变分推断和 EM 实现贝叶斯推断, 以避免手动参数挑选. 最近, 考虑到众包问题中标注之间常常存在相关性, 对标注者之间的相关性建模及利用吸引了研究者的关注. 文献[10]利用无向的马尔科夫网络描述任意两个标注之间的依赖关系; 文献[11]基于标注者聚类假设, 从理论上对基于混淆矩阵的众包模型进行了最小最大错误率分析; 文献[12]通过为真实类别建立子类混合模型, 反映众包标注的张量结构, 以描述标注者的相关性.

为了利用样本特征以帮助众包学习, 文献[3]提出使用特征到真实标记的逻辑回归分类器作为 DS 模型中真实标记的先验假设. 这一思路随后被扩展到其他类型的分类器模型, 如高斯过程分类器<sup>[13]</sup>. 随着深度学习的兴起, 使用深度神经网络作为分类器模型的深度众包学习成为众包领域的研究趋势<sup>[15-17]</sup>. 文献[15]利用卷积神经网络作为分类器, 并基于 EM 进行迭代式优化, 每次迭代中求解当前标注者参数下最优的神经网络分类器. 为了避免 EM 算法的计算开销, 文献[16,17]从神经网络构造的角度, 在分类器输出之后增加一层系数作为标注者能力参数, 从而将分类器参数和标注者能力参数看作网络的不同层参数, 端到端地进行随机梯度下降更新. 文献[16,17]虽然避免了 EM 优化的高计算复杂度, 但一方面, 其神经网络缺乏概率图模型的可解释结构; 另一方面, 不能保证最大化标注似然或者标注似然的下界.

本文借鉴深度生成式模型及其优化技术发展, 主要是变分自编码器模型<sup>[18]</sup>及结构变分自编码器<sup>[20]</sup>, 提出了本文的深度生成式众包学习方法. 变分自编码器<sup>[18]</sup>是一种典型的深度生成式模型, 其使用神经网络学习样本隐空间表示并从隐空间重构样本原空间. 通过使用概率分布描述隐空间, 变分自编码器可以在隐空间进行数据随机采样并重构生成新的样本, 在数据生成方面具有广泛的应用价值, 被视为无监督学习领域的重要研究方法. 使用重参数化技巧, 变分自编码器用神经网络拟合概率分布参数, 并基于变分推断优化实现参数的端到端高效优化, 为深度生成式模型提供了优化框架. 结构变分自编码器<sup>[20]</sup>设计并利用了概率图模型的共轭结构, 将分布表示为指数族分布形式, 并使用神经网络拟合指数族分布的参数, 从而可以使用传统生成式模型, 如话题模型<sup>[28]</sup>基于共轭分布结构的快速贝叶斯推断方法. 注意到, 文献[21-25]也基于变分自编码器及其扩展模型提出了深度生成式的众包学习方法. 例如: 文献[21]基于变分自编码器<sup>[18]</sup>最小化众包标注的重构错误; 文献[22-24]基于半监督变分自编码器<sup>[19]</sup>分别提出了利用无标记数据的半监督众包分类及聚类学习方法; 文献[25]基于结构化变分自编码器优化<sup>[20]</sup>提出了一种全贝叶斯的深度生成式众包分类方法, 但其均基于标注者条件独立假设, 并未考虑标注者相关性建模. 本文基于结构变分编码器的优化技术<sup>[20]</sup>, 将文献[12]的子类混合模型扩展到利用样本特征学习分类器的深度生成式模型, 实现了高效的贝叶斯参数推断.

上述相关工作主要集中于本文关注的单标记分类任务的标注融合, 其他领域的广泛需求也催生了众多的众包研究问题, 例如多标记众包学习<sup>[29]</sup>、基于众包学习的交互式特征选择<sup>[30]</sup>、可信众包机制研究<sup>[31]</sup>等, 这里不做详细阐述.

## 2 问题形式化及相关背景

### 2.1 问题形式化

本文使用  $X=\{x_1, \dots, x_N\}$  表示  $N$  个样本的集合, 其中,  $x_i \in \mathbf{R}^d$  表示第  $i$  个样本,  $d$  表示样本特征空间维度. 用  $W$  表示标注者个数, 样本集  $X$  上的众包标注结果可以表示为  $L \in \{0, 1, \dots, K\}^{N \times W}$ , 其中,  $K$  表示类别个数,  $L_{ij}$  表示第  $j$  个标注者在样本  $x_i$  上的标注.  $L_{ij}=k$  ( $k \neq 0$ ) 表示标注者  $j$  将  $x_i$  标注为属于第  $k$  个类别;  $L_{ij}=0$  表示样本  $x_i$  没有被第

$j$  个标注者标注, 即标注缺失. 在众包数据中, 通常每个标注者仅标注部分样本数据, 所以标注往往存在缺失. 众包学习目标为从标注  $L$  中估计样本  $X$  的真实类别  $Y=\{y_1, \dots, y_N\}$ .

接下来介绍经典的基于标注者独立假设的众包学习模型, 然后提出本文的模型.

### 2.2 标注者独立模型

图 1 所示为经典的基于标注者独立性假设的众包生成过程. 为描述方便, 与本文贝叶斯框架保持一致, 这里以文献[10]中的独立贝叶斯分类器组合模型(independent Bayesian classifier combination, iBCC)为例.

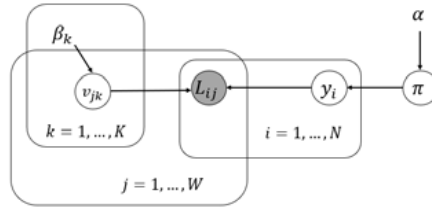


图 1 独立贝叶斯分类器组合模型的概率图表示

该模型不考虑特征利用, 由两部分组成: 标注生成过程  $p(L_{ij}|y_i, V_j)$  及真实标记先验  $p(y_i | \pi)$ ,  $V_j = \{v_{jk}\}_{k=1}^K$  表示标注者  $j$  对应的标注过程参数,  $v_{jk}=[v_{jk1}, \dots, v_{jkK}] \in [0,1]^{K*1}$  表示真实标记  $y_i=k$  时样本被标注为各个类别的概率,  $\pi$  表示真实标记的先验分布参数.  $\beta_k, \alpha$  分别对应  $v_{jk}, \pi$  的分布参数. 该模型假设标注者条件独立, 即给定某个样本  $x_i$  及其真实类别  $y_i$ , 众包标注互相独立, 以  $y_i=k$  为例:

$$p(L_{i1}, \dots, L_{iW} | y_i = k, \{V_j\}_{j=1}^W) = \prod_{j=1}^W \mathcal{I}(L_{ij} \neq 0) p(L_{ij} | y_i = k, v_{jk}) \tag{1}$$

