Partial Label Learning with a Partner

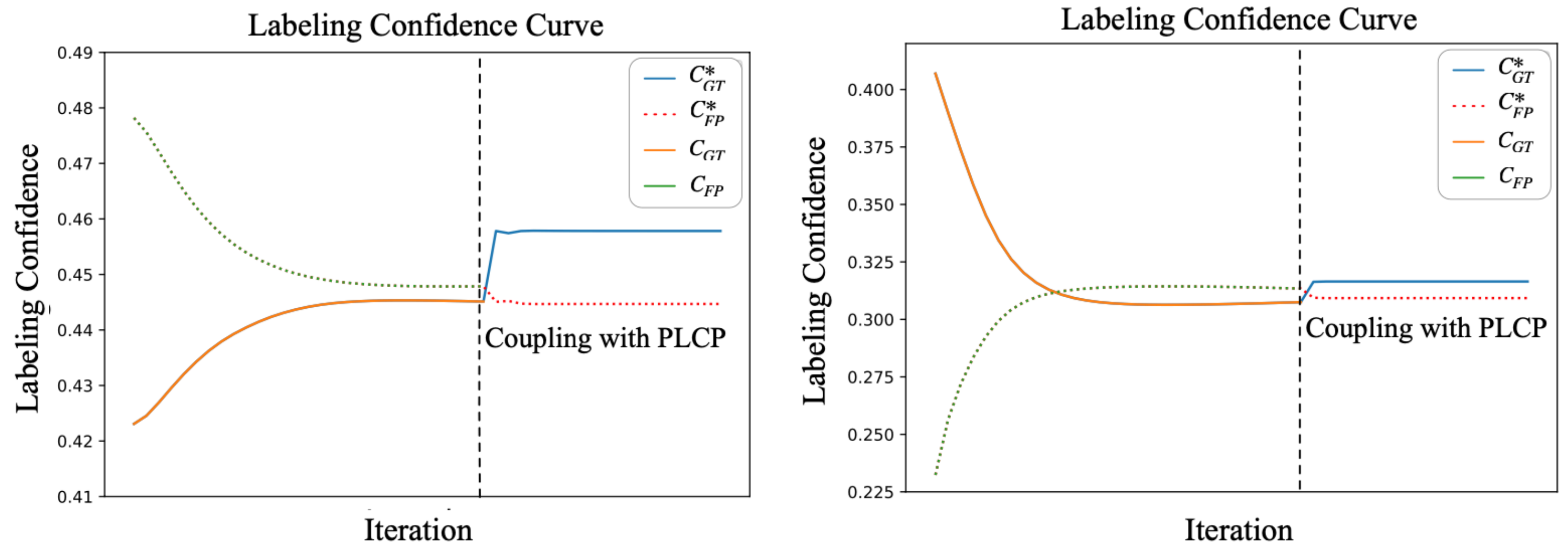
Chongjie Si[1], Zekun Jiang[1], Xuehui Wang[1], Yan Wang[2], Xiaokang Yang[1], Wei Shen[1]
[1]MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University
[2]Shanghai Key Lab of Multidimensional Information Processing, ECNU

AAAI-24 / IAAI-24 / EAAI-24

Introduction

Task: Partial label learning, each sample has several labels, among which only one is ground truth.

Intriguing Phenomena:



- Each candidate label's labeling confidence is likely to continually increase or decrease until convergence.
- A high-confidence false positive label may still have significant confidence later, risking incorrect identification of the true label.

Motivation:

