Uncertainty Based Loss Reweighting for Noisy Labeled Dataset

Anonymous Author(s)
Submission Id: 123-A56-BU3*

ABSTRACT

In practical usage of computer vision datasets, label noise is often inevitably introduced by crowd-sourcing labeling process. Although deep neural networks are somewhat robust against small amount of noise, convolutional networks are shown to easily overfit due to the large amount of network parameters. Assigning weights to data samples is a well developed idea and multiple approaches with solid theory background have been proposed. Most of the reweighting techniques propose different mapping from training loss to sample weight, while we propose a novel reweighting framework based on well calibrated aleatoric and epistemic uncertainty of network prediction. In short, these two kinds of uncertainty correspond to the inherent heteroscedastic noise in the dataset and the lack of confidence within the model respectively. Compared to loss based or other reweighting methods, proposed method based on uncertainty provides more robust performance on various experiment settings and datasets including CIFAR10, CIFAR100. In addition, it gives us a more comprehensive insight on the current status of the model throughout training process, which we can utilize to achieve better performance, and we leave this for future work.

CCS CONCEPTS

• **Computing methodologies → Computer vision**; **Cost-sensitive learning**.

KEYWORDS

noisy datasets, sample reweighting, uncertainty quantification

ACM Reference Format:
Anonymous Author(s). 2018. Uncertainty Based Loss Reweighting for Noisy Labeled Dataset. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

As the scale of datasets in research and industries grows tremendously, the problem of label noise becomes harder and harder to avoid. Simply because it is not feasible for experts to produce high quality label for each instance, some of the large datasets has automatically extracted label by web crawlers, and others gain label by crowd-sourced labeling. It is apparent that both labeling process inevitably introduces label noise, which has been proved to severely downgrade the model accuracy[30][5].

One of the most well developed class of tackling label noise is to assign sample weights when computing loss. Intuitively, clean data should make more effect on the trained model than corrupted data, but the distribution of clean and noisy data is often inaccessible. Thus, some of the previous works focused on relying on the confusion matrix or other statistics to estimate the how the noise is distributed, and assign sample weights based on its probability of being a clean data[2]. Others try to figure out the relationship between loss and sample weight[23]. Shu et al. trained a mapping from loss to sample weight in a meta learning manner, but same training loss may imply different situations. For example hard cases and corrupted cases may have the same loss.

Inspired by these recent developments, we set our aim to finding statistics that gives us the tendency of the data sample(note that we do not need the full distribution because our desired output is the sample weight) being corrupted, and reflects the confidence of the model as well. Fortunately, we found out that the two kinds of uncertainty estimation derived by Bayesian inference would do the job[12]. Previous work have utilized predicted uncertainty to reduce the weight of overconfident pseudo label in self training settings[21], and identify the potential noisy samples for relabeling[15]. We depart from these existing work by diving into the definition of the above two kinds of uncertainty. We proposed a novel loss reweighting framework based on the different properties of the two kinds of uncertainties. It is worth a mention that our framework has no need for clean validation set and is highly explainable.

Our contribution mainly includes:

(1) Finding the most suitable statistics to account for the uncertainties by following the definition, analysing experiments results by various previous works, doing our own experiments for validation.
(2) Proposing a novel reweighting scheme using the above statistics and achieve state of art results in standard datasets for training with presence of label noise including CIFAR10, CIFAR100.
(3) Incidentally provide solution to some frequently faced problems in dataset besides label noise utilizing properties of predicted uncertainty.
(4) apply to curriculum learning(derive the difficulty using uncertainty)

2 MODEL FRAMEWORK

Need math derivation, only qualitative is not enough

3 EXPERIMENTS

4 RELATED WORK

In this section, recent research in two most related areas will be introduced briefly.