这里,  $\mathcal{I}(\cdot)$  为指示函数, 括号内条件满足时取值为 1, 不满足时为 0. 假设样本独立, 则众包标注  $L$ , 真实标记  $Y$  以及参数  $V=\{v_{jk}\}, \pi$  的联合分布可以写为

$$p(L, Y, V, \pi) = p(L | Y, V) p(Y | \pi) p(\pi) p(V) = p(\pi) \prod_{i=1}^N p(y_i | \pi) \prod_{j=1}^W \mathcal{I}(L_{ij} \neq 0) p(L_{ij} | y_i, v_{jk}) \cdot \prod_{k=1}^K p(v_{jk}) \tag{2}$$

图 1 的模型可以看作 DS 模型<sup>[6]</sup>的贝叶斯扩展, 对比 DS 模型求解参数  $V, \pi$  的点估计, 贝叶斯模型估计参数  $V, \pi$  的后验分布, 可以描述更多不确定性. 针对图 1 模型, 现有工作分别提出基于吉布斯采样<sup>[10]</sup>、贝叶斯变分均值场推断<sup>[26]</sup>及 EM<sup>[27]</sup>等推断技术的参数估计方法.

现有大多数众包学习算法基本均是基于标注者条件独立性假设. 为了刻画标注者相关性, 文献[10]提出利用无向的马尔科夫网描述任意两个标注之间的依赖关系, 但其无法处理样本标注存在缺失的情况; 文献[12]提出为真实类别建立子类混合模型, 标注者在给定子类上条件独立, 以描述标注者的相关性. 本文基于文献[12]的子类混合模型, 将其扩展到利用样本特征学习分类器的深度生成式模型, 具体实现见下节.

### 3 利用标注者相关性的深度生成式众包学习方法

图 2 所示为本文所提方法的概率图模型结构.

与图 1 类似, 该模型主要包含两个组成部分: 标注生成过程  $p(L_{ij}|y_i, g_i, V_j)$  及真实标记先验  $p(y_i|x_i, \gamma, \pi)$ . 与图 1 模型不同: 该模型真实标记先验依赖于样本特征, 本文使用参数为  $\gamma$  的深度神经网络分类器刻画依赖过程; 此外, 标注生成过程引入了额外的隐变量  $g_i \in \{1, \dots, M\}$  以实现标注者相关性刻画,  $M$  为隐变量  $g_i$  的取值范围. 参数  $v_{jkm} \in [0,1]^{K*1}$  对应隐变量  $\{y_i=k, g_i=m\}$  时样本被标注者  $j$  标注为各个类别的概率,  $V_j = \{v_{jkm}\}_{k,m}, k=1, \dots, K; m=1, \dots, M$ . 下面给出模型各部分的具体含义.

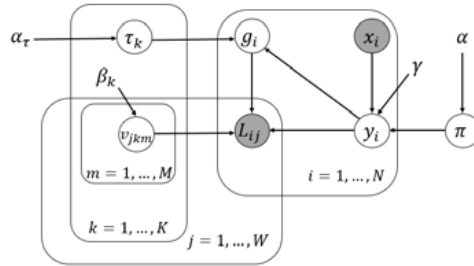


图 2 本文所提方法的概率图表示

• 标注生成过程

与公式(1)中假设真实标记给定时,各标注者互相独立不同,文献[12]将每个类别刻画为  $M$  个子类的混合模型,即数据集共被划分为  $K * M$  个子类,在每个子类上,各标注者条件独立.以类别  $k$  为例,其子类分布  $p(g|y=k)$  使用参数为  $\tau_k \in [0,1]^{M \times 1}$  的  $M$  类离散概率分布表示,即:

$$p(g|y=k) \triangleq p(g|y=k, \tau_k) = \text{Categorical}(g|\tau_k) \tag{3}$$

对每个样本  $x_i$ ,使用两个隐变量  $y_i \in \{1, \dots, K\}$ ,  $g_i \in \{1, \dots, M\}$  分别指示样本的真实类别  $y_i$  及其某个子类  $g_i$ ,以  $y_i=k, g_i=m$  为例,使用参数  $v_{jkm} \in [0,1]^{K \times 1}$  表示第  $j$  个标注者将样本标注为各个类别的概率,则标注  $L_{ij}$  的概率可以通过下式得到:

$$p(L_{ij}=l|y_i=k, g_i=m, V_j) \triangleq p(L_{ij}=l|y_i=k, g_i=m, v_{jkm}) = v_{jkl} \tag{4}$$

则样本  $x_i$  的多个标注的联合分布可以表示为

$$p(L_{i1}, \dots, L_{iW} | y_i = k, g_i = m, \{V_j\}_{j=1}^W) = \prod_{j=1}^W \mathcal{I}(L_{ij} \neq 0) p(L_{ij} | y_i = k, g_i = m, v_{jkm}) \tag{5}$$

结合公式(3)和公式(5),得到公式(6):

$$p(L_{i1}, \dots, L_{iW} | y_i = k, \{V_j\}_{j=1}^W) = \sum_{m=1}^M p(g_i = m | y_i = k) p(L_{i1}, \dots, L_{iW} | y_i = k, g_i = m, v_{jkm}) \tag{6}$$

对比公式(6)与上节标注者独立模型中的公式(1)可以看到:给定样本真实标记,众包标注之间不再互相独立.假设样本互相独立,标注  $L$  的生成过程可以写为

$$P(L|Y, G, V) = \prod_{i=1}^N \prod_{j=1}^W \mathcal{I}(L_{ij} \neq 0) p(L_{ij} | y_i, g_i, V_j) \tag{7}$$

• 真实标记先验

对于真实标记的先验模型  $p(y_i|x_i, \gamma, \pi)$ ,本文分别使用  $p(y_i|\pi)$  和  $p(y_i|x_i, \gamma)$  表示特征无关的先验与特征相关先验,这里使用参数为  $\gamma$  深度神经网络分类器  $f(\cdot)$  实现  $p(y_i|x_i, \gamma)$ ,具体定义如下:

$$p(y_i|\pi) = \text{Categorical}(y_i|\pi), p(y_i|x_i, \gamma) = \text{Categorical}(f(x_i; \gamma)) \tag{8}$$

假设样本互相独立,则关于  $Y$  的先验可以表示为

$$p(Y|X, \gamma, \pi) = p(Y|X, \gamma) p(Y|\pi) = \prod_{i=1}^N p(y_i|\pi) p(y_i|x_i, \gamma) \tag{9}$$

• 参数共轭先验

除了上述的标注生成过程及真实标记先验,本文假设参数  $v_{jkm}, \pi, \tau_k$  服从共轭的狄利克雷先验分布,定义如下:

$$p(v_{jkm}) = \text{Dir}(v_{jkm}|\beta_k), p(\pi) = \text{Dir}(\pi|\alpha), p(\tau_k) = \text{Dir}(\tau_k|\alpha_\tau) \tag{10}$$

• 联合分布

由以上定义,众包标注  $L$ 、隐变量  $Y, G$  以及参数  $\Theta = \{V = \{v_{jkm}\}, \pi, \tau = \{\tau_k\}\}$  的联合分布可以写为

$$p(L, Y, G, \Theta|X, \gamma) = p(L, Y, G, V, \pi, \tau|X, \gamma) = p(L|Y, G, V) p(Y|\pi) p(Y|X, \gamma) p(G|Y, \tau) p(V) p(\pi) p(\tau) \tag{11}$$

- 学习目标

对于公式(11)所表示的模型, 本文目标为通过最大化众包标注的似然  $p(L)$ , 估计样本真实标记  $Y$ 、子类  $G$  及参数  $\Theta=\{V, \pi, \tau\}$  的后验分布  $p(Y|L, X)$ ,  $p(G|L, X)$ ,  $p(\Theta|L, X)$ .

