Partial Label Learning via Label Influence Function

Xiuwen Gong
Faculty of Engineering, University of Sydney, Australia

ICML 2022

Joint work with Dong Yuan, Wei Bao

Problem Setting

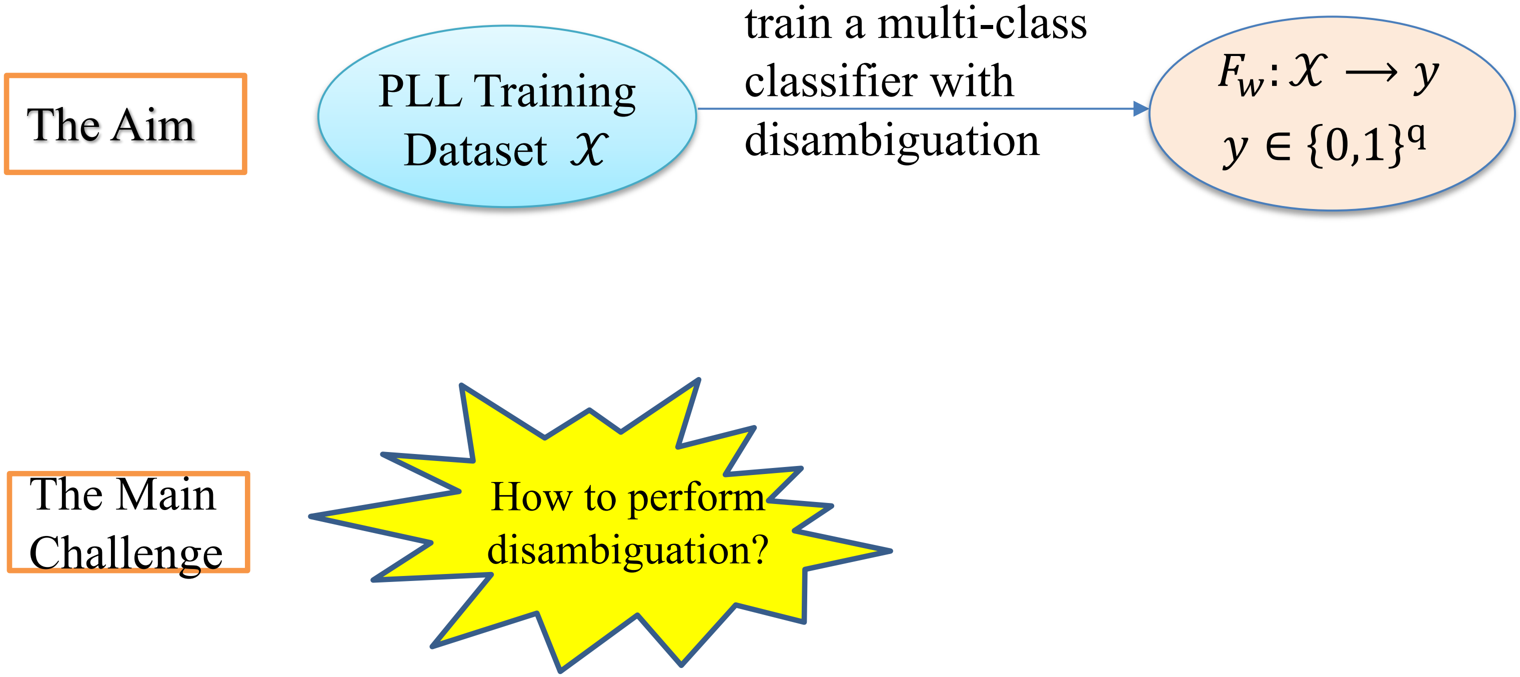
Partial Label Learning (PLL)

- Each instance is associated with a set of candidate labels, but only one is the ground-truth label, while others are false-positive labels.

Applications

- Document annotation;
- Image annotation;
- Video annotation.

