Uncertainty-aware Learning Against Label Noise on Imbalanced Datasets

Yingsong Huang[1*], Bing Bai[1*], Shengwei Zhao[1], Kun Bai[1], Fei Wang[2]

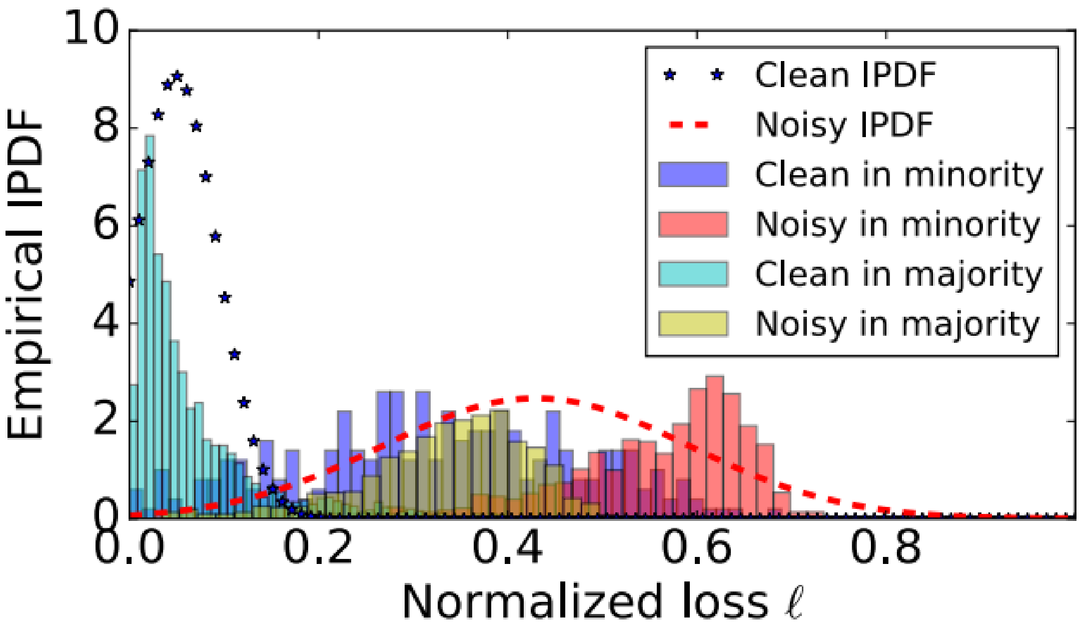
[1]Tencent Security Big Data Lab, Tencent Inc., China
[2]Department of Population Health Sciences, Weill Cornell Medicine, USA
{hudsonhuang,icebai,swzhao,kunbai}@tencent.com
few2001@med.cornell.edu

AAAI2022

Introduce

- **Inter-class loss distribution discrepancy**

As we can see, the loss distributions of noisy samples in the majority class overlaps with clean samples in the minority class.



➡ class-agnostic noise modeling may not work well when the loss distributions of the different classes vary significantly

under 50% symmetric noise and 1:10 class imbalance

Introduce

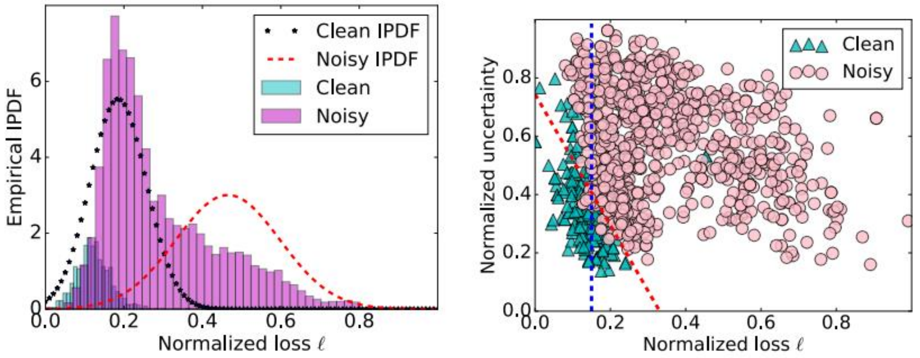
- **Misleading predictions due to uncertainty**

epistemic uncertainty: accounts for uncertainty in the model parameters, which could be reduced by observing more data

