Unifying semi-supervised and robust learning by mixup
(2019)

Ryuichiro Hataya, Hideki Nakayama
Resume

May 7, 2019

1 Introduction

In this work the authors consider learning from coruped data, using both semi-supervised and robust learning methods, in SSL (semi supervised learning) setting, the training data consists of a small amount of labeld exampled and a large amount of unlabeld data, SSl uses the large unlabeled data to enhance the performances on a limited number of labeled data, for robust learning to label noise (RLL), all the data is labeld but some of them are mislabeled, in a RLL setting, learners need to enhance their performances using corrupted labels anad avoid performance deterioration cause by it.

In this paper, the authors joint both setting by the concept of trusted data, they assume that some labels are guaranteed to be clean, and the rest is noisy, and the two frameworks are unified by controlling the ratio of corrupted data to all the labels.

2 Learning from Bi-Quality data

We assume that the given data consist of two parts: trused data $\mathcal{D}_T$ and untrusted data $\mathcal{D}_U$, and the ratio of trusted and untrusted data to the entire data is:

$$p = \frac{|\mathcal{D}_T|}{|\mathcal{D}_T| + |\mathcal{D}_U|} \left(\text{thus } 1 - p = \frac{|\mathcal{D}_U|}{|\mathcal{D}_T| + |\mathcal{D}_U|} \right)$$

We also introduce the quality of the untruted data as:

$$q = 1 - \frac{\mathbf{D}\left(\mathbf{p}_U(y|x) || \mathbf{p}_T(y|x)\right)}{\mathbf{D}\left(\mathbf{p}(y) || \mathbf{p}_T(y|x)\right)}$$

So we have $q = 0$ if the lables are compleletly independent of the inputs (random) and $q = 1$ if the labels are correct, to we're in a SSL setting if $q = 0$ where unstrusted data is considered unlabeled data, and RLL is when we don't have any trused data $p = 0$, and the untrusted data are somehow accurate to a given degree: $0 < q < 1$, so if q is relatively high, we can use the untrusted labels with RLL given that they are somehow informative, but if q if close to zero, it is better to use a SLL method, but in practive we can't decide what policy to used given that we can't find how accurate the trusted data is, so we need an adaptative approch between the two:

$$\mathcal{L}_U = \gamma \mathcal{L}_{\text{robust}} + (1 - \gamma) \mathcal{L}_{\text{semi}}$$

So the untrusted data loss function will utilize both SSL and RLL settings, and the loss is calculated using a mix of mixup (mixmixpup)

3 mixmixup

For the trusted data, we get the loss between the mixup version of two inputs and their labels, for the untrusted data, we have two parts, one for RLL where we use mixup between the two untrusted labels of the two samples, and SLL where we use the two predictions of the model as our mixup labels:

Algorithm 1 mixmixup: learning from bi-quality data

Prepare trusted set: $\mathcal{D}_T$, untrusted set: $\mathcal{D}_U$
Initialize neural network f
Set hyper parameters: α, β, γ
Set loss function: $\mathcal{L}(\cdot, \cdot)$: categorical cross entropy

for $k \in \{0, 1, \ldots, K-1\}$:
  Sample $\lambda_\alpha \sim \mathrm{Beta}(\alpha, \alpha)$, $\lambda_\beta \sim \mathrm{Beta}(\beta, \beta)$
  Sample (x_i, y_i), (x_j, y_j) from $\mathcal{D}_T$
  $\mathcal{L}_T = \mathcal{L}(f(\lambda_\alpha x_i + (1-\lambda_\alpha)x_j), \lambda_\alpha y_i + (1-\lambda_\alpha)y_j)$
  Sample (x'_i, y'_i), (x'_j, y'_j) from $\mathcal{D}_U$
  Predict $y''_i = f(x'_i)$ and $y''_j = f(x'_j)$
  $\mathcal{L}_{\text{robust}} = \mathcal{L}(f(\lambda_\alpha x'_i + (1-\lambda_\alpha)x'_j), \lambda_\alpha y'_i + (1-\lambda_\alpha)y'_j)$
  $\mathcal{L}_{\text{semi}} = \mathcal{L}(f(\lambda_\beta x'_i + (1-\lambda_\beta)x'_j), \lambda_\beta y''_i + (1-\lambda_\beta)y''_j)$
  $\mathcal{L}_U = \gamma\mathcal{L}_{\text{robust}} + (1-\gamma)\mathcal{L}_{\text{semi}}$
  Update parameters of f with $\mathcal{L}_T + \sigma(k)\mathcal{L}_U$ ▷ For σ, see Section 3

4 Resutls

Table 1: **mixmixup can handle bi-quality data effectively.** Test accuracy of WRN-28-2 trained on bi-quality CIFAR-10. In the columns under TRUSTED and UNTRUSTED, we list the number of samples of trusted and untrusted data with quality q.

	METHOD	TRUSTED	UNTRUSTED	ACCURACY
A	Basic	4,000	N/A	0.72
	input mixup (Zhang et al. (2017b))	4,000	N/A	0.78
B	Basic	4,000	41,000 ($q = 0.0$)	0.29
	Basic	4,000	41,000 ($q = 0.6$)	0.78
C	input mixup (Zhang et al. (2017b))	4,000	41,000 ($q = 0.0$)	0.43
	input mixup (Zhang et al. (2017b))	4,000	41,000 ($q = 0.6$)	0.89
D	mixmixup (ours)	4,000	41,000 ($q = 0.0$)	0.88
	mixmixup (ours)	4,000	41,000 ($q = 0.6$)	0.90

Table 2: **Semi-supervised methods can surpass robust learning methods under shared settings.** Test accuracy of WRN-28-2 trained on CIFAR-10 with other state-of-the-art methods for semi-supervised learning and robust learning. Here, † and †† refer to the scores in the tables are from Oliver et al. (2018) and our re-implementations, respectively.

	METHOD	TRUSTED	UNTRUSTED	ACCURACY
A	input mixup (Verma et al. (2018))	4,000	41,000 (no label)	0.89
	Mean teacher† (Tarvainen & Valpola (2017))	4,000	41,000 (no label)	0.84
	VAT† (Miyato et al. (2018))	4,000	41,000 (no label)	0.86
B	MentorNet DD-MLP†† (Jiang et al. (2018))	4,000	41,000 ($q = 0.6$)	0.87
	GLC††(Hendrycks et al. (2018))	4,000	41,000 ($q = 0.6$)	0.84
C	mixmixup (ours)	4,000	41,000 ($q = 0.0$)	0.88
	mixmixup (ours)	4,000	41,000 ($q = 0.6$)	0.90