Bidirectional Adaptation for Robust Semi-Supervised Learning with Inconsistent Data Distributions

Lin-Han Jia[1], Lan-Zhe Guo[1], Zhi Zhou[1], Jie-Jing Shao[1], Yu-Ke Xiang[2], Yu-Feng Li[1]

[1]National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China
[2]Consumer BG, Huawei Technologies, Shenzhen, China

Presented by: Lin-Han Jia
2023.06.21

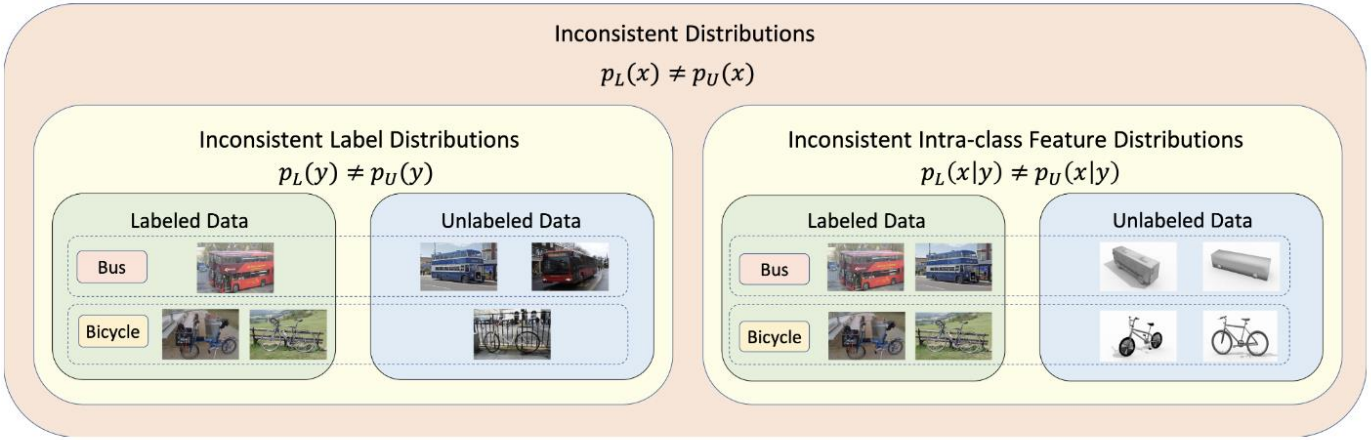
□Background

- **Semi-Supervised Learning**
  - Machine learning relies on a large amount of labeled data.
  - Semi-supervised learning can effectively utilize unlabeled data.
  - It has a limited scope and relies on the same distribution between labeled and unlabeled data.
  - There is a risk of significant performance degradation in real-world applications.

- **Robust Semi-Supervised Learning**
  - Utilizing a large amount of unlabeled data that has a different distribution from the current labeled data for learning.
  - The goal is to ensure that the semi-supervised algorithm does not perform too poorly in real-world applications.
  - This extends the applicability of classical semi-supervised learning.
  - It reduces the risk associated with using semi-supervised learning algorithms.

□Background

- **Inconsistent Distributions**
  - We can only observe that the feature distributions of unlabeled and labeled data are inconsistent.
  - Inconsistent feature distributions are equivalent to a combination of inconsistent class distributions and inconsistent intra-class feature distributions.
  - Inconsistent distributions between labeled and unlabeled data lead to low quality of pseudo-labels.
  - Inconsistent distributions between unlabeled data and target data result in poor performance and weak robustness of the learner.



□Theoretical Research

- **Generalization Error**
  - Bias of pseudo-label predictions.
  - Variance of pseudo-label predictions.
  - Distribution distance caused by pseudo-label predictions.
  - Bias of target predictions.
  - Variance of target predictions.
  - Distribution distance caused by target predictions.
- **Optimization Object**
  - Bias and distribution distance of pseudo-label predictions.
  - Bias and distribution distance of target predictions.
  - Objective conflict.
  - Objective can be decoupled.