当不使用深度神经网络分类器作为真实标记先验时, 本文模型等价于文献[12]中的非深度标注者关系利用模型, 可以直接使用图 1 中假设标注者独立模型的优化方法. 比如, 文献[12]采用了贝叶斯变分均值场消息传递方法<sup>[26]</sup>以实现快速推断. 但当引入神经网络分类器  $p(Y|X, \gamma)$  时, 吉布斯采样<sup>[10]</sup>和 EM<sup>[27]</sup>的迭代式优化效率非常低; 而变分均值场消息传递方法<sup>[26]</sup>需要标注似然  $p(L)$  符合共轭指数族分布, 非线性神经网络分类器先验并不满足要求. 下节中, 本文将基于结构化变分自编码器<sup>[20]</sup>的优化进展, 为本文模型提出一种随机变分推断优化算法, 对模型中参数实现高效的端到端优化, 保证在最大化标注似然的对数变分下界.

#### 4 基于自然梯度的随机变分推断

作为一种代表性的深度生成式模型, 变分自编码器<sup>[18]</sup>基于重参数化技巧, 通过使用神经网络拟合概率分布的参数并端到端得进行梯度下降. 文献[20]将其扩展到贝叶斯框架下的结构化变分自编码器, 设计并利用了概率图模型的共轭结构, 从而可以使用传统生成式模型如话题模型<sup>[28]</sup>基于共轭分布结构的快速贝叶斯推断方法. 具体来说, 文献[20]基于文献[28]的自然梯度随机变分推断框架(natural gradient stochastic variation inference, SVI), 将分布表示为指数族分布形式, 并使用神经网络拟合真实标记指数族分布的参数, 从而进行均值场消息传递和自然梯度计算, 实现高效二阶优化. 基于文献[20]本文提出利用标记关系的深度生成式众包模型优化方法, 实现细节如下.

假设真实标记  $Y$ 、子类  $G$  及参数  $\Theta=\{V, \pi, \tau\}$  的后验分布服从变分均值场分布, 即  $q(Y, G, \Theta)=q(Y)q(G)q(\Theta)$ . 类似变分自编码器, 针对公式(11), 其对数标注似然  $\log p(L)$  的变分证据下界(evidence lower bound, ELBO)可以表示如下:

$$\log p(L) \geq \mathcal{L}(Y, G, \Theta, \gamma) \triangleq E_{q(Y, G, \Theta)} \left[ \log \frac{p(L, Y, G, \Theta | X, \gamma)}{q(Y)q(G)q(\Theta)} \right] \quad (12)$$

为了利用分布的共轭结构性质, 用  $\eta$  表示指数族分布的自然参数,  $t(\cdot)$  表示充分统计量,  $\log Z(\cdot)$  表示对数分割函数, 本文将公式(3)、公式(8)、公式(10)中关于  $p(v_{jkm})$ ,  $p(\pi)$ ,  $p(\tau_k)$ ,  $p(Y|\pi)$ ,  $p(g|y=k, \tau_k)$  的定义重写成指数族分布的形式:

$$p(v_{jkm}) = \exp\{\langle \eta_{v_{jkm}}, t(v_{jkm}) \rangle - \log Z(\eta_{v_{jkm}})\} \quad (13)$$

$$p(\pi) = \exp\{\langle \eta_{\pi}, t(\pi) \rangle - \log Z(\eta_{\pi})\} \quad (14)$$

$$p(\tau_k) = \exp\{\langle \eta_{\tau_k}, t(\tau_k) \rangle - \log Z(\eta_{\tau_k})\} \quad (15)$$

$$p(y|\pi) = \exp\{\langle \eta_y(\pi), t(y) \rangle - \log Z(\eta_y(\pi))\} = \exp\{\langle t(\pi), (t(y), 1) \rangle\} \quad (16)$$

$$p(g|y=k, \tau_k) = \exp\{\langle \eta_g(\tau_k), t(g) \rangle - \log Z(\eta_g(\tau_k))\} = \exp\{\langle t(\tau_k), (t(g), 1) \rangle\} \quad (17)$$

对于公式(13)–公式(17), 其  $\eta$ ,  $t(\cdot)$ ,  $\log Z(\cdot)$  的具体取值为

$$\eta_{v_{jkm}} = \begin{bmatrix} \beta_{k1} - 1 \\ \vdots \\ \beta_{kK} - 1 \end{bmatrix}, \eta_{\pi} = \begin{bmatrix} \alpha_1 - 1 \\ \vdots \\ \alpha_K - 1 \end{bmatrix}, \eta_{\tau_k} = \begin{bmatrix} \alpha_{\tau_1} - 1 \\ \vdots \\ \alpha_{\tau_M} - 1 \end{bmatrix}, \eta_y(\pi) = \begin{bmatrix} \log \pi_1 \\ \vdots \\ \log \pi_K \end{bmatrix}, \eta_g(\tau_k) = \begin{bmatrix} \log \tau_{k1} \\ \vdots \\ \log \tau_{kM} \end{bmatrix} \quad (18)$$

$$t(v_{jkm}) = \begin{bmatrix} \log v_{jkm1} \\ \vdots \\ \log v_{jkmK} \end{bmatrix}, t(\pi) = \begin{bmatrix} \log \pi_1 \\ \vdots \\ \log \pi_K \end{bmatrix}, t(\tau_k) = \begin{bmatrix} \log \tau_{k1} \\ \vdots \\ \log \tau_{kM} \end{bmatrix}, t(y) = \begin{bmatrix} y_1 \\ \vdots \\ y_K \end{bmatrix}, t(g) = \begin{bmatrix} g_1 \\ \vdots \\ g_M \end{bmatrix} \quad (19)$$

$$\log Z(\eta_{v_{jkm}}) = \sum_{l=1}^K \log \Gamma(\beta_{kl}) - \log \Gamma(\sum_{k=1}^K \beta_{kl}), \log Z(\eta_{\pi}) = \sum_{k=1}^K \log \Gamma(\alpha_k) - \log \Gamma(\sum_{k=1}^K \alpha_k) \quad (20)$$

$$\log Z(\eta_{\tau_k}) = \sum_{m=1}^M \log \Gamma(\alpha_{\tau_m}) - \log \Gamma(\sum_{m=1}^M \alpha_{\tau_m}), \log Z(\eta_y(\pi)) = 0, \log Z(\eta_g(\tau_k)) = 0 \quad (21)$$

