Class-Imbalanced Semi-Supervised Learning with Adaptive Thresholding

(ICML2022)

Lan-Zhe Guo and Yu-Feng Li

Nanjing University, China

Learning And Mining from DatA

http://www.lamda.nju.edu.cn

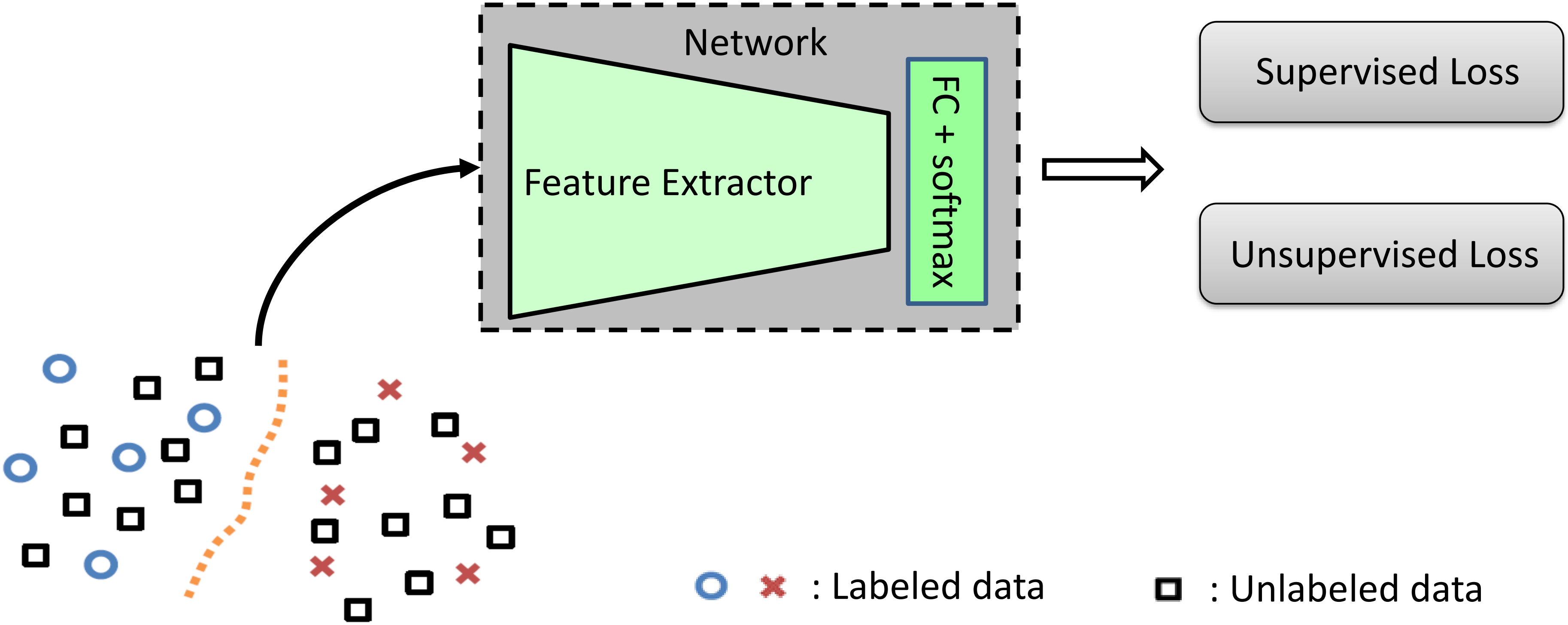
Outline

- **Motivation**
- Proposed method
- Experiments
- Conclusions

Semi-Supervised Learning

Learning And Mining from DatA
http://www.lamda.nju.edu.cn

SSL: learning from **labeled** and **unlabeled** data with a learning **model**



Assumption

Labeled and unlabeled training data are class balanced

Class Imbalanced Datasets

Real-world datasets are often class-imbalanced

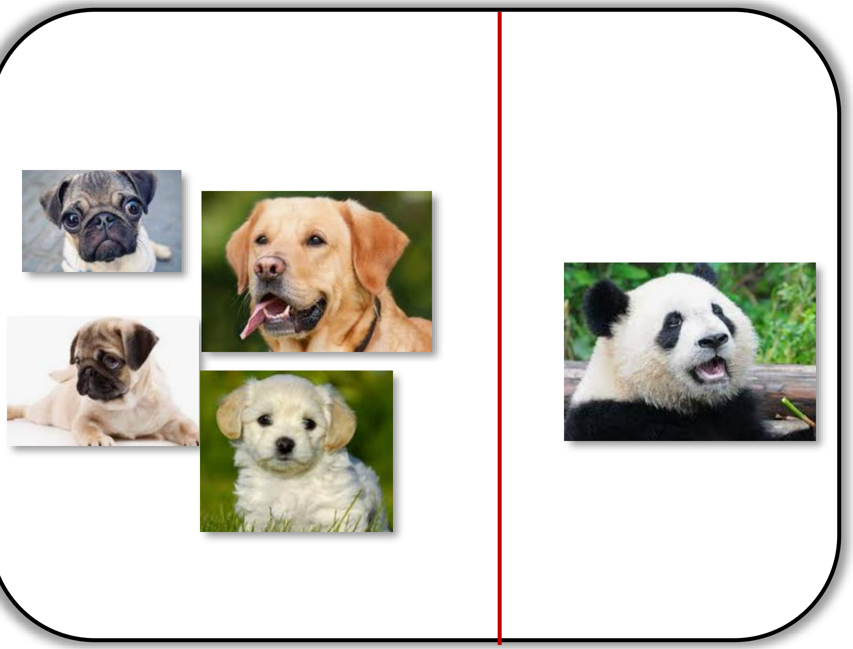


