

分散型深層学習における複数のデータ圧縮技術を利用した 最適通信スケジューリングの提案

福田 竜大[†] 橘 拓至^{††}

[†] 福井大学 工学部 〒910-8507 福井県福井市文京 3-9-1

^{††} 福井大学 大学院工学研究科 〒910-8507 福井県福井市文京 3-9-1

E-mail: [†]ryudai-f@network.fuis.u-fukui.ac.jp, ^{††}takuji-t@u-fukui.ac.jp

あらまし 複数のプロセッサを使用する分散型深層学習では、各プロセッサで学習処理を実行することで学習時間を大幅に短縮できる。しかしながら、プロセッサ間で学習結果を通信する必要があり、通信データサイズが総学習時間に大きく影響する。それゆえ、データ圧縮技術を利用して通信データサイズを低減することが期待されるが、各層の学習処理や通信方式によって適切な圧縮率や圧縮処理時間が異なり使用する圧縮技術によって総学習時間が変化してしまう。本稿では、分散型深層学習において、複数のデータ圧縮技術を利用する最適通信スケジューリングを提案する。提案法では、データ圧縮率や処理時間が異なる複数の圧縮技術を各層で適切に使用することで総学習時間の最小化を実現する。提案法の性能をシミュレーションによって評価し、提案法の有効性を評価する。数値例から、提案法によって各層で適切なデータ圧縮技術を利用することで総学習時間を低減できることがわかる。

キーワード 分散型深層学習, 誤差逆伝播法, データ圧縮, 通信スケジューリング

Optimal Data Communication Scheduling Considering Multiple Data Compression Techniques in Distributed Deep Learning

Ryudai FUKUDA[†] and Takuji TACHIBANA^{††}

[†] School of Engineering, University of Fukui 3-9-1 Bunkyo, Fukui, Fukui 910-8507, Japan

^{††} Graduate School of Engineering, University of Fukui 3-9-1 Bunkyo, Fukui, Fukui 910-8507, Japan

E-mail: [†]ryudai-f@network.fuis.u-fukui.ac.jp, ^{††}takuji-t@u-fukui.ac.jp

Abstract In distributed deep learning, which uses multiple processors, the training time can be greatly reduced by executing the training process on each processor. However, it is necessary to communicate the training results between processors, and the size of the communication data greatly affects the total training time. Although it is expected that the data size is decreased using a data compression technique, the appropriate compression ratio and time vary depending on the learning process of each layer and the communication method and the total training time varies depending on the compression technique. In this paper, we propose an optimal communication scheduling method that uses multiple data compression techniques in distributed deep learning. The proposed method minimizes the total training time by appropriately using multiple compression techniques with different data compression ratios and processing times in each layer. The performance of the proposed method is evaluated by simulation to investigate the effectiveness of the proposed method. The numerical examples show that the proposed method can reduce the total training time by using appropriate data compression techniques in each layer.

Key words Distributed deep learning, Back propagation, Data compression, Communication scheduling