这里,  $\Gamma(\cdot)$  代表 Gamma 函数. 类似地, 变分后验分布  $q(Y)$ ,  $q(G)$ ,  $q(\Theta)$  也可以写为指数族分布的形式:

$$q(\theta) = \exp\{\langle \eta_\theta, t(\theta) \rangle - \log Z(\theta)\}, \theta \in Y \cup G \cup \Theta \tag{22}$$

使用上述的指数族分布表达形式, 则公式(12)的变分证据下界可以重写为关于自然参数的目标:

$$\mathcal{L}(\eta_Y, \eta_G, \eta_\Theta, \gamma) \triangleq E_{q(Y, G, \Theta)} \left[ \log \frac{p(L, Y, G, \Theta | X, \gamma)}{q(Y)q(G)q(\Theta)} \right] \tag{23}$$

为了利用模型的共轭性质, 类似结构变分自编码器<sup>[20]</sup>, 本文使用深度神经网络 $\gamma(\cdot)$ 的输出构造先验 $p(y_i|x_i, \gamma)$ 的共轭模型势函数:

$$\psi(y_i|x_i, \gamma) \triangleq \langle \gamma(x_i), t(y_i) \rangle \tag{24}$$

使用势函数 $\psi(y_i|x_i, \gamma)$ 的形式替换 $p(y_i|x_i, \gamma)$ , 得到变分下界表示:

$$\hat{\mathcal{L}}(\eta_Y, \eta_G, \eta_\Theta, \gamma) \triangleq E_{q(Y, G, \Theta)} \left[ \log \frac{p(L, Y, G, \Theta) \exp\{\psi(Y | X, \gamma)\}}{q(Y)q(G)q(\Theta)} \right] \tag{25}$$

至此, 变分证据下界 $\hat{\mathcal{L}}$ 中的分布为具有共轭结构性质的指数族分布形式, 其参数为指数族分布的自然参数 $\eta_Y, \eta_G, \eta_\Theta$ 及神经网络参数 $\gamma$ , 接下来就可以采用文献[28]的自然梯度随机变分推断进行参数求解, 具体求解如下.

- 给定其他变量, 求解 $\eta_Y$

其他变量固定时, 隐变量 $Y$ 的最优解 $q^*(Y)$ 关于样本独立, 即 $q^*(Y) = \prod_{i=1}^N q^*(y_i)$ , 各样本分布 $q^*(y_i)$ 存在如下闭式解:

$$\log q^*(y_i) = E_{q(\pi)} \log p(y_i | \pi) + \langle \gamma(x_i), t(y_i) \rangle + E_{q(g_i)q(V)} \log p(\{L_{ij}\}_{j \in W_i} | y_i, g_i, V) + const \tag{26}$$

$$\eta_{y_i}^* = E_{q(\pi)} t(\pi) + \gamma(x_i) + \sum_{j=1}^W \mathcal{I}(L_{ij} \neq 0) [E_{q(v_{jL_{ij}})} t(v_{jL_{ij}})] * [E_{q(g_i)} t(g_i)] \tag{27}$$

这里,  $v_{jL_{ij}}$  中的  $L_{ij}$  指示下标. 以  $L_{ij}=k$  时为例,  $v_{jk}$  是以  $v_{jkm}$  为列组成的矩阵, 具体形式为  $v_{jk} = [v_{jk1}, \dots, v_{jkK}] \in [0, 1]^{K \times M}$ .  $E_{q(v_{jL_{ij}})}, t(v_{jL_{ij}})$  分别对矩阵  $v_{jL_{ij}}$  的各列运算, \*表示矩阵相乘.

- 给定其他变量, 求解 $\eta_G$

其他变量固定时, 隐变量 $G$ 的最优解 $q^*(G)$ 关于样本独立, 即 $q^*(G) = \prod_{i=1}^N q^*(g_i)$ , 各样本分布 $q^*(g_i)$ 存在如下闭式解:

$$\log q^*(g_i) = E_{q(y_i)q(\tau_{y_i})} \log p(g_i | y_i, \tau_{y_i}) + E_{q(y_i)q(V)} \log p(\{L_{ij}\}_{j \in W_i} | y_i, g_i, V) + const \tag{28}$$

$$\eta_{g_i}^* = [E_{q(\tau)} t(\tau)] * [E_{q(y_i)} t(y_i)] + \sum_{j=1}^W \mathcal{I}(L_{ij} \neq 0) [E_{q(v_{jL_{ij}})} t(v_{jL_{ij}})]^\top * [E_{q(y_i)} t(y_i)] \tag{29}$$

这里,  $\tau$ 是以 $\tau_k$ 为列组成的矩阵, 具体形式为 $\tau = [\tau_1, \dots, \tau_K] \in [0, 1]^{M \times K}$ ;  $E_{q(\tau)}, t(\tau)$ 分别对矩阵 $\tau$ 的各列运算;  $\top$ 表示矩阵转置; \*表示矩阵相乘.

- 给定 $\eta_Y, \eta_G$ 求解 $\eta_\Theta, \gamma$

将 $\eta_Y^*, \eta_G^*$ 带入公式(23), 得到如下关于 $\eta_\Theta, \gamma$ 的优化目标:

$$\mathcal{J}(\eta_\Theta, \gamma) \triangleq \mathcal{L}(\eta_Y^*, \eta_G^*, \eta_\Theta, \gamma) \tag{30}$$

针对公式(30), 文献[20]中证明了 $\mathcal{J}(\eta_\Theta, \gamma)$ 为公式(23)的最优下界, 即:

$$\max_{\eta_Y, \eta_G} \mathcal{L}(\eta_Y, \eta_G, \eta_\Theta, \gamma) \geq \mathcal{J}(\eta_\Theta, \gamma) \tag{31}$$

根据文献[20], 可以推导 $\mathcal{J}(\cdot)$ 关于 $\eta_\Theta$ 的梯度如下:

$$\tilde{\nabla}_{\eta_\Theta} \mathcal{J} = [\eta_\Theta^0 + E_{q^*(Y)q^*(G)} (t(Y, G, X, L), 1) - \eta_\Theta] + \nabla_{\eta_Y, \eta_G} (\mathcal{L}(\eta_Y^*, \eta_G^*, \eta_\Theta, \gamma), 0) \tag{32}$$