→ minority classes

aleatoric uncertainty: captures noise inherent in the observations, which cannot be reduced even if more data were to be collected

→ noisy samples



(a) Distribution of loss values (b) Joint distribution of loss values and epistemic uncertainty

Figure 2: Loss values and epistemic uncertainty distribution on CIFAR-10 with 90% symmetric noise, a warm-up of 30 epochs. (a) Loss distribution of clean samples overlaps with noisy samples due to misleading predictions induced by epistemic uncertainty. (b) Label noise modeling considering epistemic uncertainty distinguishes clean samples from noisy ones more accurately.

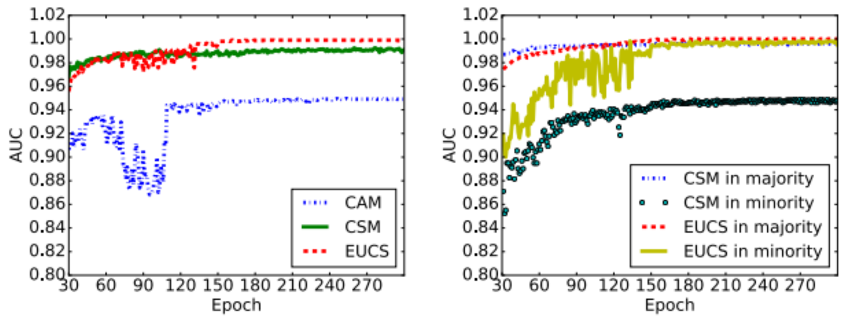
[1]What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?

Method

Epistemic Uncertainty-aware Class-specific Noise Modeling Adapted for Class-imbalanced Data(EUCS)

epistemic uncertainty

we place a prior distribution over the parameters $p(W)$, infer the posterior $p(W|X,\tilde{Y})$, and finally obtain the marginal probability $p(\tilde{Y}|X)$. We then can estimate the epistemic uncertainty as the entropy of $p(\tilde{Y}|X)$.



(a) Comparison of different label noise modeling methods

(b) Effects of different label noise modeling methods on minority and majority classes

Figure 3: AUC for clean/noisy sample classification on synthetic CIFAR-10 under 1:10 imbalance with 50% symmetric noise. Train for 300 epochs with different label noise modeling strategies. (a) In contrast to Class-agnostic modeling (CAM), Class-specific modeling (CSM) is more robust to class imbalance and separates clean and noisy samples more accurately under class imablance. (b) EUCS mainly contributes to identifying noise in minority classes compared to CSM.

clean probability

$$\omega_i = p(y_i^* = \tilde{y}_i|\ell_i, \epsilon_i)$$

placing a threshold τ on the probabilities ω

$$\omega_i = p(y_i^* = \tilde{y}_i|\ell_i, \epsilon_i) = (1-\epsilon_i)^r \boxed{p(y_i^* = \tilde{y}_i|\ell_i)}^{1-r} \dashrightarrow p(\mu_{j0}|\ell_i)$$

$$\mathcal{L}_j = \{\ell_i|\ell_i \in \mathcal{L}, \tilde{y}_{ij} = 1\} \Leftarrow \boxed{\mathrm{GMM}(\mu_{j0}, \mu_{j1}) \quad \mu_{j0} \leq \mu_{j1}}$$

$$\Rightarrow \; y_i = \omega_i \tilde{y}_i + (1-\omega_i)\hat{y}_i$$

Method

Aleatoric Uncertainty-aware Learning Against Label Noise(AUL)

aleatoric uncertainty

aleatoric uncertainty can be modeled by logit corruption with Gaussian noise, and the formulation leads to new loss functions, which can attenuate the effect from corrupted labels and making the loss more robust to noisy data.

