

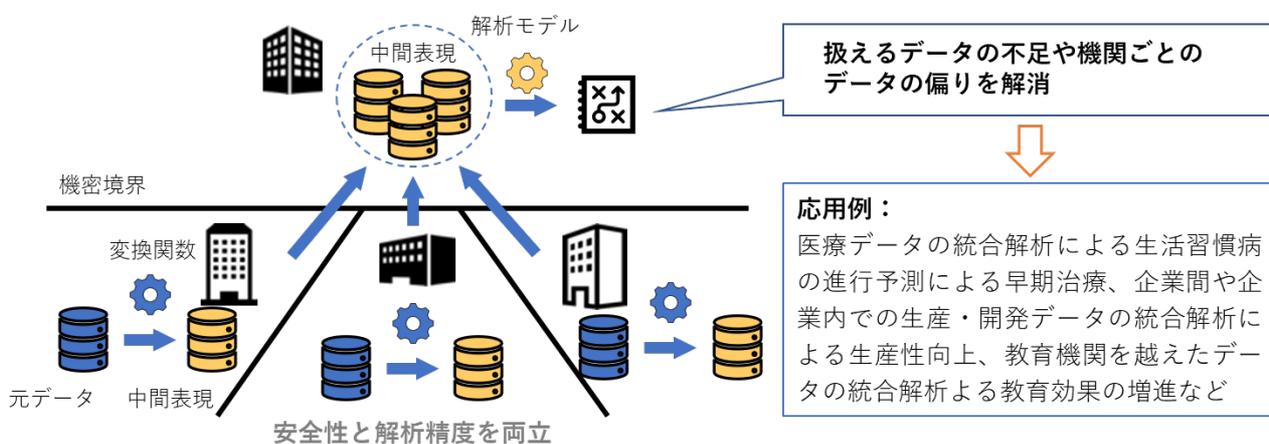
複数の企業・機関が保有するデータを統合解析できる AI 技術を開発 — 秘匿データの安全性担保と解析精度の向上を実現 —

NEDOと筑波大学 システム情報系 櫻井鉄也教授らは、複数の企業・機関がそれぞれ保有するデータに対して、互いにアクセスせずに統合的に解析する人工知能(AI)技術「データコラボレーション解析」を開発しました。

AIによる解析の精度を上げるには十分な数のデータを集めることが必要です。本技術は、各企業・機関が保有するプライバシー情報などを含む元データの代わりに、元データをAI技術により変換した「中間表現データ」のみを共有する仕組みです。これにより、元データに含まれる秘匿性の高い情報の安全性を担保しつつ、多数のデータの取り扱いが可能となることでAIの解析精度の大幅な向上を実現します。

具体的な応用例として、医療分野における生活習慣病データの解析による疾患予測をはじめ、企業間や企業内での生産・開発データ解析による生産性向上、複数教育機関による学生データの統合解析に基づく教育効果増進があります。また、将来的には、様々な機関にある質の高いデータを、匿名性を維持したまま収集し、AIによる分析を行う新たなプラットフォームの確立も期待されます。

なお、本技術は8月10日からマカオで開催されるAI分野のトップカンファレンスである「IJCAI-19」で発表されます。



1. 概要

近年、データ収集・蓄積が容易になったことで、さまざまな企業・機関が生産性向上や医療分野での早期治療などに向けて、独自でデータを蓄積し、人工知能(AI)による解析に取り組んでいます。ここで、AIによる解析の精度を上げるには十分な数のデータを集めることが必要となります。それを効率よく推進するために、これら複数の企業・機関が保持するデータを解析する方法として、単独の企業・機関のデータのみを用いる「個別解析」と、各者の元データを1カ所に集める「集中解析」があります。ただ、個別解析はデータ数の不足や分布の偏りによる解析性能の

低さが課題となっている一方、集中解析も医療データなどプライバシー情報を含んだデータに対して行うことは困難でした。そのため、元データを共有せずに企業・機関の垣根を越えた統合的な解析が行える技術が必要とされています。