Image Classification task

There are far more "dogs" than "pandas"

Credit Card Fraud Detection

There are far few “fraud” than “normal”

Performance Degradation

Previous SSL methods suffer **performance degradation problem** with imbalanced dataset

Take the SOTA FixMatch method as an example:

Pseudo-Label Generation

$$\mathbf{q}_b = f(\mathbf{y}|\alpha(\mathbf{x}_b^u);\theta)$$

$$\hat{\mathbf{y}}_b^u = \arg\max(\mathbf{q}_b)$$

Generate pseudo-labels for $\mathcal{A}(x)$ using the prediction on $\alpha(x)$

Pseudo-Label Selection

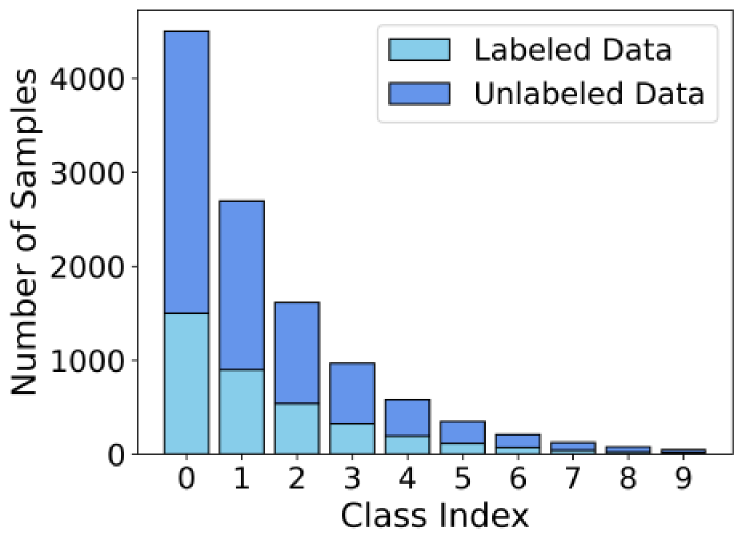
$$\mathbb{I}(\max(\mathbf{q}_b) \geq \tau)$$