instance-dependent noise factor $\delta^{x_i} \sim \mathcal{N}(0, \sigma^{x_i})$

class-dependent noise factor δ^{y} ➡ $I+\delta \sim \mathcal{N}(0, \sigma)$

$$\widehat{v}_i(W) = \delta^{x_i}(W) + \delta^{y} f_W(x_i) \quad \Rightarrow \quad \widehat{y}_i = \mathrm{softmax}((I+\delta) f_W(W) + \delta^{x_i}(W))$$

$$\ell_x(W) = -\frac{1}{|\widehat{\mathcal{X}}|} \sum_{x_i \in \widehat{\mathcal{X}}} (y_i)^{\top} \log \frac{1}{T} \sum_{t=1}^{T} (\widehat{y}_{it}(x_i; W, \sigma^{x_i}, \sigma))$$

Method

EUCS

warm up → estimate the clean probabilities ω_i and the corrected labels → divide the training data into clean samples and noisy samples with ω_i

↓

use MixMatch to augment $\hat{\mathcal{X}} \cup \hat{\mathcal{U}}$ into $\mathcal{X}' \cup \mathcal{U}'$

↓

SSL (AUL)

- $\ell_c = \ell_x + \lambda_u \ell_u$

- $$\ell_x = -\frac{1}{|\mathcal{X}'|} \sum_{x_i, y_i \in \mathcal{X}'} (y_i)^\top \log \frac{1}{T} \sum_{t=1}^{T} (\widehat{y}_{it}(x_i; W, \sigma^{x_i}, \sigma))$$

- $$\ell_u = \frac{1}{|\mathcal{U}'|} \sum_{x_i, y_i \in \mathcal{U}'} \left\| y_i - \frac{1}{T} \sum_{t=1}^{T} (\widehat{y}_{it}(x_i; W, \sigma^{x_i}, \sigma)) \right\|_2^2$$

Experiments

Dataset		CIFAR-10 Sym.				CIFAR-10 Asym.	CIFAR-100 Sym.			
Noise Rate		20%	50%	80%	90%	40%	20%	50%	80%	90%
CE	best	86.8	79.4	62.9	42.7	85.0	62.0	46.7	19.9	19.9
	last	82.7	57.9	26.1	16.8	72.3	61.8	37.3	8.8	3.5
CoT+	best	89.5	85.7	67.4	47.9	-	65.6	51.8	27.9	13.7
	last	88.2	84.1	45.5	30.1	-	64.1	45.3	15.5	8.8
PENCIL	best	92.4	89.1	77.5	58.9	88.5	69.4	57.5	31.1	15.3
	last	92.0	88.7	76.5	58.2	88.1	68.1	56.4	20.7	8.8
ML	best	92.9	89.3	77.4	58.7	89.2	68.5	59.2	42.4	19.5
	last	92.0	88.8	76.1	58.3	88.6	67.7	58.0	40.1	14.3
M-correction	best	94.0	92.0	86.8	69.1	87.4	73.9	66.1	48.2	24.3
	last	93.8	91.9	86.6	68.7	86.3	73.4	65.4	47.6	20.5
DivideMix	best	**96.1**	94.6	93.2	76.0	93.4	77.3	74.6	60.2	31.5
	last	95.7	94.4	92.9	75.4	92.1	76.9	74.2	59.6	31.0
ELR+	best	95.8	94.8	93.3	78.7	93.0	**77.6**	73.6	60.8	33.4
	last	94.6	93.8	91.1	75.2	92.7	**77.5**	72.4	60.2	30.8
ULC	best	**96.1**	**95.2**	**94.0**	**86.4**	**94.6**	77.3	**74.9**	**61.2**	**34.5**
	last	**95.9**	**94.7**	**93.2**	**85.8**	**94.1**	77.1	**74.3**	**60.8**	**34.1**

Table 1: Comparison with state-of-the-art methods in test accuracy on balanced CIFAR (%).