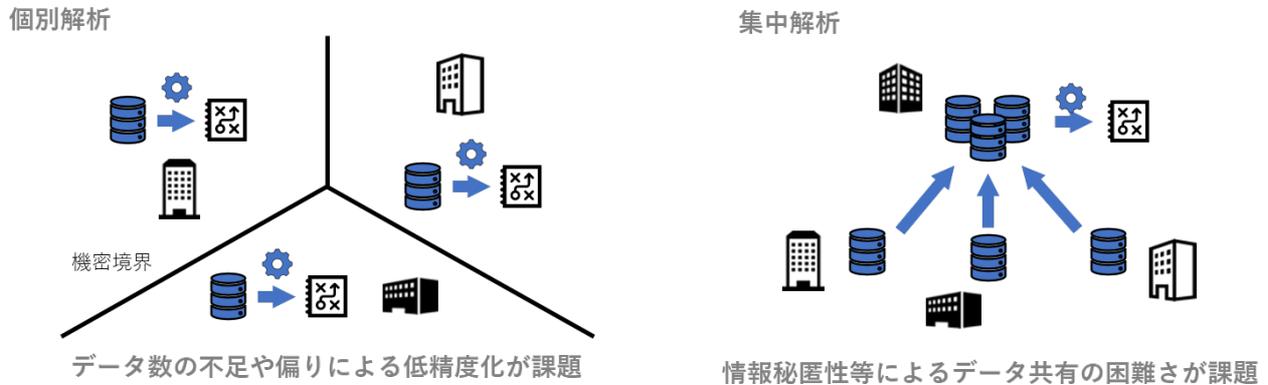


図2 個別解析(左)と集中解析(右)のイメージ

そこで、国立研究開発法人新エネルギー・産業技術総合開発機構(NEDO)と国立大学法人筑波大学 システム情報系 櫻井鉄也教授らは2018年度から、企業・機関が保有する各データの統合解析技術の研究開発^{※1}に取り組んでいます。

そして今般、各企業・機関が分散して保有する元データに対して、互いにアクセス(共有)することなく統合的に解析するAI技術「データコラボレーション解析」を開発しました。本技術は、各企業・機関が保有するプライバシー情報などを含む元データの代わりに、元データをAI技術により変換した「中間表現データ^{※2}」のみを共有する仕組みです。これにより、元データに含まれる秘匿性の高い情報の安全性を担保しつつ、多数のデータの取り扱いが可能となることでAIの解析精度の大幅な向上を実現します。

医療データや企業データ、教育データなどの秘匿性の高い情報を含むデータは、複数の企業・機関で共有することが情報セキュリティの理由から困難であり、これまでは単独の企業・機関での解析が行われてきました。これに対して、本技術では、元データを共有することなく統合的に解析を行うことで、各企業・機関の持つデータ分布の偏りなどによりこれまででは得られなかったような解析結果の抽出と、解析性能の飛躍的な上昇が期待されます。具体的な応用先としては、医療データ解析では生活習慣病の進行予測による早期治療やメンタル疾患の予兆発見・早期対策、企業データ解析では企業間や企業内での生産・開発データ統合解析による生産性向上、教育データ解析では教育機関を超えた学生データの統合解析による教育効果の増進などが挙げられます。また、将来的には、様々な機関にある質の高いデータを、匿名性を維持したまま収集し、AIによる分析を行う新たなプラットフォームの確立も期待されます。

現在、本技術の実用化に向け、筑波大学附属病院が保有するデータの解析、および学校法人聖路加国際大学と株式会社島津製作所から得られるデータの解析を行うための協議を進めています。また、医療以外の分野での解析を目指して、様々な業界へ適用を呼び掛けていく予定です。

なお、本技術は8月10日からマカオで開催されるAI分野のトップカンファレンスである「IJCAI-19^{※3}」で発表されます。

2. データコラボレーション解析の詳細

開発したデータコラボレーション解析は、プライバシー情報などを含むデータをそのまま共有する代わりに、各企業・機関が独自の変換関数によって元のデータに戻すことができない形式に変換(不可逆変換)し、「中間表現データ」として共有します。共有された中間表現データは各機関がそれぞれの変換関数で変換したデータであり、単純には統合解析ができません。この解決策として、各企業・機関で、共有可能なデータ(アンカーデータ^{※4})と元データのそれぞれの中間表現を構築し、アンカーデータを目印に各企業・機関の中間表現を統合可能なデータコラボレーション形式に変換することで統合解析を実現します。また、中間表現データは各機関による元データの推測が困難な不可逆変換で生成されたものであり、その変換関数も共有しないことから、元データに含まれるプライバシー情報などの安全性が担保されます。

