Week 13

— Paper Reading

An Embarrassingly Simple Baseline for Imbalanced Semi-Supervised Learning

Challenge: Imbalanced SSL (both $\mathcal{L}$ and $\mathcal{U}$ exhit class-imbalanced distribution)
the issue of confirmation bias

Problem: K-way classification problem

$D_L := \{(x_i, y_i)\}_{i=1}^{N}$

$\Downarrow$

N_k: number of labeled sample for class k

$$N = \sum_{k=1}^{K} N_k$$

Imbalanced ratio $\gamma_l = \frac{N_1}{N_K}$

$N_1 > N_2 > \cdots > N_K$

$D_U := \{u_j\}_{j=1}^{M}$

$\Downarrow$

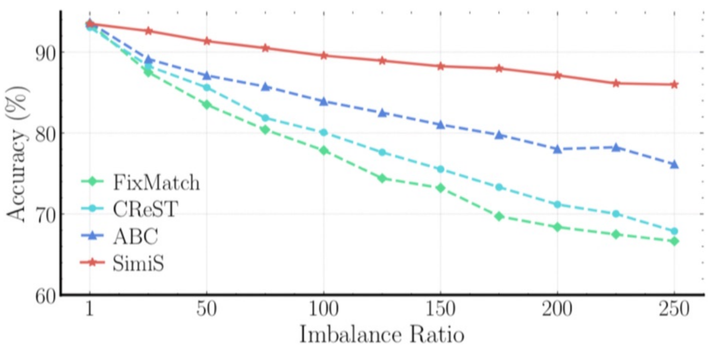
class distribution is usually unknown in practice

$\gamma_u > 0$

(also consider reversely ordered class distribution in experiment)

SimiS: A Simple Baseline for Imbalanced SSL

the test accuracy of imbalanced algorithms increases as γ_l decreases.



$\Downarrow$

Would simply reducing the dataset imbalance be an effective way to improve the performance for imbalanced SSL?

$\Downarrow$

We supplement the infrequent classes with more pseudo-labels and frequent classes with less pseudo-labels after each training epoch.

$D_L \to P_L(y) \in \mathbb{R}^K, \quad P_L(y=k) = \frac{N_k}{N}$ $\qquad D_U \to P_U$

(1) generate pseudo-labels and corresponding confidence scores for all U

$$\underbrace{\hat{y}_j}_{\in\{1,2,\cdots,K\}} = \arg\max f_\theta(u_j). \qquad c_j = \max f_\theta(u_j)$$

(2) the number of pseudo-labels added for class k

$$S_k = \underbrace{\beta}_{\frac{N}{K}\ \text{hyper-parameter}} \big(\underbrace{P_L(\arg\max_k N_k)}_{\text{class marginal of the most frequent class}} - P_L(k) \big)$$

(3) Sample the first S_k most confident pseudo-labels to expand the labeled set and start the next training epoch with expaned set.

$\Downarrow$

① outperform existing methods

② show great robustness against a wide range of imbalance ratio and unlabeled data distribution.

Experiment: construct long-tailed $N_k = N_1 \gamma_l^{\frac{k-1}{K-1}}$, $M_k = M_1 \gamma_u^{\frac{k-1}{K-1}}$

