4 パーソナルデータの安全な利活用

パーソナルデータのプライバシー保護と安全な利活用に向けた秘密計算技術の展開

NTT セキュアプラットフォーム研究所（以下、SC 研）では、プライバシーを保護しつつパーソナルデータなどの利活用を実現する秘密計算技術・匿名化技術の実社会での利用に向けて、医療データを用いた実証実験などに取り組んでいる。以下では、秘密計算技術の最近の研究状況について述べる。

――データを暗号化したままで計算・分析できる技術

昨今、さまざまな分野のデジタルトランスフォーメーションにより、分野横断的なデータの利活用がイノベーションを促進し、経済成長などさまざまな分野の発展につながることが期待されている。

一方、データの管理に伴うインシデントリスクや社会的責任の大きさ、企業戦略などの保護の観点によるデータのセキュリティ対策の必要性などがデータ利活用促進を阻害する要因となっている。

SC 研では、そのような要因の解消に貢献するため、世界に先駆けて、データを暗号化したまま分析可能な秘密計算技術の研究開発に取り組んできた。秘密計算技術の利点は、計算対象の元データをデータ提供者やサーバ運用者に一切見せることなく、分析の結果だけを得られることにある（図1）。

これにより、これまで競合する同業他社や異業種企業など他組織に開示することが難しかったデータを持ち寄った新しい統合分析を、データ漏洩の心配なく安全に実現できる。

組織横断的なデータ分析が実現すると、複数の組織や業界をまたがるサプライチェーンや顧客データの有効活用が促進され、今まで得られなかった知識発見など、新たな社会的価値の創出が期待される。

NTT セキュアプラットフォーム研究所
セキュアデータ流通基盤プロジェクト
（左）主任研究員　**高橋 誠治**氏
（右）主任研究員　**廣田 啓一**氏

SC 研では、これまで医療分野をはじめ、さまざまな分野への技術適用実験で知見を蓄えつつ、演算機能の充実や高速化などの改良を加え、秘密計算システム「算師®」の開発を進めてきた。

