論 文 内 容 の 要 旨

博士論文題目

An Infrastructure for Collaborative Machine Learning on Resource- Constrained Heterogeneous Environments
（資源制約を有する異種混合環境のための協調的機械学習基盤）

氏　　名

THONGLEK KUNDJANASITH

（論文内容の要旨）

ソフトウェア産業の急速かつ成功的な成長には、コラボレーションが不可欠である。ソースコードにおける GitHub やコンテナイメージにおける DockHub のようなソフトウェア開発基盤においては、多様なバックグラウンドや組織を持つ個人が協力し、大手テック企業でさえ開発や維持が困難と感じる複雑で大規模なソフトウェアの構築を可能としている。しかし、機械学習モデルについては、このような連携インフラはまだ存在しない。そのため、本研究では、データプライバシーの制限や既存のリソース制約などの障壁を機械学習モデルの共同開発から取り除くためにの基盤である LiberatAI を提案する。

LiberatAI は、本研究が提案する、機械学習モデルを共同開発するためのインフラストラクチャであり、研究者が協力することで単独の企業がよりも優れたモデルを構築できる可能性を提供する。LiberatAI は、データのプライバシーを守りながらモデルを訓練するために連合学習を利用しており、複数開発者が、様々に異なる各自の環境において、共同でモデルを訓練することができる。LiberatAI では以下に示すの 3 つのモジュールにより、多様なストレージ、コンピューティング、通信リソース上でのモデルの学習をサポートする。

(1) Compressor モジュールは、環境のストレージ容量に合わせてモデルサイズを縮小する。

(2) Aggregator モジュールは、異種計算資源で学習されたモデルを集約する。

3) Sparsifier モジュールは、サーバとクライアント間でモデルを交換するために、モデルをスパース化する。

評価実験として、LiberatAI において、胸部 X 線画像から COVID-19 症例を検出するための最先端のニューラルネットワークモデルの構築を行なった。その結果、LiberatAI の 6 つの異なるハードウェア環境で構築された異る構造を持つアンサンブルモデルは、単一環境下で訓練された COVID-NET よりも 5.39%高い精度を得ることができた。

氏　名	THONGLEK KUNDJANASITH

（論文審査結果の要旨）

複数の計算機を用いた分散機械学習は現在活発に研究が行われている領域であり、訓練データのプライバシーを確保可能とする連合学習（Federated Learning）等の技術開発が進められている。連合学習はデータの多様性を確保するために、多様なユーザが訓練データを直接交換することなく協調して機械学習モデルを構築可能とする技術であるが、多様なユーザを受け入れるに当たって考慮すべき、各ユーザ側における利用可能な計算資源の異質性を考慮した設計になっていない。そのため、連合学習を用いて多様なユーザによる協調機械学習システムの構築は現実的には困難である。そこで、本論文は、ユーザ側で利用可能な計算機資源の制約に応じて、利用する資源量を調節可能な協調機械学習プラットホームである LiberatAI を構築した。具体的には、LiberatAI では、多様なストレージ、コンピューティング、通信リソース上でのモデルの学習をサポートするために、1）機械学習モデルを圧縮する Compressor モジュール、2）計算量が異なる複数のモデルを統合可能とする Aggregator モジュール、3）連合学習におけるサーバ・クライアント間で交換されるモデルをスパース化することで通信量を削減する Sparsifier モジュールを提案した。また、評価実験として、プライバシー性が高いデータに基づいた機械学習アプリケーションのシナリオを想定し、胸部 X 線画像から COVID-19 症例を検出するための機械学習モデルの構築を LiberatAI 上で行なった。その結果、LiberatAI 上において 6 つの異なるハードウェア環境で構築されたモデルを統合した機械学習モデルは、単一環境下で訓練された COVID-NET よりも 5.39%高い精度を達成することができた。

このように、本論文は、連合学習をより現実的で多様な計算機環境が混在する環境においても実行可能な手法に拡張するために、様々なリソース制約に対処する技術を研究開発し、その有効性を定量的に示している。したがって、本論文は博士（工学）の学位論文として認めるに値すると判断する。