Label Aggregation with Instance Grouping Model

Li'ang Yin, Jianhua Han, Yong Yu
Shanghai Jiao Tong University
No.800 Dongchuan Road
Shanghai, 200240, China
{yinla,hanjianhua44,yyu}@apex.sjtu.edu.cn

ABSTRACT

Label aggregation is one of the key topics in crowdsourcing research. Most researchers make their e orts in modeling ability of users and di culty of instances. In this paper, we consider label aggregation from the view of grouping instances. We assume instances are sampled from latent groups and they share the same true label with their corresponding groups. We construct a graphical model named InGroup(**In**stance **Group**ing model) to infer latent group assignment as well as true labels. The experimental results show the advantages of our model compared with baselines.

Keywords

Crowdsourcing, Label Aggregation, Instance Grouping, Graphical Model

1. INTRODUCTION

With the growing of online crowdsourcing platforms, crowdsourcing has become more and more popular. In the research of utilizing crowdsourcing data, label aggregation is one of the key topics. In many labeling scenarios, one instance may receive multiple labels from di erent users. For example, in order to identify the main object in an image, the image is shown to several users and each user gives his/her label. Having multiple labels, we want to know the most proper label(true label) of the image. Traditionally majority vote is applied to reduce user noise and infer a true label.

Recent research works have proposed sophisticated models for label aggregation. Most works design a generative model and introduce ability of users or di culty of instances(or both) in labeling process [1, 4]. Additionally, some works group users to deal with the sparsity of labels [2].

We consider label aggregation from the view of grouping instances. A real-world example(from CUB-200-2010) is used to illustrate the idea. Users are asked to label some local attributes(such as bill shape) of bird images. Images are clawed from websites and the key di culty is that some images do not show the corresponding attribute. One user can hardly label bill shape when the bird's bill is covered by a tree branch. Though it seems impossible to label that attribute, we observe that birds belong to the same category have the same bill shape. For example, black footed albatross(category) has hooked bill, then we can infer that an image(instance) of black footed albatross contains hooked bill even the bird's bill is covered. We introduce `latent group' to represent the concept of category since the prior knowledge about category is usually unknown. We construct a graphical model named InGroup(**In**stance **Group**ing model) to infer latent group assignment as well as true labels.

2. METHOD

Given M users and N instances, let $l_{u,i}$ denote the label user u assigns to instance i. We use binary labels in this paper, i.e. $l_{u,i} \in \{0, 1\}$. Furthermore, let x_i be a column vector to present feature of instance i. Our goal is to infer true label t_i for each instance.

InGroup(**In**stance **Group**ing model) is similar to Gaussian mixture model. Figure 1 illustrates its generative process. Assume there are K latent groups, each instance i is assigned to group k with probability $p(z_i = k)$, where z_i is the partition index. To simplify our model and avoid over t-ting, we use hard partition as in classical k-means algorithm. That is $p(z_i = k) = 1$ if instance i belongs to group k, and $p(z_i = k) = 0$ otherwise. We also use $z_i = k$ to denote instance i is assigned to group k in the rest paper.

The generative process of InGroup is described as follows:

For each group k, its center c_k has a prior multivariate normal distribution

$$p(c_k) \sim \mathcal{N}(\mu_0, I),$$

where I is the identity matrix and μ_0 can be set to the average value of all x_i. We use q_k to denote the probability of group k's label being `1'. q_k has a prior beta distribution

$$p(q_k) \sim \text{Beta}(\alpha, \beta).$$

For each instance i, its feature x_i is sampled from group k where $z_i = k$,

$$p(x_i) \sim \mathcal{N}(c_k, I).$$

Each label $l_{u,i}$ is also sampled from group k where $z_i = k$,

$$p(l_{u,i}) \sim \text{Bernoulli}(q_k).$$

Let X, L, C, Q, Z denote sets $\{x_i\}, \{l_{u,i}\}, \{c_k\}, \{q_k\}$ and $\{z_i\}$ respectively. Our goal is to nd proper C, Q and Z which maximize the posterior probability

$$p(C, Q \mid X, L; Z) \propto p(X, L \mid C, Q; Z)p(C)p(Q). \quad (1)$$

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2889353.

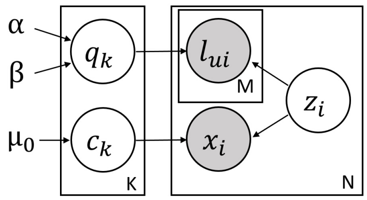

Figure 1: Graphical model presentation of InGroup.

