【基盤研究（S）】

大区分Ｊ

研究課題名　圧縮秘匿計算による大規模データ処理

東京大学・大学院情報理工学系研究科・教授

定兼（さだかね）　邦彦（くにひこ）

研究課題番号：21H05052　研究者番号：20323090

研究期間：令和３年度－令和７年度　研究経費（期間全体の直接経費）：122,300 千円

キーワード：秘匿計算、簡潔データ構造

【研究の背景・目的】

人間中心社会（Society 5.0）において様々な知識や情報を共有する際に発生する問題には次のものがある。

・大量のデータを学習・解析する際の計算コスト

・個人情報を共有する際のプライバシー

プライバシーを保護しつつ計算を行うためにはデータを匿名化・暗号化したまま学習・解析を行う必要がある。また、学習・解析をする際の計算コストとしては、計算時間と計算機のリソース（処理速度やメモリ量）がある。プライバシーを保護しつつ計算を行う技術として秘匿計算（secure computation）がある。秘匿計算とは、データを暗号化したまま計算を行う技術である。秘匿計算の応用としては、DNA 配列のデータベースがある。多くの患者の DNA 配列情報を集め、解析することで遺伝性疾患の治療に役立てることができる。しかし、DNA 配列情報は「究極の個人情報」であるため、患者のプライバシーを保護するために暗号化しておく必要がある。通常の暗号化では、暗号化されているデータに対して計算を行う際には，いったん復号し、計算後に再暗号化する必要がある。つまり、計算を実行する側に秘密が漏れてしまう。一方、秘匿計算では，データを暗号化したまま計算でき，計算を実行する側は個々のデータの情報は得られず計算結果のみが得られる。つまり個人のプライバシーを保ったまま学習・解析処理が行える。

計算コストについては、スパコンを使えば解決するというものではない。データのサイズを削減するためにはデータを圧縮すれば良いが，通常の圧縮では，圧縮されているデータに対して計算を行う際には、いったん復号する必要があり、多くのメモリを持つ計算機が必要になる。このような問題を解決するための技術として、簡潔データ構造がある. 簡潔データ構造とは、データを圧縮したまま計算を行う技術である。これにより、大規模データを省メモリ量の計算機で高速に処理することが可能になっている、応用としては、DNA 配列のアセンブリなどがある。

【研究の方法】

このように、情報共有のための基盤技術として秘匿計算と簡潔データ構造が存在するが、これらを同時に達成できるかは自明ではない。本研究では、様々なデータを暗号化・圧縮化したまま処理することを目指す。特に、以下のテーマを扱う。

1. 秘匿検索

DNA データベース等においては、複数のユーザがサーバにデータ検索の問い合わせを行うが、次のことを実現したい：（a）ユーザが何を検索したかを秘匿する、（b）高速な検索（サーバのデータ数 n の対数多項式時間）、（c）サーバの使用するメモリ領域を最小化する。

2. 秘匿学習・解析

データを圧縮したまま計算を行うという考えをさらに発展させた、圧縮学習についても研究を行う。一般に、学習精度を上げるためには、学習モデルを複雑にする、つまりモデルのパラメタ数を増やす必要がある。しかし、パラメタを増やした場合、その値を学習するために必要なデータ数も増大する。深層学習では大量の学習パラメタを用いるため、過学習を防ぐためには大量の学習データが必要であり、学習速度も低下する。しかし、圧縮によりデータ量を減らせば、少ないパラメタ数で学習が行えるようになり、学習に必要なデータ量も削減できる。

具体的には、自然言語等の非定型データを秘匿分析可能にする計算モデルを開発する。暗号化されたデータを復号せずにデータ処理可能にする準同型暗号に基づく様々な秘匿計算法が提案されているが、自然言語を含む非定型データに対する秘匿計算の研究は進んでいない。本研究では，口コミや評価等のプライバシーを含むデータ群を匿名化せずにそのまま暗号化し、復号することなく分析できる技術を構築することで、データ提供者と分析者が互いにメリットを享受できる仕組みを提案する。

【期待される成果と意義】

本研究で提案する「圧縮秘匿計算」は、データを圧縮、暗号化して格納し、なおかつ圧縮したまま、暗号化したまま学習・解析を行うための技術である。さらに、データを圧縮することで学習・解析にかかる時間を短縮し、かつ省資源の計算機でも実行可能とする. このように、圧縮秘匿計算は Society 5.0 のための欠かせない基盤技術となる。

【当該研究課題と関連の深い論文・著書】

・K. Shimizu, K. Nuida, G. Rätsch. Efficient privacy-preserving string search and an application in genomics. Bioinformatics, 32(11):1652-1661, 2016.

・定兼邦彦. 簡潔データ構造. アルゴリズム・サイエンスシリーズ 8, 共立出版, 2018.

【ホームページ等】

https://researchmap.jp/sada/