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

令和　元 年　 ６ 月 １２ 日現在

機関番号：１４３０１
研究種目：挑戦的萌芽研究
研究期間：2016～2018
課題番号：１６Ｋ１２４３７

研究課題名（和文）交渉による時系列データのプライバシ保護に関する研究

研究課題名（英文）A study on privacy protection of time-series data through negotiation

研究代表者
吉川　正俊（Yoshikawa, Masatoshi）

京都大学・情報学研究科・教授

研究者番号：３０１８２７３６

交付決定額（研究期間全体）：（直接経費）　　2,600,000円

研究成果の概要（和文）：プライバシ情報は適切に保護すると共にそれらを収集，解析し公益に資することも重要である．本研究では，パーソナルデータの売買を行う市場機構において各個人が自己のパーソナルデータを開示する程度の上限を指定できる取引機構を開発した．
また，個人の日常生活により生成され有用性を持つパーソナルデータの多くは時系列データであるため，そのプライバシ保護に関する研究を行った．現在，プライバシ情報漏洩リスクを数学的に厳密に表現可能な差分プライバシが広く研究されている．しかし，差分プライバシは，静的な個別データを対象として開発されたため，差分プライバシを時系列データに適用可能できるように拡張した，

研究成果の学術的意義や社会的意義
差分プライバシはプライバシ漏洩の程度を数学的に定量化できるため，2006年に提案されて以来，活発な研究が行われてきた．しかし，時系列データの場合はデータ間に相互依存関係があるため，差分プライバシの考え方を適用することは困難であった．我々の研究では隣接する時刻のデータの依存性に応じて実際のプライバシ漏洩の程度を始めて数量的に表現した．また，パーソナルデータ市場でこれまで考えられていなかった各個人がパーソナルデータ公開に対する許容度を指定できる機構を始めて導入した．これらの成果はパーソナルデータ市場実現のための理論的支柱を与える意義を持つ．

研究成果の概要（英文）：It is important to collect, analyze, and utilize personal data for pubic welfare as well as protecting privacy. In this research, we have developed a market mechanism for personal data exchange which allows each individual to specify an upper limit of the degree of disclosure of her personal data.
Recently, differential privacy, which can express privacy information leakage risk mathematically, is widely studied. However, since differential privacy was developed for static data, We have extended the notion of differential privacy to be applicable to time series data.

研究分野：データベース

キーワード：プライバシ保護　経路情報　パーソナルデータ市場　差分プライバシ

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９、ＣＫ－１９（共通）

１．研究開始当初の背景

プライバシ情報の保護は重要な研究課題であり，数学的に厳密な扱いが可能な差分プライバシ (differential privacy) の研究が盛んに行われている．しかし，差分プライバシ技術を一般化し，日常生活において生成される時空間系列データなどの実応用データに適用する研究はほとんど見られない．我々は，従来，データベース内の個別データを対象として開発されて来た差分プライバシを移動軌跡データに適用する研究を行った．しかし，この技術をより実用的なものとするためには，攻撃者の事前知識に対する現実的な仮定の設定，プライバシ保護レベルと保護データの効用のトレードオフ，実データへの適用によるフィードバックなどいくつかの非連続的な飛躍が必要であった．

２．研究の目的

プライバシ情報は適切に保護すると共にそれらを収集,解析し公益に資することも重要である．本研究では，各個人が日常的に生成する時系列プライバシ情報の価値評価を課題とする．プライバシ情報漏洩リスクを数学的に厳密に表現可能な差分プライバシを拡張する形で基礎研究を進めると共に，応用の具体的な対象として歩行時の情報保護を設定することにより，基礎と応用の相互フィードバックの形で研究を推進する．統計データ公開時の個人ごとのプライバシリスク評価やデータの依存関係を考慮した情報の価値評価を行うことにより，各個人の要求に応じた保護基準でプライバシ情報を収集，解析し，公共的利益を創出するための基盤技術を開発することができる．

３．研究の方法

プライバシ保護の理論的研究を補完する形で必要なデータを実験で取得することにより研究を進めた．

①　時系列データのための差分プライバシ技術の開発

個人の日常生活により生成されるプライバシデータは多くが時系列データである．そこで，生年月日，血液型などの静的な個別データを対象として開発された差分プライバシを半無限の時系列データに適用可能できるように拡張した.現実の状況を反映するため,特に次の点の拡張,精緻化を行った．(a)各利用者のプライバシ保護レベルに対する要求には個人差があるため，それを反映したプライバシ保護機構とする．(b) 時系列データには通常，マルコフ性など時間的依存関係が存在するため，ある対象期間の統計値の開示が，その対象期間より過去や未来のデータに対する情報をある程度漏洩することになる．このような時間依存性によるプライバシ漏洩リスクを明らかにする．