Theorem 3.4. *Assuming that the probabilities of the pseudo-label predictor making wrong predictions for each sample are equal without considering the difference among them, for any target predictor* $f \in \mathcal{F}$*, pseudo-label predictor* $h \in \mathcal{H}$, $0 \le \delta_1 \le 1$, $0 \le \delta_2 \le 1$ *and* $0 \le \delta_3 \le 1$*, with the probability of at least* $(1-\delta_1)(1-\delta_2)(1-\delta_3)$*:*

$$\begin{aligned} E(f, \mathcal{D}_T|h, D_L, D_U) \le & \frac{n_l}{n_l + n_u^w}\hat{E}(f, D_L) \\ & + \frac{n_u^w}{n_l + n_u^w}\hat{E}(f, \tilde{D}_U^w) + var(\mathcal{F}, n_l + n_u^w, k, \delta_1) \\ & + Disc(f, \mathcal{D}_T, Mix_{\frac{n_l}{n_l+n_u^w}}(\mathcal{D}_L, \mathcal{D}_U^w)) \\ & + \frac{n_u^w}{n_l + n_u^w}(\hat{E}(h, D_L) + var(\mathcal{H}, n_l, k, \delta_2) \\ & + var(\mathcal{H}, n_u, k, \delta_3) + Disc(h, \mathcal{D}_L, \mathcal{D}_U)) \end{aligned} \tag{7}$$

where $\hat{E}(f, \tilde{D}_U^w)$ *is the weighted disagreement rate between the noisy pseudo-labels and the prediction results of* f *on the unlabeled dataset* $\tilde{D}_U$*.*

$$\begin{aligned} \min_{f\in\mathcal{F}, h\in\mathcal{H}} [& \frac{n_l}{n_l + n_u^w}\hat{E}(f, D_L) + \frac{n_u^w}{n_l + n_u^w}\hat{E}(f, \tilde{D}_U) \\ & + Disc(f, \mathcal{D}_T, Mix_{\frac{n_l}{n_l+n_u^w}}(\mathcal{D}_L, \mathcal{D}_U^w)) \\ & + \frac{n_u^w}{n_l + n_u^w}\hat{E}(h, D_L) + \frac{n_u^w}{n_l + n_u^w}Disc(h, \mathcal{D}_L, \mathcal{D}_U)] \end{aligned}$$

☐Analysis of Semi-Supervised Learning Algorithms

- **Pseudo-labeling**
  - The pseudo-label predictor is a combination of target predictor and mapping function.

$$\forall f \in \mathcal{F}, h = p \circ f \in \mathcal{H}.$$

- **Consistency**
  - The pseudo-label predictor is a combination of augmentation and target predictor.

$$\forall f \in \mathcal{F}, h = f \circ a \in \mathcal{H}.$$

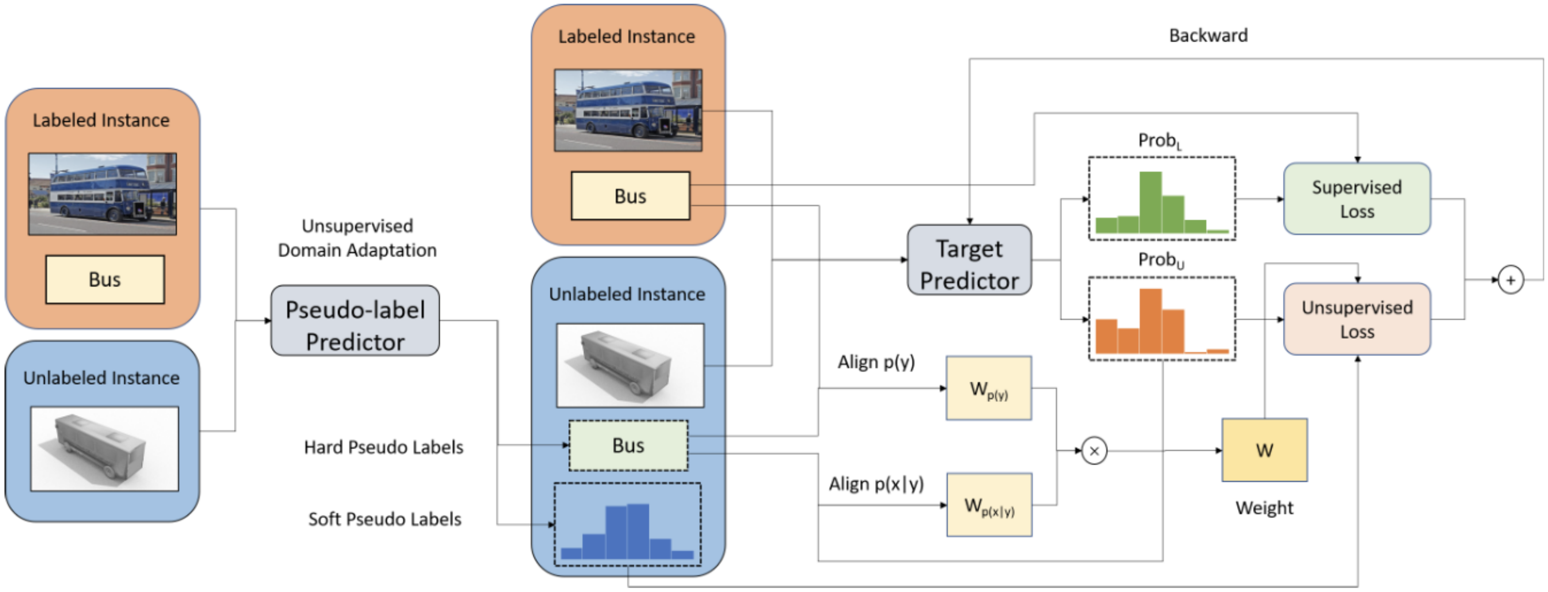
- **Mixed Methods**
  - The pseudo-label predictor is a combination of augmentation, target predictor and mapping function.

