西田レポート：秘密計算システム「算師®」の活用法（1）

算師® の実用化に向けた取組み
～医療分野の適用実験を通じて～

NTT セキュアプラットフォーム研究所
データセキュリティプロジェクト
西田 祥子

パーソナルデータや企業秘密など守るべきデータの安心・安全な利活用に向け、データを暗号化したまま、実用的な速度で安全に集計・統計処理できる秘密計算システム 算師® の解説及び、本システムの実験の事例を紹介する。

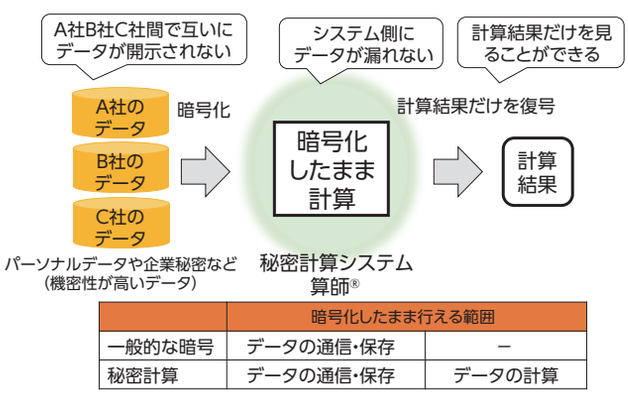
1. はじめに

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、企業戦略等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。

NTT 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（図 1）。秘密計算技術の利点は、計算対象の元データや計算過程をデータ提供者やサーバ運用者に一切見せることなく、統合分析の結果を得られることにある。これにより、これまで同業の競合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 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等の改良を加え、秘密計算システム算師® の開発を進めてきた。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、回帰や主成分分析など用途に応じた分析を実現できる。更に異業種データを横断分析するため、複数の表を結合するキーを漏らすことなくデータを結合して分析できる機能を具備している。秘密計算技術の最大の課題であった性能面においても、通常のコンピュータ処理との差は「一桁レベル」に迫る世界最高レベルの計算速度を達成している。

算師® のデータの暗号化の仕組みには、NTT がエディタとして貢献した ISO 標準（ISO/IEC19592-2）に準拠する秘密分散を採用している。秘密分散とはデータを複数のシェアと呼ばれる断片に分散する方式で、分散された個々のシェアからは元データおよび計算結果の情報は漏れることは



	暗号化したまま行える範囲	
一般的な暗号	データの通信・保存	−
秘密計算	データの通信・保存	データの計算

図 1　秘密計算の特徴と利点

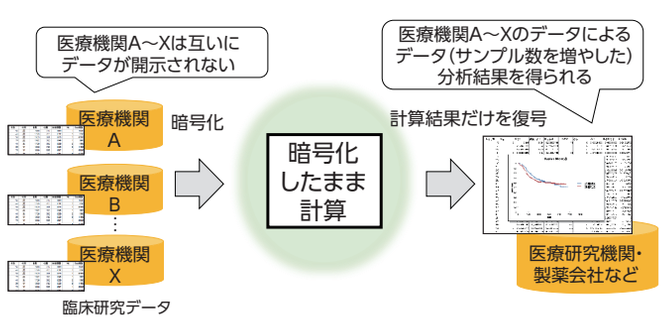


図 2　臨床研究データへの適用と効果

ない。さらに、もしサーバの故障などでシェアが消失した場合でも、一定数のシェアが残っていれば、データを復元することが可能である。このように算師® が採用する秘密分散はデータの機密性と可用性を実現している。

本稿ではこれまでに取り組んだ、医療分野における秘密計算の実験の事例を 2 つ紹介する。

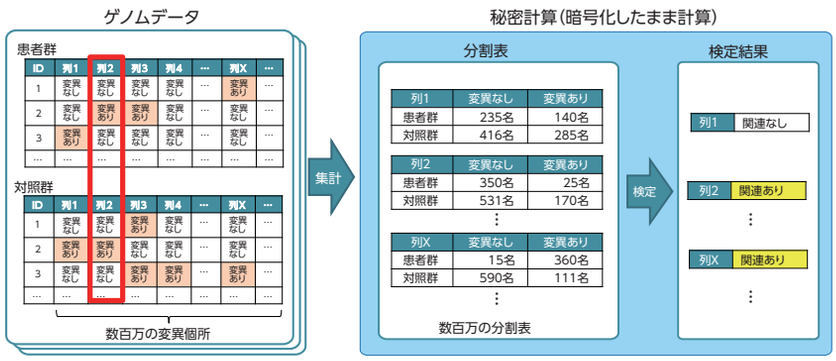


図 3　ゲノムワイド関連解析による疾病情報と遺伝子の関連性分析

2. 臨床研究データへの適用

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が課題であり、データを安全に取り扱いつつ、同時に医療統計分析を実施可能とする技術が求められる。

NTT と日本成人白血病治療共同研究グループは、データの更なる安心・安全な活用を目的とし、世界で初めて臨床研究データに秘密計算の適用を実証した（図 2）。本実験で求める医療統計分析の演算は、秘密計算の処理コストが高いデータのソートという処理が含まれていたが、独自のアルゴリズムのデータ操作演算の開発を行い、10 万件のデータのソートを 20 秒で行うという実用的な時間での処理を可能にした。これにより、臨床研究データを暗号化したまま、実用に資する速度で、研究者が求める演算の実現を確認することができた。秘密計算の活用により、安心・安全な臨床研究データの環境を実現し、医療の質のさらなる向上が見込まれる。

3. ゲノムワイド関連解析への適用

ゲノムデータは、個人の遺伝情報という機微性が高い性質のデータであるため、慎重な取り扱いが求められる。そのため、複数の研究機関の間で相互に制限なくゲノムデータを開示し合うことは容易ではなく、また、高精度の分析が求められる。

これらの課題に取り組むため、NTT と国立大学法人東北大学　東北メディカル・メガバンク機構は、世界で初めて、正確性を期したフィッシャー正確確率検定というゲノムワイド関連解析を秘密計算で実現した。

ゲノムワイド関連解析では疾病情報と遺伝子の関連性を発見するために、ゲノムデータを患者群と対照群に分け、数百万の変異個所について遺伝子変異の有無を集計した分割表を作成する。それぞれの分割表に対し、検定という手法を用いて患者群と対照群に遺伝子変異の有無との関連性があるかどうかを調べ、有意的な関連性がある場合は、その遺伝子個所と疾病情報に何らかの関連があることが分かる（図 3）。

これまで近似による分析はされていたが、分割表に現れる数字が小さい時には、統計検定の結果が正しくなくなる恐れがあり、このような場合にはフィッシャーの正確確率検定という手法で分析する必要がある。しかし、当分析には大きな数の階乗計算が必要となり、かつ数百万もの遺伝的多様性を 1 つずつ解析する必要があることから、従来の秘密計算技術にて当分析を行うためには 1 年以上の時間がかかることが想定されたが、アルゴリズムを工夫することにより、従来の 1 年以上から約 20 分程度にという現実的な時間に短縮することに成功した。これにより、複数の研究機関が安全にゲノムデータを持ち寄った、より高い信頼性の下で疾病情報と遺伝子の関連性を分析することができ、医療のさらなる発展が期待される。

4. おわりに

本稿では、医療分野における秘密計算技術の適用実験の取り組みと、そこから得られた知見に基づく算師® の実用化に向けた取り組みを解説した。次回以降は医療分野以外での事例や、今後期待される算師® の利活用方法を紹介する予定である。

〈秘密計算「算師®」のことなら下記へ〉
seg-product-p-ml@hco.ntt.co.jp