Towards Semi-supervised Learning with Non-random Missing Labels

Yue Duan[1], Zhen Zhao[2], Lei Qi[3], Luping Zhou[2], Lei Wang[4], and Yinghuan Shi[1]

[1] Nanjing University, China [2] University of Sydney, Australia [3] Southeast University, China [4] University of Wollongong, Australia

Introduction

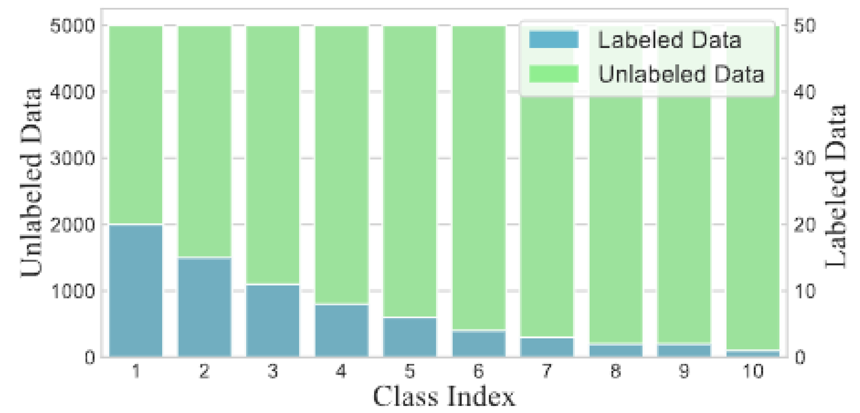


Fig. 1 The class distribution of total data is balanced whereas labeled data is unevenly distributed across classes.

Current prevailing SSL methods utilize the model trained on the labeled data to impute pseudo-labels for the unlabeled data, thereby boosting the model performance. Although these methods have made exciting advances in SSL, they only work well in the conventional setting, i.e., the labeled and unlabeled data fall into the same (balanced) class distribution. [1] originally terms the scenario of the **labeled and unlabeled data belonging to mismatched class distributions** as label *Missing Not At Random* (MNAR). During the same period, [2, 3] also independently explored the issue of mismatched distributions. For example, a typical MNAR scenario is shown in Fig. 1. The pseudo-rectifying ability of the SSL model could be severely perturbed in MNAR.

Motivation

◆**It is feasible to guide pseudo-rectifying from the class level, i.e., pointing out the latent direction of class transition based on its current class prediction only.**

+

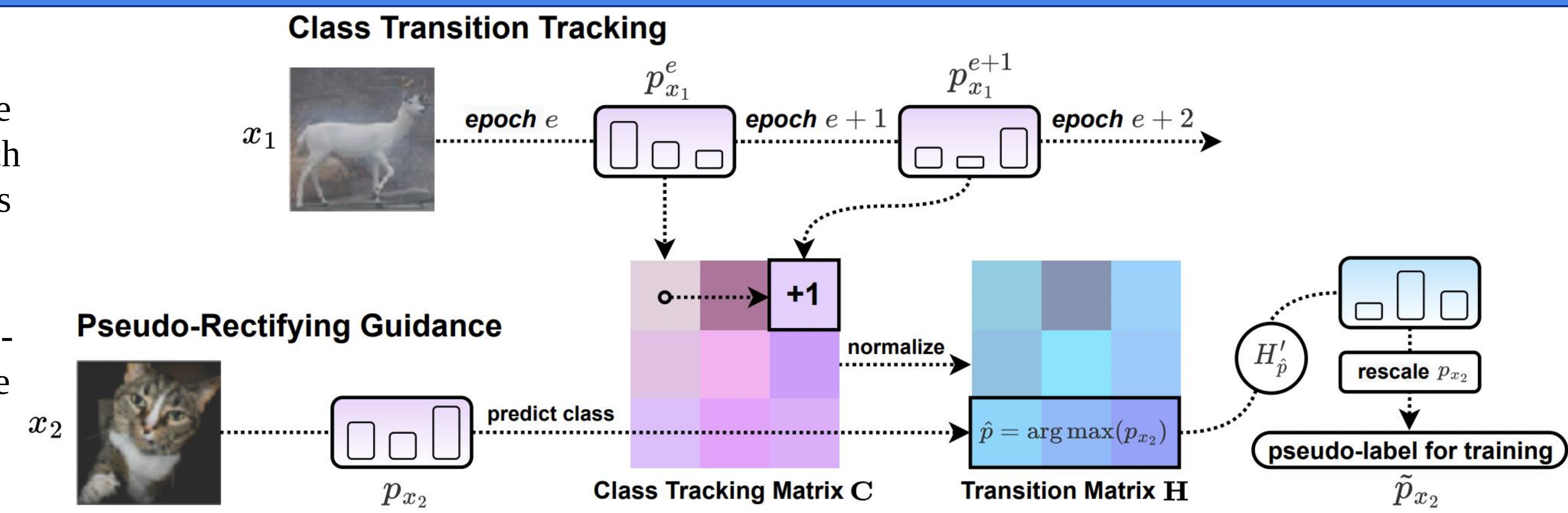
◆**Our intuition could be regarded as perturbations on some confident class predictions to preserve the pseudo-rectifying ability of the model. Such a strategy does not rely on the matched class distributions assumption and therefore is amenable to MNAR.**