这里,  $\eta_\Theta^0$ 为使用模型时参数 $\Theta$ 先验分布的自然参数值. 对于本文的模型, 关于 $\eta_{v_{jkm}}, \eta_\pi, \eta_{\tau_k}$ 的自然梯度推导得到:

$$\tilde{\nabla}_{\eta_{v_{jk}}} \mathcal{J} = \eta_{v_{jk}}^0 + \sum_{i=1}^N \mathcal{I}(L_{ij} \neq 0) [E_{q^*(y_i)} t(y_i) \otimes \bar{L}_{ij}] * [E_{q^*(g_i)} t(g_i)]^\top - \eta_{v_{jk}} \tag{33}$$

$$\tilde{\nabla}_{\eta_\pi} \mathcal{J} = \eta_\pi^0 + \sum_{i=1}^N E_{q^*(y_i)} t(y_i) - \eta_\pi \tag{34}$$

$$\tilde{\nabla}_{\eta_\tau} \mathcal{J} = \eta_\tau^0 + \sum_{i=1}^N [E_{q^*(g_i)} t(g_i)] * [E_{q^*(y_i)} t(y_i)]^\top - \eta_\tau \quad (35)$$

这里,  $\eta_{v_{jk}}$  是以  $\eta_{v_{jkM}}$  为列组成的矩阵, 具体形式为  $\eta_{v_{jk}} = [\eta_{v_{jk1}}, \dots, \eta_{v_{jkM}}] \in [0, 1]^{K \times M}$ ;  $\eta_\tau$  是以  $\eta_{\tau_k}$  为列组成的矩阵, 具体形式为  $\eta_\tau = [\eta_{\tau_1}, \dots, \eta_{\tau_K}] \in [0, 1]^{M \times K}$ .  $\tilde{\nabla}_{\eta_{v_{jk}}} \mathcal{J}$ ;  $\tilde{\nabla}_{\eta_\tau} \mathcal{J}$  分别对矩阵  $\eta_{v_{jk}}$ ,  $\eta_\tau$  的各列求导运算;  $\bar{L}_{ij}$  为  $L_{ij}$  独热编码表示. 对于神经网络参数  $\gamma$ , 其梯度  $\nabla_\gamma \mathcal{J}$  可以基于现有的深度神经网络后向传播方式计算.

模型训练完成之后, 样本的真实标记概率、标注者的能力参数可以通过对应后验分布充分统计量的期望  $E_{q^*(y_i)} t(y_i)$ ,  $E_{q^*(v_{jkm})} t(v_{jkm})$  得到:

$$E_{q^*(y_i)} t(y_i) = \begin{bmatrix} \pi_{y_{i1}} \\ \vdots \\ \pi_{y_{iK}} \end{bmatrix}, E_{q^*(v_{jkm})} t(v_{jkm}) = \begin{bmatrix} \varphi(\beta_{k1}^*) \\ \vdots \\ \varphi(\beta_{kK}^*) \end{bmatrix} - \varphi(\sum_{l=1}^K \beta_{kl}^*) \quad (36)$$

算法 1 中给出了总体的实现流程. 可以看到: 与普通神经网络训练过程相比, 每轮迭代在参数  $\eta_\theta$ ,  $\gamma$  的梯度更新之外, 增加了关于  $\eta_{y_i}^*$ ,  $\eta_{g_i}^*$  的闭式解计算, 算法整体计算复杂度主要取决于梯度更新次数, 与普通神经网络训练过程相当.

**算法 1.** 利用标注者相关性的深度生成式众包学习算法.

输入: 样本  $X = \{x_1, \dots, x_N\}$ , 标注  $L \in \{0, 1, \dots, K\}^{N \times W}$ , 参数  $\theta$  的先验分布(对应自然参数  $\eta_\theta^0$ );

输出: 真实标记预测  $Y = \{y_1, \dots, y_N\}$ .

初始化: 初始化参数  $\eta_\theta$ ,  $\gamma$ ,  $\eta_Y$ ,  $\eta_G$

**Repeat:**

固定  $\eta_\theta$ ,  $\gamma$ ,  $\eta_G$ , 对各样本  $x_i$ , 使用公式(27)计算其真实标记后验分布的自然参数  $\eta_{y_i}^*$

固定  $\eta_\theta$ ,  $\gamma$ ,  $\eta_Y$ , 对各样本  $x_i$ , 使用公式(29)计算其子类后验分布的自然参数  $\eta_{g_i}^*$

固定  $\eta_{y_i}^*$ ,  $\eta_{g_i}^*$ , 使用公式(33)–公式(35)计算  $\eta_\theta$  的自然梯度  $\tilde{\nabla}_{\eta_\theta} \mathcal{J}$ , 使用后向传播计算神经网络参数  $\gamma$  的梯度  $\nabla_\gamma \mathcal{J}$

对  $\eta_\theta$ ,  $\gamma$  进行随机梯度上升更新

**Until** 变分下界  $\mathcal{J}(\eta_\theta, \gamma)$  收敛或者达到最大迭代次数

预测: 使用公式(36)得到真实标记预测

## 5 实验

### 5.1 实验设置

- 实验数据

本文使用文献[32]收集的两个多标记众包图片数据集, 记为 *dataset1*, *dataset2*. 两个数据分别包含 700 张、1 495 张图片, 对应 6 个、16 个类别. 原始数据中包含了来自 18 个、15 个标注者的标注, 对各标注者对应的样本子集计算标注准确率, 其 MacroF1 结果主要分布在 [0.700, 0.800], 表明多数标注者可靠, 使用这两个数据集进行众包学习验证具有可行性. 文献[32]的实验结果表明: 当标注者数目达到 10 时, 大多数方法的效果会趋于一致. 因此, 为了提高实验效率, 本文使用标注样本最多的 9 个标注者, 并使用原始的 1 248 维 Fisher 向量特征. 本文在各个类别上独立进行单分类众包学习, 得到 22 个二分类任务数据集.

- 对比方法

本文对比了 3 组代表性的众包学习方法, 包括: (1) 仅利用标注信息的多数投票法 MV (majority voting) 方法、DS 模型<sup>[6]</sup>、基于最小最大化熵原则的 MaxEn 模型<sup>[9]</sup>; (2) 使用逻辑回归分类器作为真实标记先验的非深度生成式方法 Yutc<sup>[3]</sup>; 3) 未考虑标注者相关性的深度生成式模型 BayesDGC<sup>[25]</sup>.

本文方法记做 BayesDGC-w, 当不考虑标注者相关性, 即各类别的子类混合模型中子类数目  $M=1$  时,



BayesDGC-w 等价于 BayesDGC. 当不使用深度神经网络分类器作为真实标记先验时, BayesDGC-w 等价于文献[12]中仅利用标注信息的 EBCC 方法. 因此, 对于深度生成式模型 BayesDGC, BayesDGC-w, 本文也对比了其不使用样本特征的非深度贝叶斯变体, 即 BayesGC, BayesGC-w, 以分别检验深度神经网络分类器和标注者相关性对于众包学习的效果.

