大規模機械学習向けクラスタにおける ネットワークバンド幅とパラメータ交換手法に関する考察

黎 明曦†,谷村 勇輔‡†,中田 秀基‡† （† 筑波大, ‡ 産総研）

背景

- 大規模データ並列機械学習
  - モデルのパラメータ（もしくはグラディエント）を定期的に交換して同期
- どの程度のネットワークが必要なのか

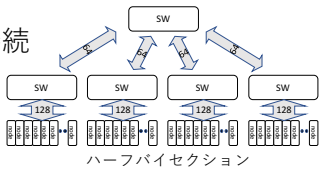
研究の目的

- バイセクションバンド幅とパラメータ交換手法の関係を調査
  - 並列シミュレータ「SimGrid」を利用

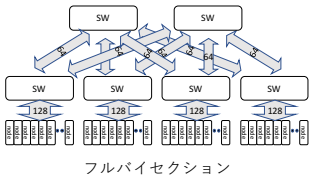
ネットワークモデル

- 各クラスタ：128ノード
- 各スイッチ：最大256接続

- 比較的大規模なスイッチで構成した複数のサブクラスタを、上位のスイッチで接続してスケールアップ



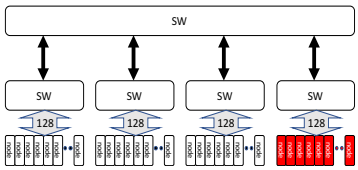
- 上位のスイッチを複数設けてファットツリーを構成



パラメータサーバによる同期

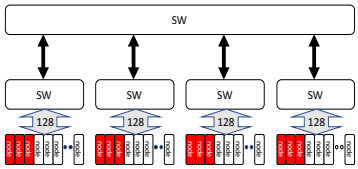
集中サーバ

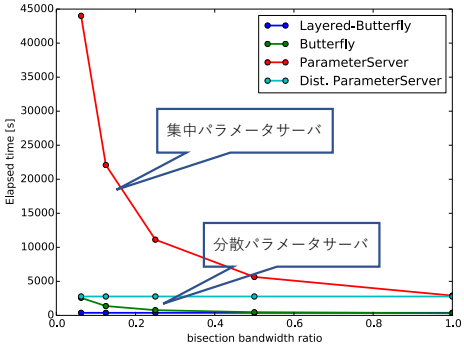
- 一つのサブクラスタをすべてパラメータサーバに



分散サーバ

- パラメータサーバをサブクラスタに均等に分散





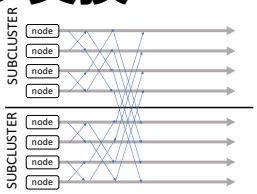
- パラメータサーバはバタフライと比較して一般に低速
- 十分なバンド幅があれば、集中型と分散型の性能の は同等
- 集中パラメータサーバは特にバイセクションバンド幅の低下に敏感

交換モデルサイズ1Gbyte
ネットワーク速度1GByte/s
パラメータ交換間隔1秒、10回の交換で測定

直接パラメータ交換

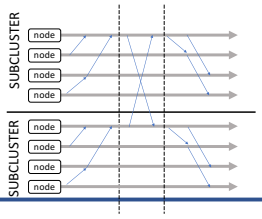
単純バタフライ

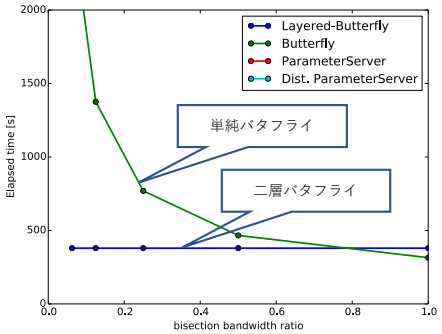
- 通信段数：Log$_2$NM
  N = クラスタ内ノード数
  M = サブクラスタ数



二層バタフライ

- 通信段数:2Log$_2$N+Log$_2$M
  N = クラスタ内ノード数
  M = サブクラスタ数





- 十分なバンド幅があれば、単純バタフライのほうが高速
- 単純バタフライはバイセクションバンド幅の低下に敏感
- 二層バタフライはバイセクションバンド幅の低下に影響を受けない

まとめ

- 2階層ファットツリーネットワークを前提にバイセクションバンド幅とパラメータ交換のレイテンシの関係をシミュレーションで調査
- パラメータ交換手法を選択することで、比較的プアなネットワークでも大規模機械学習が可能

今後の課題

- 非同期実行の評価
  - レイテンシを隠蔽する事が可能
  - 学習に及ぼす影響の定量的な評価

- この成果の一部は、国立研究開発法人新エネルギー・産業技術総合開発機構 （ＮＥＤＯ）の委託業務の結果得られたものです。
- 本研究はJSPS科研費 JP16K00116の助成を受けたものです。