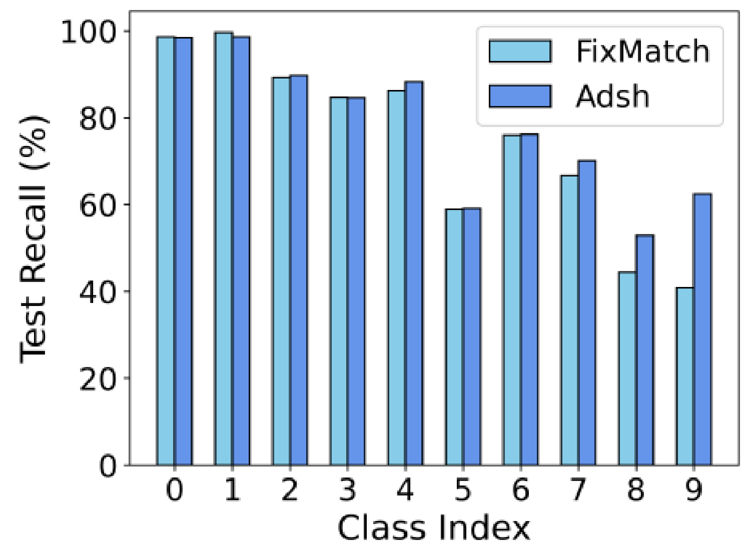
Select pseudo-labels based on a fixed confidence threshold τ $(\mathrm{e.g.}, \tau = 0.95)$

Performance Degradation

LAMDA
Learning And Mining from DatA
http://www.lamda.nju.edu.cn

The prediction confidence is biased towards the majority classes
Adopting a fixed threshold results in low recall rates for minority classes





Can we design an SSL algorithm that selects pseudo-labels based on an adaptive class-dependent threshold?

Outline

- Motivation
- **Proposed method**
- Experiments
- Conclusions

Our Proposal: Adsh

Learning And Mining from DatA
http://www.lamda.nju.edu.cn

We explicitly encodes the number of pseudo-labels to be selected for each class into the learning objective

a larger S_k indicates a larger number of pseudo labels would be selected for class k

$$\begin{aligned}
\min_{\hat{\mathbf{y}},\mathbf{s},\theta} \quad & \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} -y_{i,k}\log f(\mathbf{y}=k|\alpha(\mathbf{x}_i^l);\theta) \\
& +\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{K}[-\hat{y}_{i,k}\log f(\mathbf{y}=k|\alpha(\mathbf{x}_i^u);\theta) \\
& \boxed{-s_k\hat{y}_{i,k}}] \\
\text{s.t.} \quad & \hat{\mathbf{y}}_i = [\hat{y}_{i,1},\cdots,\hat{y}_{i,K}] \in \{0,1\}^K \\
& 0 \le \mathbf{1}^\top \hat{\mathbf{y}}_i \le 1 \\
& s_k > 0, \ \ \forall 1 \le k \le K
\end{aligned}$$

Similar techniques have also been applied in domain adaptation [Zou et al., 2018] and curriculum learning [Zou et al., 2018]

How to set S

☐ If the ground-truth label distribution is known

we solve S_k to make the pseudo-label $\hat{y}$ has the same class distribution as the ground-truth y^*

$$\sum_{i=1}^{M} \mathbb{I}(f(\mathbf{y} = k|\alpha(\mathbf{x}_i^u);\theta) \geq \exp(-s_k))$$
$$= \frac{\sum_{i=1}^{M} \mathbb{I}(f(\mathbf{y} = 1|\alpha(\mathbf{x}_i^u);\theta) \geq \exp(-s_1))}{\gamma_k}$$

☐ If the ground-truth label distribution is unknown

We solve S_k to make sure the same percentage of pseudo-labels are selected for each class

$$\rho = \frac{\sum_{i=1}^{M} \mathbb{I}(f(\mathbf{y} = 1|\alpha(\mathbf{x}_i^u);\theta) \geq \tau_1)}{length(C_1)}$$

Adaptive Threshold

Given a learning model f, we have:

$$\hat{y}_{i,k} = \begin{cases} 1, & \textbf{\textit{if}}\ k = \operatorname{argmax} \dfrac{f(\mathbf{y}=k|\alpha(\mathbf{x}_i^u);\theta)}{\exp(-s_k)}, \\ & \dfrac{f(\mathbf{y}=k|\alpha(\mathbf{x}_i^u);\theta)}{\exp(-s_k)} \geq 1. \\ 0, & \textbf{\textit{otherwise}}. \end{cases}$$