Method	Test Accuracy
CE	69.2
M-correction	71.0
Meta-Learning	73.5
PENCIL	73.5
DivideMix	**74.8**
ELR+	**74.8**
ULC	**74.9±0.2**

Table 3: Comparison with state-of-the-art methods in test accuracy on Clothing1M (%).

Experiments

Dataset		CIFAR-10 Sym. 20%		CIFAR-10 Sym. 50%		CIFAR-100 Sym.20%		CIFAR-100 Sym.50%	
Resampling Ratio		1:5	1:10	1:5	1:10	1:5	1:10	1:5	1:10
CE	best	87.1	85.4	72.0	68.4	64.9	61.4	48.2	43.0
	last	86.8	85.0	70.9	67.8	64.5	60.9	47.7	42.2
CoT+	best	82.5	76.7	71.9	53.2	49.2	39.9	30.6	29.2
	last	82.3	76.7	71.5	50.9	49.0	39.6	30.5	29.2
PENCIL	best	80.5	74.5	73.8	65.4	52.1	45.5	33.2	28.4
	last	80.0	73.0	73.5	65.2	51.1	43.3	31.9	25.2
ML	best	75.9	70.0	73.8	62.2	54.2	47.5	41.2	34.8
	last	74.6	68.0	61.6	54.0	51.1	44.4	35.2	30.3
M-correction	best	88.1	80.1	83.5	77.5	62.3	55.4	51.3	44.4
	last	87.0	77.3	83.1	76.8	62.2	54.7	50.0	44.1
DivideMix	best	93.9	74.8	85.5	66.9	65.0	51.2	56.9	44.2
	last	93.9	74.6	85.4	66.8	64.9	50.9	56.4	44.0
ELR+	best	88.2	79.8	82.7	65.5	59.6	49.9	53.4	44.7
	last	87.0	78.4	82.5	64.6	59.3	49.6	52.5	43.9
ULC	best	**95.0**	**93.8**	**94.9**	**92.5**	**75.5**	**72.8**	**71.6**	**57.2**
	last	**94.9**	**92.6**	**94.7**	**92.2**	**75.1**	**72.3**	**70.9**	**56.7**

Table 2: Comparison with state-of-the-art methods in test accuracy on imbalanced CIFAR (%).

imbalance: we randomly choose half the classes and randomly sub-sample 1/5 and 1/10 examples in these classes while other classes remain the same

Experiments

Dataset		CIFAR-10				CIFAR-100			
		Sym. 20%		Sym. 50%		Sym.20%		Sym.50%	
Resampling Ratio		1:5	1:10	1:5	1:10	1:5	1:10	1:5	1:10
ULC	best	**95.0**	**93.8**	**94.9**	**92.5**	**75.5**	**72.8**	**71.6**	**57.2**
	last	**94.9**	**92.6**	**94.7**	**92.2**	**75.1**	**72.3**	**70.9**	**56.7**
ULC w/o CSM	best	94.2	82.5	88.5	70.5	67.1	59.4	61.3	50.2
	last	93.8	81.3	87.9	69.8	66.2	58.7	60.5	49.8
ULC w/o EUM	best	94.1	90.7	91.3	89.9	68.0	67.2	65.9	53.2
	last	93.8	90.1	90.8	88.7	67.5	66.9	65.4	52.6
ULC w/o AUL	best	94.2	91.8	92.5	91.2	70.2	66.5	68.3	52.2
	last	94.0	91.5	91.7	90.5	69.7	66.1	67.2	51.6

Table 4: Ablation study results on imbalanced CIFAR (%). CSM refers to class-specific noise modeling, EUM refers to epistemic uncertainty-aware noise modeling, and AUL refers to aleatoric uncertainty-aware learning.

Thank you