公募シンポジウム
公募シンポジウム3
セキュアかつフェアなデータ流通プラットホームとセキュリティ基盤技術
2019年11月23日(土) 14:40 ～ 16:40 C会場 (国際会議場 2階国際会議室)

[3-C-2-02] プライバシを保護した医療データ利活用システムについて

○宮地 充子[1]（1. 大阪大学）

キーワード：big data, PSI, PDDI, privacy-preserving classification protocol

各機関で独立に保管される医療情報を新薬開発や治療法の改善、健康診断結果の追跡による保健指導などに利活用することは重要である。その際、異なる機関が保管する同一患者のデータを突合できることは必須である。しかしながら、医療機関において独立に管理された医療情報を患者の名前、住所などの機微情報を漏らすことなく、突合することは容易ではない。我々の提案するデータ突合システム Privacy-preserving Distributed Data Integration（ PDDI）はプライバシ情報の漏洩を懸念することなく、複数の機関で独立に管理されたデータの突合を可能とする。本方式では、どの医療機関も共通の患者以外の情報は一切入手することができない特徴を持ち、対象がわかっているデータベースのクエリとは本質的に異なる。また、医療機関がデータベースを統合する必要がないため、独立・分散してデータを保管可能である。

一方、プライバシ保護秘匿データ解析プロトコルでは、データ解析の対象である患者の医療情報、および解析モデルの両者をお互い秘匿しながら医療情報の解析結果を患者が得ることができる。このプロトコルは誤り訂正符号理論に基づいており、量子計算機の攻撃に対しても安全に線形関数によるデータ解析を行うことができる。既存の誤り訂正符号理論に基づいた線形関数の秘匿データ解析プロトコルは紛失通信というコストの大きい構成要素を必要としていたが、本プロトコルでは紛失通信を利用しない効率的な構成となっている。

また、今回、 PDDIソフトウェアの全面的な設計見直しと、 Pythonによる再実装を行ったので報告する。従来ソフトウェアは、通信部分がボトルネックであり、ネットワークを介して突合を行う場合にパフォーマンスの劣化が見られたが、本実装ではそれらを大幅に見直し全体的なパフォーマンス向上を行った。また、ソフトウェアのインストールも、 Dockerコンテナ上に配置可能としたため簡易に環境構築が可能となった。

[公募シンポジウム]
セキュアかつフェアなデータ流通プラットホームとセキュリティ基盤技術
プライバシを保護した医療データ利活用システムについて

宮地充子*1*2*3、高野祐輝*1、河内亮周*4
*1 大阪大学, *2 北陸先端科学技術大学院大学,
*3 独立行政法人科学技術振興機構 CREST, *4 三重大学

Privacy-preserving Multi-party Medical Data Utilizing System

Atsuko Miyaji*1*2*3, Yuki Takano*1, Akinori Kawachi*4
*1 Osaka University, *2 Japan Advanced Institute of Science and Technology,
*3 Japan Science and Technology Agency CREST, *4 Mie University

It is an important issue to match and analyze medical records managed at different medical institutions independently without leakage of privacy information such as patient's name, address, etc. In this paper, we introduce two technologies to realize them. The first technology, Privacy-preserving Distributed Data Integration (PDDI), is a data integration system that enables to match and extract records managed by multiple organizations while preserving privacy. The second is a privacy-preserving classification protocol that classifies client information based on a classification model held by a server without knowing partner's information each other.

Keywords: big data, PSI, PDDI, privacy-preserving classification protocol

1 はじめに

ビッグデータの解析結果は新製品開発など様々な活用が期待され，そのデータ収集・解析・利用の促進・定着は重要である．医療分野においては，患者のプライバシを確保しつつ，カルテ情報を医療分野の発展に利活用できることが望まれる．さらに，データ所有者である患者がデータの利活用に合意できる枠組みの構築が必須である．

