

データを秘匿したまま連携できる 連合学習技術とLLMへの適用可能性

荒木 俊則

要 旨

深層学習に基づくAIの発展は目覚ましく、大規模言語モデル（LLM）は人間との自然なやり取りすら可能です。しかし、このようなAIを作成するためには多くのデータが必要であり、データ確保は非常に重要な課題です。連合学習は、その課題の一部である複数組織に散らばっている機密性の高いデータの活用を助ける技術です。本稿では、基本的な連合学習方式3種類（水平、垂直、転移）を解説したうえで、近年の注目の高いLLMなどの生成AIに対する連合学習の適用可能性について述べます。



連合学習／データ連携／異業種連携／LLM／ファインチューニング

1. はじめに

近年の生成AIの発展は目覚ましく、2020年に発表された大規模言語モデル（Large Language Model、以下、LLM）GPT-3は人間との自然なやり取りすら可能とし、その後も急速に発展を続けています。このようにAIに高度な機能を持たせるためには大量のデータが必要となるため、データ確保はAI開発の最も重要な課題の1つです。

本稿で紹介する連合学習は、データ確保に関する問題の一部を解決し、複数の組織に散らばったデータを効果的に活用可能とします。単にデータを1箇所に集めるだけでは、提供したデータがAIの学習以外にも使われる可能性や、個人情報の漏えいとなる可能性があり、データの機密性やプライバシーは確保できません。しかし、連合学習を用いれば、データ自体を集めることなく分散したデータのすべてをAIの作成に利用することができます。

本稿では、さまざまなシーンに向けた3種類の連合学習を紹介し、先進的なトピックとしてLLMのような生成系AIへの連合学習の適用可能性について述べます。

2. 連合学習とは

連合学習は、2017年にGoogle¹⁾によって提唱されたAIの学習技術です。これを用いれば、複数箇所に散らばったデータを集積することなく、すべてのデータを用いたAIの学習ができます。

図1は基本的な連合学習の枠組みで、AIを学習したい参加者3人が各々の保有するデータを用いて1つのグローバルモデルを学習する様子を表しています。中央サーバを介してグローバルモデルを作成していることが、特徴的です。

学習は、グローバルモデルを共有する参加者で次の

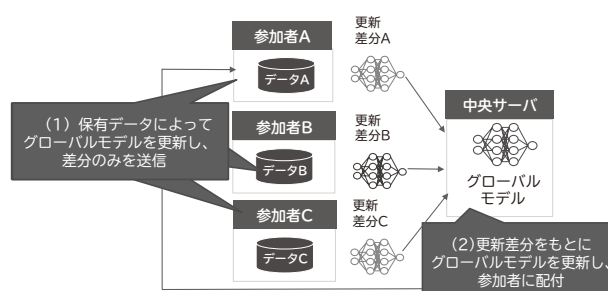


図1 基本的な連合学習の枠組み

(1)、(2) を繰り返して行います。

(1) 各参加者は、それぞれが保有するデータを用いてグローバルモデルを更新し、その更新差分を中央サーバへ送付します。

(2) 中央サーバは、受け取った更新差分をもとにグローバルモデルを更新して、参加者へ配付します。中央サーバがグローバルモデルを更新する方法はさまざまですが、例えば更新差分の平均値でグローバルモデルを更新します。

前述した方法においては、データを一度も中央サーバへ集めることなく、参加者すべてのデータを反映したグローバルモデルを作ることができます。こうして作られたグローバルモデルには、それぞれの参加者が保有するデータだけで学習するよりも高い性能が期待できます。

また、この方法では中央サーバがグローバルモデルを入手してしまいがちですが、これを秘匿したい場合には「秘密計算」²⁾と呼ばれるデータを暗号化した状態で計算できる技術が有効です。

2.1 水平連合学習

第2章の冒頭で説明したような、すべての参加者が同じ形式のデータを所有して同じ用途のAIを学習する方法は、「水平連合学習」と呼ばれます。同様のデータを保有する参加者が同機能のAIを学習するケースで用いられ、例えば、金融機関が連携して不正取引検知AIを開発するようなケースが当てはまります(図2)。