本技術に基づくデータ統合解析は、具体的に以下のステップで行われます。

- (1) 企業・機関ごとに元データとアンカーデータを一緒に中間表現を作成
- (2) 作成した中間表現を解析機関のサーバで共有
- (3) 中間表現をデータコラボレーション形式に変換し、統合
- (4) 統合データをAI技術で解析した学習済みモデルを構築
- (5) 解析機関から学習済みモデルを受領
- (6) 学習済みモデルを病状診断や薬効予測、商品開発・生産などに活用

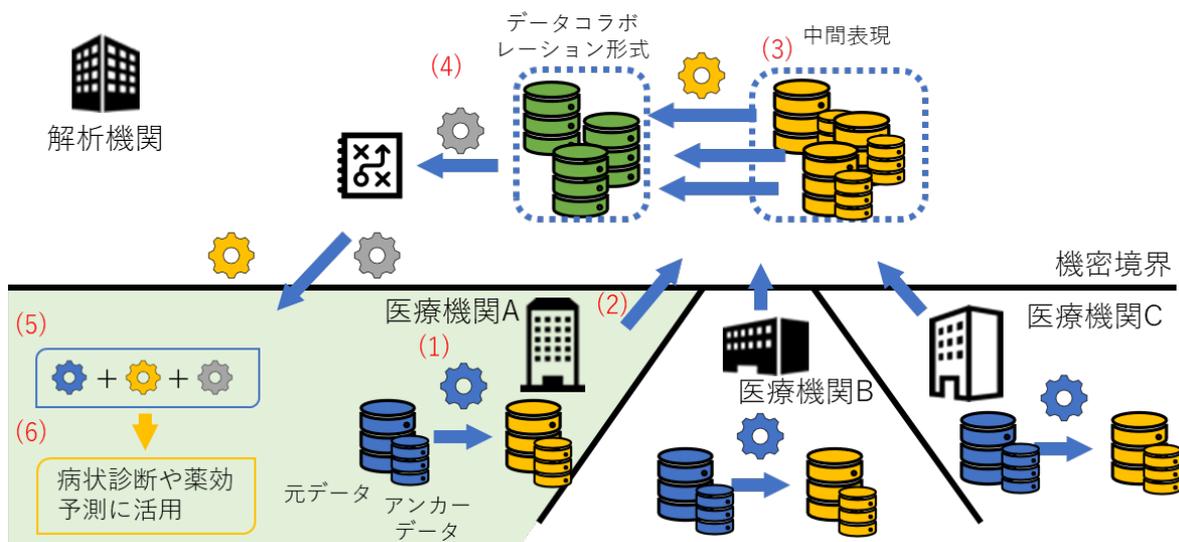


図3 複数の医療機関によるデータコラボレーション解析の手順例

【注釈】

※1 データの統合解析技術の研究開発

事業名：次世代人工知能・ロボット中核技術開発／次世代人工知能技術の日米共同研究開発研究開発／データコラボレーション解析による生産性向上を目指した次世代人工知能技術の研究開発

実施期間：2018年度～2019年度

※2 中間表現データ

各機関の持つデータから元データの推測が困難な不可逆変換で生成されたデータ。本解析技術では、元データではなく各機関で独立に生成した中間表現を共有する。

※3 IJCAI-19

The 28th International Joint Conference on Artificial Intelligence

<https://ijcai19.org/>

※4 アンカーデータ

各機関で独立に生成された中間表現を統合可能なデータコラボレーション形式に変換するための目印となるデータ。各機関で共有可能なダミーデータを利用する。

3. 問い合わせ先

(本ニュースリリースの内容についての問い合わせ先)

NEDO ロボット・AI部 担当: 渡邊、鈴木 TEL: 044-520-5242

筑波大学 システム情報系 担当: 教授 櫻井鉄也 TEL: 029-853-5501

(その他NEDO事業についての一般的な問い合わせ先)

NEDO 広報部 担当: 佐藤、中里、坂本 TEL: 044-520-5151 E-mail: nedo_press@ml.nedo.go.jp