WOLT Poster

Author: Guo Jun-Lin

Advisors: Wai Kin Chan / Ercan Engin Kuruoglu

Date: 2020.07.07

Complexity reduction in neural networks using simulated annealing algorithm

Tsinghua-Berkeley Shenzhen Institute

Guo Jun-Lin ; Ercan Engin Kuruoglu ; Victor Chan

Abstract & Introduction

Artificial neural networks have proved to be very successful in multiple applications including image classification, voice analysis, and language translation. Deep learning architectures such as AlexNet, VGG16 and GoogleNet have showed outstanding accuracy on image classification task. However, this performance is achieved by millions of parameters. To make networks more light weight, real-time and accessible to individual users, we need to reduce the parameters, i.e. links between nodes which would decrease the accuracy. The goal of our research is to show that significant part of performance loss can be recovered by optimizing partially connected network architectures using Simulated Annealing algorithm.

Masked Artificial Neural Network without Training

In order to construct neural networks, architecture elements, such as weight, bias, and an activation function need to be specified. Sometimes a normalization or regularization method is also applied to enhance robustness. The output of the network will be put into a loss function along with labels and the whole model will be updated by back propagation layer by layer.

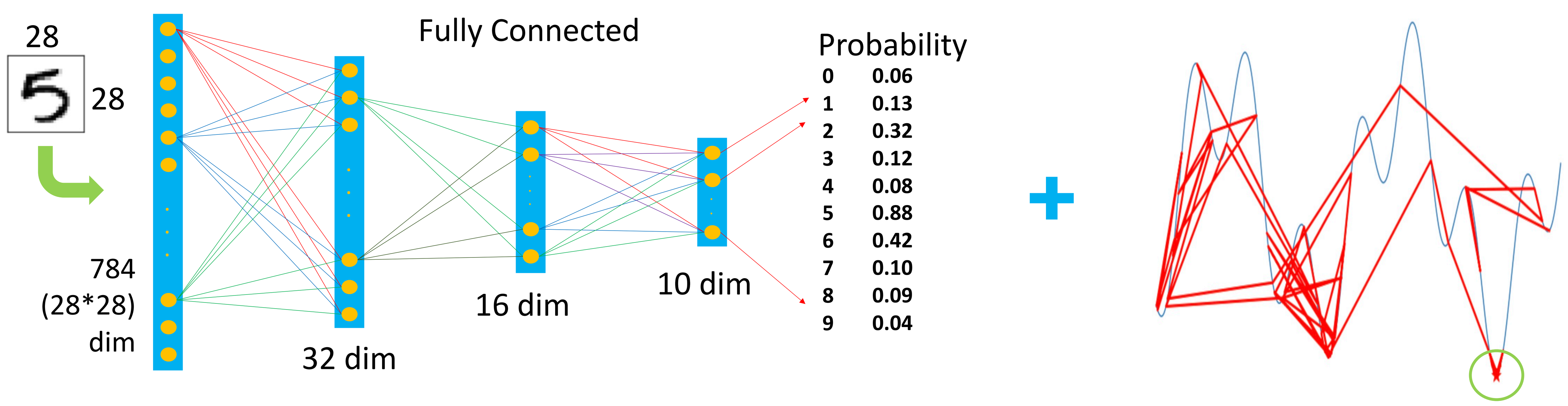

Figure 1. An illustration of artificial neural network and simulated annealing algorithm

To study the effect of the partially-connected architecture, we utilize a network with 2 hidden layers. Due to the local optimality of backpropagation algorithm, in 50% cases of initial conditions in both fully connected and partially connected cases, BP does not converge if there is no normalization method included during the training. Therefore, we split both the successful and unsuccessful model into 2 cases.

