Combining DropConnect and Feedback Alignment for Efficient Regularization in Deep Networks

Haiping Huang
RIKEN Brain Science Institute
Wako-shi, Saitama 351-0198, Japan

Alireza Goudarzi
RIKEN Brain Science Institute
Wako-shi, Saitama 351-0198, Japan

Taro Toyoizumi
RIKEN Brain Science Institute
Wako-shi, Saitama 351-0198, Japan

Abstract

DropConnect is a powerful regularization technique for deep networks, which improves generalization accuracy. It works by temporarily removing links from the network with probability $(1 - p_{dc})$ during training. However, the optimal probability of keeping the links p_{dc} has to be adjusted by trial and error. In addition, storing the exact configuration of the weight matrix to be used during backward pass becomes demanding for large networks. We have introduced a theoretical model that gives a principled way of choosing p_{dc}. Finally, we combine DropConnect and random feedback weights to perform efficient regularization in deep networks and avoid overfitting.

Overview

Overfitting remains a challenging problem in supervised learning. When training data is limited, a model may observe false relationships in the training data that are missing in the test data even when they are from the same underlying distribution. This problem is specially pronounced in deep networks owing to a large number of parameters in the model. Various regularization techniques have been used to combat this challenge (Goodfellow, Bengio, and Courville 2016). More recently, Dropout (Srivastava et al. 2014) and DropConnect (Wan et al. 2013) have been proposed as effective ways of regularizing the network. In Dropout, one randomly switches off some of the nodes in the forward pass of the network during training. The algorithm effectively searches through exponentially many models and combine their results, approximating a Bayesian sampling (Srivastava et al. 2014). DropConnect generalizes this by temporarily removing some of the links during the forward pass. This allows DropConnect to perform finer grain search between models than Dropout (Wan et al. 2013). In practice, DropConnect is implemented by applying a different binary mask to the network connection matrix for each input pattern. The mask is generated by sampling a binomial distribution. However, the probability with which the connections are kept must be adjusted experimentally. Moreover, storing the mask to use during the backward pass becomes demanding in large models.

Here we have developed a theoretical model that gives us first insight into choosing probability of keeping links in a principled way. We then combined DropConnect with a newly introduced technique called feedback alignment to reduce memory demand of DropConnect. In the forward pass the weights are set to zero with probability $(1 - p_{dc})$ and a forward pass through the model is performed. The backward pass is carried out by replacing the weight matrices by a predefined fixed random matrices (feedback alignment) and the original matrices are then updated using the computed gradients. This avoids the requirement for storing the masks and achieves powerful regularization in the network. Our current results show that this algorithm improves generalization error over standard backpropagation by 5%, especially for limited data where models are more susceptible to overfitting ($<$ 10k samples).

Theoretical background for this algorithm is based on the observation that owing to redundancies in the connection matrix, neural networks are robust in the absence of a fraction of connections between layers to some extent, i.e., the test error does not change until a critical fraction of connections have been removed. This redundancy can be qualitatively captured by a simple random active path model, in which we construct randomly and independently each active path from the input at the bottom layer to the output at the top layer, and each path serves as a constraint in a graphical model. Increasing the number of path amounts to increasing the weight's mean degree. The ground state energy of the model decreases with the mean degree until the mean degree exceeds a critical value. This model qualitatively explains that the dropconnect probability less than a critical value must be avoided for good learning performance.

References

Goodfellow, I.; Bengio, Y.; and Courville, A. 2016. *Deep Learning*. Cambridge, MA: MIT Press.

Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; and Salakhutdinov, R. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* 15:1929–1958.

Wan, L.; Zeiler, M.; Zhang, S.; Cun, Y. L.; and Fergus, R. 2013. Regularization of neural networks using dropconnect. In Dasgupta, S., and Mcallester, D., eds., *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, volume 28, 1058–1066. JMLR Workshop and Conference Proceedings.