Are Anchor Points Really Indispensable in Label-Noise Learning ?

Xiaobo Xia[1,2] Tongliang Liu[1] Nannan Wang[2] Bo Han[3] Chen Gong[4]
Gang Niu[3] Masashi Sugiyama[3,5]

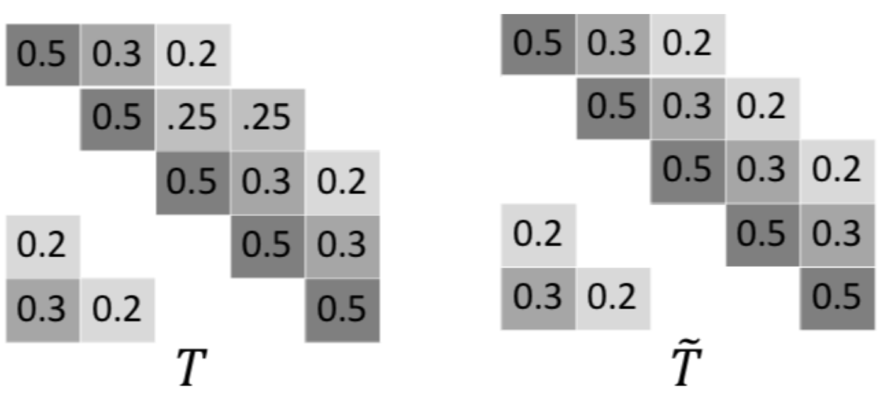
[1]University of Sydney [2]Xidian University [3]RIKEN
[4]Nanjing University of Science and Technology [5]University of Tokyo

OVERVIEW

- A new paradigm called T-Revision combating with noisy labels is presented, which effectively learn transition matrices, leading to better classifier.
- The generalization error bound is derived to theoretically prove the effectiveness of our method.
- Experiments on simulated noisy dataset and real noisy dataset demonstrate the excellence of algorithm.

MOTIVATION

Figure 3. Illustrative experimental results (using a 5-class classification problem as an example).



- Let an example have $P(\bar{Y}|X) = [0.141; 0.189; 0.239; 0.281; 0.15]$;
- With the true transition matrix T, $P(Y|X) = (T^\top)^{-1} P(\bar{Y}|X) = [0.15; 0.28; 0.25; 0.3; 0.02]$
- If the transition matrix is not accurately learned as $\tilde{T}$, $P(Y|X) = (\tilde{T}^\top)^{-1} P(\bar{Y}|X) = [0.1587; 0.2697; 0.2796; 0.2593; 0.0325]$

IMPORTANCE REWEIGHTING

$$\bar{R}_{n,w}(T,f) = \frac{1}{n}\sum_{i=1}^{n} \frac{g_{\bar{Y}_i}(X_i)}{(T^\top g)_{\bar{Y}_i}(X_i)} l(f(X_i), \bar{Y}_i)$$

QR CODE

Paper

Github

ALGORITHM DESCRIPTION

Figure 1. An overview of the proposed method. The proposed method will learn a more accurate classifier because the transition matrix is renovated.

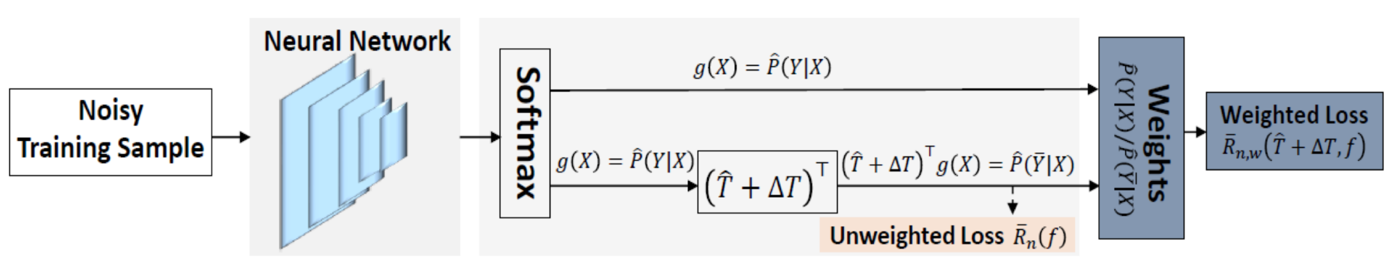


Figure 2. The process of Reweight-R Algorithm.

Algorithm 1 Reweight T-Revision (Reweight-R) Algorithm.

Input: Noisy training sample $\mathcal{D}_t$; Noisy validation set $\mathcal{D}_v$.