ビッグデータに関する既存研究では，ビッグデータの効率的な解析手法を改良する研究が多い．一方，本研究課題ではデータ所有者（医療の場合は患者に相当する）に着目し，データのプライバシ保護を実現しつつ，解析結果の適切なデータ所有者への還元・フィードバックを実現し，データ所有者，解析，その利用という 3 つの機能を信頼の環で連結することを目標としている． [MOSFH16]では耐サイバー攻撃の観点から医療データなどの安全な管理方法について提案し，[MNK17]では医療データの安全な利活用を促進する 2 技術であるプライバシ保護付き共有データ抽出手法と秘匿分類プロトコルについて発表を行った．本稿ではこの 2 つの技術についての進展及び医療データへの応用について説明する．

第一は，プライバシを保護しつつ共通データを抽出する方法でプライバシ保護付き共有データ抽出手法 (Privacy Set Intersection)と呼ばれる．医療では特に，癌治癒後，次の病気になった際に同じ病院に通わないケースが多く考えられる．このようなケースにおいて，異なる機関で管理される同一の患者のデータ突合の重要性は非常に高い [Gon17]. PSI では異なる機関が保管するデータのうち各機関が保管する共通データをそれ以外の情報は漏らさずに求める手法である． 2 機関のデータの共有データの抽出だけでなく，一般に複数機関のデータの抽出にも利用可能である．データを突合する簡易的な方法に本研究課題では複数機関における共有データの抽出方法について議論する．

第二は，サーバとクライアントがお互いの情報を秘匿したまま，サーバの持つ分類モデルに基づいてクライアントが持つ個人情報を分類する秘匿分類プロトコル(Privacy-Preserving Classification Protocol)である．本研究課題については既存研究における分類モデルの一般性とその効率について考察を行い，本研究で提案する新たなプロトコルについて概説する．

さらに医療の分野で上記研究成果を利用する状況を説明し，その具体的な成果について検討する．

本稿の構成は次の通りである．第 2 章では各研究の特徴および今後の展開について述べる．第 3 章で医療分野における適用事例について述べる．第 4 章で結論をまとめる．

2 各研究課題

2.1 プライバシを保護したデータ抽出手法

2.1.1 プライバシを保護した集合演算(PSI)

我々を取り巻く情報社会では，多種多様なデータが多機関で収集される．例えば，小学校で児童がブランコでけがをした事例を考える．このとき，事故が起こった遊具に関するデータは学校，病院への救急搬送データは消防署，傷害・後遺症に関するデータは病院に管理される．つまり，学校での生徒の事故に関する情報では，学校，消防署，病院がそれぞれ同じ事故で異なるデータを管理する．

学校における事故の予防安全の実現には，事故の統計的因果モデルの作成が重要である．これにはこのように異なる機関に分散した関連データの統合が必須である．つまり，異なる機関が独立に収集したデータから生徒の名前などの機微情報は洩れることなく，同じ生徒の事故の情報を突合（分散多機関データ突合）できると，事故の詳細なデータの収集が可能になる．

ここで，異なる機関がもつ医療データの突合方法とプライバシの関係について考える．単純な方法は，1つの医療機関が全データを別の医療機関に渡せば，同じ患者を検索することで，データを突合することができる．しかし，この場合，本来もつはずでなかった患者の情報である名前，住所などの機微情報を別の医療機関が入手することになる．別の方法として，第 3 の機関(データ預託機関)にそれぞれの病院が医療情報を渡し，その第 3 の機関で突合することもできる．しかし

この場合には，第 3 の機関に患者の機微情報が移動することになる．つまり，単純な突合方法は突合に用いる情報が必要となるため，突合を実施する機関に機微情報が移動し，プライバシ保護を実現することが困難になる．

そこで相互の持つ重要な情報を外部に漏らすことなく，必要な情報のみのやり取りを行うための有効な技術が必要である．近年 Private Set Intersection Protocol(PSI) と呼ばれる各機関が持つデータの積集合などの集合演算を，プライバシを保護しつつ実現するプロトコルが注目されている．一般的な PSI は次のような特性を持つ． S と C をデータの集合として持つサーバとクライアントの存在を考える．サーバとクライアントはそれぞれ S と C を入力としてプロトコル通りに通信をおこなうと，最終的にサーバは|C| のみを，クライアントは|S|, C ∩ S を得ることができる．上記の病院と学校の例では，病院側はウイルス感染者の人数以外は学校に知られることなく，ウイルスに感染した学生の情報を得ることができる(図 1 参照)．また学校側はウイルスに感染していない学生の情報は学生数以外知られることはない．それぞれのデータ内容がプライバシ，ウイルスに感染した学生情報が病院の求める情報と考えると，互いのプライバシを守りつつ，クライアントは目的の情報を得ることができる．

