研　究　報　告　書

「実社会情報ネットワークからのプライバシ保護データマイニング」

研究タイプ：通常型
研究期間：　平成21年10月～平成25年3月
研 究 者：　佐久間　淳

1.　研究のねらい

ネットワークベースのサービスや多様な携帯デバイス等の高度化により，多種多様かつ詳細な個人・組織の実社会情報が蓄積されつつある。物品やサービスの購買・販売履歴，ヒトや車両等の精細な地理的移動・行動履歴などの粒度の細かい実社会情報からの知識発見は新しい価値を持つサービスを生み出す源泉となりうるが，知識発見の対象となる個人情報の詳細度が増せば増すほど、それを開示・活用することへの心理的抵抗感は大きくなる。多様な知識資源が計算機にとって活用可能な形式で蓄積されつつある中、プライバシ保護と知識獲得を両立させる技術は知識獲得技術の発展のためには不可欠であるといえる。

このような社会的背景を鑑み、分散秘密情報源からの安全な知識獲得を目指すプライバシ保護データマイニング(privacy-preserving data mining, PPDM)が近年活発に研究されている. PPDM とは，二人(以上)のエージェントがそれぞれ秘密の入力データセットを持つときに、それらが互いにデータ見せ合うことなくデータマイニングを実行し，得られた知識のみの共有を目指す技術である。従来、PPDM では、入力データはデータベーステーブルに格納されていることを想定し，多くは 2 エージェント間において，属性単位やレコード単位で秘密情報を定義することが想定されてきた(プライバシモデル)。しかしながら、「何処へ行き何をしたか」、「何に興味を持ち何を購入したか」、「誰は誰と協力関係にあるか」といった実社会情報から、プライバシ保護を考慮しつつ知識発見を実現するためには、情報の構造、流通経路、情報の開示可能な範囲、計算主体の物理的所在や通信・計算能力等を考慮してプライバシモデルの定義をし、また知識発見アルゴリズムもそれに応じて設計する必要がある.

本研究では、情報の構造、流通経路、計算主体の性質等を考慮したネットワーク構造によるプライバシモデルである private graph を中心概念とし、これに基づく実社会情報ネットワークからの安全な知識獲得を実現する多様なアルゴリズム群と、その実行・開発フレームワークである Fairy Ring を構築した。これにより、実社会のネットワーク構造情報が持つ秘匿性を侵害することなく、安全に知識発見および発見知識の流通を実現するための理論的基盤との構築と開発技術の提供を実現した。

2.　研究成果

（1）概要
本研究課題の研究テーマは大きく下記の3テーマで計画された。 研究テーマ A プライベートグラフにおける秘密計算 研究テーマ B プライベートグラフにおけるデータ集約･匿名化と開示の理論 研究テーマ C プライベートグラフにおける分散データマイニング開発環境

研究テーマ A については、ノードランキング、ノードラベル予測、オンライン予測等のアルゴリズムを実現し、研究テーマ C において、これらをアンドロイド上で実現可能なシステム Fairy Ring を開発した。研究テーマ B については、ネットワーク構造情報を表現する行列を含む、半正定値行列を、安全に開示(差分プライバシー)するための理論的枠組みと、それを実現するアルゴリズムを提案した。以下、各テーマについて詳細を説明する。

(2)詳細

研究テーマ A1. プライベートグラフにおけるランキング問題: ,

”誰が誰から買った”, ”誰が誰に電話 (メール) をした”,”誰が誰を評価した”, などのリンクや社会的紐帯は, 機密性やプライバシの問題のため, 公にされることが好まれない場合が多い. この研究では、秘密のリンクを含むネットワークを対象として, その秘匿性を損なうことなく安全にリンク解析を適用可能な方法を開発し、これまで秘密性のために解析対処とならなかった現実の多様なネットワークからの情報抽出を可能にした。具体的には、この研究で提案したアルゴリズム PrivateRank は Web リンク解析において著名な PageRank 法をプライベートグラフ上で安全に実現する方法を構築した。(受賞1、論文投稿予定)

研究テーマ A1 プライベートグラフにおけるラベル予測問題

このテーマでは、ノードがラベル情報を持つプライベートグラフにおいて, 秘密を保護しつつノードラベルを予測するアルゴリズムを提案した。例として、ノードが個人, リンクが物理的な接触関係, ノードのラベルが感染状態であるような感染症ネットワークが考えられる。ラベル予測のための従来の半教師付き学習方法はリンクやラベルの公開を前提としている. このアルゴリズムでは、加法準同型性公開鍵暗号を用いたマルチパーティーコンピュテーションによって半教師付き学習の一方式であるラベル伝播法をプライベートグラフ上で安全に実行する方法を開発し、ラベルとリンクが各ノードの秘密情報であっても、その秘密を保護しつつ、ラベル予測を可能にした。(論文2, 受賞1)