本文方法 BayesDGC-w 使用单隐层(节点数为 100)感知机作为深度神经网络分类器, 子类混合模型中子类数目设置为  $M=3$ , 参数  $v_{jkm}$ ,  $\pi$ ,  $\tau_k$  分别使用狄利克雷先验  $Dir(v_{jkm} | \beta_k^0)$ ,  $Dir(\pi | \alpha^0)$ ,  $Dir(\tau_k | \tau^0)$ . 对于标注者能力参数  $v_{jkm}$ , 其先验参数  $\beta_k^0 \in R^K$  设置为  $\beta_{kk}^0 = 5$ ,  $\beta_{kk'}^0 = 2$ ,  $k \neq k'$ , 以反映标注者能力优于随机猜测. 对于  $\alpha^0 \in R^K$  和  $\tau^0 \in R^M$ , 其各元素值为 1.1. 优化过程采用 Adam 优化器<sup>[33]</sup>, 学习率设为 0.001, 迭代轮数  $epochs=400$ .  $M=1$  时, 其他参数使用同样的设置, 得到 BayesDGC 方法实现. 不使用神经网络分类器, 得到 BayesGC, BayesGC-w 实现. 其他对比方法中, 除了 DS 模型采用混淆矩阵刻画标注者能力, 其他方法采用原论文中的参数设定.

为测试众包标注数据量对各学习方法的影响, 本文以 10% 的间隔随机保留 10%–100% 比例的标注, 记录 10 次重复实验的均值与标准差. 由于原本数据为多标记任务, 类别严重不平衡, 如 *dataset1* 对应的 6 个类别中, 平均每张图片有 1.24 个正标记; *dataset2* 对应的 16 个类别中, 平均每张图片具有 1.80 个正标记. 本文采用 ROC 曲线下面积 AUC (area under the ROC curve) 作为评价指标.

## 5.2 实验结果

图 3(a)展示了 BayesDGC-w 与对比方法在 22 个数据集上 10 种标注比例下(220 个实验)的平均 AUC 结果, 可以看到: 未考虑标注质量建模的 MV 方法效果最差; 深度众包方法 BayesDGC-w 与 BayesDGC 远远优于其他的非深度众包方法; 而考虑了标记关系利用的 BayesDGC-w (BayesGC-w)的平均 AUC 比未考虑标记关系利用的 BayesDGC (BayesGC)有明显提升. 图 3(b)展示了 Nemenyi 检验结果. Nemenyitest 是一种常用的检验, 用来比较多个方法在多个数据集上的整体表现<sup>[34]</sup>. 图 3(b)上部水平线上的数字指示各方法在 220 次实验中的平均排名. 两个方法平均排名之差大于某个阈值(critical difference, CD)时, 说明二者存在统计显著差异; 否则不存在. 阈值 CD 与对比方法个数、实验次数、显著性  $p$  值有关. 设置  $p=0.05$ , 得到本文  $CD=0.70789$ . 图中红线连接排名之差小于 CD 的算法. 可以看到: 深度生成式的方法 BayesDGC-w (BayesDGC)显著优于其他的非深度方法, 而考虑标记关系利用的 BayesDGC-w 与 BayesDGC 效果可比.

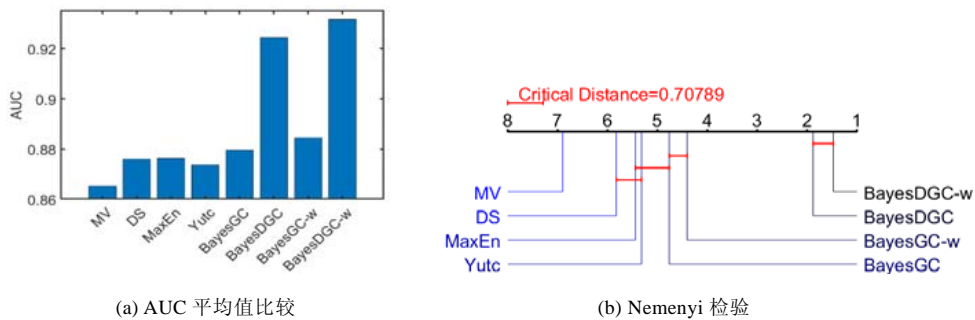


图 3 BayesDGC-w 与其他 7 种对比方法在 22 个真实数据集上的 AUC 总体效果比较

用  $l$  表示二分类任务中两个数据集对应的第  $l$  个类别, 图 4 给出了 22 个数据集上的具体实验效果. 可以看到: 在大多数情况下, 本文的 BayesDGC-w 方法稳定地优于其他方法. 随着标注量的增加, 各方法的 AUC 效果呈稳定上升趋势. 与使用样本特征信息的 BayesDGC-w, BayesDGC, Yutc 方法对比, 尤其在标注量较少的情况下 ( $\leq 40\%$ ), 仅利用标注的几种方法(MV, DS, MaxEn, BayesGC, BayesGC-w)效果非常接近, 并与 BayesDGC-w, BayesDGC 方法存在明显差距. 这说明样本特征包含重要的互补信息. 对于使用逻辑回归分类器的 Yutc 方法, 其在某些数据集如 *dataset1*  $l=6$  以及 *dataset2*  $l=3, 14, 16$  上取得可比甚至更好的效果; 但在某些数据集, 如 *dataset2*  $l=5, 10, 13$  上结果并不稳定. 可能是因为线性模型复杂度低于数据拟合所需要的模型复

杂度, 并且 Yutc 方法的非贝叶斯实现使得参数设置对效果影响很大. 而 BayesDGC-w, BayesDGC 的深度学习提供了足够的学习能力, 其贝叶斯推断自动拟合最适合数据的参数, 从而不需要手动参数选择. 对比 BayesDGC-w (BayesGC-w) 与 BayesDGC (BayesGC) 可以看到, 利用标注者相关性有助于改善众包学习效果. 这与文献[12]中的结果一致. 下节将给出子类数目  $M$  的不同取值时本文方法的效果.