典型的な手法である既存のプロトコル [KS05]では全参加者の入力データ数を一致させなければならないという制限やデータサイズや機関数に依存する処理時間，通信量が大きな課題となる．[MBD12]においてはデータを所有する参加者以外に第三者機関（データ預託機関）を導入する必要があり，その機関もデータを入手する．また，PSI の応用として，共通データを抽出後，データの加算結果などの処理結果を出力する手法[IKNPRSSSY]が提案されている．しかしながら，どの既存方法も，単一の属性を前提としており，複数の属性を統合することができない．

本研究では PSI を用いて突合する属性とそれに付随する属性のデータ統合を行う方式であり，既存研究とは考え方が異なる．

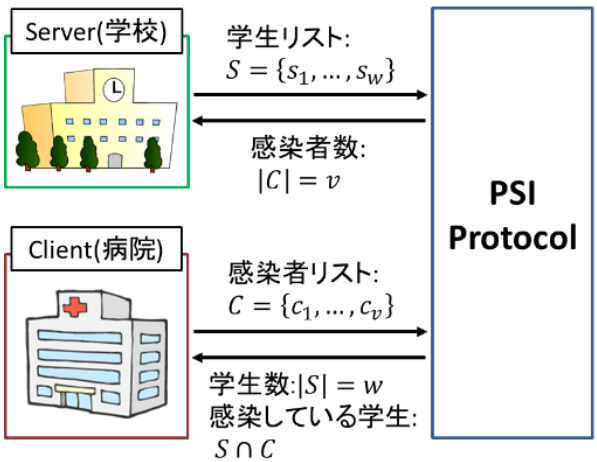

図 1 PSI

2.1.2 特徴

本研究室で提案された方法は，機微情報を他の機関に移動することなくデータ突合を実現する方式である． この方式は，データ統合システム Privacy-preserving Distributed Data Integration(PDDI)においてプログラム実装されている．具体的には，本方式を用いると，データベースから突合する属性と統合する属性を選択し，突合属性をPSIを利用して突合し，突合された属性と付随する必要属性を統合することが可能である．

本方式の持つ特徴について列挙する．

1. どの機関も共通に含まれるユーザ以外の情報について何も入手することはない．(Query ベースとは異なる)
2. 突合に利用する情報は，計算機サーバを含めて，どの機関にも移動しない．
3. 各機関の処理時間は機関数に依存しない．各機関のデータ数に制限や条件はない．
4. 第三者機関によるデータ収集・管理が不要で，容易に導入可能．

2.1.3 PDDI システム

本システムを利用することによって，ユーザは情報漏洩を懸念することなく，複数機関が所有するデータの統合を実現することができる．. 図 2 では，全機関に存在する同一ユーザを突合し，その後ユーザに関する情報で各機関に独立に存在するデータを統合する．

ステップ①

各データ集合の突合したいデータ x(図 2 の場合，名前)をハッシュ関数 H 等で圧縮する．例えば機関1のデータの名前「竹田」は

竹田→H(竹田)

と一意に不可逆な情報に変換される．その後，さらに準同型暗号 E で暗号化する．

H(竹田)→E (H(竹田))

同様の処理を機関 2, 3 で行い，暗号化データを秘匿計算機サーバに送付する．

ステップ②

準同型暗号は，暗号文の和が，元の文の和の暗号文に一致する性質を持つ．この性質を用いて秘匿計算機サーバでは入手した暗号化データをそれぞれ加えることで，各機関のデータの総和の暗号結果を計算できる．図 2 の場合，下記のように，各機関で計算された E (H(竹田))のような名前の暗号文は秘匿計算機サーバで加えられ，名前の和の暗号文となる．

