様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２９　年　　６　月　２５　日現在

機関番号：　１２１０２
研究種目：　若手研究(A)
研究期間：　2012～2016
課題番号：　２４６８００１５

研究課題名（和文）大規模非構造型秘密情報のためのアウトソース型プライバシ保護データマイニング基盤

研究課題名（英文）Outsourcing of privacy preserving data mining for large-scale non-structured information

研究代表者

佐久間　淳（SAKUMA, Jun）

筑波大学・システム情報系・教授

研究者番号：９０３７６９６３

交付決定額（研究期間全体）：（直接経費）　　20,200,000円

研究成果の概要（和文）：大規模データを扱うアウトソース型プライバシ保護データマイニングにおいては、計算過程の秘密を保護する秘密計算と、計算結果からの入力データの推測を防ぐ差分プライバシ等の統計的プライバシの保証の二つが必要である。この研究では、様々な統計やデータマイニングのタスクにおいて、秘密計算と統計的プライバシーを保証する基礎研究を行った。具体的には、秘密計算についてはベクトル・行列積、積集合サイズ、統計的検定の秘密計算などを開発した。統計的プライバシーについては外れ値検出の差分プライバシーや出力の区間化による安全性保証などの研究を行った。

研究成果の概要（英文）：For outsourcing of privacy-preserving data mining with large-scale data, we consider two tasks: secure computation and statistical privacy. When the data contains private information and is distributed over multiple locations, the former provides a methodology that computes a specified function with the distributed data sources with keeping the secrecy of data in the process of computation. The latter aims to prevent inference of input from the computed results. In this research project, we developed studies on secure computation and statistical privacy (including differential privacy) for various statistics and data mining tasks. More specifically, we developed secure computation of vector matrix multiplication, set intersection cardinality, statistical testing, and so on. For statistical privacy, we studied differential privacy of outlier detection and privacy protection by interval release.

研究分野：　プライバシー，機械学習

キーワード：　準同型暗号　差分プライバシー　プライバシ保護データマイニング　アウトソーシング

１．研究開始当初の背景
個人のプライバシ情報や組織の機密情報等、公開されないが潜在的には知識発見に利用可能な秘密情報の規模は、公開情報よりもはるかに大きい。 知識発見は対象データが解析者に対して閲覧可能であることを前提とするが、公開されない秘密情報も、暴露のリスクなしに知識発見の対象とすることができれば、その利益は極めて大きい。

プライバシ保護データマイニング(privacy-preserving data mining、PPDM)とは、秘密情報が複数の主体に分散しているときに、これを他に開示せずに、安全に知識発見するための秘密計算技術として発展してきた。

２．研究の目的
本研究は、情報提供者が文書、画像、センサデータ等の大規模非構造型秘密情報を保持し、情報利用者が秘密の解析条件/クエリを保持する時に、両者がこれを互いに秘密にしたまま、情報利用者のリクエストに基づきデータを解析し、結果を実用的な時間で応答する計算基盤を構築することを目的とする。

非構造型情報解析の基礎的演算として、集合演算、類似度評価、パターン照合に焦点をあてる。 これらは単独ではキーワード検索、類似検索等の情報検索に必須の計算であり、またこれらをビルディングブロックとすることで、大規模非構造型秘密情報を対象とする多様な PPDM アルゴリズムの構築が可能になるためである。

３．研究の方法
大規模データを扱うアウトソース型プライバシ保護データマイニングにおいては、計算過程の秘密を保護する秘密計算と、計算結果からの入力データの推測を防ぐ差分プライバシ等の統計的プライバシの保証の二つが必要である。ここでは、様々な統計やデータマイニングのタスクにおいて、秘密計算と統計的プライバシを保証する基礎研究を、様々な側面から行った。

４．研究成果

(1) 秘密計算に関する研究:
多様なデータ構造を用いた基礎的な計算について秘密計算を開発した。ここでは、代表的な二つの成果を説明する。

(1)-1 ベクトル･行列積の秘密計算について
行列とベクトルの積は様々なデータ解析において広く用いられている基本的な演算の 1 つである。 例えば各種統計量の評価、固有値計算、文字列パターン照合などの計算は行列とベクトルの積のみで評価できる。
準同型暗号はあらかじめ与えられた回数の暗号文の演算(加算および乗算)を許すが乗算の回数が大きくなった場合、その計算コストが増加することが知られている。入力データが大規模である場合その計算コストは無視できないことから、効率化が必要である。

