ICAI 2023
12th International Conference on Applied Informatics
Eger, Hungary, 2023

Deep learning from noisy labels with some adjustments of a recent method

István Fazekas[a], **László Fórián**[b], **Attila Barta**[a]

[a]University of Debrecen, Faculty of Informatics
fazekas.istvan@inf.unideb.hu
barta.attila@inf.unideb.hu

[b]University of Debrecen, Doctoral School of Informatics
forian.laszlo@inf.unideb.hu

Abstract

Deep neural networks have excellent performance in image classifications tasks, but they are in need of large sets of training data with correct labels. This is a drawback, since labeling is difficult or too expensive in many cases. The available datasets are often contaminated by label noise, that is why the challenge of learning with noisy labels has become an important research topic with several directions [1]. Even though deep neural networks tend to learn the simple, consistent patterns first, they can easily overfit to noisy labels [2]. If we are able to prevent this overfitting and treat the label noise during the training process, we can obtain models with good generalization ability.

In this work, we have investigated the possibilities of the improvement of a recent method in the topic of learning with label noise. We have applied some modifications to various points of the training process, evaluated those adjusted models and drawn conclusions from the results. We outline some details about this method and our contributions in the following paragraphs.

JoCor [4] is one of the recent state-of-the-art techniques for learning with label noise. It uses the idea of the selection of small-loss samples along with the utilization of two neural networks and it gradually increases the agreement between them. This model is trained with two classifiers in the background and a joint loss function which contains an additional term to reduce the divergence of the two networks, they are forced to make similar predictions. This scheme has a regularization effect during the training, it plays an important role to prevent overfitting.

JoCor uses convolutional neural networks (CNNs) with several convolutional and batch-normalization layers in the background, but it can be changed to any other neural network. The parameters of these two classifiers are updated simultaneously by the joint loss function, which is a weighted sum of the supervised loss of the two networks (two cross-entropy terms) and a contrastive loss term. The latter quantity is a symmetric Kullback-Leibler divergence (the sum of two KL terms: $D_{KL}(p||q) + D_{KL}(q||p)$ if p and q are the two distributions obtained from the softmax outputs). Images considered as clean are selected with the small loss criterion using the joint loss function. At the start of the training procedure, the whole training dataset is used, then fewer training examples are selected in the upcoming epochs until it gradually reaches the ratio $1 - \tau$, where τ is the known or estimated ratio of the noisy labels in the training dataset. JoCor shows very impressive performance on several datasets with label noise, including CIFAR-10 and CIFAR-100 with two types of synthetic label noise. In the case of symmetric label noise, a given proportion of the labels is flipped to one of the other classes according to a discrete uniform distribution. On the other hand, asymmetric label noise is generated by taking pairs of classes (which are similar to each other, for which humans make some mistakes, too), and a proportion of the data labels are flipped between these class pairs.

The method of JoCor can be considered as a special ensemble of the two classifiers. Unlike the techniques using a disagreement strategy ([3], [5]), JoCor can be naturally extended to more than two networks. This raises the question: is it worth to use JoCor with three neural networks if we have the computational capacity?

One of our results is that the answer for the above question is yes; we were able to make a significant improvement in the considered symmetric and asymmetric noise cases on CIFAR-10 and CIFAR-100 using three networks and totally six Kullback-Leibler terms (for every possible pair of softmax outputs). Similar results were obtained by using only three KL terms in a circular manner, but the improvement of the model over the training process was slightly slower and the test performance seemed to have a larger variance. Cross-entropy contrastive losses were also applied, however they led to moderately weaker performance with larger variance as well. Various other adjustments were made to the model with the JoCor technique and the conclusions of those investigations are also going to be presented.

References

[1] G. Algan, I. Ulusoy: *Image classification with deep learning in the presence of noisy labels: A survey*, Knowledge-Based Systems 215 (2021), p. 106771, issn: 0950-7051, doi: https://doi.org/10.1016/j.knosys.2021.106771.

[2] D. Arpit, S. Jastrzębski, N. Ballas, D. Krueger, E. Bengio, M. S. Kanwal, T. Maharaj, A. Fischer, A. Courville, Y. Bengio, S. Lacoste-Julien: *A Closer Look at Memorization in Deep Networks*, in: ed. by D. Precup, Y. W. Teh, vol. 70, Proceedings of Machine Learning Research, International Convention Centre, Sydney, Australia: PMLR, Aug. 2017, pp. 233–242, url: http://proceedings.mlr.press/v70/arpit17a.html.

[3] B. Han, Q. Yao, X. Yu, G. Niu, M. Xu, W. Hu, I. Tsang, M. Sugiyama: *Co-teaching: Robust training of deep neural networks with extremely noisy labels*, in: NeurIPS, 2018, pp. 8535–8545.

[4] H. Wei, L. Feng, X. Chen, B. An: *Combating Noisy Labels by Agreement: A Joint Training Method with Co-Regularization*, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 13723–13732, doi: 10.1109/CVPR42600.2020.01374.

[5] X. Yu, B. Han, J. Yao, G. Niu, I. Tsang, M. Sugiyama: *How does Disagreement Help Generalization against Label Corruption?*, in: International Conference on Machine Learning, 2019, pp. 7164–7173.