

エッジコンピューティング環境を想定した非同期分散型深層学習の実現

NTTは、エッジコンピューティング上の機械学習を想定した非同期分散型深層学習技術 (Edge-consensus Learning) を実現しました。

現在の機械学習、特に深層学習では、1カ所（クラウド）にデータを集約し、画像・音声認識等のモデルを学習することが一般的です。しかし、あらゆるモノがネットワークに接続するIoT時代において、膨大なデータをクラウドに集約することは困難です。またプライバシー保護の観点で、データをローカルにあるサーバ・機器にとどめたいという需要も増加しています。関連して、EUの一般データ保護規則 (GDPR) のようなプライバシー保護のための法的規則も強化されつつあります。こうした時代において、データを蓄積・分析・処理するサーバを分散化し、上位システム（クラウド）や通信網の処理負荷を低減させ、応答速度やプライバシー保護の観点でユーザの利便性を高めるエッジコンピューティングへの期待が高まっています。

本研究の目的は、エッジコンピューティングのように分散配置されたサーバ群に分散してデータが蓄積されていく環境でも、あたかも1カ所にデータを集約して学習したかのようなグローバルモデルを得るための学習アル

ゴリズムを開発することです。今回開発した技術は、①統計的に非均一なデータがサーバ群に蓄積されていて、かつ②サーバ群がモデルに関連する変数を非同期に通信・交換していても、全部のデータを1カ所に集めて学習したのと同等のモデルを得られることを確認したという点で、学術性・実用性が共に高い学習アルゴリズムだといえます。

本成果は、2020年8月23日から開催されたアメリカ計算機学会 (ACM) 主催の国際会議KDD 2020 (Knowledge Discovery and Data Mining, 採択率16.9%) にて発表されました。また本成果についての多角的な検証を目的に、関連したコードをGithubにて公開しました (<https://github.com/nttclab/edge-consensus-learning>)。

■研究の背景

現在の機械学習、特に深層学習では、1カ所にデータを集約し、1カ所でモデルを学習するのが一般的です。しかし、データ量の激増やプライバシー保護の観点から、近い将来データは分散蓄積されるようになります。例えば、エッジコンピューティング構想では、データ蓄積や

- ・リング状ネットワーク (V=8)
- ・統計的に非均一なデータ (各ノード: 10クラス中5クラス)
- ・非同期通信 (ランダムなタイミングで変数を交換)

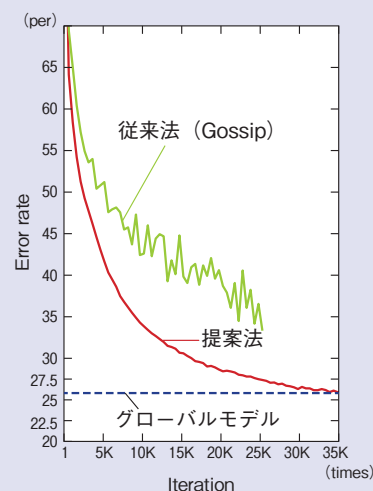
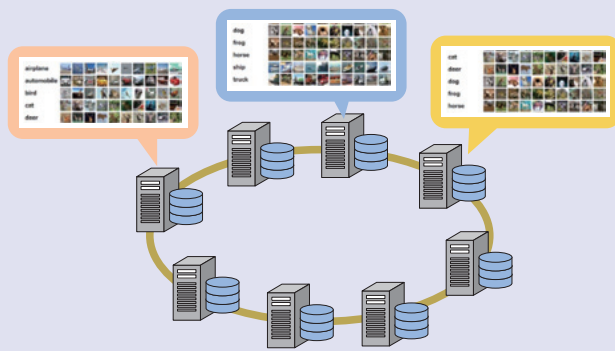


図 実験結果 (抜粋)

