Layer-wise Asynchronous Training of Neural Network with Synthetic Gradient

Xupeng Tong, Hao Wang, Ning Dong, Griffin Adams
Part One

BACKGROUND
Back Propagation of Training CNN

1 BACKGROUND
Asynchronous SGD

\[ w' = w - \eta \Delta w \]

Training of CNN by Synthetic Gradient

Each Layer can be trained independently
Part Three
SYNCHRONOUS TRAINING
Minimizing the synthetic input/gradient simultaneously with the general loss

\[ L_M = \sum_i \| \delta_i - \hat{\delta}_i \|^2 \]

\[ L_g = \sum_i \| h_i - \hat{h}_i \|^2 \]

\[ L(w_1, b_1, \ldots, w_i, b_i, \ldots, w_n, b_n) = L(W, B) \]
Part Four
ASYNCHRONOUS TRAINING
4-1 ONE POSSIBLE ARCHITECTURE
4-1 INFRASTRUCTURE - BASIC
4-2 INFRASTRUCTURE - LIGHT
• Synthetic gradient and synthetic input as a new alternative of batch normalization / Dropput

• M and I auxiliary network introduces noises in the input/gradient of each layer

• No exact update is required!

• The model will learn how to reduce the variance within each batch

• while keeping the flavor of that specific batch.
Thank You