Multi-Class Multi-Label Independent Model for Crowdsourcing

Jing Zhang
School of Computer Science and Engineering
Nanjing University of Science and Technology
Nanjing, Jiangsu, China
jzhang@njust.edu.cn

Xindong Wu
School of Computing and Informatics
University of Louisiana at Lafayette
Lafayette, Louisiana, U.S.
xwu@louisiana.edu

ABSTRACT

Multi-Class Multi-Label Independent (MCMLI) model is a probabilistic generative model for truth inference in crowdsourcing. This document provides the details of this model.

CCS CONCEPTS

• **Computing methodologies** → **Maximum likelihood modeling**; **Mixture models**; **Latent variable models**; • **Information systems** → **Crowdsourcing**;

KEYWORDS

Crowdsourcing, label aggregation, maximum likelihood estimation, mixture models, probabilistic graphical models

1 THE MCMLI MODEL

We first present a novel probabilistic model for the MCMLI model. Then, we solve the model with an EM algorithm.

1.1 Label-Independent Model

We first propose a novel probabilistic generative model, namely *multi-class multi-label independent* (MCMLI) model, under an assumption that the labels are mutually independent. The probabilistic graphical model representation of MCMLI is illustrated in Figure 1.

Generation of true labels. For multi-class classification, each true label (supposed to be the mth label in multi-label annotation) is independently drawn from a multinoulli distribution with parameters $\boldsymbol{\theta}^{(m)} = [\theta_1^{(m)}, ..., \theta_K^{(m)}]$, where $\sum_{k=1}^{K} \theta_k^{(m)} = 1$. That is, for the mth label of an instance $\mathbf{x}_i$, we have $\mathrm{P}(y_i^{(m)}|\boldsymbol{\theta}^{(m)}) = \prod_{k=1}^{K} \left(\theta_k^{(m)}\right)^{\mathbb{I}(y_i^{(m)}=k)}$. Consequently, the probability of the M-dimensional true label vector $\mathbf{y}_i$ of the instance is

$$\mathrm{P}(\mathbf{y}_i|\Theta) = \prod_{m=1}^{M} \prod_{k=1}^{K} \left(\theta_k^{(m)}\right)^{\mathbb{I}(y_i^{(m)}=k)}, \tag{1}$$

where $\Theta = [\boldsymbol{\theta}^{(1)}, ..., \boldsymbol{\theta}^{(M)}]$ is a set of parameters for all M multinoulli distributions.

This research has been supported by the National Natural Science Foundation of China under grant 61603186, the Natural Science Foundation of Jiangsu Province, China, under grant BK20160843, the China Postdoctoral Science Foundation under grants 2017T100370 and 2016M590457, and the Science Foundation (for Youth) of the Science and Technology Commission of the Central Military Commission (CMC), China, and the US National Science Foundation (NSF) under grant IIS-1613950.

SIGKDD'18, August 2018, London, United Kingdom

ACM ISBN 123-4567-24-567/08/06...$15.00
https://doi.org/10.475/123_4

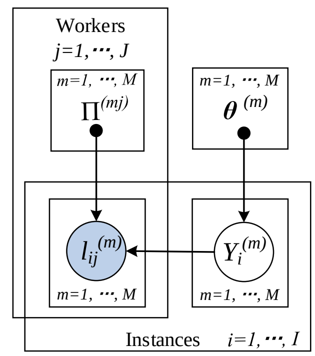

Figure 1: Probabilistic graphical model representation of the MCMLI model.

Generation of crowdsourced labels. In multi-class classification, a confusion matrix is a powerful tool that can comprehensively depict the distribution of a classifier's capability over all pairs of classes, providing fine-grained information to upper-layer applications. In MCMLI, each worker j independently provides value to each label of an instance $\mathbf{x}_i$. We use a set of confusion matrices $\mathbf{\Pi}^{(j)} = [\Pi^{(1j)}, ..., \Pi^{(Mj)}]$ to model the reliability of worker j with respect to M labels. We denote all sets of confusion matrices of totally J workers by a parameter set $\widetilde{\mathbf{\Pi}} = \{\mathbf{\Pi}^{(1)}, ..., \mathbf{\Pi}^{(J)}\}$. In matrix $\Pi^{(mj)}$, each element $\pi_{kd}^{(mj)} (1 \leq k, d \leq K)$ represents the probability of worker j labeling (true) class k as class d on the mth label, which derives $\mathrm{P}(l_{ij}^{(m)} = d | y_i^{(m)} = k) = \pi_{kd}^{(mj)}$. That is, $l_{ij}^{(m)}$ conditioning on $y_i^{(m)} = k$ obeys a multinoulli distribution with parameters $[\pi_{kd}^{(mj)}]_{d=1}^{K}$ and $\sum_{d=1}^{K} \pi_{kd}^{(mj)} = 1$.