Stage 1: Learn $\hat{T}$

1: Minimize the unweighted loss to learn $\hat{P}(\bar{Y}|X)$ without a noise adaption layer

2: Initialize $\hat{T}$ according to Eq. (1) by using instances with the highest $\hat{P}(\bar{Y}=i|X)$ as anchor points

Stage 2: Learn the classifier f **and** ΔT

3: Initialize the neural network by minimizing the weighted loss with a noisy adaption layer $\hat{T}^\top$

4: Minimize the weighted loss to learn f and ΔT with a noisy adaption layer $(\hat{T}+\Delta T)^\top$;

//Stopping criterion for learning $\hat{P}(\bar{Y}|X)$, f and ΔT: when $\hat{P}(\bar{Y}|X)$ corresponds the minimum classification error on the noisy validation set $\mathcal{D}_v$

Output: $\hat{T}$, ΔT, and f.

RESULTS

Figure 4. The estimation error of the transition matrix by employing classifier-consistent and risk- consistent estimators on CIFAR10 dataset.
(a) Sym-20 label noise. (b)Sym-50 label noise.

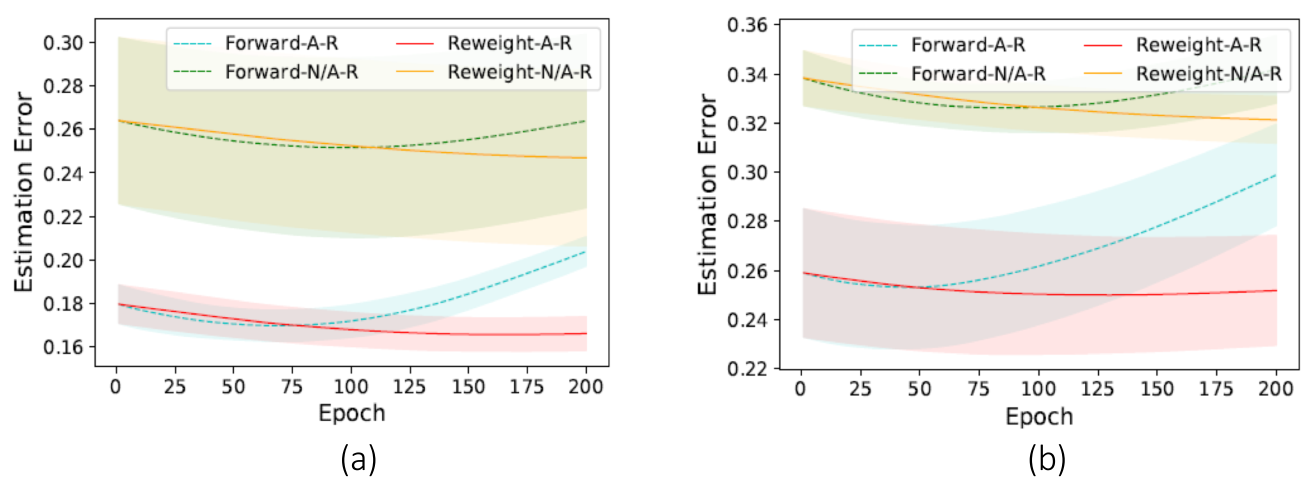


Table 1. Classification accuracy on synthetic noisy dataset and real noisy dataset.
(a) / (b) Means and Standard Deviations(Percentage) of classification accucacy on *CIFAR10* dataset with / without anchor points.
(c) Classification Accuracy (Percentage) on *Clothing1M*.

Models	Sym-20	Sym-50
Decoupling-A	79.85±0.30	52.22±0.45
MetorNet-A	80.49±0.52	70.71±0.24
Co-teaching-A	82.38±0.11	72.80±0.45
Forward-A	85.63±0.52	77.92±0.66
Reweight-A	86.77±0.40	80.16.±0.46
Forward-A-R	88.10±0.21	81.11±0.74
Reweight-A-R	**89.63±0.13**	**83.40±0.65**

(a)

Models	Sym-20	Sym-50
Decoupling-N/A	75.37±1.24	47.19±0.19
MetorNet-N/A	78.51±0.31	67.37±0.30
Co-teaching-N/A	81.72±0.14	70.44±1.01
Forward-N/A	84.75±0.81	74.32±0.69
Reweight-N/A	85.53±0.26	77.70.±1.00
Forward-N/A-R	86.93±0.39	77.14±0.65
Reweight-N/A-R	**88.90±0.22**	**81.55±0.94**

(b)

Decoupling	MentorNet	Co-teaching	Forward	Reweight	Forward+R	Reweight+R
53.98	56.77	58.68	71.79	70.95	72.25	**74.18**

(c)