Ⅱ 医療AIを加速させる研究開発の動向

2. 連合学習による医療AIの開発

阮（ゲン） 佩穎（ハイエイ） エヌビディア（同）シニアディープラーニングソリューションアーキテクト

背景：なぜ医療AIに連合学習（FL）が必要か？

人工知能（AI）の研究，特に機械学習と深層学習の進歩は，放射線医学，解剖学，病理学，ゲノミクス，創薬，その他の分野における破壊的イノベーションにつながっている。機械学習は，現代の医療システムが収集する膨大な医療データから，正確かつ堅牢な統計モデルを構築する有望なアプローチとして登場した。最新の深層学習モデルは，臨床グレードの精度を達成するために，十分に大きなキュレーションされたデータセットから学習する必要がある。しかし，大量の医療データへのアクセスは，プライバシーの問題で困難である。

医療データは非常に機密性が高く，その使用は厳しく規制されている。たとえ，データの匿名化がこれらの制限を回避できたとしても，患者名や生年月日などのメタデータを削除するだけでは，プライバシーを保護するのに十分ではない場合が多いことが，現在では通念となっている。例えば，CTやMRIのデータから患者の顔を復元することが可能である。

医療分野でのデータ共有が組織的に行われていないもう一つの理由は，高品質のデータセットを収集，管理，維持するにはかなりの時間，労力，費用がかかるからである。その結果，そのようなデータセットには大きな商用価値があり，自由に共有される可能性は低くなる。

十分なデータへのアクセスがなければ，機械学習はその潜在能力を十分に発揮することができず，最終的には研究から臨床への移行に失敗してしまう可能性が高い。本稿では，この課題を解決するための技術である連合学習（以下，FL）と医療分野での応用，さらに，その技術動向と課題について説明する。

連合学習（FL）とは？

FLは，データそのものを共有せずにアルゴリズムを協調的に学習することにより，データ管理とプライバシーの問題に対処しようとする学習パラダイムである。もともとはモバイルやエッジデバイスのユースケースなど，異なるドメイン向けに開発されたものだが，最近では医療アプリケーションでも注目されている。FLは，患者データを医療機関の外に出すことなく，学習プロセスは各参加機関でローカルに行われ，**図1 a**に示すように，モデルのパラメータ，勾配などのみが転送される。結果として，安全性と高精度モデル作成の両方を実現する効率的なプライバシー保護ソリューションとなる。最近の研究では，**図1 b**に示すように，FLによって学習されたモデルは，中央集中型データで学習されたモデルと同等の性能を達成し，単一機関のデータのみを参照するモデルよりも優れていることが示されている。

図2に示すように，FLワークフローはさまざまなトポロジーで実現することができる。医療アプリケーションで最も一般的なものは，クライアントサーバを介したもの（**図2 c**）と，ピアツーピアアプローチ（b）の2つである。いずれの

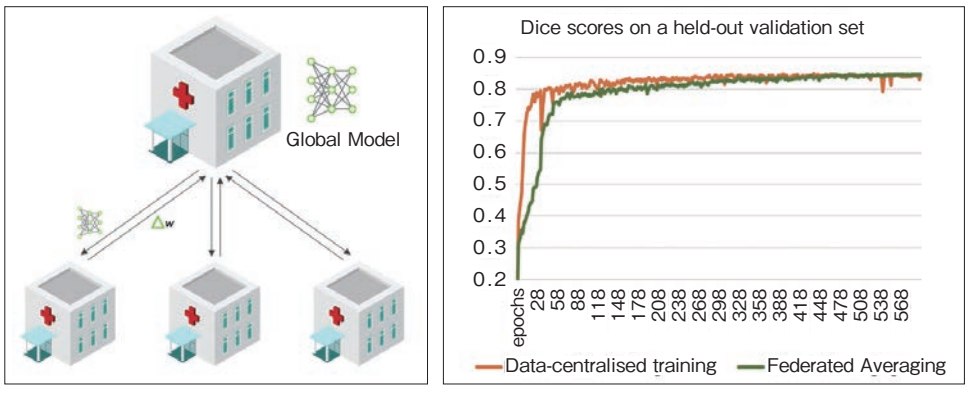

図1 連合学習のワークフロー（a），中央集中型データのトレーニングと連合学習性能の比較（b）

〈0913-8919/22/￥300/ 論文 /JCOPY〉