Dataset	CIFAR10-LT						CIFAR100-LT			STL10-LT	
N_1		1500			500			150		150	
M_1	3000	30	3000	4000	40	4000	300	30	300	M≈100k	
γ_l	100	100	150	100	100	150	10	10	15	10	20
γ_u	100	1/100	150	100	1/100	150	10	1/10	15	NA	NA
Supervised	63.62±0.40	63.62±0.40	59.82±0.32	47.62±0.87	47.62±0.87	43.88±1.61	48.01±0.45	48.01±0.45	45.37±0.54	46.85±1.65	41.6±0.77
FixMatch [31]	76.49±0.72	68.92±0.79	72.15±0.94	73.14±1.03	62.52±0.93	65.68±0.67	57.76±0.39	57.56±0.47	53.97±0.17	66.56±1.02	56.29±4.0
w/ DARP [19]	77.37±0.50	70.36±1.55	74.02±0.06	71.12±0.82	62.16±1.10	65.63±0.63	56.14±0.46	56.40±0.21	52.81±0.50	63.74±0.54	56.03±1.81
w/ CReST [38]	79.90±0.33	86.71±0.39	74.70±0.53	77.69±0.71	76.37±3.84	68.20±0.33	58.56±0.34	60.07±0.24	55.43±0.17	65.52±1.01	61.38±1.7
w/ CReST+ [38]	79.60±0.06	73.98±1.34	75.39±0.42	78.70±0.40	63.72±0.87	72.73±2.26	58.19±0.37	59.53±0.34	55.39±0.23	66.27±0.59	62.63±1.69
w/ ABC [25]	84.01±0.15	83.45±0.54	80.94±0.85	79.40±0.88	79.21±0.44	69.50±1.86	58.25±0.20	59.24±0.17	55.38±0.47	70.64±0.89	65.68±3.06
w/ DASO [28]	78.87±0.80	74.47±0.60	74.92±0.36	73.63±0.46	65.08±0.90	67.13±1.06	58.16±0.21	59.25±0.23	54.82±0.53	69.31±0.91	62.45±2.23
w/ CoSSL [12]	82.35±0.79	77.47±0.56	79.00±0.41	75.82±0.61	73.26±0.78	70.56±0.55	58.00±0.39	57.77±0.31	55.49±0.43	71.44±0.45	69.01±0.80
w/ SAW [22]	80.93±0.31	76.73±0.66	77.67±0.14	75.20±1.01	70.04±1.58	68.51±0.16	57.55±0.45	58.12±0.34	54.00±0.65	69.30±0.69	65.80±1.22
w/ Adsh [14]	78.43±0.54	70.45±0.70	73.96±0.47	75.97±0.68	65.64±1.82	66.55±2.94	58.65±0.36	55.49±0.55	54.55±0.40	69.35±1.12	64.82±1.41
w/ DePL [34]	80.65±0.52	74.53±0.61	76.58±0.12	76.98±1.70	69.67±1.34	71.95±2.54	57.08±0.29	57.31±0.55	53.89±0.44	69.46±0.62	65.93±1.22
w/ SimiS	**88.31±0.08**	**87.96±0.29**	**86.27±0.23**	**83.51±0.25**	**82.83±0.40**	**81.51±0.35**	**68.45±0.31**	**68.79±0.03**	**66.25±0.09**	**77.48±0.31**	**73.69±0.42**
w/ SimiS + LA	**89.78±0.23**	**90.99±0.08**	**88.36±0.26**	**85.69±0.44**	**86.15±0.53**	**83.75±0.16**	**71.28±0.06**	**71.47±0.12**	**69.64±0.11**	**79.10±0.15**	**75.85±0.08**
ReMixMatch [2]	78.86±0.77	75.05±0.07	74.61±0.82	74.76±0.52	70.80±0.05	71.10±0.95	60.78±0.05	61.08±0.33	57.54±0.34	75.33±0.37	67.43±0.86
w/ DARP [19]	79.50±0.53	76.04±0.89	75.26±0.10	76.47±0.40	71.20±0.58	66.15±0.88	62.12±0.40	62.26±0.39	58.63±0.51	73.51±0.11	67.72±1.84
w/ CReST [38]	79.12±2.12	**88.45±0.17**	73.13±2.60	72.12±0.93	76.68±0.38	66.74±0.57	61.18±0.42	65.79±0.55	58.89±0.13	72.38±0.58	67.79±1.71
w/ CReST+ [38]	81.02±0.20	62.70±0.26	76.85±0.61	79.76±0.55	58.28±0.55	73.76±1.18	64.21±0.34	58.68±0.49	61.84±0.33	63.94±0.55	58.91±1.84
w/ ABC [25]	82.27±0.89	80.62±0.88	79.43±0.31	77.61±0.89	73.85±0.52	69.55±2.06	62.86±0.12	64.14±0.19	59.21±0.22	74.10±0.60	70.26±0.57
w/ DASO [28]	75.04±1.58	81.89±0.54	68.83±0.22	72.00±2.07	77.44±1.28	66.37±1.69	62.12±0.24	62.80±0.69	58.97±0.29	75.74±0.13	73.49±0.32
w/ CoSSL [12]	79.31±0.23	76.26±0.60	75.63±0.24	75.36±0.64	72.44±0.37	72.91±0.26	62.19±0.55	61.78±0.45	58.79±0.75	73.48±0.20	68.44±0.18
w/ SimiS	**86.91±0.23**	85.94±0.08	**81.67±0.19**	**84.62±0.23**	**79.68±0.38**	**82.50±0.35**	**67.41±0.45**	**69.29±0.42**	**64.75±0.41**	**80.84±0.41**	**79.20±0.29**
w/ SimiS + LA	**87.34±0.39**	**86.94±0.18**	**82.67±0.37**	**85.60±0.18**	**79.95±0.18**	**82.95±0.20**	**68.42±0.23**	**70.30±0.14**	**65.92±0.08**	**81.91±0.08**	**79.62±0.18**