E (H(竹田)) + E (H(田中)) + E (H(竹田)) + E (H(田中))
+ E (H(竹田)) + E (H(山田))
= E (3H(竹田) + 2H(田中) + H(山田))

この暗号化データの総和を各機関に送付する．

ステップ③

受信した各機関のデータの総和の暗号文を復号して，突合したデータ，図 2 の場合，「竹田」が突合属性であることをわかる．次に，各機関に要求される「竹田」の必要属性を出力する．

2.1.4 PDDI システムの医療への応用事例

ここで PDDI システムの医療への適用事例について説明する．がん治療の進歩に伴い，がん生存者（がんサバイバー）は増加している．また，人口の高齢化によって，がんサバイバーも高齢化している．高齢化にともなって増加すると予想される脳卒中や認知症等の神経疾患とがんとの関係については，不明な点が多い．この大きな理由ががんサバイバーの脳卒中や認知症の症状を把握する病院と，がん治療の症状を把

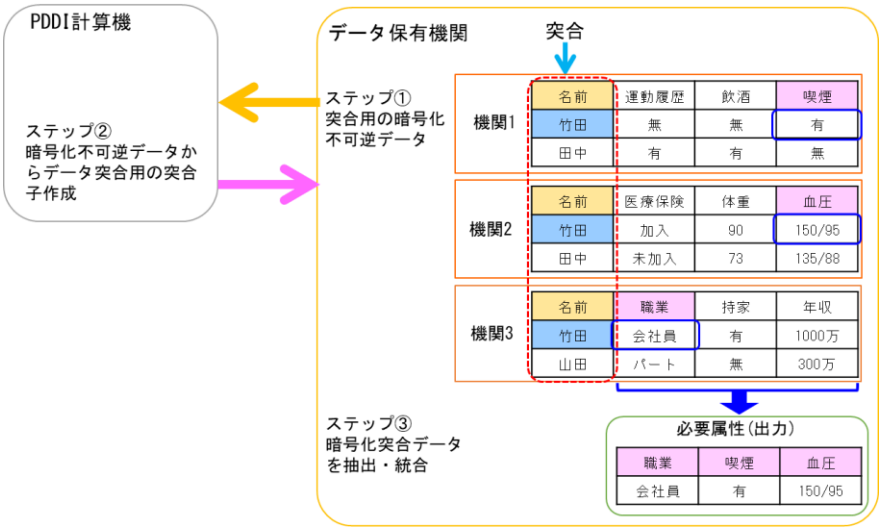

図 2 システムモデル

図 3 がんデータと脳卒中データの突合事例

握する病院が異なることが理由として挙げられる. 本 PDDI システムでは, 異なる病院が保管する同一の患者のデータを突合し, 同一患者のがん発症の日時と脳卒中発症の日時のみを抽出することが可能である. (図 3) 重要なことは, 突合する患者の名前などの情報はそれぞれの医療機関から外部に出ることがなく, 発症日時というプライバシ情報がないデータが突合されて出力されるということにある.

2つの医療機関に適用し, 医療データを突合する仮想実験が行われており, 実用化に十分な機能を持つことが確認されている [MYGNMM18].

2.2 プライバシ保護付き共有データシステム設計と実装および実験環境構築例

本節では PDDI ソフトウェアの実装と実験環境について述べる. 我々は, 従来までも PDDI ソフトウェアの実装を行ってきており, その動作を確認してきていたが, 従来ソフトウェは次の 3 点で問題があることが明らかとなった. すなわち, 1. ソフトウェアデプロイメントの困難さ, 2. ネットワークを介した際のパフォーマンス低下, 3. モジュラリティの低さである.

ソフトウェアデプロイメントの困難さとなった原因は, ソフトウェアが複数のライブラリ, アプリケーションに依存しており, その結果としてマニュアルが煩雑になり理解するのが難しくなってしまったのが原因である. その結果, 人手で設定すべき項目が多くなり, 全体を把握するのが難しくなってしまっていた.