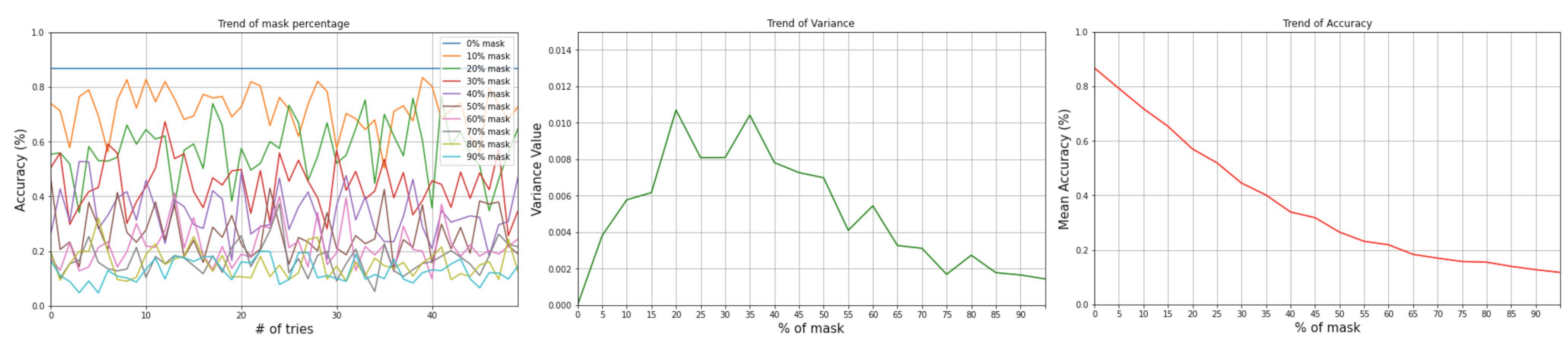

Figure 2. Accuracy Trend of different % of mask under 50 trials of independent training

We randomly disconnect different 50% links and the variance of 50 trials implies that some links are more important than the others, making the need to optimize the architecture.

Masked Artificial Neural Network with Back Propagation

For the successful case, the network parameters are robust enough to train the model successfully again regardless of how many percentage of links are disconnected. Even though sometimes crucial

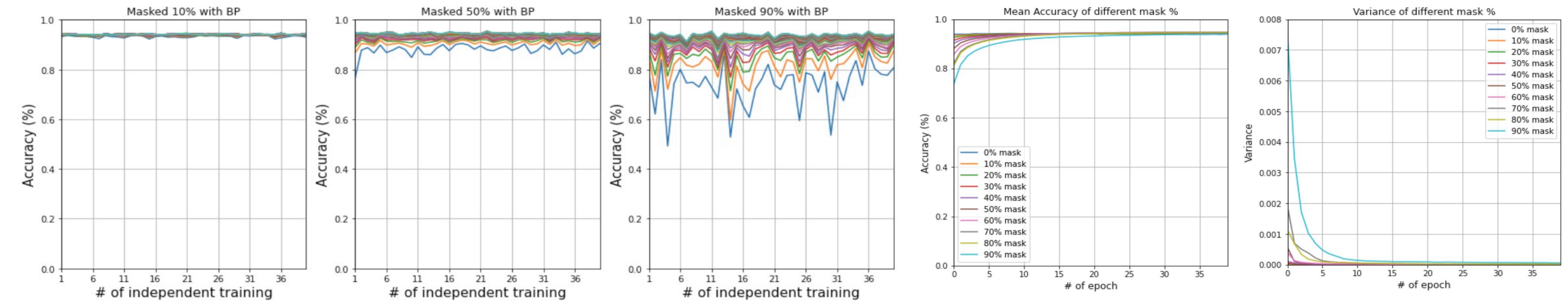

Figure 3. Accuracy Trend of different % of mask after 40 epochs training (successful case)

links are shut down, causing significant accuracy drop at that round, back propagation can still find its own way to improve the accuracy back to the original level again. As for the unsuccessful case, we