研究テーマ A2エキスパート間の秘密のアドバイスに基づくオンライン予測

このテーマでは、学習者が複数のエキスパートの忠告(予測)を各時刻において得て、これに基づいて次の時刻の系列の予測を行うオンライン予測問題において、エキスパートの忠告に秘密情報が含まれるために、他のエキスパートや学習者に開示できない状況におけるオンライン予測を実現する方法を構築した。エキスパートが保持する予測情報を他者と共有することができない場合の学習者の予測精度は、それが共有できた場合と比べ悪化するように思えるが、構築した secure exponential weighting 法では,暗号理論的ツールを利用することによって, 予測/損失情報を学習者や他のエキスパートと共有しなくても,それらがすべて共有された場合と同じ予測精度が達成可能であることを理論的に示すことに成功した。このことから、オンライン予測において、エキスパートの忠告のプライバシ保護は、オンライン予測の妨げにならないことを理論的に示した。(論文5、受賞3)

研究テーマ A3 プライベートグラフにおけるラベル予測問題

この研究ではラベル付きプライベートグラフにおいて、ノードラベルが秘密情報である場合に、ラベルやリンクの秘密を保護したままラベルを予測するプロトコルを構築した。例として、ノードが個人，リンクが接触，ノードのラベルが感染状態である感染症ネットワークにおける感染状態予測ン度が挙げられる。ラベル予測のための従来の半教師付き学習方法はリンクやラベルの公開を前提としていたが、本方式では、加法準同型性公開鍵暗号を用いた秘密計算を半教師付き学習の一方式であるラベル伝播法に導入して、プライベートグラフ上で、安全にラベル予測が実行可能になった。（論文1、受賞2）

研究テーマ A4 プライベートグラフにおける安定結婚問題
安定結婚問題とは，男女同数のグループについて，それぞれの選好に基づいた安定した一対一マッチングを求める問題である．ここで，選好とは，ある人の好みに基づく異性のグループに対する順位付けであり，安定したマッチングとは，どの二つのマッチングを組み換えても，組み換えられた男女の選好の順位が低くなることを指す。この選好を表す情報は，個々人は個々人のセンシティブな情報の一つであるケースも多くプライバシの問題が生じる．この研究では、全員の選好のプライバシを保護しつつ，安定マッチングのみを計算可能なアルゴリズムを構築した。（発表 1、論文投稿予定）

研究テーマ A における代表的な成果は以上4つであるが、これに加え、この領域において以下のアルゴリズム開発を行った。

A5 PPDM におけるパーティ同士の結託を防ぐ基礎的プロトコル （論文 4）
A6 複数プライベートグラフにおけるノードランキング問題（論文2）

研究テーマ B プライベートグラフにおけるデータ集約･匿名化と開示の理論
B1 半正定値行列の差分プライバシと低ランク近似
ソーシャルネットワークや、ユーザ･アイテムの購入関係を表すネットワーク、人の移動履歴が構成するネットワークなど、サービスを行う特定の事業者が収集可能なネットワークについて、これが外部公開可能であれば有益なデータ解析が可能であると考えられるが、その公開がプライバシ侵害になる可能性がある。
研究テーマ A はプライベートグラフ上のノードが計算主体となり、各ノードが分散し保持する秘密情報を入力として、秘密計算を行うアルゴリズム群を中心とした研究を行った。これに対して、研究テーマ B は、このようなプライベートグラフに関する情報を収集するエンティティがプライベートグラフ自体を外部に公開する際に発生するプライバシの問題に関する理論的考察と、安全に公開するためのアルゴリズムの開発を行った。
具体的には、プライベートグラフを半正定値行列として表現し、これを公開するときに、行列自体ではなく、その低ランク近似生成し、これに、そのランクに比例するスケールでノイズを加えることによって、差分プライバシとよばれる安全性を保証できることを理論的に示した。またそのようなランダム化を実現するためのアルゴリズムを構築し、プライベートグラフを安全に公開するための方法を示した。（発表3、4、論文投稿中）

B2　秘密計算における差分プライバシのための指数メカニズム
研究テーマ A で開発したような様々なアルゴリズムの秘密計算は、入力を秘匿しつつある定められた計算を行い，その出力のみを共有することを可能にするが、出力は入力の情報が正しく反映されるために，秘密計算の出力から入力の情報が漏れる場合が存在する．この研究では、秘密計算の出力に差分プライバシを保証するメカニズムを構築した。具体的には、差分プライバシを与えるメカニズムのクラスを一般的に捉えることができる指数メカニズムの概念を用いる．具体的には，ある秘密計算を指数メカニズムになるよう変形し，その指数メカニズムとしての振る舞いから，与えられる差分プライバシを評価する．本稿では，指数メカニズムの実行に必要なサンプリングを加法準同型暗号を用いた重み秘匿紛失ルーレットにより行う方法について述べる．（発表2、論文投稿予定）