Method

Class Transition Tracking

Pseudo-rectifying is defined as the change of the label assignment decision made by the SSL model for the same sample according to the knowledge learned at each new epoch. This process may cause **class transition**, i.e., given a sample, its class prediction at the current epoch is different from that at the last epoch. The design of class transition tracking (CTT) is desirable for the following reasons:

- Class transitions can reflect the **similarities** between classes and provide class-level guidance. We believe that two classes are conceptually similar if they are frequently misclassified to each other by the classifier.
- Class transitions contain the model's efforts in **identifying rare classes.**



Pseudo-Rectifying Guidance (PRG)

Class tracking matrix C is obtained by tracking the class transitions of pseudo-labels (e.g., p_{x_1} for sample x_1) between epoch e and epoch $e+1$ caused by pseudo-rectifying procedure. The Markov random walk defined by transition matrix H (each row H_i represents the transition probability vector corresponding to class i) is modeled on the graph constructed over C. Generally, given a pseudo-label, e.g., p_{x_2} for sample x_2, class- and batch-rescaled H (i.e., H') is utilized to provide the class-level pseudo-rectifying guidance for p_{x_2} according to its class prediction $\hat{p} = \mathrm{argmax}(p_{x_2})$. Finally, the rescaled pseudo-label $\tilde{p}_{x_2}$ is used for the training.

Experiments

Method	CIFAR-10			CIFAR-100			mini-ImageNet	
	$\gamma = 20$	50	100	50	100	200	50	100
Π Model*	21.59	27.54	30.39	24.95	29.93	33.91	11.77	15.30
MixMatch*	26.63	31.28	28.02	37.82	41.32	42.92	13.12	18.30
ReMixMatch*	41.84	38.44	38.20	42.45	39.71	39.22	22.64	23.50
FixMatch*	56.26	65.61	72.28	50.51	48.82	50.62	23.56	26.57
+ Crest*	51.10 ↓5.16	55.40 ↓10.21	63.60 ↓8.68	40.30 ↓10.21	46.30 ↓2.52	49.60 ↓1.02	–	–
+ DARP*	63.14 ↑6.88	70.44 ↑4.83	74.74 ↑2.46	38.87 ↓11.64	40.49 ↓8.33	44.15 ↓6.47	–	–
+ CADR*	79.63 ↑23.37	93.79 ↑23.37	93.97 ↑21.69	59.53 ↑9.02	60.88 ↑12.06	63.30 ↑12.68	29.07 ↑5.51	32.78 ↑6.21
+ PRG (Ours)	**94.04** ↑37.78 ±0.18	**94.09** ↑28.48 ±0.18	94.28 ↑22.00 ±0.22	59.11 ↑8.60 ±0.54	61.84 ↑13.02 ±0.45	63.41 ↑12.79 ±4.08	44.28 ↑20.72 ±0.54	44.99 ↑18.42 ±1.25
+ PRG$^{\text{Last}}$ (Ours)	93.81 ↑37.55 ±0.98	93.44 ↑27.83 ±1.05	93.48 ↑21.20 ±0.79	59.54 ↑9.03 ±0.99	62.36 ↑13.54 ±0.23	60.56 ↑9.94 ±1.86	40.73 ↑17.17 ±1.27	43.89 ↑17.32 ±0.14
SimMatch	83.45 ±2.32	86.77 ±2.15	90.12 ±1.90	60.06 ±1.17	60.35 ±0.59	61.14 ±0.24	39.49 ±1.04	40.37 ±0.96
+ PRG (Ours)	86.87 ↑3.42 ±2.38	91.68 ↑4.91 ±1.59	**94.59** ↑4.47 ±0.32	**65.65** ↑5.59 ±0.61	**65.89** ↑5.54 ±0.30	66.50 ↑5.36 ±0.17	**44.61** ↑5.12 ±0.24	**46.48** ↑6.11 ±0.15
+ PRG$^{\text{Last}}$ (Ours)	86.46 ↑3.01 ±1.75	90.48 ↑3.71 ±0.96	94.22 ↑4.10 ±0.14	65.10 ↑5.04 ±0.40	65.52 ↑5.17 ±0.29	**66.62** ↑5.48 ±0.19	42.06 ↑2.57 ±1.81	44.86 ↑9.95 ±4.49