他にも、参加者間で異なる種類のデータを保有する場合に対応した連合学習も開発されています。代表例として、「垂直連合学習」と「転移連合学習」を次に紹介します。

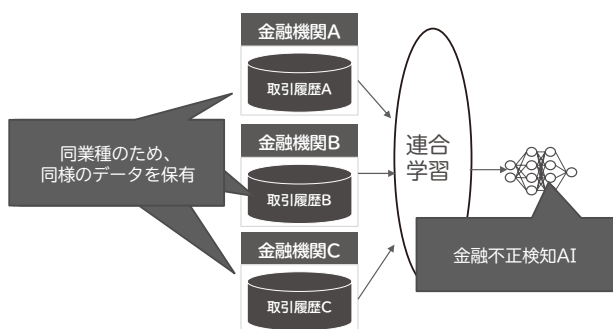


図2 水平連合学習の例

2.2 垂直連合学習

第2章2節では、参加者間で保有するデータの種類の異なるが、同一の標本(ユーザーなど)に関する情報である場合に有効な「垂直連合学習」を紹介します。

例えば、カード会社はユーザーの収入、資産などを保有し、ECサイトはユーザーの閲覧履歴や嗜好に関する情報を保有しているといったケースです。ユーザーごとの情報を突合すれば各ユーザーに関してリッチな情報となるので、両方のデータを使って予測するAIには単独で開発する場合に比べて高い性能が見込めます。

垂直連合学習の「学習」の特徴は、その学習過程において参加者がAIを分割して保有し、部分的に計算する点です。各参加者はそれぞれが保有するデータを入力してAIの計算の一部を行い、部分計算結果を受け取った中央サーバが残りを計算します(図3、順方向の→)。その後、結果を逆向きにフィードバックしてAIのパラメータを更新していきますが、その過程も分散して行います(図3、逆方向の←)。

このような方法で学習されたAIモデルは、カード会社とECサイトで共用することになりますが、カード会社やECサイトは単独で予測に必要なデータを揃えられないため、AIを利用する際には協力が必要となります。また、カード会社とECサイト間でデータの対応が取れていないと意味のある学習ができません。不要な情報共有を避けつつ対応だけを取るためには、秘密計算が有効です。

2.3 転移連合学習

第2章3節では「転移連合学習」を紹介します。垂直連合学習は、特定ユーザーに関して保有するデータが参加者

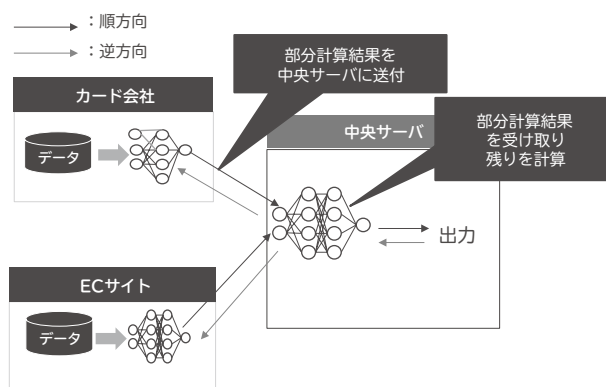


図3 垂直連合学習の概要

間ですべて異なるという設定でしたが、転移連合学習では保有データのいくつかが重複しているケースで有効な技術です。この技術は共通するデータを糊として用いて、あるドメイン向けに作られたAIを別のドメイン向けに転移させることができます。異業種連携に有用とされており、保険会社と不動産会社の連携ならば、保険会社の顧客から不動産の見込み顧客を発見して送客に活用することができます。逆に、不動産会社から保険会社への送客も可能です(図4)。

転移連合学習の「学習」は、2つのステップからなります。最初のステップでは、それぞれが保有するデータを揃える前処理を行います。具体的には、共通して保有する属性から、個別に保有しているデータを補うAIを連携して学習し交換します(図5)。