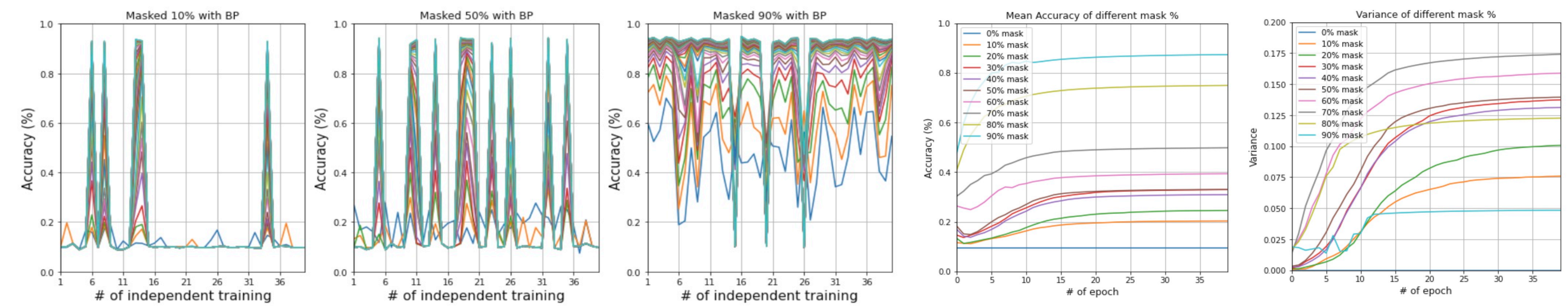

Figure 4. Accuracy Trend of different % of mask after 40 epochs training (unsuccessful case)

found that when there are more links being disconnected, it is more possible that the network can escape from the local optimum and achieve high accuracy again. This phenomenon implies that the simulated annealing algorithm can guide the network to connect and disconnect links under certain rules so that the final accuracy can approach the fully connected network using far less links.

Masked Artificial Neural Network with Back Propagation + Simulated Annealing

Simulated annealing algorithm is motivated by the annealing process in solid state physics. It can cool down the electrons in their lowest energy states. In simulated annealing, we want the parameters to achieve their lowest cost constellation. This is achieved by a Markov chain. The stationary distribution should be a Boltzmann law. Convergence to a stationary distribution is achieved by acceptance rejection scheme which ensures that detailed balance condition is satisfied.

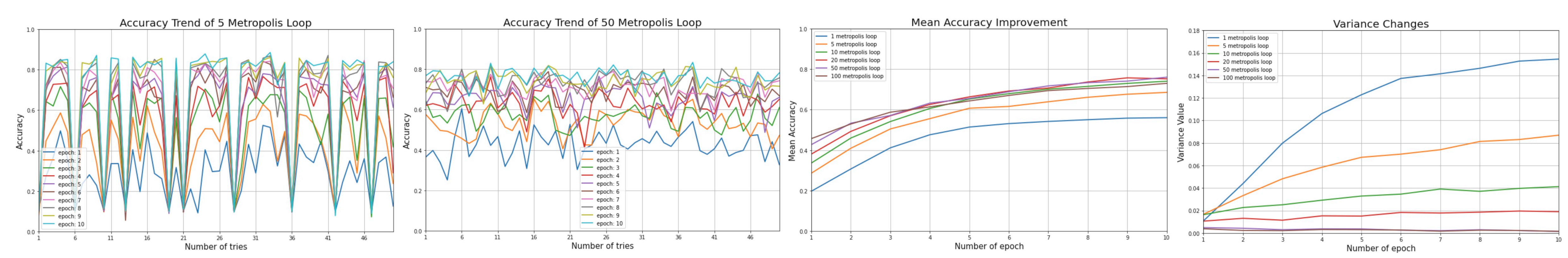

Figure 5. Accuracy Trend of different metropolis loop length under 50 times of independent training

In our experiment, the results show that simulated annealing algorithm can robustly help to train a partially connected neural network when the metropolis loop length is big enough. In the case of our simple network, we found that the network will always converge and get good results when the metropolis loop length equals to 50. Because of the randomness, we tried 50 times independently to confirm the robustness of simulated annealing algorithm under 1/10/20/50 metropolis loop length.

Conclusion

By integrating simulated annealing algorithm to a neural network, the training process can be robust. Although it takes more time on examine configurations, it make sure that the loss value is converged.

Contact: Guo Jun-Lin – guojl19@mails.tsinghua.edu.cn Ercan Engin Kuruoglu – kuruoglu@sz.tsinghua.edu.cn Victor Chan – chanw@sz.tsinghua.edu.cn