$$\forall f \in \mathcal{F}, h = p \circ f \circ a \in \mathcal{H}.$$

- **Three Shortcomings**
  - The coupling of pseudo-label predictor and target predictor leads to conflicting optimization objectives.
  - The distribution bias between labeled and unlabeled data leads to low quality of pseudo-labels.
  - Sample weights cannot effectively align the distribution of unlabeled data with the target distribution.

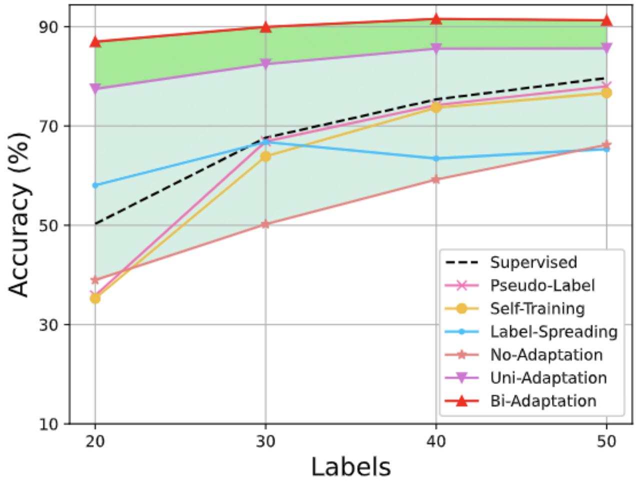
◻Algorithm Framework

- **Bidirectional Adaptation Algorithm**
  – Decoupling the pseudo-label predictor and target predictor avoids optimization conflicts.
  – Improving the accuracy of pseudo-labels through domain adaptation.
  – Aligning the target distribution by weighting unlabeled samples.
    - Aligning p(x|y) with intra-class weights.
    - Aligning p(y) with inter-class weights.



◻Experiments

- **Theoretical Arguments**



- **Performance Robustness**

Table 3. Experiments on VisDA-2017 with 150 labels, 300 labels and 600 labels.

Methods	150 labels		300 labels		600 labels	
	S/R	R/S	S/R	R/S	S/R	R/S
Supervised	85.33 ± 1.54	78.50 ± 0.68	89.64 ± 0.73	81.81 ± 0.62	92.20 ± 0.45	84.13 ± 0.36
Mean Teacher	84.15 ± 1.08	73.68 ± 1.00	86.90 ± 0.61	76.90 ± 0.46	89.05 ± 0.48	79.86 ± 0.30
FixMatch	78.46 ± 4.15	67.10 ± 9.46	82.88 ± 0.85	71.74 ± 0.45	87.68 ± 1.15	79.54 ± 1.88
FlexMatch	83.43 ± 1.74	67.90 ± 1.77	88.09 ± 0.53	75.17 ± 1.34	90.11 ± 1.09	79.28 ± 0.38
UASD	85.58 ± 1.55	78.59 ± 0.41	89.58 ± 0.79	81.82 ± 0.68	92.29 ± 0.45	84.04 ± 0.31
CAFA	83.95 ± 1.79	72.89 ± 1.03	87.81 ± 0.47	76.48 ± 0.72	89.84 ± 0.62	78.63 ± 0.44
Ours	$\mathbf{85.92 \pm 1.16}$	$\mathbf{79.15 \pm 0.39}$	$\mathbf{89.85 \pm 0.71}$	$\mathbf{82.27 \pm 0.60}$	$\mathbf{92.46 \pm 0.38}$	$\mathbf{84.28 \pm 0.36}$

Thanks!