ネットワークを介した際のパフォーマンス劣化については, 同時にまとめて通信すべきところを, 複数の通信で行っていたためである. PDDI では, 内部的に Bloom Filter と呼ばれるデータ構造を用いているが, Bloom Filter 利用時に計算される key を複数回計算しその都度転送処理を行っている箇所が特にボトルネックになっていると考えられた.

モジュラリティの低さは, GUI 部分, 認証, ネットワーク通信, 暗号化などが全て同一のソフトウェアとして実装されていたため, 一部を改変すると他に影響しやすいという問題があった.

そこで我々はこれらを解決すべく, 以下の方針で PDDI ソフトウェアの再設計を行い, 実装を全面的に刷新した.

設計方針

1. PDDI ソフトウェアでは核となるビジネスロジックのみ実装すること
2. 設定はすべて設定ファイルで行えること
3. デプロイメントは極力自動で行えるようにすること
4. Web による連携を用意にするために, Restful API を用いて通信を行うこと
5. 通信は極力まとめて行うこと

ビジネスロジックのみを実装することで, PDDI を利用する機関は各々にカスタマイズして利用することが可能となる. しかし, これはつまり, 認証などの PDDI 外のメカニズムは独自に実装する必要があるということになるが, 一般的な認証用の

ソフトウェアを使えば良いため実際的な問題にはならない．また，設定ファイルをファイル変更のみで行えるようにすることで，ソフトウェアデプロイメントに対するハードルを緩和させた．従来 PDDI ソフトウェアでは，複数アプリケーション，ライブラリを用いていたため設定方法が煩雑になってしまっていたが，新規ソフトウェアでは容易に設定することができるようになった．さらに，実験環境の構築は Docker を用いて行えるようにしたため，本番環境でも Docker を用いたデプロイが行えるようになる．そのため，デプロイメント時に設定すべき項目が非常に少なくなった．

ネットワーク通信のインターフェース部分は Restful API を定義し，本 API 経由で通信を行うようにした．これにより，他のウェブアプリケーションとの連携が用意になり，ネットワーク利用時のデバッグも用意になった．

極力データをまとめて送受信することで，ネットワーク利用時のパフォーマンスを改善した．また，前回実装では Java と C++言語を用いて行っていたが，今回実装ではソフトウェア的な問題点の改善に注力するために，Python 言語で全面的に書き直したため，ソースコードレベルでのメンテナンス性も向上していると考えられる．一方，暗号化部分では Python 実装の方が遅くなっている箇所がある可能性があるため，今後，より詳細なパフォーマンス評価を行う予定である．

以下では実験環境の構築方法について述べる．PDDI ソフトウェアでは Docker と Docker Compose を用いているため，実行するための環境は基本的に以下の 2 回で終了する．

```
$ docker-compose build
$ docker-compose up -d
```

コンテナ起動後に docker コマンドで確認してみると，以下のように 4 つのコンテナが動作していることがわかる．

```
$ docker ps
CONTAINER ID        IMAGE
a59c5548665f        docker_party2
2f7cb5081da1        docker_dealer
620438d4ecd1        docker_party1
8f067f27bd6f        docker_client
```

ここで，IMAGE が docker_party であるコンテナが，データを提供する機関，docker_dealer がデータ統合を行う機関，docker_client がデータを利用する機関を想定している．コンテナ起動後は，コンテナにアタッチして実際に PDDI を動作させることができる．例えば，docker_dealer によりデータ統合用機関を実行するには以下のようにする．

```
$ docker attatch pddi-dealer
root@dealer #> python3 /pddi/server.py
 * Running on http://0.0.0.0:5000/
```

従来ソフトウェアではデータ統合用機関用のソフトウェアを実行させることすらハードルが高かったが，今回ソフトウェアではこのように非常に簡単に実行させることが可能となった．

2.3 秘匿分類プロトコル

医療情報処理への機械学習の応用は機械学習分野の発展に伴ってより一層重要になってきている．医療における機械学習の応用例としては例えば，青木ら[ASH+09]はサポートベクトルマシンという基本的な分類器を甲状腺機能異常のスクリーニングに応用している．またこのサポートベクトルマシンを応用して Yu ら[YLV+10]は糖尿病の有無の分類を行っている．