次に、交換したAIを用いて各自が保有していないデー

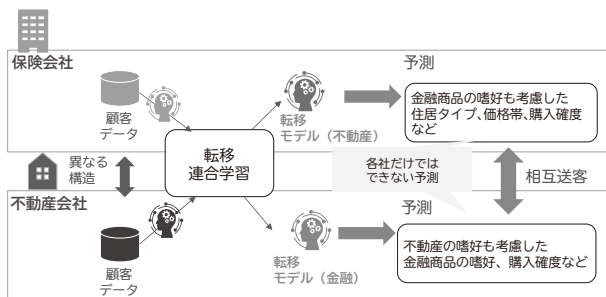


図4 転移連合学習の利用イメージ

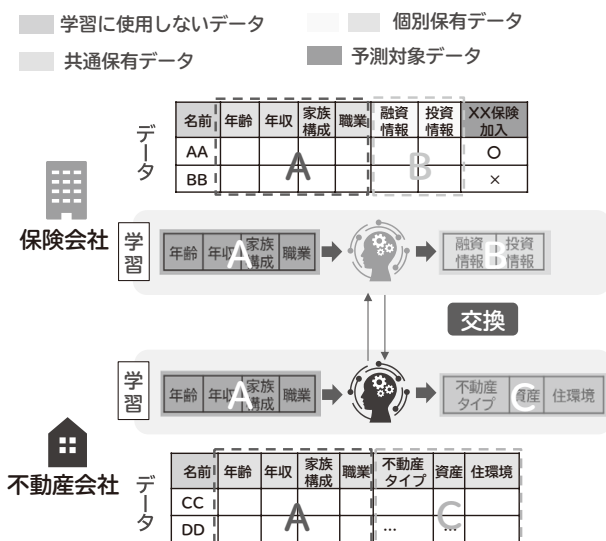


図5 保有データを揃える前処理

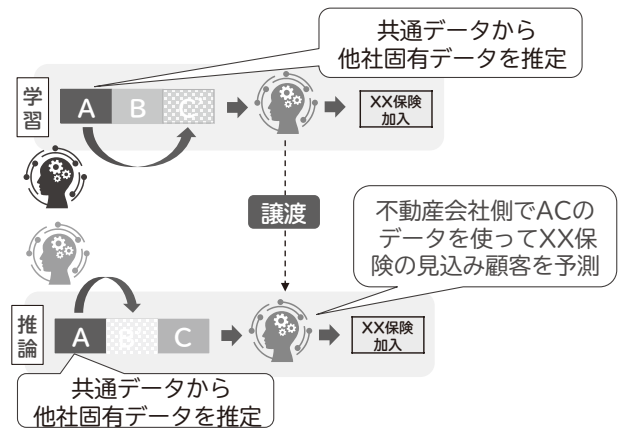


図6 揃えたデータを用いたAI学習と予測

タをAIの予測に基づいて埋めます。その後、予測対象のデータ（XX保険への加入）を保有する側（保険会社）がAIを学習して、不動産会社に譲渡します。不動産会社は、譲渡されたAIを用いて自社の顧客のなかからXX保険への興味を評価します。評価結果をもとにDMなどを用いて見込み顧客に対してコンタクトすれば、保険会社への送客が実現できます。(図6)。

前述した例では、予測対象であるXX保険に関する加入状況の正例（加入者）・負例（非加入者）の両方が手に入ることを前提としています。しかし、例えば、保険自体への興味を測りたい場合、保険会社の保有するデータは少なくとも問い合わせや来店があった人のデータとなり、正例（興味のある人のデータ）と言えます。負例がない場合、AIの学習はより困難になりますが、NECでは正例だけしかない場合にも有効な転移連合学習方式³⁾を開発しています。

3. 連合学習の生成AIへの適用可能性

第3章では、今後の課題として急激に多くのシーンで利用されるようになった生成AIにおける連合学習の活用可能性とその課題について述べます。

3.1 水平連合学習の生成AIへの適用可能性

生成AIの一種であるLLMは、利用者が各自の保有するデータによって調整（ファインチューニング、以下、FC）してから活用されるケースが増え、さまざまな箇所にFC済みのLLMが散らばっています(図7)。散在する複数の

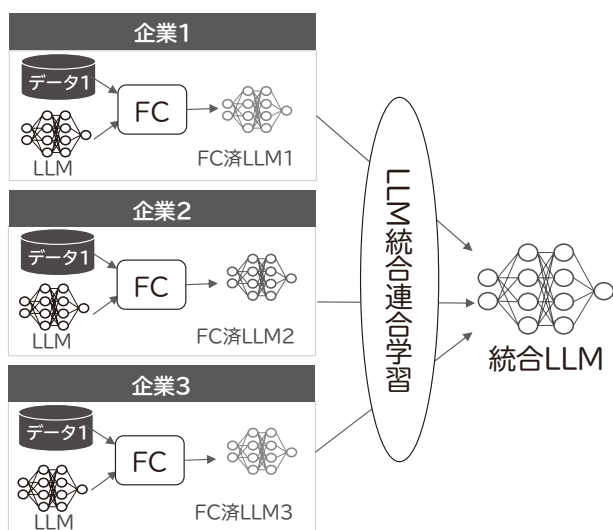


図7 FC済みLLMの統合

LLMを連合すれば、更に高性能のLLMが得られる可能性があります。FC済みLLMや追加の訓練データは企業秘密となって容易に共有できない可能性があります。ここで連合学習の知見を活用すれば、FC済みのLLMやその訓練データを秘匿したまま統合LLMを生成できると考えます。どうすれば効果的に多くの情報を統合LLMに集積できるか、どれだけ効率的（通信量や計算量）に統合できるかといった点に技術的課題がありそうですが、更なるLLMの活用範囲の拡大に水平連合学習は貢献できると考えられます。

