

クラウド・エッジ連携による DNN モデル運用方式の提案と評価

大越 淳平† 浜 直史† 近藤 伸和†

(株)日立製作所 研究開発グループ†

1 序論

ビッグデータ活用に代表される高度なデータ処理を実現するため、中央拠点にデータを集約し潤沢な計算機リソースを活用するクラウドコンピューティングや、各拠点の計算機リソースを用いてリアルタイム性の高い処理を実現するエッジコンピューティングが新たなコンピューティング形態として提案され、社会に浸透しつつある[1]。

近年では、IoT デバイスが生成する膨大なデータに機械学習を適用し高度な分析を行うことで、業務効率や生産効率を向上させるデータ分析のニーズが高まっている。特に、機械学習の一分野である深層学習は、従来人手で定義していた特徴量を自動で学習することが可能であり、IoT と連携した産業応用の機運が高まっている[2]。

しかし、深層学習における DNN (Deep Neural Network) モデルの学習には、膨大な学習データと計算機リソースを必要とする。そのため、クラウドコンピューティングでは学習データを IoT デバイスからクラウドに集約するための通信帯域の不足が、エッジコンピューティングでは DNN モデルの学習のためにエッジにおける計算機リソースの不足がそれぞれ問題となる (図 1)

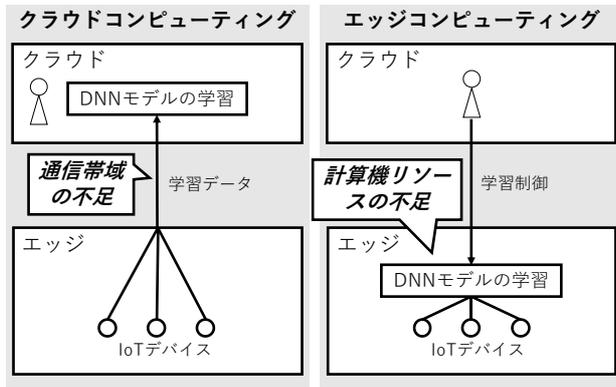


図 1 従来のコンピューティング形態

これら従来のコンピューティング形態で生じる種々の問題に対し、我々はエッジとクラウドが連携した新たなコンピューティングコンセプト「Embodied Computing」を提案している[3]。本研究の目的は、このコンピューティングコンセプトにおいて、深層学習の適用により生じる

前述の通信帯域と計算機リソースの問題を解決する DNN の運用方式を提案することにある。

2 提案システム

提案システムにおける DNN の運用方式をシステムアーキテクチャとともに図 2 に示す。基本的なアイデアは、汎用的なデータを用いて DNN モデルの学習をクラウドで事前に施し、残りの学習をエッジに展開後にエッジのデータを用いて行う点にある。これは、fine-tuning と呼ばれる、別のデータセットで学習済みのパラメータを目的とするデータセットの学習のための初期値として利用する技術に着想を得ている[4]。

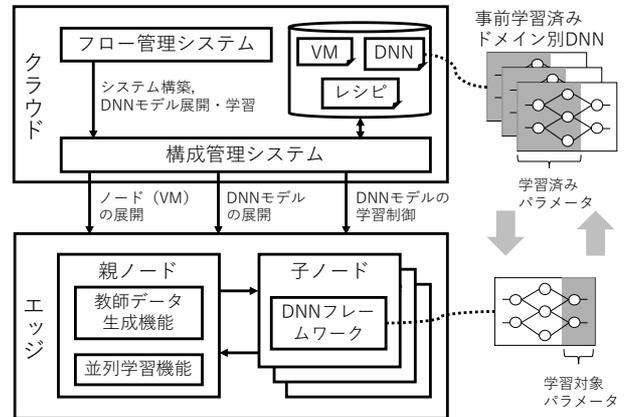


図 2 提案システム

この基本的なアイデアにより、計算機リソースを必要とする初期の学習をクラウドで実現できるため、エッジで必要とする計算機リソースを大幅に削減することが可能となる。また、クラウドでの学習には汎用的なデータセットを用いるためエッジのデータを必要とせず、クラウドへのデータ集約に伴う通信帯域の問題も回避される。

前述の DNN モデルの効率的な運用を実現するため、提案システムは以下の機能およびシステムを備える。

- (1) **フロー管理システム (@クラウド)**  
クラウドよりエッジのシステム構築、DNN モデルの展開・学習をフローとして管理する。
- (2) **構成管理システム (@クラウド)**  
計算ノードの展開、DNN モデルの展開・学習をフロー管理システムと連携して管理する。
- (3) **教師データ生成機能 (@エッジ)**  
DNN モデルの学習に必要な教師データを各加工処理 (画像であれば反転、コントラスト調整など) により自動で生成する。

Proposal and Evaluation of DNN Model Operation Method with Cloud/Edge Collaboration

†Junpei Okoshi, Naofumi Hama, Nobukazu Kondo, Hitachi, Ltd. Research & Development Group