機会学習は一般的に二つのフェイズで構成される．訓練フェイズでは大量の訓練用学習データを用いて分類モデルを生成し，その後，分類フェイズにおいては，分類対象の実データを与えられて，生成された分類モデルを使って実際に分類を行う．この際に訓練データや分類対象データがプライバシ情報である場合，そのプライバシ保護が重要になってくる．暗号基盤技術の応用として機械学習におけるプライバシ保護技術が近年盛んに研究されている．多くの従来研究においては訓練用データのプライバシ保護を議論しているが，最近では分類フェイズにおけるプライバシ保護の研究も急速に進んできている．

秘匿分類プロトコルの概要は以下の通りである．このプロトコルは分類モデルを保持するサーバと分類対象となるプライバシデータを保持するクライアントの間で実行される．サーバはクライアントに知的財産である分類モデルの内容をクライアントに漏らしたくはなく，クライアントは自身のプライバシデータをサーバに漏らしたくはないが，自身のプライバシデータの分類結果は知りたい場合，この秘匿分類プロトコルに則ってサーバ・クライアント間の通信を行えば，互いの情報のプライバシは保護したままクライアントは自身のプライバシデータの分類結果を得ることができる．

このようなプライバシ保護機能は分類モデルの歪曲化(garbling)技術，紛失通信プロトコル(oblivious transfer)，準同型公開鍵暗号方式(homomorphic encryption)といった暗号基盤技術を巧みに利用することで実現できる．概要を説明すると，歪曲化はデータ分類機能を保ったまま分類モデル自身を暗号化する技術であり，紛失通信プロトコルはクライアントが要求する情報がどれかをサーバに知られることなくクライアントが入手できる技術，さらに準同型公開鍵暗号は暗号文同士の演算によって暗号化された情報を操作することが可能な技術である．

これらの技術を用いて既に秘匿分類プロトコルがいくつか提案されている．Barni ら[BFL+09][BFL+11]は線形分岐プログラムと呼ばれるシンプルな計算モデルに対する歪曲技法および紛失通信プロトコルから線形分岐プログラムの秘匿分類プロトコルを構成している．彼らはさらにそのプロトコルに基づいて心電図のプライバシ保護分類器を実装している．またさらにパーセプトロンの一般化となるニューラルネットワークモデルの秘匿分類プロトコルを研究している．

Bost らの研究[BAT+15]では Barni らの結果の一般化としてさらに汎用的な学習モデルにおける秘匿分類プロトコルを提案し，その安全性および効率を解析している．Bost らは Paillier 公開鍵暗号方式および Goldwasser-Micali 公開鍵暗号方式といった数論に基づいた加法準同型性公開鍵暗号を利用してパーセプトロン，Fisher 線形判別，サポートベクトルマシンを含む線型判別器およびナイーブベイズ分類器の秘匿分類プロトコルを提案し，またレベル付き完全準同型性暗号（二つの暗号文の加算だけでなく回数制限付きで乗算も可能な暗号化技術）を応用することにより決定木の秘匿分類プロトコルの構成を与えている．

Wu ら[WFN+16]は紛失通信プロトコルおよび加法準同型性暗号を利用した巧みな構成によって効率の良くないレベル付き完全準同型公開鍵暗号を用いずに決定木の秘匿分類プロトコルを実現し，既存結果よりも非常に高速であることを実証している．また彼らは提案プロトコルを決定木だけでなく機械学習において著名な学習モデルの一つであるランダム

フォレストモデルにも拡張できることを示している.

前年度までの研究においては Aguilar-Melchor ら[AAB+17]の誤り訂正符号ベースの公開鍵暗号プロトコル HQC を基にした秘匿線形関数計算プロトコルを提案していた. その研究の延長として, 本研究では, さらに同じ公開鍵暗号プロトコル HQC を基に秘匿大小比較計算プロトコルを構成した. そしてそれらのプロトコルを組合わせることによって秘匿サポートベクトルマシンの提案を行った. この公開鍵暗号プロトコル HQC は米国立標準技術研究所(NIST)が実施しているポスト量子暗号標準化コンペティションへ提出されており, 量子計算機の攻撃に耐えうる高い安全性を持つと考えられている. Bost ら[BAT+15]では数論に基づく加法準同型公開鍵暗号から秘匿計算プロトコルを提案しているが, このプロトコルの安全性は量子計算機によって容易に破られてしまう. したがって本研究のプロトコルは既存研究に比べて量子計算機の攻撃に耐えうるという利点を持つ.