3.2 転移連合学習生成AIへの適用可能性

LLMのFCに付随して、また別の課題があります。LLMサービスを提供する会社が独自のLLMを保有し、LLMをFCしてから利用したいユーザー企業が追加の訓練データを保有している状況を考えます。このとき、LLMサービス提供企業は独自LLMを秘匿し、ユーザー企業は追加の訓練データに含まれた機密情報を秘匿したいケースもあり得ます（図8）。

企業の競争力の源となるような独自LLMを他社へ開示できない場合や、ユーザー企業が生の機密情報は他社に秘匿したい場合が想定されます。このような設定でユーザー企業が自身の保有する訓練データをLLMに記憶（転移）させるために、転移連合学習の知見が利用できると考えます。どこまで訓練データを秘匿できるか、どこまで効

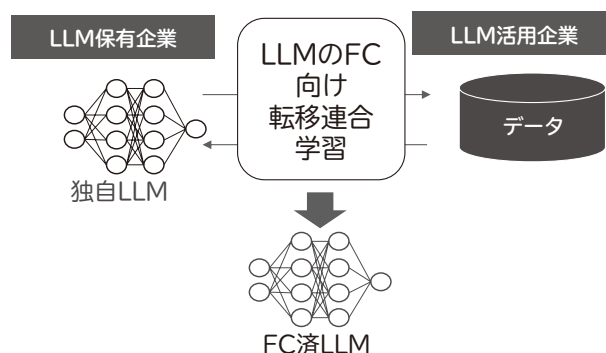


図8 FC向け転移連合学習

率的（通信量や必要となるGPUなどの計算資源の量）にFCできるかといった点に、技術的課題があります。

4. むすび

本稿では、主要な連合学習方式（水平・垂直・転移）の概要と特徴を解説し、近年の発展著しい生成AI領域への連合学習の適用可能性について述べました。

* Googleは、Google LLC.の商標です。

* その他記述された社名、製品名などは、該当する各社の商標または登録商標です。

参考文献

- 1) H.Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas: Communication-Efficient Learning of Deep Networks from Decentralized Data, PMLR 54, pp.1273-1282, 2017
<https://arxiv.org/abs/1602.05629>
- 2) NEC：高秘匿連合学習 プライバシー・機密情報保護とAI活用を両立, 2021.9
<https://jpn.nec.com/rd/special/202103/index.html>
- 3) Junki Mori, Ryo Furukawa, Isamu Teranishi and Jun Sakuma: Heterogeneous Domain Adaptation with Positive and Unlabeled Data, IEEE BigData, 2023
<https://arxiv.org/abs/2304.07955>

執筆者プロフィール

荒木 俊則

セキュアシステムプラットフォーム
研究所
ディレクター

NEC 技報のご案内

NEC 技報の論文をご覧くださいありがとうございます。
ご興味がありましたら、関連する他の論文もご一読ください。

NEC技報WEBサイトはこちら

NEC技報（日本語）

NEC Technical Journal（英語）

Vol.75 No.2 ビジネスの常識を変える生成AI特集 ～社会実装に向けた取り組みと、それを支える生成AI技術～

ビジネスの常識を変える生成AI特集によせて
生成AI技術への取り組み ～基盤から応用、ルール作りまで～

◇ 特集論文

急速に広まる生成AIの市場適用

NECにおける生成AIの取り組みについて
電子カルテと医療文書の作成支援による医師業務効率化
映像×LLMによる報告書作成業務の自動化
映像分析と生成AIによるリアル世界の行動理解
サイバー脅威インテリジェンス生成自動化
生成AIの社内活用を進めるNEC Generative AI Service (NGS)
ソフトウェア・システム開発への生成AIの活用
LLMとMIで革新する素材開発プラットフォーム
LLMと画像分析を活用した被災状況の把握

生成AIの可能性を高める基盤技術

日本語性能に優れたNECの大規模言語モデル (LLM)
NECの生成AIを支える国内企業で最大規模のAIスーパーコンピュータ
より安全な大規模言語モデル (LLM) を目指して
データを秘匿したまま連携できる連合学習技術とLLMへの適用可能性
大規模言語モデル (LLM) によるFew-shotクラスタリングの可能性
オープンドメイン常識推論のための知識拡張型プロンプト学習
AI連携とオーケストレーションのための基盤ビジョンLLM
クエリを考慮した新規手法により関連する企業データを減らしLLM APIの使用コストを最適化

AI技術が社会へ浸透するために

AI標準化・ルールメイクに関する動向とNECの取り組み
人権尊重に向けたNECのAIガバナンスの取り組み
RCModelを用いたAIリスクマネジメントのための人材育成事例

◇ NEC Information

2023年度C&C賞表彰式典開催



Vol.75 No.2
(2024年3月)

特集TOP