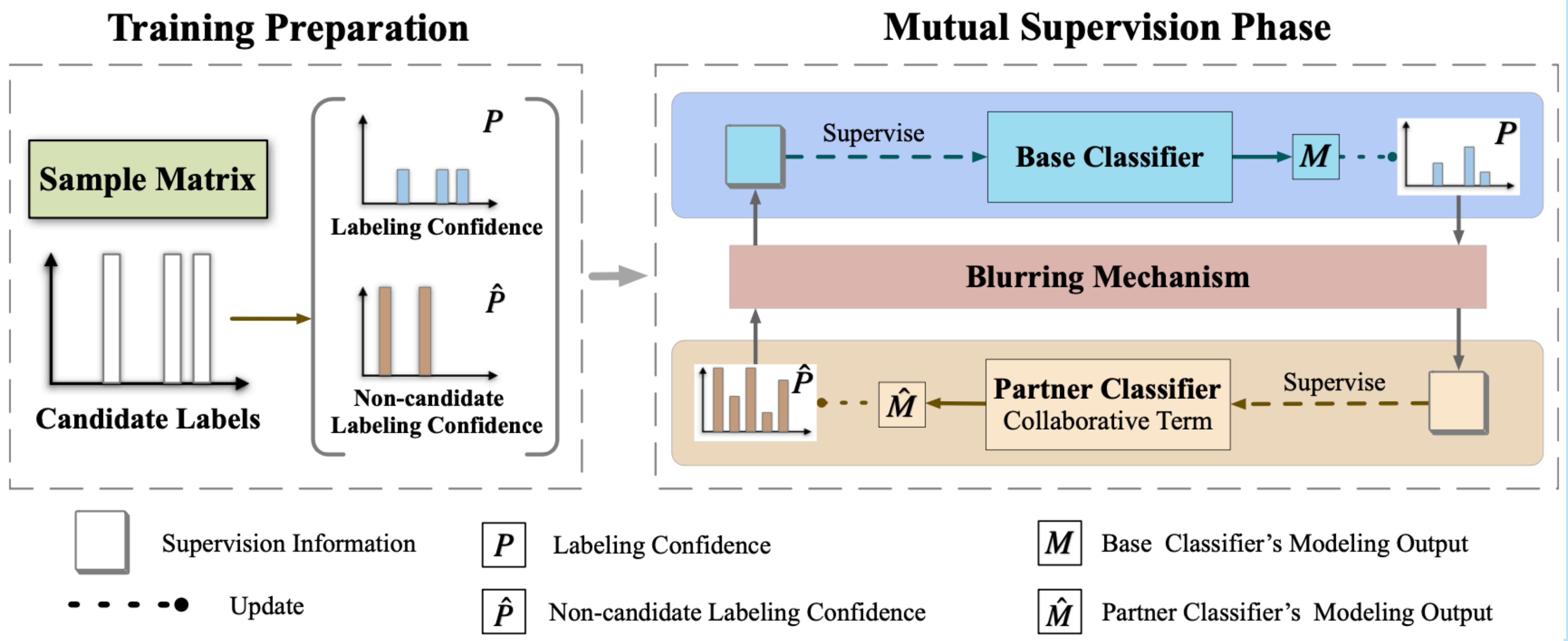
- Correcting mislabeled samples for a PLL classifier itself is hard.
- Non-candidate label information is rarely investigated in PLL.

Contributions:

- We highlight two representative errors a PLL classifier may make, which has not been previously investigated in PLL.
- We introduce a partner classifier based on non-candidate labels to better identify and correct mislabeled samples of a base classifier through a mutual supervision framework, which is applicable to all types of PLL approaches.
- We propose a novel collaborative term in the partner classifier, which links the base classifier and itself. Additionally, a blurring mechanism is introduced to add uncertainty to the outputs.

Method

Framework:



Component:

- Base classifier: any existing PLL approach.
- Update strategy: $\mathbf{P}_1 = \mathcal{T}_0\left(\mathcal{T}_{\mathbf{Y}}(\alpha\mathbf{P} + (1-\alpha)\mathbf{M})\right)$
- Blurring mechanism:
  - $\mathbf{Q}_1 = \phi\left(e^k \mathbf{P}_1\right) \odot \mathbf{Y}$, $\phi(x) = e^x$
  - Normalize $\mathbf{Q}_1$ to obtain $\mathbf{O}_1$
- Partner classifier with the collaborative term:

$$\min_{\hat{\mathbf{W}}, \hat{\mathbf{b}}, \mathbf{C}} \left\| \mathbf{X}\hat{\mathbf{W}} + \mathbf{1}_n \hat{\mathbf{b}}^\top - \mathbf{C} \right\|_F^2 + \gamma \operatorname{tr}\left(\mathbf{O}_1 \mathbf{C}^\top\right) + \lambda \|\hat{\mathbf{W}}\|_F^2$$

$$\text{s.t. } \hat{\mathbf{Y}} \leq \mathbf{C} \leq \mathbf{1}_{n \times l}, \ \mathbf{C}\mathbf{1}_l = (l-1)\mathbf{1}_n$$

Pipeline:

A partner classifier, built on non-candidate label data, facilitates mutual supervision with the base classifier. Each supervision stage updates labeling confidence using modeling output and applies a blurring mechanism. This output serves as supervision for interacting with the partner classifier, which operates similarly to the base classifier.

