IB-002

【既発表論文紹介】

Involving CPUs into Multi-GPU Deep Learning

レドゥック・トゥン† 関山 太朗‡ 根岸 康† 今井 晴基† 河内谷 清久仁†
Tung D. Le Taro Sekiyama Yasushi Negishi Haruki Imai Kiyokuni Kawachiya

出典：The 9th ACM/SPEC International Conference on Performance Engineering (ICPE 2018), pp. 56–67

本講演では，国際会議 ICPE 2018 にて発表した，複数 GPU を用いた深層学習を高速化する手法について概説する．ニューラルネットワークを用いた深層学習トレーニングは非常に計算量の多い処理であり，GPU を用いても数日を要することがある．これを短縮するため，複数の GPU を用いることが一般的になってきており，その典型的な手法として各 GPU に異なる入力データを与える「データ並列」方式があげられる．しかしこの方式では各 GPU で求めた勾配値を定期的に交換・集計する必要があり，この処理がスケーラビリティを下げる原因となっている．我々は，この集計処理を，各 GPU での学習処理と並行して，従来は活用されていなかった CPU に行わせることで高速化する「CPU-GPU データ並列（CGDP)」学習方式を提案する．そのコストモデルと，Caffe に実装し性能評価した結果についても述べる．

† 日本アイ・ビー・エム(株) 東京基礎研究所, IBM Research - Tokyo
‡ 国立情報学研究所, National Institute of Informatics（現所属）