The Aim and Main Challenge for PLL



Motivation

perform disambiguation

the averaging-based method [1] [2]

It can be easily misled by the the noisy false-positive labels in the candidate set, which fails to generalize well in testing.

the identification-based method [3] [4]

labeling confidence-based method [5] [6]

neural network-based method [7] [8]

The identification-based method usually update the model parameter and latent label/confidence variable via an iterative procedure. As the ground-truth variable is discrete, the optimization is often NP-hard. Existing approaches usually take the label that incurs a minimal loss as the ground-truth label or use the weight to represent which label has a high likelihood to be the ground-truth label.

Little work has been done to investigate from how a candidate label changing a predictive model. Motivated by influence function (IF) that characterizes how a model's predictive loss changes when a small fraction of data points being perturbed, this paper first attempts to apply IF to deal with PLL.

Partial Label Learning via Label Influence Function

Our Contributions

We provide a new insight into partial label learning (PLL) from the perspective of influence function, and develop a novel framework called Partial Label Learning via Label Influence Function (PLL-IF),

We first define a quantity called Label Impact to quantify how a candidate label changes a predictive model, which can be further employed as an indicator to identify the most influential candidate label with highest impact on a model optimizer.

We then introduce Label Influence Function (LIF) to efficiently approximate the label impact, which avoids retraining the model after each label is removed or perturbed, and largely reduces the heavy computation of the label impact. We further propose a novel ground-truth label identification method called Ground-truth Label Identification via Label Influence Function (GLI-LIF)

Lastly, we implement the PLL-IF framework through a representative non-neural network model (i.e., SVM model) and a basic neural network model respectively. Extensive experiments are conducted on synthetic and real-world datasets to demonstrate the superiorities of the proposed PLL-IF framework.

Partial Label Learning via Label Influence Function

- In order to identify the ground-truth labels from the candidate label set, we propose a novel ground-truth label identification method called Ground-truth Label Identification via Label Influence Function (GLI-LIF), the objective function can be expressed as follows:

$$y = \arg \max_{j \in Y, l \in Y \setminus j} \left\{ \left\| \Delta \hat{\mathbf{w}}_j \right\|_1, \left\| \Delta \hat{\mathbf{w}}_l \right\|_1 \right\}$$

$\|\cdot\|_1$ denotes the ℓ_1-norm.

$$\Delta \hat{\mathbf{w}}_j = \epsilon \mathcal{I}(\mathbf{z})$$

$$\Delta \hat{\mathbf{w}}_l = \varepsilon (\mathcal{I}(\mathbf{z}_\delta) - \mathcal{I}(\mathbf{z}))$$

$$\mathcal{I}(\mathbf{z}) \triangleq \left. \frac{d \hat{\mathbf{w}}_{\epsilon, -\mathbf{z}}}{d \epsilon} \right|_{\epsilon = 0} = -H_{\hat{\mathbf{w}}}^{-1} \nabla_{\mathbf{w}} \ell(\mathbf{y}, f(\mathbf{x}; \hat{\mathbf{w}}))$$

where $H_{\hat{\mathbf{w}}} = \frac{1}{n} \sum_{i=1}^{n} \nabla_{\mathbf{w}}^2 \ell(\mathbf{y}_i, f(\mathbf{x}_i; \hat{\mathbf{w}}))$ is the Hessian matrix and is positive definite by assumption.

Experiment Results

Table 3: Win/tie/loss counts of pairwise t-test (at 5% significance level) between PLL-IF+NN and each baseline.

Config. \ Method	PLL-IF+NN vs -					
	PL-SVM	CLPL	M3PL	SURE	PRODEN	PL-BLC
varying p, r = 1	54\0\0	54\0\0	54\0\0	53\1\0	53\1\0	47\7\0
varying p, r = 2	54\0\0	54\0\0	54\0\0	54\0\0	47\4\3	42\6\6
varying p, r = 3	54\0\0	54\0\0	54\0\0	54\0\0	50\3\1	46\3\5
Total	162\0\0	162\0\0	162\0\0	161\1\0	150\8\4	135\16\11

Table 4: Win/tie/loss counts of pairwise t-test (at 5% significance level) between PLL-IF+SVM and each baseline.