Tab. 1 Accuracy (%) in MNAR under CADR's protocol.

Method	CIFAR-10 ($n_L = 40$)		CIFAR-10 ($n_L = 250$)		CIFAR-100 ($n_L = 2500$)		mini-ImageNet ($n_L = 1000$)	
	$N_1 = 10$	20	100	200	100	200	40	80
FixMatch	85.72 ±0.93	76.53 ±3.03	69.76 ±5.57	46.53 ±8.12	61.31 ±3.67	41.38 ±2.84	36.20 ±0.36	28.33 ±0.41
+ CADR	85.54 ↓0.18 ±2.07	75.11 ↑1.42 ±3.41	92.25 ↑22.49 ±1.61	63.92 ↑17.39 ±5.47	**61.62** ↑0.31 ±0.93	46.16 ↑4.78 ±1.45	36.08 ↓0.12 ±0.84	30.52 ↑2.19 ±0.99
+ PRG (Ours)	**91.87** ↑6.15 ±1.05	77.44 ↑0.91 ±15.96	**93.93** ↑24.17 ±0.16	**67.86** ↑21.33 ±16.98	61.49 ↑0.18 ±3.93	**49.84** ↑8.46 ±1.37	**39.99** ↑3.79 ±0.76	**35.39** ↑7.06 ±0.47
+ PRG$^{\text{Last}}$ (Ours)	85.66 ↓0.06 ±5.93	**77.85** ↑1.32 ±1.86	92.80 ↑23.04 ±1.44	64.00 ↑17.47 ±5.02	60.41 ↓0.90 ±1.01	43.80 ↑2.42 ±1.71	39.84 ↑3.64 ±0.05	33.17 ↑4.84 ±0.52

Tab. 2 Accuracy (%) in MNAR under our protocol.

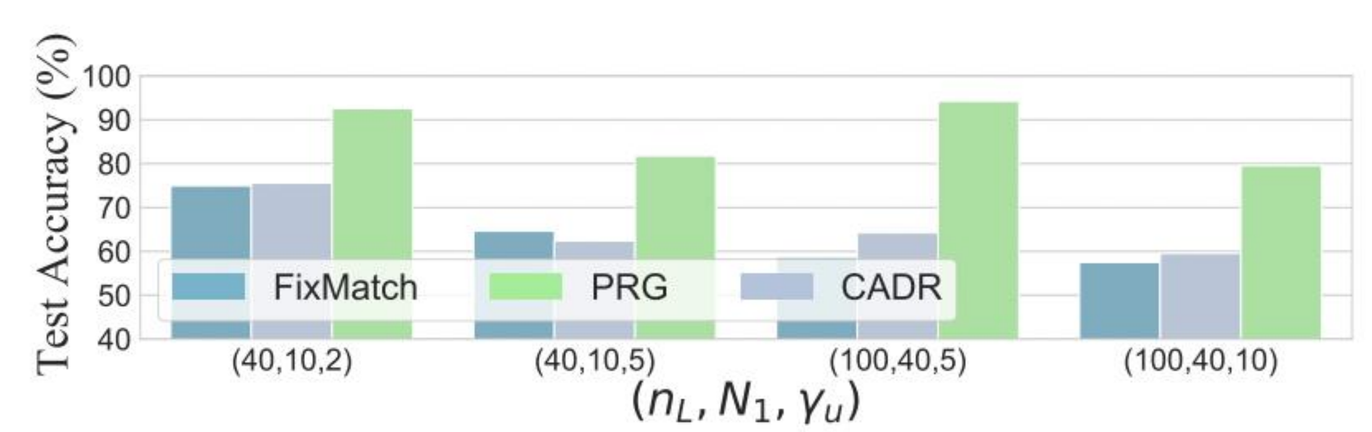


Fig. 2. Our protocol with imbalanced and mismatched labeled and unlabeled data.