If $\exp(s_k - s_{k'}) > \frac{f(\mathbf{y}=k'|\alpha(\mathbf{x}_i^u);\theta)}{f(\mathbf{y}=k|\alpha(\mathbf{x}_i^u);\theta)}$, we obtain the adaptive threshold:

$$\boxed{\mathbb{I}(\max(\mathbf{q}_i) \geq \exp(-s_{\hat{\mathbf{y}}_i^u}))}$$

where $\mathbf{q}_i = f(\mathbf{y}|\alpha(\mathbf{x}_i^u);\theta)$ and $\hat{\mathbf{y}}_i^u = \operatorname{argmax}(\mathbf{q}_i)$

Outline

- Motivation
- Proposed method
- **Experiments**
- Conclusions

Same Class Distribution

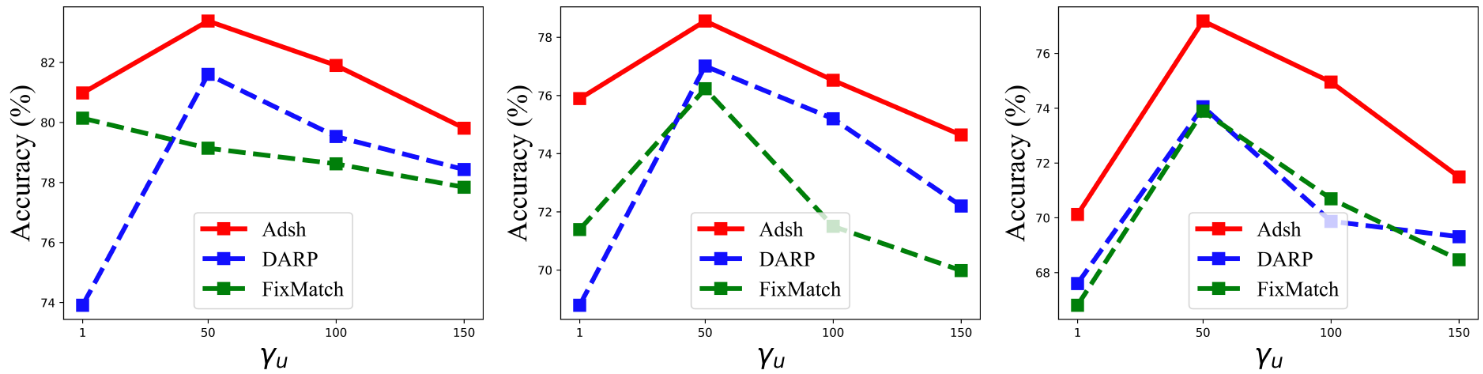
Learning And Mining from DatA
http://www.lamda.nju.edu.cn

	Imbalanced CIFAR-10 Dataset					
	$N_1 = 1500, M_1 = 3000$			$N_1 = 500, M_1 = 4000$		
Algorithm	$\gamma = 50$	$\gamma = 100$	$\gamma = 150$	$\gamma = 50$	$\gamma = 100$	$\gamma = 150$
Supervised	65.23 ± 0.05	58.94 ± 0.13	55.63 ± 0.38	51.31 ± 0.34	45.82 ± 0.41	40.90 ± 0.39
CBL	65.52 ± 0.31	58.52 ± 0.45	52.36 ± 0.58	51.94 ± 0.71	46.22 ± 0.92	41.58 ± 1.24
Re-Sampling	64.53 ± 0.39	56.34 ± 0.42	53.21 ± 0.51	51.96 ± 0.65	48.13 ± 1.25	40.26 ± 1.88
cRT	67.82 ± 0.14	63.43 ± 0.45	59.56 ± 0.44	56.28 ± 1.45	48.11 ± 0.79	45.02 ± 1.08
LDAM	68.91 ± 0.10	63.15 ± 0.24	58.68 ± 0.30	56.41 ± 0.92	49.27 ± 0.88	45.10 ± 0.75
Mean-Teacher	68.84 ± 0.82	61.33 ± 0.28	54.79 ± 0.31	56.34 ± 1.68	48.55 ± 0.77	45.32 ± 1.20
MixMatch	73.59 ± 0.46	65.03 ± 0.26	62.71 ± 0.29	65.32 ± 1.20	56.41 ± 1.96	52.38 ± 1.88
ReMixMatch	78.96 ± 0.29	72.88 ± 0.12	68.61 ± 0.40	76.83 ± 0.98	70.12 ± 1.23	59.58 ± 1.30
FixMatch	79.10 ± 0.14	71.50 ± 0.31	68.47 ± 0.15	77.34 ± 0.96	68.45 ± 0.94	60.10 ± 0.82
DARP	81.60 ± 0.31	75.23 ± 0.14	69.31 ± 0.26	76.72 ± 0.46	69.41 ± 0.50	61.23 ± 0.31
CReST	82.03 ± 0.26	75.08 ± 0.41	69.84 ± 0.39	76.18 ± 0.36	69.50 ± 0.70	60.81 ± 0.55
Adsh	$\mathbf{83.38 \pm 0.06}$	$\mathbf{76.52 \pm 0.35}$	$\mathbf{71.49 \pm 0.30}$	$\mathbf{79.27 \pm 0.38}$	$\mathbf{70.97 \pm 0.46}$	$\mathbf{62.04 \pm 0.51}$