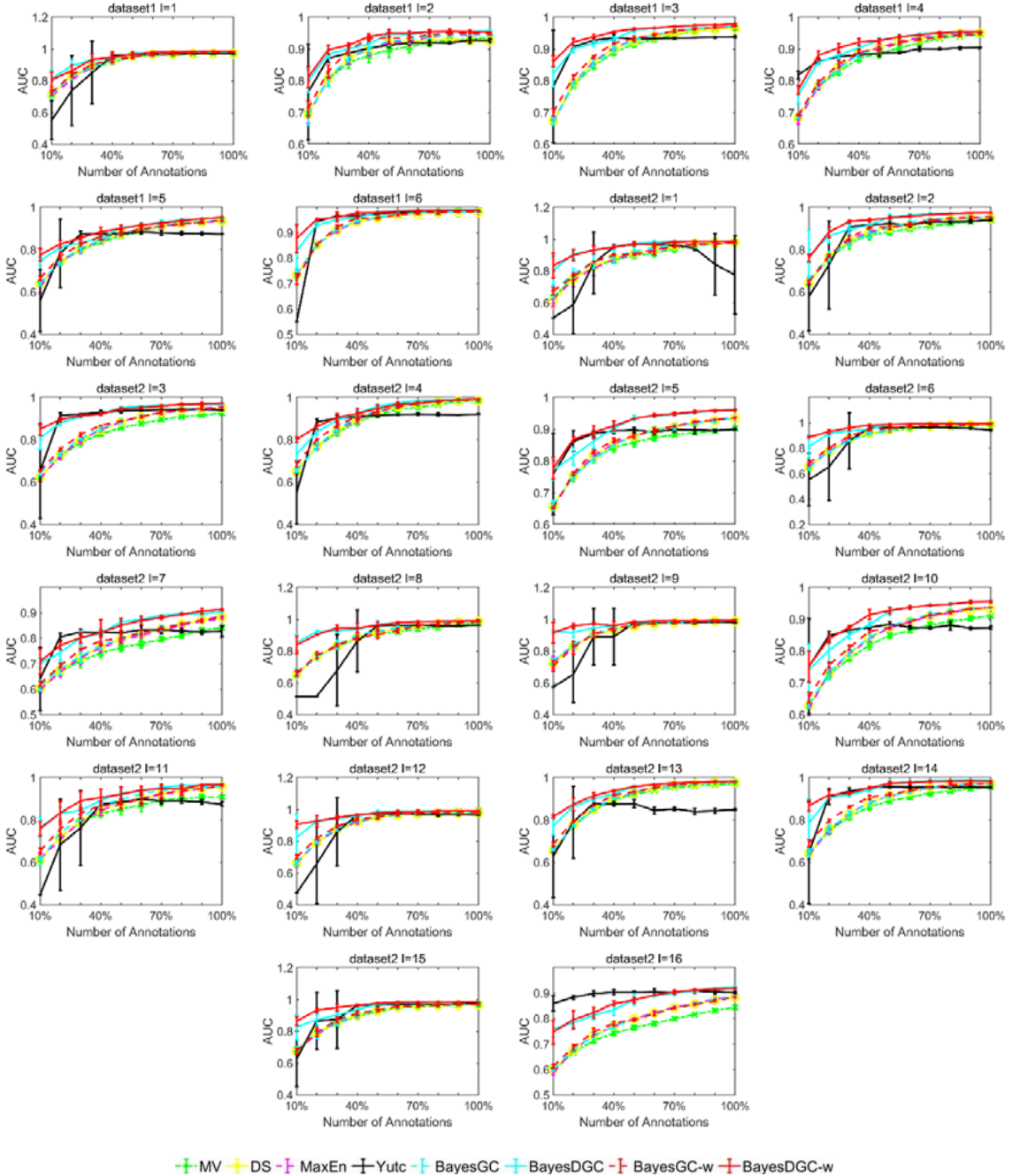


图 4 各方法在 22 个真实数据集上的 AUC 结果

### 5.3 参数讨论

本节讨论子类混合模型中子类个数  $M$  对 BayesDGC-w 模型的影响. 图 5 所示为在 4 个数据集上使用 100% 标注,  $M=1, 3, 5, 10, 20, 30$  这 6 种不同取值时, BayesDGC-w 的 AUC 结果.  $M=1$  等价于不考虑标注相关性的情形. 这里选择了两组代表性的结果: (1) *dataset2*  $l=1, 12$  对应考虑标注者相关性( $M>1$ )有助于改善学习的情形; (2) *dataset2*  $l=3, 11$  对应子类数目过多( $M=20, 30$ )时不利于学习的情形, 这是因为标注者能力参数  $\{v_{jkm}\}$  假设空间过大, 导致优化时更容易陷入局部最优解. 因此, 本文实验中从优化稳定性角度考虑, 设置子类数目为  $M=3$ , 之后将考虑从标注者能力参数正则化或者非参贝叶斯学习角度对这一问题进行探索.

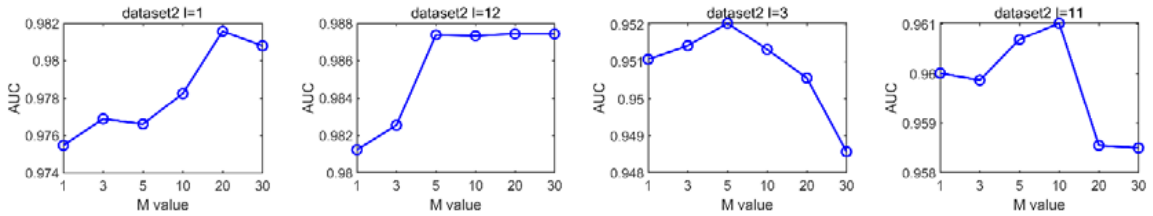


图 5 参数  $M$  不同取值对 BayesDGC-w 学习效果的影响

## 6 结束语

本文提出了一种利用标注者相关性的深度生成式众包学习方法, 通过在标注生成过程中引入各类内标注者能力的混合模型, 实现标注相关性刻画. 为了实现贝叶斯推断, 基于结构化变分自编码器优化技术, 本文利用概率分布的共轭结构性质, 将变分消息传递与神经网络参数随机梯度下降结合实现端到端地高效优化, 避免了 EM 算法与吉布斯采样方法的迭代式计算开销. 实验中发现, 混合模型中混合成分的个数对于模型效果影响较大. 后续工作将考虑从参数正则化或者非参贝叶斯学习角度对这一问题进行探索, 并尝试将标注者相关性建模思路扩展到多标记众包学习问题中的标记关系建模.

