Optimal Memory-aware Backpropagation of Deep Join Networks

Julien Herrmann*[1]

[1]Inria Bordeaux – Inria Bordeaux – France

Abstract

Deep Learning training memory needs can prevent the user to consider large models and large batch sizes.
In this work, we propose to use techniques from memory-aware scheduling and Automatic Differentiation (AD) to execute a backpropagation graph with a bounded memory requirement at the cost of extra recomputations.
The case of a single homogeneous chain, the case of a network whose all stages are identical and form a chain, is well understood and optimal solutions have been proposed in the AD literature.
The networks encountered in practice in the context of Deep Learning are much more diverse, both in terms of shape and heterogeneity.

In this work, we define the class of backpropagation graphs, and extend those on which one can compute in polynomial time a solution that minimizes the total number of recomputations. In particular we consider join graphs which correspond to models such as Siamese or Cross Modal Networks.

Keywords: backpropagtion, neural network

*Speaker