Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss

Abraham Jose

1 Summary

The author developes two novel techniques to tackle the problem of class imbalance in the dataset during training, which can also be applied for the transfer learning problems. The first contribution is the loss function itself. The loss function is LDAM (label-distribution-aware margin) which is a class dependent margin for multiple classes which ensures better accuracies of the frequent classes and the minority classes with hinge loss and enforced margin. The next contribution is an effective training scheduler that defers re-weighting until after the initial stage, when the model learns some aspects of the data to reduce complications. The model performed really well in CIFAR-10, Tiny ImageNet and iNaturalist 2018 dataset and has shown state of the results in class-wise performance for both text and image dataset.

2 Good points

The proposed method 1, explaining LDAM with theoretically-principled justifications of LDAM and how it optimizes a uniform label generalization error bound, which is important in the field, reducing the ambiguity and increasing the explainability. Also, the author the author does ablation study for training scheduling including Empirical Risk Minimization(ERM), Re-weighting(RW), Re-sampling(RS) and to prove his point on various datasets proves the fact that the rescheduler is robust against dataset.

3 Weak points

However, in the ablation study for training scheduler has been accounted based on the cross-entroy loss and not LDAM loss that is proposed by the author. The effects and performance might be affected with a different loss function, even if LDAM is a derivative of the cross-entropy given the performance boost from LDAM. If the DRW works well because the deferred re-weighting provides the model time to learn a good initial representation on the network, it should be prominent in the DRS, deferred re-scheduling.

4 Questions

Author says that the LDAM reminiscent the re-weighting and both complements the learning process. While LDAM affects the multiclass nature of the problem in a more involved way, the re-weighing affects on the class probabilities. So the LDAM-DRW approach is addressing the same problem and it is possible that there might be other approaches that might not be complimenting this method and can you explain the possible such scenarios, for example with a certain loss function with which the DRW wont be working?

5 Ideas