Consider each instances $\mathbf{x}_i$ is independently labeled by J workers, the likelihood of all observed noisy labels on the instance can be calculated as follows:

$$\mathrm{P}\left(L_i|\widetilde{\mathbf{\Pi}}\right) = \sum_{k^{(1)}=1}^{k^{(1)}=K} \cdots \sum_{k^{(M)}=1}^{k^{(M)}=K} \Big[\mathrm{P}\left(y_i^{(1)} = k^{(1)}, ..., y_i^{(M)} = k^{(M)} | \Theta\right) \cdot \mathrm{P}\left(L_i | y_i^{(1)} = k^{(1)}, ..., y_i^{(M)} = k^{(M)}, \widetilde{\mathbf{\Pi}}\right)\Big]. \tag{2}$$

Because we assume that the true labels of an instance are mutually independent, we have

$$\mathrm{P}\left(y_i^{(1)}=k^{(1)},...,y_i^{(M)}=k^{(M)}\middle|\Theta\right)=\theta_{k^{(1)}}^{(1)}\cdots\theta_{k^{(M)}}^{(M)}, \tag{3}$$

$$\begin{aligned}&\mathrm{P}\left(L_i\middle|y_i^{(1)}=k^{(1)},...,y_i^{(M)}=k^{(M)},\widetilde{\mathbf{\Pi}}\right)\\&=\prod_{j=1}^{J}\mathrm{P}\left(l_{ij}^{(1)},...,l_{ij}^{(M)}\middle|y_i^{(1)}=k^{(1)},...,y_i^{(M)}=k^{(M)},\widetilde{\mathbf{\Pi}}\right)\\&=\prod_{j=1}^{J}\left(\prod_{m=1}^{M}\prod_{d^{(m)}=1}^{K}\left(\pi_{k^{(m)}d^{(m)}}^{(mj)}\right)^{\mathbb{I}(l_{ij}^{(m)}=d^{(m)})}\right).\end{aligned} \tag{4}$$

Here, because we need to present the prior of $\mathbf{y}_i$ under the conditions that each of its element is assigned with a different value, we use the superscript (m) to distinguish these ks on different labels. Similarly, the variable d is also decorated by the superscript (m).

Finally, the log-likelihood of all crowdsourced labels of the entire datasets is

$$\begin{aligned}\ln\mathrm{P}(\boldsymbol{L}|\Theta,\widetilde{\mathbf{\Pi}})=\sum_{i=1}^{I}\ln\Bigg(\sum_{k^{(1)}=1}^{k^{(1)}=K}\cdots\sum_{k^{(M)}=1}^{k^{(M)}=K}\Bigg[\prod_{m=1}^{M}\theta_{k^{(m)}}^{(m)}\\ \cdot\prod_{j=1}^{J}\prod_{d^{(m)}=1}^{K}\left(\pi_{k^{(m)}d^{(m)}}^{(mj)}\right)^{\mathbb{I}(l_{ij}^{(m)}=d^{(m)})}\Bigg]\Bigg).\end{aligned} \tag{5}$$

1.2 Inference with EM

Our optimization objective is to maximize the likelihoods of the observed data $\mathbf{L}$ defined by Eq.(5), which can be achieved by expectation-maximization (EM) algorithm, which iteratively applies E-step and M-step.

E-step. We calculate the expected value of the log likelihood function, with respect to the conditional distribution of $\mathbf{Y}$ given the observed noisy labels $\mathbf{L}$ under the current estimates of parameters $\mathbf{\Psi}_1^{old}$:

$$\mathcal{Q}_1\left(\mathbf{\Psi}_1,\mathbf{\Psi}_1^{old}\right)=\mathbb{E}_{\mathbf{Y}|\mathbf{L},\mathbf{\Psi}_1^{old}}\left[\ln\mathrm{P}\left(\mathbf{L},\mathbf{Y}\middle|\mathbf{\Psi}_1\right)\right], \tag{6}$$

where $\mathbf{\Psi}_1=\{\Theta,\widetilde{\mathbf{\Pi}}\}$. In fact, this will simply result in the calculation of the joint posterior probability of $\mathbf{y}_i$ for each instance $\mathbf{x}_i$ by applying Bayes' theorem as follows:

$$\begin{aligned}&\mathrm{P}\left(y_i^{(1)}=k^{(1)},...,y_i^{(M)}=k^{(M)}\middle|\mathbf{L},\mathbf{\Psi}_1\right)\\&=\prod_{m=1}^{M}\mathrm{P}\left(y_i^{(m)}=k^{(m)}\middle|\mathbf{L},\mathbf{\Psi}_1\right)\\&\propto\prod_{m=1}^{M}\theta_{k^{(m)}}^{(m)}\prod_{j=1}^{J}\prod_{d^{(m)}=1}^{K}\left(\pi_{k^{(m)}d^{(m)}}^{(mj)}\right)^{\mathbb{I}\left(l_{ij}^{(m)}=d^{(m)}\right)}.\end{aligned} \tag{7}$$