#### (4) 並列学習機能 (@エッジ)

DNN フレームワークを並列動作させ、結果をアンサンブル学習 [5] により統合する。また、教師データ生成機能と連携し、各 DNN フレームワークには異なる教師データを与えることにより、各 DNN モデルの統計的独立性を向上させる。

(1) および (2) を備えることにより、エッジのシステム構築・運用をクラウドより一括管理可能となり、エッジ数が増加した際に顕在化するシステム構築・運用工数の増大を回避することが可能となる。また、(3) により工数を要する教師データの生成を一部自動化し、(4) によりエッジの計算機リソースを効率的に活用することが可能となる。

### 3 評価

本研究では、画像データ分析を想定し、入力画像を特定のクラスに分類することを目的としたプロトタイプを開発した。また、開発したシステムを用いて、(1) システム構築・運用に要する工数、(2) DNN モデルの学習に要する期間の2つの観点で提案手法を定量的に評価した。通信帯域の不足に関しては、提案システムにより学習データのクラウド集約が生じないため、定性的に問題が解決されている。表 1 に評価環境およびプロトタイプのシステム構成を示す。

表 1 評価環境およびシステム構成

CPU	Intel(R) Core i7-6700
Memory	16GB
OS	CentOS 7.3 1611
GPU	GeForce GTX 1080
DNN フレームワーク	Tensorflow-gpu 1.0.0 <sup>1</sup>
フロー管理システム	Pentaho DI 7.0 <sup>2</sup>
構成管理システム	Itamae 1.9.10 <sup>3</sup>
データセット	cifar10 <sup>4</sup>
DNN モデル	Alex Net[6]

(1) に関しては、クラウド：1 拠点、エッジ：10 拠点の想定で、実行されるコマンドや設定ファイルの書き換えに要するステップ数の削減効果を、人手による場合を基準に以下のシナリオにて評価した。その後、その削減率を用いて一般的な IT システムの構築に要する工数の削減効果を算出した。

#### 【シナリオ】

1. クラウドの環境構築 (OS, VM など) : 1 拠点
2. エッジの環境構築 (OS, VM など) : 10 拠点
3. 計算ノード・学習済み DNN モデルの展開
4. 学習済み DNN モデルの fine-tuning

ただし、4. の fine-tuning に関しては、学習の実行 (設定ファイルの書き換えやスクリプトの実行) のみが対象で、学習期間や試行錯誤による工数は含まれない。

(2) に関しては、シナリオの 4. において、DNN モデルの学習を開始してから収束するまでの期間の削減効果を、人手による試行錯誤を含む場合を基準に評価した。その後、その削減率を用いて一般的なデータ分析に要する工数の削減効果を算出した。

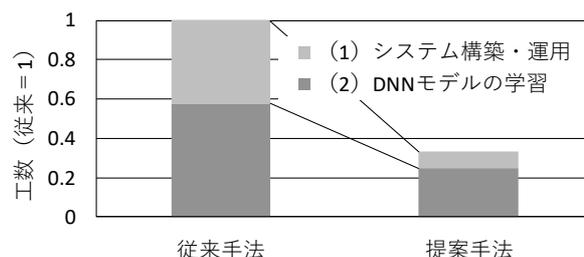


図 3 工数の削減効果

図 3 に、(1) システム構築・運用と (2) DNN モデルの学習に要する工数の削減効果 (従来手法の工数を 1 とした相対値) を示す。提案手法により、システム構築・運用の工数の 79% を、DNN モデルの学習の期間の 58% をそれぞれ削減し、全体の工数の 67% を削減できることがわかった。

### 4 結言

本研究では、クラウドとエッジの利点を活かした DNN モデルの運用方式を提案した。これにより、潤沢な計算機リソースを有するクラウドと大量データを有するエッジの特性を活かした DNN モデルの学習が可能となった。また、プロトタイプを用いた評価により、標準的なシステム構築・運用、および DNN モデルの学習を想定した環境で 67% の工数を削減可能であることを示した。

今後の課題として、実環境 (実データ) におけるフィージビリティ検証が挙げられる。

#### 参考文献

- [1] W. Shi et al., "Edge Computing: Vision and Challenges," in IEEE IoT Journal, 2016.
- [2] 総務省, "情報通信白書平成 28 年度版," 2016.
- [3] 近藤 伸和, "コンピューティング方式の進化と身体モデル," 日本経営システム学会 第 58 回 全国研究発表大会, 2017.
- [4] G. E. Hinton et al., "Reducing the Dimensionality of Data with Neural Networks," science, 2006.
- [5] T. G. Dietterich, "Ensemble Methods in Machine Learning," Multiple Classifier Systems, 2000.
- [6] A. Krizhevsky et al., "ImageNet Classification with Deep Convolutional Neural Networks," NIPS, 2012.

\*本書に掲載されている会社名・製品名は一般に各社の登録商標または商標です。

1 <https://www.tensorflow.org/>

2 <http://www.pentaho.com/>

3 <https://github.com/itamae-kitchen/itamae>

4 <https://www.cs.toronto.edu/~kriz/cifar.html>