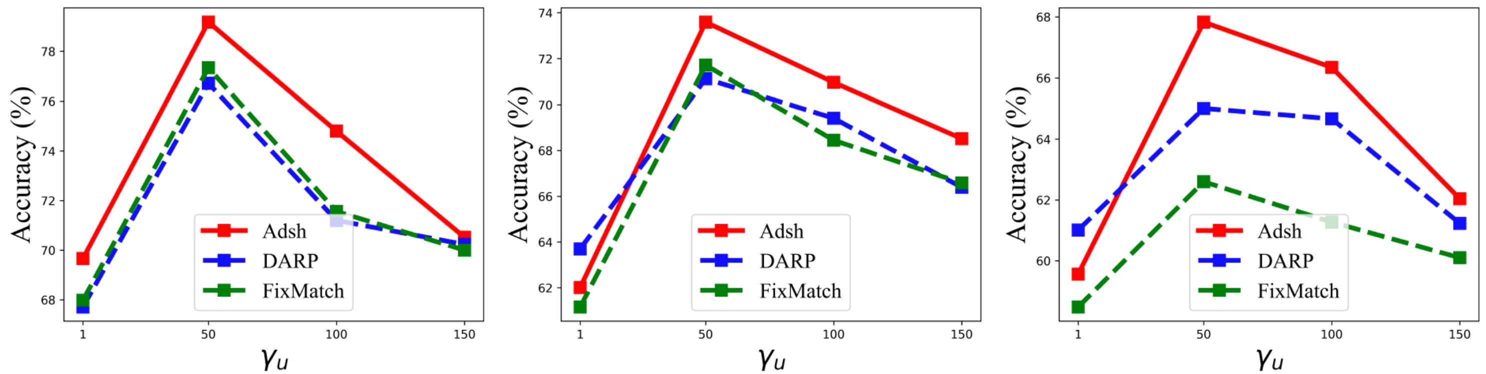
Our proposal achieves significant performance gain compared with SOTA SSL methods and Imbalanced SSL methods

Different Class Distribution

http://www.lamda.nju.edu.cn



(a) $N_1 = 1500, \gamma_l = 50$ (b) $N_1 = 1500, \gamma_l = 100$ (c) $N_1 = 1500, \gamma_l = 150$



(d) $N_1 = 500, \gamma_l = 50$ (e) $N_1 = 500, \gamma_l = 100$ (f) $N_1 = 500, \gamma_l = 150$

Our proposal achieves significant performance gain on more than twenty settings

Outline

- Motivation
- Proposed method
- Experiments
- **Conclusions**

Conclusions

How to develop a robust SSL for class-imbalanced distribution?

- ✓ In this work, we propose **an adaptive class-dependent threshold for pseudo-label selection in semi-supervised learning**
- ✓ By combining the adaptive threshold with FixMatch, we improve the performance significantly

Future work

- Robust SSL for other types of data beyond image
- Theoretical Analysis for class-imbalanced SSL

Thank you!