Eq.(7) directly provides the expected value of an indicator function (calculated as a marginal probability) that will be used in M-step as follows:

$$\begin{aligned}\mathbb{E}\left[\mathbb{I}\left(y_i^{(m)}=k\right)\right]&=\mathrm{P}\left(y_i^{(m)}=k\right)\\&=\sum_{y_i^{(m')}(\forall m'\in\{1,...,M\}\backslash m)}\mathrm{P}\left(y_i^{(m)}=k,\cdots\middle|\mathbf{L},\mathbf{\Psi}_1\right)\end{aligned} \tag{8}$$

M-step. We determine the revised parameter estimate $\mathbf{\Psi}_1$ by maximizing the objective function $\mathcal{Q}_1$ formed as Eq.(6), i.e., $\mathbf{\Psi}_1^{new}=\mathrm{argmax}_{\mathbf{\Psi}_1}(\mathbf{\Psi}_1,\mathbf{\Psi}_1^{old})$. The parameters are updated as follows: (Derivation details are in Section 1.3.)

$$\hat{\theta}_k^{(m)}=\sum\nolimits_{i=1}^{I}\mathbb{E}\left[\mathbb{I}\left(y_i^{(m)}=k\right)\right]\Big/I, \tag{9}$$

$$\hat{\pi}_{kd}^{(mj)}=\frac{\sum_{i=1}^{I}\left\{\mathbb{E}\left[\mathbb{I}\left(y_i^{(m)}=k\right)\right]\mathbb{I}\left(l_{ij}^{(m)}=d\right)\right\}}{\sum_{i=1}^{I}\mathbb{E}\left[\mathbb{I}\left(y_i^{(m)}=k\right)\right]}. \tag{10}$$

After convergence. The integrated labels of instance $\mathbf{x}_i$ should be the sequence of class values with the maximum posterior probability. That is,

$$\hat{y}_i^{(1)},...,\hat{y}_i^{(M)}=\underset{k^{(1)},...,k^{(M)}}{\mathrm{argmax}}\left\{\mathrm{P}\left(y_i^{(1)}=k^{(1)},...,y_i^{(M)}=k^{(M)}\middle|\mathbf{L},\mathbf{\Psi}_1\right)\right\}. \tag{11}$$

1.3 Derivation Details in M-Step

M-step finds new values of the parameters that maximize the objective functions $\mathcal{Q}_1$ defined by Eq.(6). Because we have obtained the posterior probabilities of the latent variables, in the following derivations, we omit the condition of the observed label $\mathbf{L}$ and old values of parameters $\mathbf{\Psi}_1$.

For the MCMLI model, the objective is to maximize the quantity as follows:

$$\begin{aligned}\mathbb{E}_{\mathbf{Y}}\left[\ln\mathrm{P}\left(\mathbf{L},\mathbf{Y}|\mathbf{\Psi}_1\right)\right]&=\sum_{i=1}^{I}\mathbb{E}_{\mathbf{Y}}\left[\ln\mathrm{P}\left(\mathbf{l}_i,\mathbf{y}_i|\Theta,\mathbf{\Pi}_i\right)\right]\\&=\sum_{i=1}^{I}\mathbb{E}_{\mathbf{Y}}\left[\ln\left(\mathrm{P}\left(\mathbf{l}_i|\mathbf{y}_i,\mathbf{\Pi}_i\right)\underbrace{\mathrm{P}\left(\mathbf{y}_i|\Theta\right)}_{constant}\right)\right].\end{aligned} \tag{12}$$