Config. \ Method	PLL-IF+SVM vs -					
	PL-SVM	CLPL	M3PL	SURE	PRODEN	PL-BLC
varying p, r = 1	54\0\0	51\3\0	45\5\4	40\4\10	24\4\26	9\3\42
varying p, r = 2	54\0\0	54\0\0	51\3\0	46\5\3	12\3\39	8\1\45
varying p, r = 3	54\0\0	52\2\0	49\4\1	38\4\12	8\3\43	16\5\33
Total	162\0\0	157\5\0	145\12\5	124\13\25	44\10\108	33\9\120

Table 5: Mean accuracy ± standard deviation via five-fold cross validation on six real-world datasets for all methods. The best results are in bold. •/○ indicates that our method (PLL-IF+NN / PLL-IF+SVM) is significantly superior / inferior than the baseline (pairwise t-test at %5 significance level).

Method	Lost	MSRCv2	Mirflickr	BirdSong	Soccer Player	Yahoo!News
PLL-IF+NN	**0.809 ± .041**	**0.538 ± .027**	**0.569 ± .030**	**0.753 ± .003**	**0.560 ± .004**	**0.683 ± .007**
PLL-IF+SVM	0.782 ± .012	0.513 ± .022	0.561 ± .003	0.723 ± .017	0.554 ± .010	0.646 ± .026
PL-SVM	0.691 ± .012•	0.481 ± .037•	0.441 ± .061•	0.661 ± .067•	0.462 ± .006•	0.615 ± .015•
CLPL	0.732 ± .032•	0.433 ± .020•	0.549 ± .017	0.635 ± .019•	0.367 ± .004•	0.471 ± .049•
M3PL	0.747 ± .031•	0.499 ± .026•	0.480 ± .016•	0.694 ± .065•	0.440 ± .005•	0.623 ± .062•
SURE	0.767 ± .026	0.508 ± .021	0.562 ± .015	0.702 ± .025•	0.531 ± .014	0.632 ± .015•
PRODEN	0.765 ± .014	0.452 ± .017•	0.524 ± .011	0.721 ± .004	0.559 ± .005	0.674 ± .005
PL-BLC	0.806 ± .032	0.536 ± .037	0.558 ± .038	0.746 ± .017	0.540 ± .008	0.679 ± .005

References

[1] Cour, T., Sapp, B., and Taskar, B. Learning from partial labels. Journal of Machine Learning Research, 12:1501-1536, 2011.

[2] Zhang, M. and Yu, F. Solving the partial label learning problem: An instance-based approach. In IJCAI, pp.4048–4054, 2015.

[3] Yu, F. and Zhang, M. Maximum margin partial label learning. Machine Learning, 106(4):573–593, 2017.

[4] Chai, J., Tsang, I. W., and Chen, W. Large margin partial label machine. IEEE Transactions on Neural Networks and Learning Systems, 31(7):2594–2608, 2020.

[5] Yan, Y. and Guo, Y. Partial label learning with batch label correction. In AAAI, pp. 6575–6582, 2020.

[6] Feng, L. and An, B. Partial label learning with self-guided retraining. In AAAI, pp. 3542–3549, 2019.

[7] Lv, J., Xu, M., Feng, L., Niu, G., Geng, X., and Sugiyama, M. Progressive identification of true labels for partial label learning. In ICML, volume 119, pp. 6500–6510. PMLR, 2020.

[8] Feng, L., Lv, J., Han, B., Xu, M., Niu, G., Geng, X., An, B., and Sugiyama, M. Provably consistent partial-label learning. In NeurIPS, 2020.

Thank you!