most of the baseline methods present a degraded performance
(1) larger γ_l
(2) $\gamma_l = 1/\gamma_u$

Perform consistently across settings with different class imbalance ratios.

Intuitively, there are more samples in the $\mathcal{U}$ of the tailed classes in $\mathcal{L}$ when $\gamma_l = 1/\gamma_u$ and the model would learn better if these samples could be utilized correctly during training. (SimiS)

Robustness to Various Imbalance Distributions:

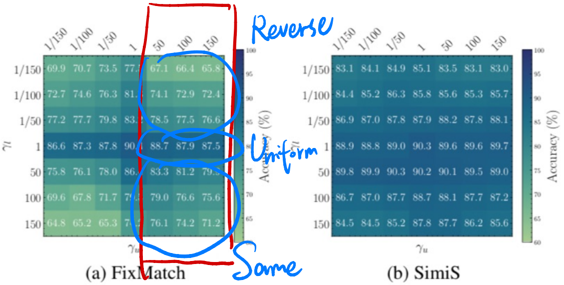


Figure 2. Accuracy of (a) FixMatch and (b) SimiS on the balanced test set of CIFAR10, trained with different combination of γ_l and γ_u. SimiS performs robustly across different training distribution.

FixMatch: $\gamma_l = \gamma_u$

SimiS: robust

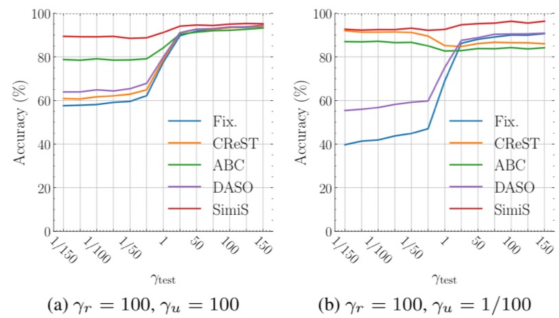


Figure 3. Accuracy on test set of different imbalance ratios. We train FixMatch with different algorithms on CIFAR10-LT using (a) $\gamma_r = 100$, $\gamma_u = 100$, and (b) $\gamma_r = 100$, $\gamma_u = 1/100$. SimiS performs robustly across different testing distribution.

Imbalanced test distribution

More Analysis:

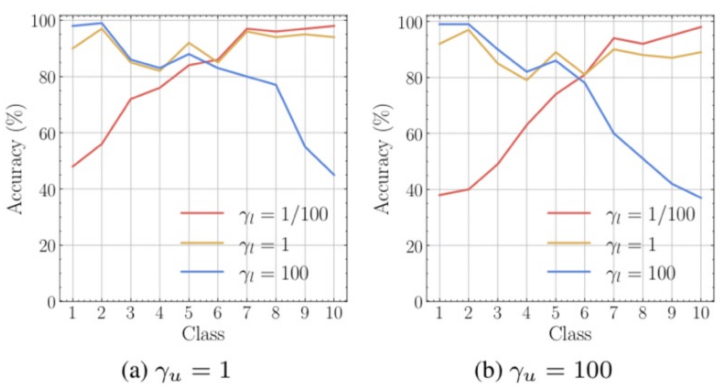


Figure 5. Class-wise accuracy on test set of CIFAR10-LT using FixMatch. We train FixMatch with different imbalance ratios in label set, i.e., $\{1/100, 1, 100\}$, and imbalance ratio in unlabeled set (a) $\gamma_u = 1$; and (b) $\gamma_u = 100$. The labeled set and unlabeled set have fixed size $N = 3420$ and $M = 7440$ respectively.

The necessity and sufficiency of reducing the imbalance ratio in the labeled set during training.