研究テーマ C プライベートグラフにおける分散データマイニング開発環境 Fairy Ring
研究テーマ C では，研究テーマ A で開発したアルゴリズム群のための開発環境および、これを実際にサービスとして　提供するための実行環境として、スマートフォンを利用した秘密計算開発・実行環境 Fairy　Ring を構築した。このフレームワークの特徴としては、従来の秘密計算の応用シナリオでは，計算能力が高く，常時接続・常時稼動を前提とした，複数の対等な高性能サーバ間での実行を想定して設計されてきたが、プライベートグラフ上での計算にあるように、日常生活におけるカジュアルな用途での秘密計算はほとんど検討されてこなかった．近年のスマートフォンは，計算性能や通信速度に優れ，従来の携帯端末では，現実的な計算時間で実行困難であった秘密計算が，実現可能になりつつことから、このフレームワークでは,準同形性公開鍵暗号を利用した少数の基礎的な秘密計算をビルディングブロックとし，これを組み合わせることで，研究テーマ A で開発したアルゴリズム群を含む、様々な個人間の意思決定を簡単に可能にする秘密計算を実装した．（発表1、5、論文投稿予定）

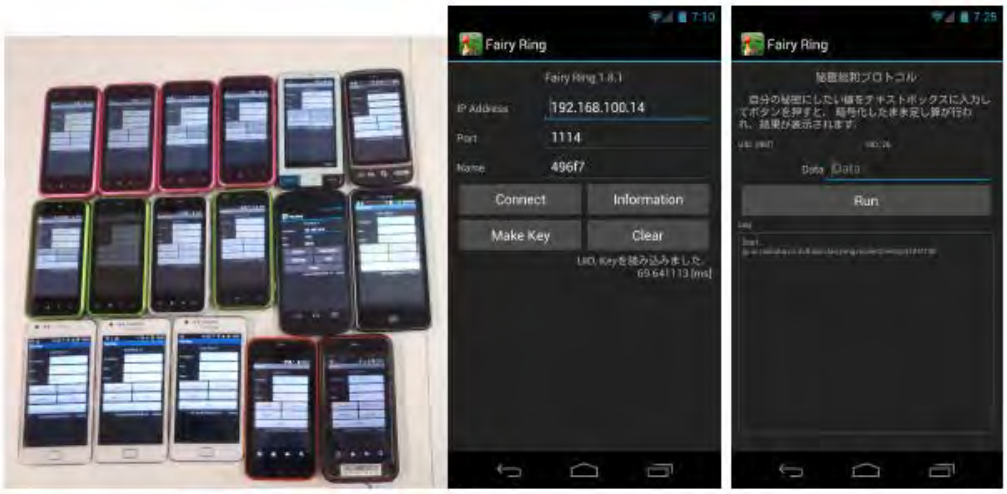

Figure 1 Fairy Ring の動作画面

3.　今後の展開

プライベートグラフを中心概念として、これにまつわる多様なアルゴリズム群とその Android 上での実行・開発環境 Fairy Ring を構築した（研究テーマ A および研究テーマ C）。Fairy Ring については、これをサービスとして利用するユーザの観点について、

- Android Market での公開
- ユーザーインターフェースの改良
- 利用可能な秘密計算の種類の増強

Fairy Ring を開発環境として利用する開発者の観点について、

- フレームワークの整理
- ライブラリの増強

を今後行う予定である。

広くユーザに秘密計算に触れる機会を提供することで、新たなニーズを探りつつ、秘密計算が提供できる新しい知識やサービスや価値について考察を深め、今後の研究開発に役立てたい。

研究テーマ B では、プライベートグラフの安全な開示を実現するための理論の構築を目指した。理論的には新規性の高い方式を提案した。一方、理論を実現するためのアルゴリズムはマルコフ連鎖モンテカルロ法に頼るため、大規模グラフにおいては計算時間に問題がある点が課題である。この点についての解決をはかるとともに、論文公表を行うことが今後の課題となる。

大局的な視点からは、ストリームデータの計算モデルにおける高速な秘密計算モデルの構築を構想している。比較的少数のエンティティを対象とし、複雑であるがデータ量自体は比較的少ないグラフデータとは異なり、大規模データを秘密計算に適用するための新しい計算モデルを開発したい。