Formula (1) can be solved by an expectation-maximization algorithm. In E-step, we need to nd best partition z_i for each instance given current $\{c_k\}$ and $\{q_k\}$. By taking logarithm for formula (1) and keeping terms related with z_i, we obtain the following equation

$$z_i^* = \arg\max_{z_i} \sum_{k=1}^{K} p(z_i = k)[-\frac{1}{2}(x_i - c_k)^T(x_i - c_k) + n_{1,i}\log q_k + n_{0,i}\log(1 - q_k)], \tag{2}$$

where $n_{1,i} = \|l_{u,i} : l_{u,i} = 1\|$ denotes the number of '1' labels of instance i and $n_{0,i} = \|l_{u,i} : l_{u,i} = 0\|$ denotes the number of '0' labels.

In M-step, we update c_k and q_k respectively,

$$c_k = \frac{\mu_0 + \sum_{z_i=k} x_i}{1 + \|i : z_i = k\|}, \tag{3}$$

$$q_k = \frac{\alpha + n_{1,k}}{\alpha + \beta + n_{1,k} + n_{0,k}}, \tag{4}$$

where $\|i : z_i = k\|$ denotes the number of instances in group k, $n_{1,k} = \|l_{u,i} : z_i = k, l_{u,i} = 1\|$ denotes the number of all '1' labels in group k, and $n_{0,k} = \|l_{u,i} : z_i = k, l_{u,i} = 0\|$ denotes the number of '0' labels.

When model converges, we obtain q_k for each group. By applying the assumption that instances belong to the same group share the same true label, we can predict instance label in a simple way by rounding q_k

$$t_i = \text{round}(q_k), \text{if } z_i = k. \tag{5}$$

3. EXPERIMENT

We conduct experiments on CUB-200-2010 dataset[3]. The dataset contains 6033 bird images. Users are asked to label local attributes for a given bird image. We regard inferring true label on a speci c attribute as a label aggregation task and construct three tasks, namely bill(has all-purpose bill or not), head(has plain head pattern or not) and shape(is perching-like or not). These attributes are challenging(user labels for an instance often disagree with each other) and their ground truth can be found from whatbird.com and other bird websites. Feature vectors are constructed from user labels. Speci cally, for an image, labels on each attribute are averaged as one element in feature vector, then a 287-dimensional vector is obtained since there are 287 attributes for each image. We have tried image feature(such as SIFT), but it seems do not work on the data.

InGroup is compared with MV(majority vote), GLAD[4], DARE[1] and CBCC[2] models. The accuracy is illustrated in Table 1. We can observe that other models except InGroup show slight or no improvement comparing with classical majority vote on the tasks. This may due to sparsity of labels and di culty of instances as we have described in the introduction section. InGroup achieves better performance because it can capture relationships between instances and latent groups.

Table 1: Accuracy Comparison ($\alpha = \beta = 2$ for InGroup)

	MV	GLAD	DARE	CBCC	InGroup
bill	0.8168	0.8140	0.8145	0.8140	**0.8588**
head	0.8705	0.8664	0.8652	0.8661	**0.8938**
shape	0.8754	0.9010	**0.9105**	0.9010	**0.9161**

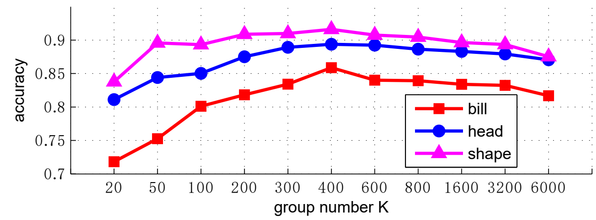

Figure 2: Accuracy varies along with group number K.

The latent group number K is a key parameter in our model. We illustrate in Figure 2 how accuracy varies along with di erent values of K. The best accuracy is usually achieved with group number ranging from 300 to 600(about 10-20 instances in each latent group). When K is too small, one group contains too many instances that cannot be distinguished from each other. When K is too large, there are not enough instances in each group to infer reliable group label. Note that setting K to the number of all the instances makes InGroup the majority vote model.

4. CONCLUSION

In this paper, we consider label aggregation from the view of grouping instances and propose a graphical model InGroup. Experimental results show that InGroup achieves high accuracy on CUB-200-2010 dataset.

5. REFERENCES

[1] Y. Bachrach, T. Graepel, T. Minka, and J. Guiver. How to grade a test without knowing the answers | a bayesian graphical model for adaptive crowdsourcing and aptitude testing. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 1183{1190, 2012.

[2] M. Venanzi, J. Guiver, G. Kazai, P. Kohli, and M. Shokouhi. Community-based bayesian aggregation models for crowdsourcing. In *Proceedings of the 23rd international conference on World wide web*, pages 155{164. International World Wide Web Conferences Steering Committee, 2014.

[3] P. Welinder, S. Branson, T. Mita, C. Wah, F. Schro , S. Belongie, and P. Perona. Caltech-ucsd birds 200.

[4] J. Whitehill, T.-f. Wu, J. Bergsma, J. R. Movellan, and P. L. Ruvolo. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in neural information processing systems*, pages 2035{2043, 2009.