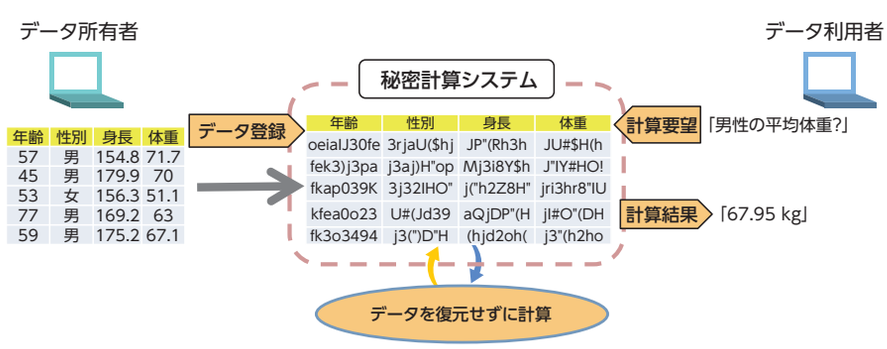

図１　秘密計算の概要

――SC研の秘密計算に関するR&Dの状況

SC 研では秘密計算技術について、研究レベルでの技術開発と実用化に向けたシステム開発を並行して実施してきた。

実用レベルでのシステム開発では 2017 年に統計演算を目的とした秘密計算システムである「算師®」を開発した。このシステムは計算機能

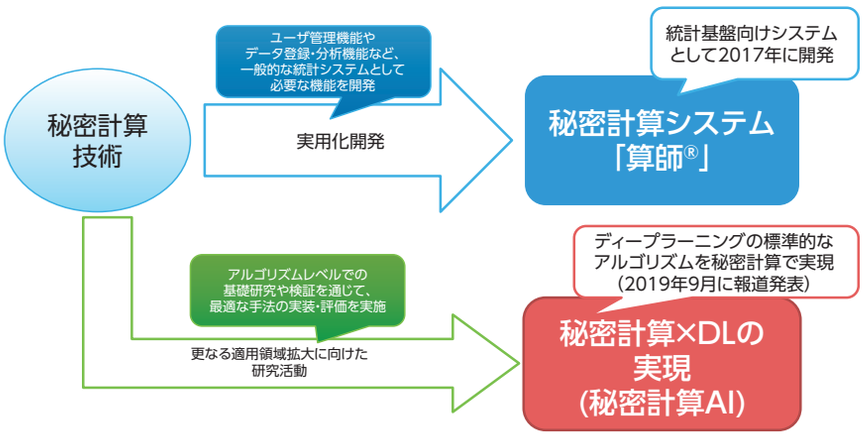

図２　SC 研の秘密計算に関する R&D

だけでなく、ユーザ管理機能やデータ登録・分析機能など、一般的な統計システムとして必要な機能を備えている。「算師®」は、平均、分散、t 検定など豊富な基本統計演算処理を具備しており、これら演算の組み合わせにより、回帰分析や主成分分析など用途に応じた分析を実現できる。また複数の表を暗号化したまま行結合、列結合することができるので、データをまたがった分析も可能となる。

研究レベルでは、更なる活用領域拡大に向けた研究活動を進めており、2019 年 9 月には秘密計算上でのディープラーニングを世界で初めて実現し、更にアルゴリズムレベルでの基礎研究や検証を通じて、最適な手法の実装・評価を実施してきた。現在も更なる適用領域拡大に向けた研究活動を継続中である（図２）。

――秘密計算システム「算師®」の特徴

秘密計算システム「算師®」は、データを暗号化したまま、実用的な速度で安全に集計・統計処理できるシステムである（図３）。複数台のサーバとクライアントソフトウェアから構成され、秘密分散ベースのマルチパーティ計算を３～４台のサーバで実行する。その特徴は、「安全」「高速」「便利」である。安全については、国際規格（ISO/IEC19592-2）に採択された秘密分散方式により安全なデータ保管を実現している。

高速については、大規模なデータの集計や統計演算を実用的な時間で処理し、世界最高レベルの高速演算を実現している。便利については、豊富な集計・基本統計機能で、用途に応じた分析を実行可能である。28 種類の基本統計演算とデータの結合を実現するとともに、R による応用分析の追加機能や Web GUI（Graphical User Interface）など多様なインターフェースを用意している。

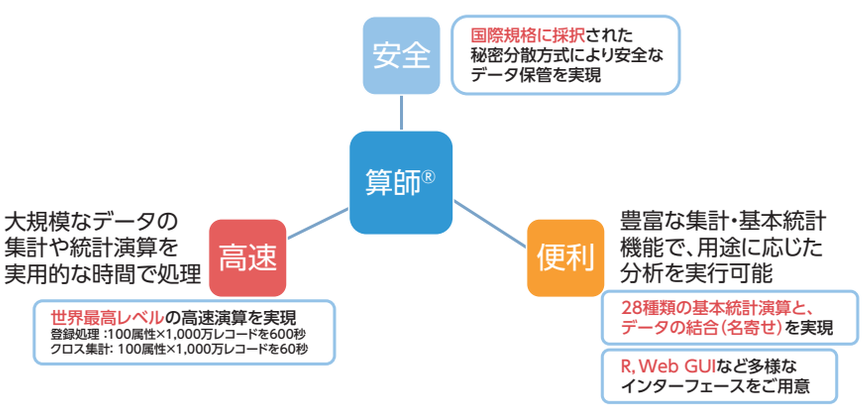

図３　秘密計算システム「算師®」

――ディープラーニングの標準的な学習処理が暗号化したまま可能に

近年進み始めている AI を使ったデータ利活用に向けて、データを暗号化したまま一度も元データに戻さずに、最適化処理を含む標準的なディープラーニングの学習処理を行う秘密計算技術を世界で初めて実現した。ディープラーニングでのデータ活用に必要な①データ提供、②データの保管、③学習処理、④予測処理、のすべてのステップを暗号化した状態で行うことができる。

サーバでは常にデータは暗号化されたままであり一度も元データに戻すことがないため、従来よりもユーザや組織が安心してデータを提供でき、学習に利用できるデータ量や種類が増え、精度の高い AI（Artificial Intelligence）の実現が可能になると考えている。例えば、個人の疾病情報、検査情報、処方情報などの医療データを秘匿しつつ、副作用実態調査や生存予測などを精度よく分析することが可能となると期待されている。

――秘密計算システムでできること・できないこと

このように、研究開発の進展により秘密計算システムでできることが

増えているが、この技術でできることとできないことを正しく認識しておく必要もある。

まず、本技術では、元データを見せることなく、データの分析結果（統計データ）だけを開示することができる。つまり、営業秘密やパーソナルデータを他者に見せることなくデータ分析の機能や結果を共有することが可能となる。