3 医療分野への応用

PDDI 技術では, 診療所や病院などの複数機関で分散管理されるデータを, 第三機関を用いずに, 互いに非開示のまま共有データに関連する情報のみを突合することが可能である. 例えば, 各医療機関が持っている希少な症例データを複数の病院で集めることにより, 希少な症例データの統計的な解析が可能となる. また, 医療現場においては, 関連するデータが複数の異なる機関で管理されるケースが頻繁に起こる. 例えば, 同じ患者が異なる疾患にかかった場合, 各疾患を専門とする複数の病院に通うことが考えられる. このように独立した 2 つの医療機関で管理された異なる疾患同士には因果関係がある可能性がある. この時, 同一の患者のデータを患者のプライバシを保護しつつ, 必要な医療データのみ突合できると, 異なる病気の因果関係に関する詳細なデータの収集が可能になる.

秘匿分類プロトコル技術は, 医療分野における機械学習の利活用のためにプライバシ保護機能を付与できる. 今回構成したプライバシ保護サポートベクトルマシンは, 例えば前述の青木ら[ASH+09]や Yu ら[YLV+10]による研究など, 医療分野における既存のサポートベクトルマシンによる分類へ応用が可能である.

さらに図 4 のように上記 2 つの技術を統合し, PDDI で収集したデータに秘匿分類プロトコルを適用することで, プライバシを保護した医療解析が可能となる.

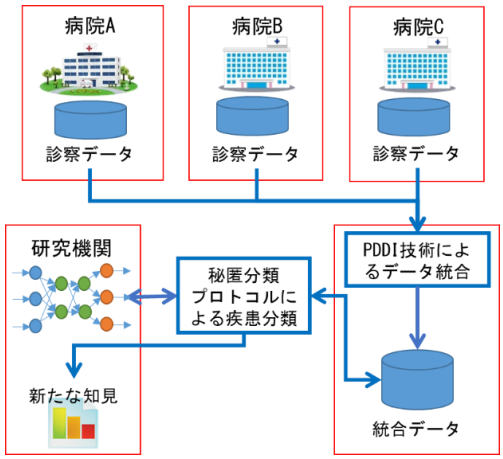

図 4 PDDI 技術と秘匿分類プロトコルを用いた研究

4 まとめ

本研究ではプライバシを保護しつつ医療データの利活用を可能にするセキュリティ技術について紹介し, 各研究課題の特徴や今後の展開について述べた. 第一はプライバシ保護付き共有データ抽出手法で, 複数機関におけるデータの突合及び抽出方法について説明した. また, 当研究室で開発した方法(PDDI)は暗号化不可逆データを用いることで, 個人を特定できる情報はどの機関にも移動せずに突合とデータ抽出を可能にし, 突合を行う医療機関数が増えても高速に処理できることを紹介した. また, 仮想実験も進められており, 今後, 実証実験を通して, 医療現場に特化した改良を行う予定である. 今後, 各種医療データへの利活用により, 異なる病気の因果関係の究明や, 稀少疾患の治療方法の発展への貢献が望まれる. 第二に秘匿分類プロトコルについてその技術を概説した. 特に既存研究においてサーバとクライアントのプライバシを保護しつつ利用可能となっている分類モデルの種類と構成要素として利用している暗号基盤技術からその効率について議論を行い, 本研究で提案している線形関数の秘匿計算プロトコルを紹介した. 次にこれらの 2 つのセキュリティ技術の医療分野における適用事例を取り上げた.

謝辞

本研究の一部は科学技術振興機構（JST）の CREST(JPMJCR1404)及び文部科学省の情報技術人材育成のための実践教育ネットワーク形成事業分野・地域を越えた実践的情報教育協働ネットワークの助成を受けています.

