Decoupled Neural Interfaces Using Synthetic Gradients

Qilei Zhang

Jun 18 2018

Abstract

Training directed neural networks typically requires forward-propagating data through a computation graph, followed by backpropagating error signal, to produce weight updates. All layers, or more generally, modules, of the network are therefore locked, in the sense that they must wait for the remainder of the network to execute forwards and propagate error backwards before they can be updated.

1. Introduction

Each layer (or module) in a directed neural network can be considered a computation step, that transforms its incoming data. These modules are connected via directed edges, creating a forward processing graph which defines the flow of data from the network inputs, through each module, producing network outputs. [1].

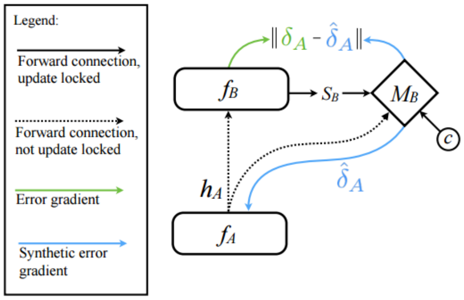
Legend:
Forward connection, update locked
Forward connection, not update locked
Error gradient
Synthetic error gradient
$\|\delta_A - \hat{\delta}_A\|$
f_B
S_B
M_B
c
h_A
$\hat{\delta}_A$
f_A

Figure 1. General communication protocol between A and B. After receiving the message hA from A, B can use its model of A, MB, to send back synthetic gradients ?A which are trained to approximate real error gradients A. Note that A does not need to wait for any extra computation after itself to get the correct error gradients, hence decoupling the backward computation. The feedback model MB can also be conditioned on any privileged information or context, c, available during training such as a label

$$x'(t) = -V'(x) + A_0 cos(wt + o) + u(t) \qquad (1)$$

2. Decoupled Neural Interfaces

We begin by describing the high-level communication protocol that is used to allow asynchronously learning agents to communicate [2]. This protocol allows A to send messages to B in a way that A and B are update decoupled C A does not have to wait for B to evaluate the true utility [3] before it can be updated Cand A can still learn to send messages of high utility to B.

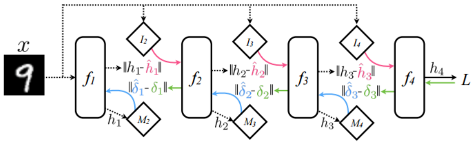

Figure 2. Completely unlocked feed-forward network training allowing forward and update decoupling of layers.

3. Conclusion

In this work we introduced a method, DNI using synthetic gradients, which allows decoupled communication between components [4], such that they can be independently updated. [5].

References

[1] A. Krogh. Neural network ensemble, cross validation and active learning. *Advances in Neural Information Processing Systems*, 7(10):231–238, 1995. 1

[2] J. Pustejovsky. The generative lexicon. *Computational Linguistics*, 17(4):409–441, 1998. 1

[3] T. D. Sanger. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2(6):459–473, 1989. 1

[4] H. H. Szu, B. A. Telfer, and S. L. Kadambe. Neural network adaptive wavelets for signal representation and classification. *Optical Engineering*, 31(9):1907–1916, 1992. 1

[5] M. C. Wittrock. Generative processes of comprehension. *Educational Psychologist*, 24(4):345–376, 1989. 1