また、データは秘密分散で暗号化されているため、安全管理措置として、分析者や秘密計算システム管理者に対しても重要な営業秘密やパーソナルデータを守ることができる。さらに、異なる事業者間のデータであっても、ID など共通のキーがあれば、データを相互に見せ合わずに、秘密計算システム内でデータ同士を連携し、横につなげた分析が可能となる。

したがって、これまで他組織に開示、共有することが難しかったデータでも、データ漏洩の心配なく複数の組織や業界をまたがって持ち寄り、組織横断的な新しい統合分析を安全に実現することができる。

一方、できないことに目を向けると、個人情報保護法の制約から、秘密計算技術を使ってもパーソナルデータを自由に他組織に提供し、利活用させることはできない。現行の法制度上、暗号化した状態でも個人情報は個人情報として取り扱う必要があるため、第三者提供などを行う場合は同意の取得など法に準じた適切な処理が必要となる。

また、分析結果である統計データから営業秘密などが推測される可能性を完全に排除するのは難しい。統計分野で一般的に存在するこの課題は、秘密計算システムにも存在するが、「算師®」では一定の基準を満たさない分析結果の出力を抑制するアウトプット・プライバシー機能を実装するとともに、安全な分析結果を生成・出力するような新たなアプローチの研究を進めている。

また現時点では、簡単な画像を除いて、音声や映像などのメディアデータを取り扱うことは難しい。演算に必要なメモリの効率化や処理速度の向上を進めつつ、定型長でない複雑なデータに対する処理方式の検討を進めている。今後の研究開発に期待いただきたい。

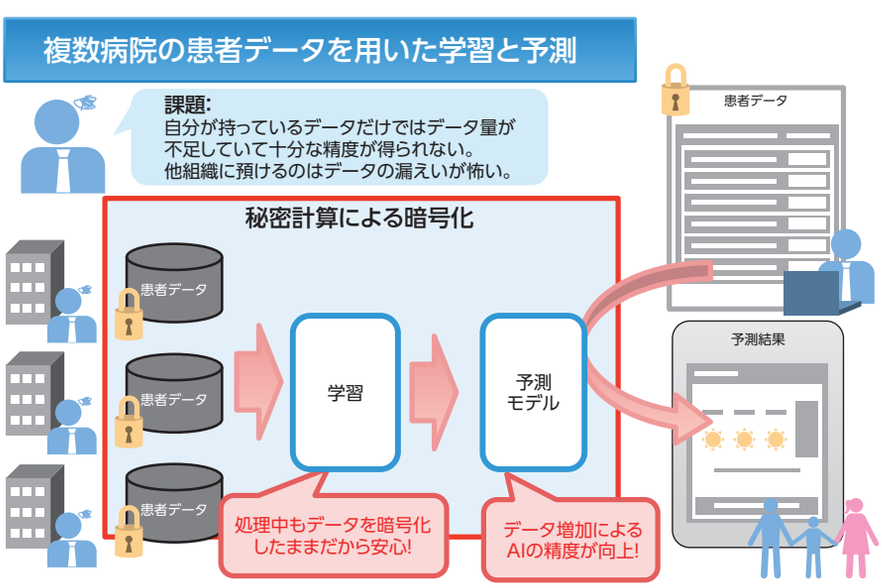

図 4　秘密計算 AI による病気リスク予測

――秘密計算AIによる病気リスクの予測

秘密計算技術の適用先として、医療分野での活用が考えられる。AI の活用が今後さらに加速すれば、臨床研究データのディープラーニングにより病気リスクの予測も可能になるだろう。しかしその場合、自病院が持っているデータだけではデータ量が不足していて十分な精度が得られない。かといって、他組織に預けるのはデータの漏えいが怖いという課題がある（図 4）。

一方、現在の医療分野では、均質な医療が受けられるよう、臨床研究で証明されたエビデンスに基づき、各種診療のガイドラインや標準治療法を定めるという取り組みが進められている。臨床研究データは個人情報であるため、臨床研究データを集約して預かる医療施設は、患者のプライバシー保護を含め、データの管理には最大限の注意を払ったセキュリティ対策をしなければならない。データを安全に取り扱いつつも、医療統計分析を実施可能とする技術も求められている。

上記のような課題に対し、秘密計算システムを用いれば、複数病院からの大量のデータを安全に利活用し、高度な医療技術の開発や評価へ貢献できる。また、データの分析精度を向上させ、治療が難しいとされる症例を早期発見・早期治療することも可能になる。