参考文献

[AAB+17] C. Aguilar Melchor et al., HQC, https://pqc-hqc.org/, 2017.
[ASH+09] 青木, 佐藤, 星, 川上, 森, 齋藤, 吉田,「医療データ解析へのサポートベクトルマシン（SVM）の応用」, 東北薬科大学研究誌, 56, 67−74（2009）
[BAT+15] Raphael Bost, Raluca Ada Popa, Stephen Tu, and Shafi Goldwasser. Machine Learning Classification over Encrypted Data. In Proc. The 2015 Network and Distributed System Security (NDSS) Symposium, 2015.
[BFL+09] Mauro Barni, Pierluigi Failla, Riccardo Lazzeretti, Annika Paus, A-R Sadeghi, Thomas Schneider, and Vladimir Kolesnikov. Efficient privacy-preserving classification of ECG signals. In Proc. 1st IEEE International Workshop on Information Forensics and Security (WIFS 2009), pages 91–95, 2009.
[BFL+11] Mauro Barni, Pierluigi Failla, Riccardo Lazzeretti, Ahmad-Reza Sadeghi, and Thomas Schneider. Privacypreserving ECG classification with branching programs and neural networks. IEEE Transactions on Information Forensics and Security (TIFS), 6(2):452–468, June 2011.
[BL13] K. Bache and M. Lichman. UCI machine learning repository, 2013.
[BLN13] Joppe W. Bos, Kristin Lauter, and Michael Naehrig. Private predictive analysis on encrypted medical data. In Microsoft Tech Report 200652, 2013.
[GNN17] S. Ghosh, J. B. Nielsen, and T. Nilges, “Maliciously Secure Oblivious Linear Function Evaluation with Constant Overhead”,
[Gon17] Gon, etc., “Validation of an algorithm that determines stroke diagnostic code accuracy in a Japanese hospital-based cancer registry using electronic medical records” BMC, Dec., 2017.
[IPS09] Yuval Ishai, Manoj Prabhakaran, and Amit Sahai, “Secure Arithmetic Computation with No Honest Majority”, TCC 2009, 294-314.
[KS05] L. Kissner and D. Song. Privacy-preserving set operations. In CRYPTO 2005, volume 3621 of LNCS, pages 241-257, Springer, 2005.
[WFN+16] David J. Wu, Tony Feng, Michael Naehrig, and Kristin Lauter. Privately Evaluating Decision Trees and Random Forests. In Proc. Privacy Enhancing Technologies (4);1-21, 2016.

[MNK17] 宮地, 中正, 河内,「プライバシを保護した多機関データ突合システムについて」, 医療情報学会 2017
[MBD12] Many, Burkhart, and Dimitropoulos. Fast private set operations with sepia. Technical Report, 345, 2012.
[MOSFH16] 宮地 充子, 面 和成, 蘇 春華, 布田 裕一, 波多野 哲也, 西田 昌平,「ビックデータ統合利活用促進のためのセキュリティ基盤技術」 医療情報学会 2016
[MYGNMM18] 宮本 潤哉, 山本 景一, 権 泰史, 中正 和久, 宮地 充子, 望月 秀樹,「Private Distributed Data Integration(PDDI)を利用した多施設臨床研究データリンケージの仮想実験」 第 22 回日本医療情報学会春季学術大会 シンポジウム 2018
[MNN17] Miyaji, Nakasho, and Nishida, ``Privacy-Preserving Integration of Medical Data A Practical Multiparty Private Set Intersection", Journal of Medical Systems, Vol. 41 No. 3, pp. 1-10, (2017).
[YLV+10] Wei Yu, Tiebin Liu, Rodolfo Valdez, Marta Gwinn, and Muin Khoury, “Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes”, BMC Med Inform Decis Mak. 2010; 10: 16.
[IKNPRSSSY] Ion, Kreuter, Nergiz, Patel, Raykova, Saxena, Seth, Shanahan, Yung, “On Deploying Secure Computing Commercially: Private Intersection-Sum Protocols and their Business Applications”, IACR Cryptology ePrint Archive 2019: 723 (2019).