Experiment

Comparison with Stand-alone Methods:

Approaches	FG-NET	Lost	MSRCv2	Mirflickr	Soccer Player	Yahoo!News
PL-CL	0.072 ± 0.009	0.710 ± 0.022	0.469 ± 0.016	0.647 ± 0.012	0.534 ± 0.004	0.618 ± 0.003
PL-CL-PLCP	0.080 ± 0.009 •	0.763 ± 0.020 •	0.493 ± 0.013 •	0.665 ± 0.011 •	0.543 ± 0.002 •	0.625 ± 0.002 •
PL-AGGD	0.063 ± 0.010	0.690 ± 0.020	0.451 ± 0.023	0.610 ± 0.012	0.521 ± 0.004	0.605 ± 0.002
PL-AGGD-PLCP	0.076 ± 0.010 •	0.717 ± 0.020 •	0.473 ± 0.017 •	0.668 ± 0.014 •	0.534 ± 0.005 •	0.609 ± 0.002 •
SURE	0.052 ± 0.007	0.709 ± 0.022	0.445 ± 0.022	0.630 ± 0.022	0.519 ± 0.004	0.598 ± 0.002
SURE-PLCP	0.076 ± 0.011 •	0.719 ± 0.019 •	0.460 ± 0.020 •	0.657 ± 0.020 •	0.527 ± 0.004 •	0.606 ± 0.002 •
LALO	0.065 ± 0.010	0.682 ± 0.019	0.449 ± 0.016	0.629 ± 0.016	0.523 ± 0.003	0.601 ± 0.003
LALO-PLCP	0.076 ± 0.010 •	0.701 ± 0.019 •	0.453 ± 0.015 •	0.647 ± 0.018 •	0.529 ± 0.004 •	0.605 ± 0.002 •
PL-SVM	0.043 ± 0.008	0.406 ± 0.033	0.389 ± 0.029	0.516 ± 0.022	0.412 ± 0.006	0.509 ± 0.006
PL-SVM-PLCP	0.081 ± 0.011 •	0.688 ± 0.029 •	0.468 ± 0.025 •	0.607 ± 0.023 •	0.526 ± 0.005 •	0.609 ± 0.002 •
PL-KNN	0.036 ± 0.006	0.300 ± 0.018	0.393 ± 0.014	0.454 ± 0.016	0.492 ± 0.003	0.368 ± 0.004
PL-KNN-PLCP	0.076 ± 0.009 •	0.662 ± 0.025 •	0.469 ± 0.016 •	0.607 ± 0.023 •	0.523 ± 0.004 •	0.593 ± 0.004 •
Improvement:	PL-CL: 3.61%	PL-AGGD: 5.10 %	SURE: 12.24 %	LALO: 4.01 %	PL-SVM: 39.26 %	PL-KNN: 53.98 %

(Column group "Data set" spans FG-NET through Yahoo!News.)

Comparison with Deep-learning Based Methods:

