

Experiments and Analysis about Variational Dropout

Bao Jinge(e0522065), Zhang Runtian (e0504862)

1 Selected Papar

"Variational Dropout Sparsifies Deep Neural Networks", Dmitry Molchanov, Arsenii Ashukha, and Dmitry Vetrov, ICML2017 [3].

2 About the Paper

As DNNs are prone to overfitting and also troubled by the computing cost, the paper as above used a modified regularization technique based on Variational Dropout named "Sparse Variational Dropout" that extends Variational Dropout to all possible values of dropout rates and leads to a sparse solution. Besides, the paper also propose a ways to greatly reduce the variance of the stochastic gradiant estimator and show that it leads to a much faster convergence and a better value of the objective function. In experiment, Sparse Variational Dropout leads to a high level of sparsity in full-connected and convolution layers of DNNs, especially on LeNet and VGG-like architectures.

3 Initial Reading

We try to divide their work into three parts.

First, Variational Dropout theory review. They introduced necessary concepts about this technique, including Variational Inference, Reparameterization Trick, and relationship between Binary Dropout and Gaussian Dropout. Then they analyzed the original Variational Dropout. Both its advantages and difficulties.

Second, they introduced Sparse Variational Dropout and explained their work. They did additive Noise Reparameterization and successfully avoid the problem of large gradient variance by original Variational Dropout. They proposed an Approximation of KL Divergence and solved the former intractable term of KL divergence. As a result, they optimized the stochastic gradient variational lower bound, and proved that Sparse Varitional Dropout leads to sparsity in case of linear regression, and can be used to DNN with large parameters.

Third, they performed experiments of Sparse Varitional Dropout on classification tasks using Lenet and VGG-like architectures, and produced Sparse Bayesian DNNs. They found that it leads to extremely sparse solutions both in fully connected and convolutional layers. They compared it with Automatic Relevance Determination effect in empirical bayes, and found it has more advantages.

The paper analyzed strange performance of this method under random label problem, but didn't provide good explanation at last.

4 Our Plan

The sparsity of DNNs is very important to solve overfitting problem when using regularization method.

We are going to learn the paper in detail, review the theories, reading the incited papers [1] [2] and truly understand both old and new Dropout technique. We will try to understand the effect of Sparse Variational Dropout mentioned in this paper from different views.

We will extend the experiment because the paper only applied the Sparse Variational Dropout to LeNets/VGG-like DNNs, and we plan to check its performance on more kinds of DNNs architectures, such as AlexNet, ResNet and so on. We will compare and analyze the results based on experiments mainly. The paper didn't solve the problem of random label. We will try our best to explore this problem by tuning our Dropout parameters and give some own explanations based on this lecture. Finally, we will give some meaningful future directions.

References

- [1] Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv preprint arXiv:1803.03635*, 2018.
- [2] Trevor Gale, Erich Elsen, and Sara Hooker. The state of sparsity in deep neural networks. *arXiv preprint arXiv:1902.09574*, 2019.
- [3] Dmitry Molchanov, Arsenii Ashukha, and Dmitry Vetrov. Variational dropout sparsifies deep neural networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2498–2507. JMLR. org, 2017.