We omit the constant factor with respect to the parameters $\mathbf{\Psi}_1$ in the partial derivatives. Thus, we only need to maximize the term $\sum_{i=1}^{I}\mathbb{E}_{\mathbf{Y}}\left[\ln\mathrm{P}\left(\mathbf{l}_i|\mathbf{y}_i,\mathbf{\Pi}_i\right)\right]$, which derives

$$\begin{aligned}&\sum_{i=1}^{I}\mathbb{E}_{\mathbf{Y}}\left[\ln\prod_{m=1}^{M}\prod_{k=1}^{K}\left(\theta_k^{(m)}\prod_{j=1}^{J}\prod_{d=1}^{K}\left(\pi_{kd}^{(mj)}\right)^{\mathbb{I}\left(l_{ij}^{(m)}=d\right)}\right)^{\mathbb{I}\left(y_i^{(m)}=k\right)}\right]\\&=\sum_{i=1}^{I}\sum_{m=1}^{M}\sum_{k=1}^{K}\mathbb{E}\left[\mathbb{I}\left(y_i^{(m)}=k\right)\right]\left[\ln\theta_k^{(m)}+\sum_{j=1}^{J}\sum_{d=1}^{K}\mathbb{I}\left(l_{ij}^{(m)}=d\right)\ln\pi_{kd}^{(mj)}\right].\end{aligned} \tag{13}$$

(1) Using a Lagrange multiplier to optimize Eq.(13) with respect to $\theta_k^{(m)}$, we construct a function

$$\mathcal{F}_1=\sum_{i=1}^{I}\mathbb{E}_{\mathbf{Y}}\left[\ln\mathrm{P}\left(\mathbf{l}_i|\mathbf{y}_i,\mathbf{\Pi}_i\right)\right]+\lambda\left(\sum_{k=1}^{K}\theta_k^{(m)}-1\right). \tag{14}$$

We let the partial derivative of Eq.(14) with respect to $\theta_k^{(m)}$ be zero. Applying sum-up-to-one condition $\left(\sum_{k=1}^{K}\theta_k^{(m)}=1\right)$, we have

$$\frac{\partial \mathcal{F}_1}{\partial \theta_k^{(m)}} = \frac{\sum_{i=1}^{I}\mathbb{E}\left[\mathbb{I}\left(y_i^{(m)}=k\right)\right]}{\theta_k^{(m)}} + \lambda = 0 \tag{15}$$

$$\Rightarrow \sum_{k=1}^{K}\sum_{i=1}^{I}\mathbb{E}\left[\mathbb{I}\left(y_i^{(m)}=k\right)\right] = -\lambda\sum_{k=1}^{K}\theta_k^{(m)} \Rightarrow I=-\lambda. \tag{16}$$

Plugging Eq.(16) into Eq.(15), we have

$$\hat{\theta}_k^{(m)} = \sum\nolimits_{i=1}^{I}\mathbb{E}\left[\mathbb{I}\left(y_i^{(m)}=k\right)\right]\Big/ I.$$

(2) Using a Lagrange multiplier to optimize Eq(13) with respect to $\pi_{kd}^{(mj)}$, we construct a function

$$\mathcal{F}_2 = \sum_{i=1}^{I}\mathbb{E}_{\mathbf{Y}}\left[\ln \mathrm{P}\left(\mathbf{l}_i\middle|\mathbf{y}_i,\mathbf{\Pi}_i\right)\right] + \lambda\left(\sum_{k=1}^{K}\pi_{kd}^{(mj)}-1\right). \tag{17}$$

We let the partial derivative of Eq.(17) with respect to $\pi_{kd}^{(mj)}$ be zero. Applying sum-up-to-one condition $\left(\sum_{d=1}^{K}\pi_{kd}^{(mj)}=1\right)$, we have

$$\frac{\partial \mathcal{F}_2}{\partial \pi_{kd}^{(mj)}} = \frac{\sum_{i=1}^{I}\left\{\mathbb{E}\left[\mathbb{I}\left(y_i^{(m)}=k\right)\right]\mathbb{I}\left(l_{ij}^{(m)}=d\right)\right\}}{\pi_{kd}^{(mj)}} + \lambda = 0 \tag{18}$$

$$\Rightarrow \sum_{d=1}^{K}\sum_{i=1}^{I}\left\{\mathbb{E}\left[\mathbb{I}\left(y_i^{(m)}=k\right)\right]\mathbb{I}\left(l_{ij}^{(m)}=d\right)\right\} = -\lambda\sum_{d=1}^{K}\pi_{kd}^{(mj)}$$

$$\Rightarrow \sum_{i=1}^{I}\mathbb{E}\left[\mathbb{I}\left(y_i^{(m)}=k\right)\right] = -\lambda. \tag{19}$$

Plugging Eq.(19) into Eq.(18), we have

$$\hat{\pi}_{kd}^{(mj)} = \frac{\sum_{i=1}^{I}\left\{\mathbb{E}\left[\mathbb{I}\left(y_i^{(m)}=k\right)\right]\mathbb{I}\left(l_{ij}^{(m)}=d\right)\right\}}{\sum_{i=1}^{I}\mathbb{E}\left[\mathbb{I}\left(y_i^{(m)}=k\right)\right]}. \qquad \blacksquare$$