処理の負荷分散が提唱されていますし、GDPRでは、国をまたぐデータの移送に制限をかけていたり、最小限のデータ収集を要請する条項も存在します。

このような状況では、データプライバシー保護と機械学習によって得られる恩恵をトレードオフとみなすのではなく、むしろデータプライバシーを保護しつつ機械学習の恩恵を受けられる世界がより望ましいといえます。そのためには、性能を犠牲にすることなくデータ収集やモデルの学習を分散化することが、技術的な課題の1つでした。

■技術のポイント・特徴

今回開発した学習アルゴリズムは、複数のサーバに異なるデータが分散して蓄積される状況でも、サーバ間で合意形成されたモデルを得ることができます。データの代わりに、モデルに関連する変数をサーバ間で非同期に通信・交換することで、合意形成されたモデルを得ます。

このアルゴリズムの有効性を検証するために行ったシミュレーション実験の結果を、以下で簡単に紹介します(図)。8台のサーバがリング状に接続されたネットワークを想定します。テスト用の画像データセットとして、一般的に用いられる物体画像認識用のデータセット(CIFAR-10)を用いました。これは、計10個のクラス(航空機、自動車、鳥、猫など)に分類可能な大量の画像で構成されています。一方本実験では、各サーバ上に統計

的に非均一となるように画像を与えます。具体的には、各サーバにはそれぞれ10クラスのうち、ある5クラス分のデータのみを与えます(ただし、8台合わせるとすべてのクラスがほぼ均等に存在するデータセットとなります)。サーバ群が非同期に通信する状況をシミュレートした結果、提案法を用いると、あたかも1カ所にデータを集約して学習したかのようなモデル(グローバルモデル)が得られることを確認しました。

■今後の展開

エッジコンピューティングを活用した大規模なAI応用が期待される分野での実用化をめざして、パートナーと連携しながら研究開発や実証実験を今後も継続していきます。コード公開を通じて、本技術のさらなる発展、アプリケーションに関するコラボレーションを促進していきます。

◆問い合わせ先

NTT先端技術総合研究所

広報担当

TEL 046-240-5157

E-mail science_coretech-pr-ml@hco.ntt.co.jp

URL <https://www.ntt.co.jp/news2020/2008/200824a.html>

エッジ端末/IoT機器が相互に通信し、自律的に知識を創発するシステムの実現に向けて

研究者紹介

丹羽 健太

NTTコミュニケーション科学基礎研究所

協創情報研究部 知能創発環境研究グループ 主任研究員

電話といえば、遠隔地にいる人と人のコミュニケーションをとるための手段です。しかし、これからの時代、エッジ端末/IoT機器といった自分の身の回りにあるモノたちが自律的に通信を介して高度な知識を学習したり、マスターとなるユーザに有用な情報を提供したりする高度AIoT(モノの人工知能)時代が訪れようとしています。私の研究は、そうしたモノたちが、自律的にモデルを学習するためのアルゴリズム数理を拘置することとそのソフトウェア実装です。

2020年8月に発表させていただいた「非同期分散型深層学習技術」は、基本的な機能の1つとして、分散したノードたちが合意形成をしながら、高品質なモデル(例えば画像認識モデル)を学習するアルゴリズムです。あたかもすべてのノードの中にあるデータを1カ所に集約したかのような深層学習モデルが得られます。人が互いに協調しながら成長を促し、最後には自分1人で成し得なかったような高度な知識を得ることに似ています。

本プロジェクトのソフトウェア実装には、何名かのエンジニアに携わっていただいています。今ではチームとして成熟しつつあります。最初のほうは、分かり合えないことや、つまづくことも多かったです。しかし、現在では、私が具体的な指示をすることなく、チームが自律的に機能し、仕事がスムーズに進んでいます。こうした協調・成長・自律といったことが現在の機械学習ではまだまだ表現されていません。自分の研究目標は、エッジ端末/IoT機器が相互に通信し、自律的に知識を創発するシステムを実現することです。