我々は準同型暗号で暗号化された行列とベクトルの積を効率的に評価する手法を考案した。 具体的には行列やベクトルの要素ごとに暗号化し、そのそれぞれついて行列ベクトル積に必要な演算を行う代わりに、行列やベクトルをそれぞれ単一の暗号文としてパッキングし、行列ベクトル積を準同型暗号上の一回の演算として処理する手法を提案した。

提案手法では、行列とベクトルはそれぞれ一つの暗号文にエンコーディングされ、これら二つの暗号文同士の積のみで行列ベクトル積の計算が可能である。 AES 80bit の安全性において、準同型暗号により暗号化された 16×16、64×64 の行列ベクトル積を評価したところ、案手法の計算時間はそれぞれ 0.01s、0.03s であり、既存手法よりも 150 倍以上高速であることを示すことができた。

(1)-2 積集合サイズの秘密計算について
二つのリストの要素を秘匿したままでその共通要素数、すなわち、積集合の大きさのみを評価する秘密計算を提案した。この 問題は、セキュアな生体認証、プライバシ保護データ　タマイニング、セキュア疫学調査などの多くの応用例に共通する基本的な要素技術の一つである。

共通の ID がない問題に対するナイーブな解は 、衝突困難性を満たした暗号学的ハッシュ関数を用いて名前をハッシュ値に変換することである。しかし、十分な精度を保証するためには、比較する集合の大 きさ n に対して $O(n^2)$ の値域のハッシュ値を用いる必要があり大規模データに適用できない。

そこで、集合の所属度を与える近似的なデータ構造 である Bloom Filter(BF) を導入し、プライバシを 保護したまま安全に共通要素数を近似するアプローチを考えた。また、小さなビット数の BF を繰返し求めて秘匿内積プロトコル を行うことで、計算効率と推定精度を向上させた。

この研究以外にも、同様の計算技法を用いて、カイ二乗検定や、正確ロジスティック回帰、ロジスティック回帰などに関する秘密計算アルゴリズムを提案した。

(2) 差分プライバシに関する研究:

外れ値検出は、異常検知などを始めとするデ

ータマイニングにおける重要なタスクのひとつである。外れ値の目的は、外れ値となるインスタンスの特定であるが、外れ値が明らかになることによって、データに個人情報が含まれる場合、外れ値の特定はプライバシの侵害につながる。この研究では、外れ値の個数を求めるクエリについて、差分プライバシを保護する手法を研究した。

差分プライバシを達成するために外れ値の個数検出クエリ(以降、外れ値個数クエリと呼ぶ)結果に加えるノイズを抑制するために平滑敏感度を用いた。この研究の貢献は次の三つである。

一つ目は、差分プライバシを達成するための外れ値個数クエリに対する大域的敏感度の下限を示し、一般的に解析が困難な局所的敏感度の上限、平滑敏感度の上限を理論的に示した。

二つ目は、計算量を改善した。外れ値個数クエリに対する局所敏感度と平滑敏感度の計算量は指数計算量となる。そこで、指数計算量を緩和するアルゴリズムを提案し、次元数をdとするとO(d^2)程度の計算量に抑えられるようにした。

三つ目は、平滑敏感度の上限を用い、高い有用性を持つ出力結果を実際に得られることを実験的に示した。

この研究以外にも、統計的プライバシの文脈において、秘密ベクトルと公開ベクトルの内積値の公開において、秘密ベクトルの値が推測されるリスクの効率的な計算法や、秘密ベクトルの値が推測されにくくなる数値の適応的区間化手法などについての研究を行った。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 8 件）

1. Wenjie Lu, Shohei Kawasaki, Jun Sakuma. Using Fully Homomorphic Encryption for Statistical Analysis of Categorical, Ordinal and Numerical Data, Proc. Of Network and Distributed System Security Symposium (NDSS2017) online proceedings (15 pages). 査読有

2. Wenjie Lu, Yoshiji Yamada, Jun Sakuma. Privacy-preserving genome-wide association studies on cloud environment using fully homomorphic encryption, BMC Medical Informatics and Decision Making 2015, 15(Suppl 5):S1. 査読有

3. Kazuto Fukuchi, Toshihiro Kamishima, Jun Sakuma, Prediction with Model-Based Neutrality. IEICE Transactions 98-D(8): 1503-1516 (2015). 査読有