4.1 Label Noise

4.2 Other Methods

Curriculum learning[3] was inspired by the fact that human and animals learn much better when learning material was organized and under some sort of order. Bengio et. al increase the ratio of difficult samples throughout the training process to spend less time on hard and probably unhelpful samples in the beginning, and finally lead the model to gain a better generalization. However, different tasks have different and sometimes hard to define metrics of difficult data, so if no modifications are made, it is a bit troublesome for practical use. Co-teaching[10] maintains two separate neural networks and in every iteration they pass the samples with the least loss in the given training data batch to the other one. Han et al. claims that this framework decreases the risk of error accumulation when training samples are selected and trained by a single network. Approaches with a noisy channel model a noise transition matrix which adds noise to the output so that the main part of the model will learn to clean produce output. When testing this channel will be removed. Loss correction[19] first estimates the noise matrix, and then modifies the loss by multiplying model predictions with noise transition matrix to match them with noisy labels, which is named 'forward', or multiplies the calculated loss with inverse of noise transition matrix to obtain unbiased estimator of loss function for clean data, which is named 'backward'. Gold Loss Correction(GLC)[11] assumed that small amount of clean data can be acquired, and estimates the noise matrix using the clean data. Other loss correction methods includes generalized mean absolute error(MAE) loss and categorical cross entropy (CCE) loss[29].

4.2.1 Sample Reweighting. Actually, sample reweighting is a spacial case for loss modification. Early approaches[17][24] tend to rely on latent parameters of the dataset that may not be easily acquired, like noise ratio, or simply the corruption matrix. Also the algorithm are rather simple for practical use.

Recently, [2] fits two beta distributions for the loss of clean and noisy data. Static bootstrapping loss is modified to apply the clean data probability derived from the above distributions.

To learn the sample weight in a meta learning manner, Learning to reweight(L2RW)[20] performs backward on backward for every mini batch with a clean validation set as meta data.

Compared to L2RW, meta-weight-net[23] has the sample weight only related to the training loss, and it learns the weight function in an explicit way by continuously updating a single-hidden-layer network as the meta parameters.

4.3 Uncertainty Quantification

5 CONCLUSION AND FUTURE WORK

REFERENCES

[1] Gorkem Algan and Ilkay Ulusoy. 2019. Image Classification with Deep Learning in the Presence of Noisy Labels: A Survey. *ArXiv* abs/1912.05170 (2019).

[2] Eric Arazo, Diego Ortego, Paul Albert, Noel E. O'Connor, and Kevin McGuinness. 2019. Unsupervised Label Noise Modeling and Loss Correction. arXiv:cs.CV/1904.11238

[3] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. In *ICML '09*.

[4] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. 2015. Weight Uncertainty in Neural Networks. arXiv:stat.ML/1505.05424

[5] Benoît Frénay and Michel Verleysen. 2014. Classification in the Presence of Label Noise: A Survey. *IEEE Transactions on Neural Networks and Learning Systems* 25 (2014), 845–869.

[6] Yarin Gal. 2016. Uncertainty in Deep Learning.

[7] Yarin Gal and Zoubin Ghahramani. 2015. Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference. arXiv:stat.ML/1506.02158

[8] Yarin Gal and Zoubin Ghahramani. 2015. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *arXiv preprint arXiv:1506.02142* (2015).

[9] Jacob Goldberger and Ehud Ben-Reuven. 2017. Training deep neural-networks using a noise adaptation layer. In *ICLR*.

[10] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Wai-Hung Tsang, and Masashi Sugiyama. 2018. Co-teaching: Robust training of deep neural networks with extremely noisy labels. In *NeurIPS*.

[11] Dan Hendrycks, Mantas Mazeika, Duncan Wilson, and Kevin Gimpel. 2018. Using Trusted Data to Train Deep Networks on Labels Corrupted by Severe Noise. arXiv:cs.LG/1802.05300

[12] Alex Kendall and Yarin Gal. 2017. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *CoRR* abs/1703.04977 (2017). arXiv:1703.04977 http://arxiv.org/abs/1703.04977

[13] Alex Kendall, Yarin Gal, and Roberto Cipolla. 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7482–7491.

