Modeling Noisy labels for Facial Expression Recognition

Mingjian Zhu, Hongxun Ding, Lanjing Yi

Computer Science and Engineering, SUSTECH, China

ABSTRACT. Facial expression recognition (FER) has a wide research prospect in human-computer interaction and emotion computing, including human-computer interaction, emotion analysis, intelligent security, entertainment, online education, intelligent medical care, etc. FER is challenging due to the class imbalance and noisy labels caused by data collection. We need to address both of these challenges in this project.

INTRODUCTION

Facial expression recognition (FER) aims to make computers understand visual emotion. Recently, the advancement of deep FER is largely promoted by large-scale labeled datasets, e.g., AffectNet. However, existing labels often exist noisy labels and the re-labeling data requires human experts. Besides, in many realistic scenarios, the assumption that the class distribution of data are balanced holds untrue and becomes the primary cause of poor model performance. Therefore, it is urgent to develop a powerful method for training models on a large amount of data with noisy labels and imbalanced data distribution.

There are two kinds of noise in the literature: feature noise and label noise. Feature noise corresponds to the corruption in observed data features, while label noise means the change of label from its actual class. Even though both noise types may cause a significant decrease in the performance, label noise is considered to be more harmful and shown to deteriorate the performance of classification systems in a broad range of problems. This is due to several factors; the label is unique for each data while features are multiple, and the importance of each feature varies while the label always has a significant impact. This work focuses on label noise; therefore, noise and label noise is used alternately throughout the article.

In real-world applications, training samples typically exhibit a long-tailed class distribution, where a small portion of classes have massive sample points but the others are associated with only a few samples. Such class imbalance of training sample numbers, however, makes the training of deep network based recognition models very challenging. The trained model can be easily biased towards head classes with massive training data, leading to poor model performance on tail classes that have limited data. Therefore, the deep models trained by the common practice of empirical risk minimization cannot handle FER applications with long-tailed class imbalance.

In this work, we propose a two-stage cycle. In the first stage, we detect the noise labels in the dataset using a Learning with Noisy Labels(LNL) method; in the second stage, we use the data with clean labels and the data with noise labels to train the model, and a Semi-Supervised Learning(SSL) method is implemented in this stage. Taking into account the long-tailed data distribution problem, we use some Long-Tailed Recognition(LTR) method as regularization. So we can create a virtuous circle.

RELATED WORK

Learning with Noisy Labels

Existing methods on LNL primarily take a loss correction approach. Some methods estimate the noise transition matrix and use it to correct the loss function. However, correctly estimating the noise transition matrix is challenging. Some methods leverage the predictions from Deep Neural Network(DNN)s to correct labels and modify the loss accordingly. These methods do not perform well under high noise ratio as the predictions from DNNs would dominate training and cause overfitting. To overcome this, some methods adopt MixUp augmentation. Another approach selects or reweights samples so that noisy samples contribute less to the loss. A challenging issue is to design a reliable criteria to select clean samples. It has been shown that DNNs tend to learn simple patterns first before fitting label noise. Therefore, many methods treat samples with small loss as clean ones. Among those methods, Co-teaching and Co-teaching+ train two networks where each network selects small-loss samples in a mini-batch to train the other.

Semi-Supervised Learning

Recent years have observed a significant advancement of SSL research. Many of these methods share similar basic techniques, such as entropy minimization, pseudo-labeling, or consistency regularization, with deep learning. Pseudo-labeling trains a classifier with unlabeled data using pseudo-labeled targets derived from the model's own predictions. Relatedly, some methods use a model's predictive probability with temperature scaling as a soft pseudo-label. Consistency regularization learns a classifier by promoting consistency in predictions between different views of unlabeled data, either via soft or hard pseudo-labels.

Effective methods of generating multiple views include input data augmentations of varying strength, standard dropout within network layers, and stochastic depth. The performance of most recent SSL methods relies on the quality of pseudolabels. However, none of aforementioned works have studied SSL in the class-imbalanced setting, in which the quality of pseudo-labels is significantly threatened by model bias.

Long-Tailed Learning

While SSL has been extensively studied, it is underexplored regarding class-imbalanced data. Recently, some methods argued that leveraging unlabeled data by SSL and self-supervised learning can benefit class-imbalanced learning. Some methods proposed a suppressed consistency loss to suppress the loss on minority classes. Some also discussed another interesting setting where labeled and unlabeled data do not share the same class distribution, while in this work we focus on the scenario when labeled and unlabeled data have roughly the same distribution, which is similar to FER.

PLAN

Current progress

Complete the survey

Complete the paper reading

Before the mid report

Implement the methods on AffectNet

Before the final report

Accomplish a relatively good accuracy rate

REFERENCES

Li, J., Socher, R., Hoi, S. C. (2020). Dividemix: Learning with noisy labels as semi-supervised learning. arXiv preprint arXiv:2002.07394.

Sun, Z., Shen, F., Huang, D., Wang, Q., Shu, X., Yao, Y., Tang, J. (2022). PNP: Robust Learning From

Noisy Labels by Probabilistic Noise Prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5311-5320).

Li, H., Wang, N., Yang, X., Wang, X., Gao, X. (2022). Towards Semi-Supervised Deep Facial Expression Recognition with An Adaptive Confidence Margin. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4166-4175).

Yi, X., Tang, K., Hua, X. S., Lim, J. H., Zhang, H. (2022). Identifying Hard Noise in Long-Tailed Sample Distribution. arXiv preprint arXiv:2207.13378.

Pham, H., Dai, Z., Xie, Q., Le, Q. V. (2021). Meta pseudo labels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11557-11568).

Wei, C., Sohn, K., Mellina, C., Yuille, A., Yang, F. (2021). Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10857-10866).

Algan, G., Ulusoy, I. (2021). Image classification with deep learning in the presence of noisy labels: A survey. Knowledge-Based Systems, 215, 106771.

Song, H., Kim, M., Park, D., Shin, Y., Lee, J. G. (2022). Learning from noisy labels with deep neural networks: A survey. IEEE Transactions on Neural Networks and Learning Systems.