### References:

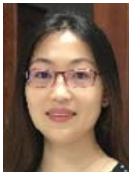
- [1] Weld D, Lin C, Bragg J. Artificial intelligence and collective intelligence. In: Malone T, Bernstein M, eds. The Collective Intelligence Handbook. 2015.
- [2] Snow R, O'Connor B, Jurafsky D, et al. Cheap and fast—But is it good? Evaluating non-expert annotations for natural language tasks. In: Proc. of the Conf. on Empirical Methods in Natural Language Processing. 2008. 254–263.
- [3] Raykar V, Yu S, Zhao L, et al. Learning from crowds. Journal of Machine Learning Research, 2010, 11: 1297–1322.
- [4] Welinder P, Branson S, Belongie S, et al. 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, Vol.23. 2010. 2024–2432.
- [5] Li Q, Li Y, Gao J, et al. A confidence-aware approach for truth discovery on long-tail data. Proc. of the VLDB Endowment, 2014, 8(4): 425–436.
- [6] 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.
- [7] Liu Q, Peng J, Ihler A. Variational inference for crowdsourcing. In: Bartlett P, Pereira F, Burges C, Bottou L, Weinberger K, eds. Proc. of the Advances in Neural Information Processing Systems, Vol.25. 2012. 692–700.
- [8] Whitehill J, Ruvolo P, Wu T, et al. Whose vote count more: Optimal integration of labels from labelers of unknown expertise. In: Bengio Y, Schuurmans D, Lafferty J, Williams C, Culotta A, eds. Proc. of the Advances in Neural Information Processing Systems, Vol.22. 2009. 2035–2043.
- [9] Zhou D, Basu S, Mao Y, et al. Learning from the wisdom of crowds by minimax entropy. In: Bartlett P, Pereira F, Burges C, Bottou L, Weinberger K, eds. Proc. of the Advances in Neural Information Processing Systems, Vol.25. 2012. 2195–2203.
- [10] Kim HC, Ghahramani Z. Bayesian classifier combination. In: Proc. of the 15th Int'l Conf. on Artificial Intelligence and Statistics. 2012. 619–627.

- [11] Imamura H, Sato I, Sugiyama M. Analysis of minimax error rate for crowdsourcing and its application to worker clustering model. In: Proc. of the 35th Int'l Conf. on Machine Learning. 2018. 2152–2161.
- [12] Li Y, Rubinstein BIP, Cohn T. Exploiting worker correlation for label aggregation in crowdsourcing. In: Proc. of the 36th Int'l Conf. on Machine Learning. 2019. 3886–3895.
- [13] Rodrigues F, Pereira F, Ribeiro B. Gaussian process classification and active learning with multiple annotators. In: Proc. of the 31th Int'l Conf. on Machine Learning. 2014. 433–441.
- [14] LeCun Y, Bengio Y, Hinton GE. Deep learning. *Nature*, 2015, 521(7553): 436–444.
- [15] Albarqouni S, Baur C, Achilles F, *et al.* Aggnet: Deep learning from crowds for mitosis detection in breast cancer histology images. *IEEE Trans. on Medical Imaging*, 2016, 35(5): 1313–1321.
- [16] Filipe Rodrigues FCP. Deep learning from crowds. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence. 2018. 1611–1618.
- [17] Tanno R, Saeedi A, Sankaranarayanan S, *et al.* Learning from noisy labels by regularized estimation of annotator confusion. In: Proc. of the 2019 IEEE Conf. on Computer Vision and Pattern Recognition. 2019. 11244–11253.
- [18] Kingma DP, Welling M. Auto-Encoding variational Bayes. In: Proc. of the 2nd Int'l Conf. on Learning Representations. 2014.
- [19] Kingma DP, Mohamed S, Rezende DJ, *et al.* Semi-Supervised learning with deep generative models. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ, eds. Proc. of the Advances in Neural Information Processing Systems, Vol.27. 2014. 3581–3589.
- [20] Johnson MJ, Duvenaud D, Wiltschko AB, *et al.* Composing graphical models with neural networks for structured representations and fast inference. In: Lee DD, Sugiyama M, von Luxburg U, Guyon I, Garnett R, eds. Proc. of the Advances in Neural Information Processing Systems, Vol.29. 2016. 2946–2954.
- [21] Yin L, Han J, Zhang W, *et al.* Aggregating crowd wisdoms with label-aware autoencoders. In: Proc. of the 26th Int'l Joint Conf. on Artificial Intelligence. 2017. 1325–1331.
- [22] Atarashi K, Oyama S, Kurihara M. Semi-supervised learning from crowds using deep generative models. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence. 2018. 1555–1562.
- [23] Shi W, Sheng VS, Li X, *et al.* Semi-supervised multi-label learning from crowds via deep sequential generative model. In: Proc. of the 26th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2020. 1141–1149.
- [24] Luo Y, Tian T, Shi J, *et al.* Semi-crowdsourced clustering with deep generative models. In: Bengio S, Wallach HM, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, eds. Proc. of the Advances in Neural Information Processing Systems, Vol.31. 2018. 3216–3226.
- [25] Li SY, Huang SJ, Chen S. Crowdsourcing aggregation with deep Bayesian learning. *Science China Information Sciences*, 2021, 64(3).
- [26] Simpson E, Roberts S, Psorakis I, *et al.* Dynamic Bayesian combination of multiple imperfect classifiers. In: Proc. of the Decision Making and Imperfection. 2013. 1–35.
- [27] Venanzi M, Guiver J, Kazai G, *et al.* Community-based Bayesian aggregation models for crowdsourcing. In: Proc. of the 23rd Int'l Conf. on World Wide Web. 2014. 155–164.
- [28] Homan MD, Blei DM, Wang C, *et al.* Stochastic variational inference. *Journal of Machine Learning Research*, 2013,14(1): 1303–1347.
- [29] Li SY, Jiang Y. Multi-Label crowdsourcing learning. *Ruan Jian Xue Bao/Journal of Software*, 2020, 31(5): 1497–1510 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5673.htm> [doi: 10.13328/j.cnki.jos.005673]
- [30] Chen CJ, Jiang L, Lei N, *et al.* An interactive feature selection method based on learning-from-crowds. *Sci Sin Inform*, 2020, 50: 794–812 (in Chinese with English abstract). [doi: 10.1360/SSI-2019-0208]
- [31] Feng Y, Wang Y, Fang CR, *et al.* An approach for developing a highly trustworthy crowd-sourced workforce. *Sci Sin Inform*, 2019, 49: 1412–1427 (in Chinese with English abstract). [doi: 10.1360/N112018-00303]
- [32] Li SY, Jiang Y, Chawla NV, *et al.* Multi-Label learning from crowds. *IEEE Trans. on Knowledge and Data Engineering*, 2019, 31(7): 1369–1382.
- [33] Kingma DP, Ba J. Adam: A method for stochastic optimization. In: Proc. of the 3rd Int'l Conf. on Learning Representations. CA, 2015.

- [34] Demsar J. Statistical comparisons of classifiers over multiple datasets. *Journal of Machine Learning Research*, 2006, 7: 1–30.

**附中文参考文献:**

- [29] 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> [doi: 10.13328/j.cnki.jos.005673]
- [30] 陈长建, 姜流, 雷娜, 等. 基于众包学习的交互式特征选择方法. *中国科学: 信息科学*, 2020, 50: 794–812. [doi: 10.1360/SSI-2019-0208]
- [31] 冯洋, 王祎, 房春荣, 等. 高可信众包群体构建方法. *中国科学: 信息科学*, 2019, 49: 1412–1427. [doi: 10.1360/N112018-00303]



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



黄圣君(1987—), 男, 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, 模式识别.



韦梦龙(1997—), 男, 硕士生, CCF 学生会员, 主要研究领域为机器学习, 数据挖掘.