[14] Diederik P. Kingma, Tim Salimans, and Max Welling. 2015. Variational Dropout and the Local Reparameterization Trick. *ArXiv* abs/1506.02557 (2015).

[15] Jan M. Köhler, Maximilian Autenrieth, and William H. Beluch. 2019. Uncertainty Based Detection and Relabeling of Noisy Image Labels. arXiv:cs.CV/1906.11876

[16] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. 2017. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 6402–6413. http://papers.nips.cc/paper/7219-simple-and-scalable-predictive-uncertainty-estimation-using-deep-ensembles.pdf

[17] Tongliang Liu and Dacheng Tao. 2016. Classification with Noisy Labels by Importance Reweighting. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 3 (March 2016), 447–461. https://doi.org/10.1109/TPAMI.2015.2456899

[18] Volodymyr Mnih and Geoffrey E. Hinton. 2012. Learning to Label Aerial Images from Noisy Data. In *ICML*.

[19] Zhen Qin, Zhengwen Zhang, Yan Li, and Jun Guo. 2019. Making Deep Neural Networks Robust to Label Noise: Cross-Training with a Novel Loss Function. *IEEE Access* PP (09 2019), 1–1. https://doi.org/10.1109/ACCESS.2019.2940653

[20] Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. 2018. Learning to Reweight Examples for Robust Deep Learning. *ArXiv* abs/1803.09050 (2018).

[21] Fabio De Sousa Ribeiro, Francesco Caliva, Mark Swainson, Kjartan Gudmundsson, Georgios Leontidis, and Stefanos Kollias. 2018. Deep Bayesian Self-Training. arXiv:cs.CV/1812.01681

[22] Mattia Segù, Antonio Loquercio, and Davide Scaramuzza. 2019. A General Framework for Uncertainty Estimation in Deep Learning. *CoRR* abs/1907.06890 (2019). arXiv:1907.06890 http://arxiv.org/abs/1907.06890

[23] Jun Shu, Qi Xie, Lixuan Yi, Qian Zhao, Sanping Zhou, Zongben Xu, and Deyu Meng. 2019. Meta-Weight-Net: Learning an Explicit Mapping For Sample Weighting. arXiv:cs.LG/1902.07379

[24] Sainbayar Sukhbaatar, Joan Bruna, Manohar Paluri, Lubomir Bourdev, and Rob Fergus. 2014. Training Convolutional Networks with Noisy Labels. arXiv:cs.CV/1406.2080

[25] Sainbayar Sukhbaatar and Rob Fergus. 2014. Learning from Noisy Labels with Deep Neural Networks. *CoRR* abs/1406.2080 (2014).

[26] Anqi Wu, Sebastian Nowozin, Edward Meeds, Richard E. Turner, Jose Miguel Hernandez-Lobato, and Alexander L. Gaunt. 2019. Deterministic Variational Inference for Robust Bayesian Neural Networks. In *International Conference on Learning Representations*. https://openreview.net/forum?id=B1l08oAct7

[27] Tong Xiao, Tian Xia, Yi Yang, Chang Huang, and Xiaogang Wang. 2015. Learning from massive noisy labeled data for image classification. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), 2691–2699.

[28] Yoel Zeldes, Stavros Theodorakis, Efrat Solodnik, Aviv Rotman, Gil Chamiel, and Dan Friedman. 2017. Deep density networks and uncertainty in recommender systems. arXiv:cs.IR/1711.02487

[29] Zhilu Zhang and Mert R. Sabuncu. 2018. Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels. In *NeurIPS*.

[30] Xingquan Zhu and Xindong Wu. 2004. Class Noise vs. Attribute Noise: A Quantitative Study of Their Impacts. *Artif. Intell. Rev.* 22, 3 (Nov. 2004), 177–210.

A RESEARCH METHODS

A.1 Part One

A.2 Part Two

B ONLINE RESOURCES