

経済産業大臣賞

「大規模深層学習のための自動並列処理ソフトウェアRaNNC」

国立研究開発法人 情報通信研究機構 田仲 正弘、鳥澤 健太郎  
 国立大学法人 東京大学 田浦 健次朗、塙 敏博

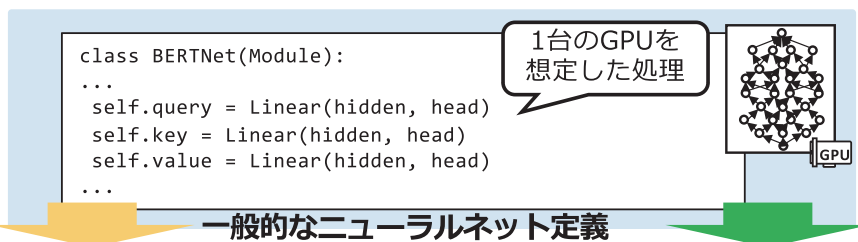
高度な知識と大きな作業コストを要する大規模な深層学習を  
 並列計算の自動化により劇的に容易にするフリーソフトRaNNCを開発

開発の背景

- 画像認識・対話システムなど、深層学習を活用する幅広いシステムにおいて、ニューラルネットワークを大規模化することで、性能を劇的に向上できる
- 大規模ニューラルネットの学習には膨大な計算が必要であり、実際の学習では、多数のGPU（深層学習等の計算を高速に行える装置）で並列に計算する処理が必須
- これまで、そうした高速な並列計算による学習を実現するには、複雑なニューラルネットの定義の書き換えが必要で、その道の専門家のいない組織では不可能

RaNNCの特長

- 大規模なニューラルネットの多数のGPUを用いた学習時の並列計算を自動化
- 並列計算のためのニューラルネット定義の書き換え等、ハイレベルな専門的知識、スキルや大量の試行錯誤が必要な作業が不要に  
 →大規模なニューラルネットの学習が飛躍的に簡単に



従来技術 大幅な書き換えや試行錯誤が必要

```
class BERTNet(Module):
    ...
    self.query_key_value =
        ColumnParallelLinear(hidden, 3*proj)
    self.dense =
        RowParallelLinear(proj, hidden)
```

計算を複数のGPUにどう分担させるかを細かく指定

特定のニューラルネットの種類にのみ適用可能

Megatron-LM(NVIDIA)での記述例

RaNNC 書き換え不要・試行錯誤等も自動化

```
class BERTNet(Module):
    ...
    self.query = nn.Linear(hidden, head)
    self.key = nn.Linear(hidden, head)
    self.value = nn.Linear(hidden, head)
    ...
```

多数のGPUでの計算の分担を自動的に決定

対象のニューラルネットの種類に制限無し

ニューラルネットの定義を人手で変更する必要のある既存のソフトウェアよりも学習も高速

成果

- 2,000億パラメータのニューラルネットの学習が可能なることを確認
- 有名なGPT-3は1,750億パラメータ
- オープンソースでGitHubに公開（商用利用も含めて無償で利用可能）
- Facebook主催のPyTorch Annual HackathonでFirst Place受賞（1947人が110ヶ国から参加、応募65件）（PyTorch Developer Tools & Libraries部門）