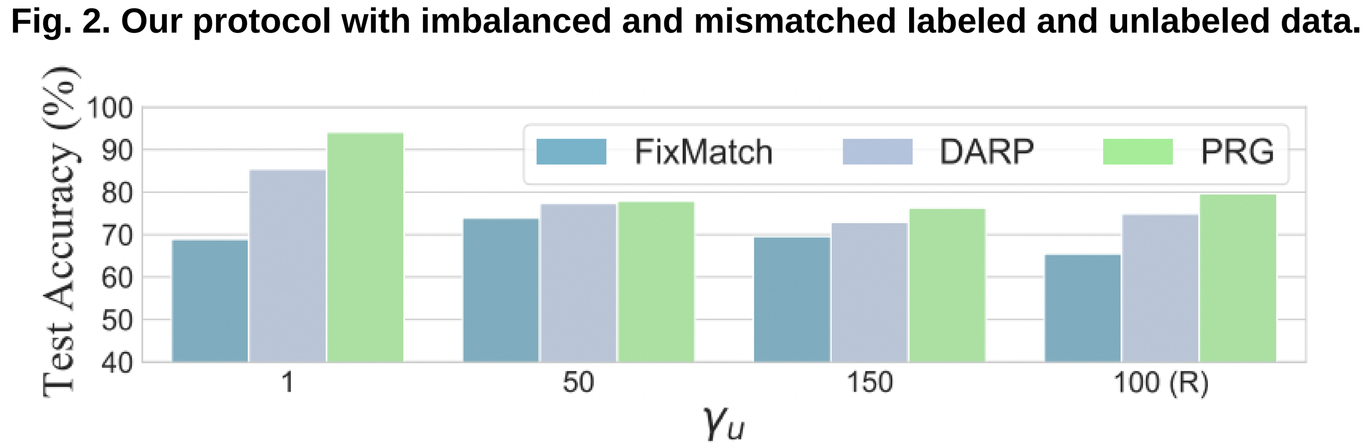


Fig. 3. DAPR's protocol with imbalanced and mismatched labeled and unlabeled data.

Method	$\gamma = 20$	50	100	$n_L, N_1 = 40, 10$	40, 20
VIME	63.38±4.42	63.75±6.10	64.80±2.76	50.13±7.56	30.73±8.69
+ PRG (Ours)	59.41±14.45	65.92±13.90	**66.60±12.58**	49.28±11.09	**34.08±16.05**
+ PRG$^{\text{Last}}$ (Ours)	**63.49±10.73**	**66.19±14.22**	66.21±10.24	**53.17±8.84**	32.45±10.10

Tab. 3 Accuracy (%) on tabular data in MNAR.

Discussion

What is the novelty and contribution? Towards addressing SSL in MNAR, we propose transition tracking based Pseudo-Rectifying Guidance (PRG) to mitigate the adverse effects of mismatched distributions via combining information from the class transition history. *We propose that the pseudo-rectifying guidance can be carried out from the class level, by modeling the class transition of the pseudo-label as a Markov random walk on the graph.* **Why does our method work for MNAR?** In MNAR, being aware of rare class plays a key role, PRG enhances the model to preserve a certain probability to generate class transition to rare classes when assigning pseudo-labels. This form of probability based on class transition history produces effective results, because we do not spare any attempt of the model to identify the rare class by class transition tracking. *Thereby, PRG helps the model to still try to identify rare classes with a certain probability while combines the class distribution information of pseudo-labels so that the model can assign labels to rare classes with a clear purpose.*

References

[1] Xinting Hu, Yulei Niu, Chunyan Miao, Xian-Sheng Hua, and Hanwang Zhang. On non-random missing labels in semisupervised learning. In *International Conference on Learning Representations, 2022.*

[2] Zhen Zhao, Luping Zhou, Yue Duan, Lei Wang, Lei Qi, and Yinghuan Shi. Dc-ssl: Addressing mismatched class distribution in semi-supervised learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.*

[3] Yue Duan, Lei Qi, Lei Wang, Luping Zhou, and Yinghuan Shi. Rda: Reciprocal distribution alignment for robust semisupervised learning. In *European Conference on Computer Vision, 2022.*