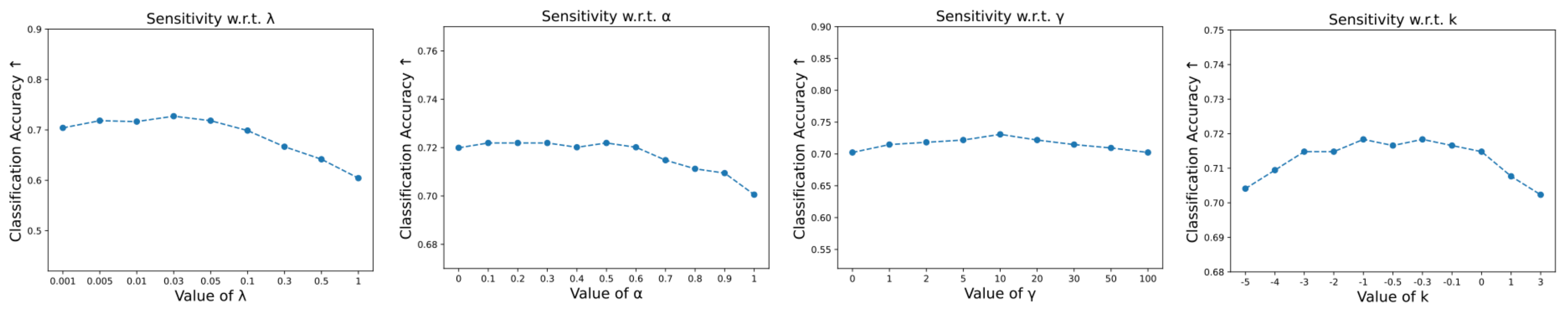
Approaches	CIFAR-10			CIFAR-100		
	$q = 0.1$	$q = 0.3$	$q = 0.5$	$q = 0.01$	$q = 0.05$	$q = 0.1$
PICO	94.39 ± 0.18 %	94.18 ± 0.12 %	93.58 ± 0.06 %	73.09 ± 0.34 %	72.74 ± 0.30 %	69.91 ± 0.24 %
PICO-PLCP	94.80 ± 0.07 % •	94.53 ± 0.10 % •	93.67 ± 0.16 % •	73.90 ± 0.20 % •	73.51 ± 0.21 % •	70.00 ± 0.35 %
Fully Supervised	$\mathcal{B}$: 94.91 ± 0.07 %	$\mathcal{B}$-PLCP: 95.02 ± 0.03 %		$\mathcal{B}$: 73.56 ± 0.10 %	$\mathcal{B}$-PLCP: 73.69 ± 0.14 %	
PRODEN	89.12 ± 0.12 %	87.56 ± 0.15 %	84.92 ± 0.31 %	63.36 ± 0.33 %	60.88 ± 0.35 %	50.98 ± 0.74 %
PRODEN-PLCP	89.63 ± 0.15 % •	88.19 ± 0.19 % •	85.31 ± 0.31 % •	64.20 ± 0.25 % •	61.78 ± 0.29 •	50.76 ± 0.90 %
Fully Supervised	$\mathcal{B}$: 90.03 ± 0.13 %	$\mathcal{B}$-PLCP: 90.30 ± 0.08 %		$\mathcal{B}$: 65.03 ± 0.35 %	$\mathcal{B}$-PLCP: 65.52 ± 0.32 %	

Ablation Study:

Kernel	Partner	Blur	FG-NET	Lost	MSRCv2	Mirflickr	Soccer Player	Yahoo!News
PL-AGGD			0.063 ± 0.010	0.690 ± 0.020	0.451 ± 0.023	0.610 ± 0.012	0.521 ± 0.004	0.605 ± 0.002
×	P	×	0.073 ± 0.011 •	0.698 ± 0.023 •	0.380 ± 0.013 •	0.542 ± 0.013 •	0.492 ± 0.003 •	0.463 ± 0.002 •
✓	P	×	0.073 ± 0.006 •	0.721 ± 0.024 ○	0.471 ± 0.016 •	0.664 ± 0.012 •	0.521 ± 0.004 •	0.608 ± 0.003 •
✓	O	✓	0.071 ± 0.001 •	0.721 ± 0.004 ○	0.470 ± 0,020 •	0.663 ± 0.011 •	0.522 ± 0.003 •	0.605 ± 0.002 •
✓	P	✓	0.076 ± 0.010	0.717 ± 0.020	0.473 ± 0.017	0.668 ± 0.014	0.534 ± 0.005	0.609 ± 0.002

(Column group "Data set" spans FG-NET through Yahoo!News.)

Sensitivity Analysis:



(a) Sensitivity w.r.t. λ. (b) Sensitivity w.r.t. α. (c) Sensitivity w.r.t. γ. (d) Sensitivity w.r.t. k.