4.　自己評価

この研究は、全体としてネットワーク構造情報にまつわるプライバシを安全に利用するための一連の理論と技術を提供とすることをねらいとしていた。この点について、プライベートグラフという基礎概念を元に、主にネットワークマイニングや多人数意思決定のための多くのアルゴリズムをマルチパーティー秘密計算として実行するためのプロトコル群を提供できたことについては、機会学習分野およびセキュリティ・プライバシ分野の両方に取って、貢献が大きかったと評価している。また、現実世界でプロトコルを動作させ、またプロトコル開発を容易にするためのフレームワークである Fairy Ring の開発は、目的からも計算方式からも、秘密計算の開発環境としてユニークであり、今後これを利用した様々な応用を見いだし，秘密計算のコモディティ化・サービス化への道を開きたい．

プライベートグラフ自体の安全な公開のための理論的枠組みについては，新しい方法論を用いるテーマだったこともあり，さきがけ期間中の論文発表に至っていないが，広い応用範囲を持つ基礎技術であり，今後は情報な安全な公開のための理論研究を進める上では，研究者自身としては重要な成果として認識している．現在論文投稿中であり，今後，この領域での貢献を目指す・

5.　研究総括の見解

プライベートなデータを扱うデータマイニングにおいて、プライバシを確保するプライベート・グ

ラフ・モデルの提案、およびその手法についての研究である。

ネットワーク環境でプライバシを安全に利用するという点について、プライベートグラフという基礎概念を元に、主にネットワークマイニングや多人数意思決定のための多くのアルゴリズムをマルチパーティー秘密計算として実行するためのプロトコル群を提供したことは評価できる。

また、現実世界でプロトコルを動作させ、またプロトコル開発を容易にするためのフレームワークである Fairy Ring を開発し、これを利用して様々な応用を見いだす体制を開発しており、今後に期待したい。

6. 主な研究成果リスト

(1)論文(原著論文)発表

1. Hiromi Arai, Jun Sakuma: Privacy Preserving Semi-supervised Learning for Labeled Graphs. Lecture Notes in Computer Science 6911 (Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2011), pp. 124-139.
2. 森井正覚, 佐久間淳, 佐藤一誠, 中川裕志:統合したグラフのプライバシ保護リンク解析. 情報処理学会論文誌, Vol.50 TOD 4(2). pp.52-60. 2011.
3. Jun Sakuma, Shigenobu Kobayashi, Large-scale k-means clustering with user-centric privacy-preservation, Knowledge and Information Systems, 2010, No. 2, pp. 253-279.
4. Bin Yang, Hiroshi Nakagawa, Issei Sato, Jun Sakuma: Collusion-resistant privacy-preserving data mining, Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 483-492.
5. Jun Sakuma, Hiromi Arai, Online Prediction with Privacy, Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 935-942.

(2)特許出願

研究期間累積件数:0件

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

主要な学会発表

(1) 照屋唯紀, 川本淳平, 佐久間淳, 強い安全性を仮定したスマートフォン向けマルチパーティー秘密計算フレームワーク, 2013 年暗号と情報セキュリティシンポジウム(SCIS2013), CDROM 予稿集 pp. 1-8.

(2)照屋唯紀, 川本淳平, 佐久間淳, 指数メカニズムによる秘密計算の出力における差分プライバシの保証, 2013 年暗号と情報セキュリティシンポジウム(SCIS2013), CDROM 予稿集 pp. 1-8.

(3)佐久間 淳, 半正定値行列の差分プライバシー, 第 15 回 情報論的学習理論ワークショップ(IBIS 2012)(第 11 回 IBISML 研究会), pp. xx-yy.

(4)佐久間 淳 , スペクトラル差分プライバシーに基づくプライバシ保護推薦アルゴリズム, 第 26 回人工知能学会全国大会, 2012, pp. CDROM 予稿集 pp. 1-4.

(5)青木 良樹, 佐久間 淳, 照屋 唯紀, どこでも秘密計算フレームワーク Fairy Ring マルチメ

ディア、分散、協調とモバイル DICOMO2012 シンポジウム, pp. 1618−1627.

受賞

（1）2009 年度(第 23 回)人工知能学会全国大会優秀賞「秘密のリンク構造を持つグラフのリンク解析」

（2）情報論的学習理論と機械学習（IBISML2010）研究会 Honorable Mention「ラベル付きグラフに対するプライバシを保護した半教師付き学習法」

（3）情報論的学習理論と機械学習（IBISML）2010 年度研究会賞 電子情報通信学会 第一種研究会「情報論的学習理論と機械学習研究会」「オンライン予測におけるプライバシ保護」

（4）コンピュータセキュリティシンポジウム 2012 優秀論文賞,「Bloom Filter を用いた積集合サイズのベイズ推定とそのプライバシー保護疫学調査への応用」

（5）平成 24 年度 日本データベース学会上林奨励賞

プレスリリース

（1）秘密計算による化合物データベースの検索技術, 2011.11.1
http://www.aist.go.jp/aist_j/press_release/pr2011/pr20111101/pr20111101.html