4. Rina Okada, Kazuto Fukuchi, Jun Sakuma, Differentially Private Analysis of Outliers, Proc. of ECML/PKDD (2) 2015: 458-473. 査読有

5. David A. duVerle, Shohei Kawasaki, Yoshiji Yamada, Jun Sakuma, Koji Tsuda: Privacy-Preserving Statistical Analysis by Exact Logistic Regression. Proc. of IEEE Symposium on Security and Privacy Workshops 2015: 7-16. 査読有

6. Wenjie Lu, Yoshiji Yamada, Jun Sakuma: Efficient Secure Outsourcing of Genome-Wide Association Studies. Proc. of IEEE Symposium on Security and Privacy Workshops 2015: 3-6. 査読有

7. Kazuto Fukuchi, Jun Sakuma: Neutralized Empirical Risk Minimization with Generalization Neutrality Bound. Proc. of ECML/PKDD (1) 2014: 418-433. 査読有

8. Hiroaki Kikuchi, Jun Sakuma: Bloom Filter Bootstrap: Privacy-Preserving Estimation of the Size of an Intersection. Journal of Information Processing 22(2): 388-400 (2014). 査読有

〔学会発表〕（計 8 件）

1. 草野 光亮 , 竹内 一郎 , 佐久間淳, 線形モデルにおける安全な予測値公開メカニズムの提案とその疾患リスク予測モデルへの適用, コンピュータセキュリティシンポジウム 2016 論文集, vol.2, pp.1207 - 1214, 2016 年 10 月, 秋田キャッスルホテル(秋田県・秋田市).

2. 陸文杰, 川崎将平, 佐久間淳, 準同型暗号による統計解析のアウトソーシング I: 記述統計量, コンピュータセキュリティシンポジウム 2015 論文集, vol.3, pp.266 - 273, 2015 年 10 月, 長崎ブリックホール(長崎県・長崎市).

3. 川崎将平，陸文杰，佐久間淳，準同型暗号による統計解析のアウトソーシング II: 予測モデリング,コンピュータセキュリティシンポジウム 2015 論文集，vol.3，pp.274 - 281，2015 年 10 月，長崎ブリックホール(長崎県・長崎市)．

4. 荒井ひろみ,津田宏治,佐久間淳，ゲノム検査結果の開示によるプライバシ侵害の評価，コンピュータセキュリティシンポジウム 2015 論文集，vol .3，pp.1258-1265，2015 年 10 月，長崎ブリックホール(長崎県・長崎市)．

5. 佐久間 淳、陸 文傑、西出 隆志、國廣 昇，プライバシポリシー執行を保証する関数評価,コンピュータセキュリティシンポジウム(CSS)2015，コンピュータセキュリティシンポジウム 2015 論文集，vol.3 pp. 40 - 47，2015 年 10 月，長崎ブリックホール(長崎県・長崎市)

6. チャンクワンカイ，福地一斗，佐久間淳 クラウドセンシングにおける差分プライバシを保証した線形回帰モデル学習 情報論的学習理論ワークショップ (IBIS2014) 信学技報，vol. 114，no. 306，IBISML2014-47，pp. 95-102，2014 年 11 月，名古屋工業大学(愛知県・名古屋市)

7. 陸文杰，佐久間 淳 Somewhat 準同型暗号上の行列ベクトル積のための効率的なパッキング手法 2015 年暗号と情報セキュリティシンポジウム(SCIS2015) 2F2-2 (8 pages)，リーガロイヤルホテル小倉(福岡県・北九州市)．

8. Shuang Wu，Junpei Kawamoto，Hiroaki Kikuchi,Jun Sakuma，Privacy-preserving Online Logistic Regression Based on Homomorphic Encryption，情報論的学習理論ワークショップ(IBIS2013) 信学技報,vol. 113，no. 139，IBISML2013-10，pp. 67-74，2013 年 7 月，東京工業大学蔵前会館(東京都・目黒区)．

NII 書誌 ID(NCID):ISSN 1882-0840

〔図書〕(計 1 件)

佐久間 淳，解析におけるプライバシ保護，講談社．

６．研究組織

(1)研究代表者

佐久間 淳（SAKUMA JUN）

筑波大学・システム情報系・教授

研究者番号：90376963