②　歩行時情報の収集と個人特性分析

歩行時の各種データを日常生活において生成される時系列情報の代表例と取り上げ，実際の歩行時の位置情報や生体情報を収集することにより，異なる時系列情報間の相関関係を明らかにする．例えば，位置情報は道路の勾配情報を決定し，勾配情報は心拍情報に影響を与えるが，その程度は十分に明らかになっていない．代表的な相関関係を持つ複数の時系列データを集めたテストデータは存在しないため，このような実際的なデータを対象とし差分プライバシ機構の理論研究を進めるためにも実データの収集を行い，生体情報については個人特性と道路混雑度や勾配などの環境特性の相互関係を明らかにする．

③　プライバシ情報利用者の意志決定支援

多数のプライバシ情報を収集し統計解析を行うことにより付加価値を提供するプライバシ情報利用者の立場からは,たとえば「平日 9 時から 5 時までの昼休み以外の位置データと心拍データ」のような条件を満足し，しかもある閾値以下の雑音が保証された統計情報が必要となる．通常，差分プライバシの理論研究ではプライバシ保護機構を提案し，プライバシ予算に対する統計値の効用を実験的に示すことが多い．プライバシ情報利用者の要求を満足するためには，逆に雑音レベルからそれを満足するプライバシ予算を決定する必要があるため従来とは逆の問題となる．この逆問題の解法を,問合せに集約関数を許すデータ起源(provenance) 問題などの拡張により開発する．

４．研究成果

まず，個人時空間データを活用するための厳密で柔軟なプライバシ保護の枠組みに関する研究を行った．具体的には，プライバシ情報漏洩リスクを数学的に厳密に表現可能な差分プライバシに基づき，時間的な相関があるデータのためのプライバシモデルを開発すると共に，データ漏洩の程度を定量化する手法の開発，データ保護手法の開発を行った．既存の差分プライバシはデータには時間的な相関関係がないと仮定する。しかし、時系列データには通常，時間的依存関係が存在することが多いため，ある対象期間の統計値の開示がその対象期間より過去や未来のデータに対する情報をある程度漏洩することになる．その上、一般に攻撃者はこのような相関関係を容易に取得できる。本研究では、攻撃者がマルコフモデルでモデル化できる時間的な相関関係の知識を持つ場合でも、差分プライバシを達成する方法を提案した。具体的には、

既存研究で提案されたプライバシ保護機構を時間的な相関があるデータを保護できる機構に転換する方法を設計した。このため，提案した枠組みは高い汎用性を持つ．この成果の論文は，IEEE ICDE 国際会議及び IEEE TKDE 論文誌に発表した．

歩行時情報の収集については，道路歩行時の歩行者の生体情報を収集するとともに，道路歩行時のビデオを視聴する模擬環境による生体情報収集も併用し，HRV，GSR，EEG などの各種生体情報をもとに感情を推定する実験を行った．

次に，差分プライバシの概念に基づいて地図上の場所を曖昧化する手法である Geo 識別不能性を拡張することにより，自宅の位置などのような移動経路の端点を曖昧化する手法を開発した．これにより，自宅などプライバシ保護が必要な場所は保護程度を高くする一方で駅の近くなど公共性が高い地域の移動経路情報は少ない誤差で収集することが可能となる．

さらに，多数の個人が自らの移動経路，購買履歴，心拍データなどの時系列データに対し個人識別ができない程度に雑音を加えた後に売却し，会社などはそのような大量のパーソナルデータの統計情報を購入できるパーソナルデータ市場に関する研究を行った．各利用者のプライバシ保護レベルに対する要求には個人差があるため，それを反映したパーソナルデータ市場の価格機構に関する研究を行った．パーソナルデータ市場の価格付けに関する研究では，個人のパーソナルデータ保護に対する要求の程度を調査するため，クラウドソーシングにより大規模アンケートを実施し，概ね保護要求の程度が異なる二つの利用者群が存在することを確認した．また，パーソナルデータ市場において各個人が市場機構とプライバシ漏洩の程度と対価に関する契約を結び，データ購買者は要求するデータに関する問合せと予算を提示する場合に，各個人のプライバシ保護要求を満たしながら，購買者に対しては低価格でしかも雑音印加の程度を低く抑えるための手法を開発した．この成果は ACM SIGIR eCommerce ワークショップで発表した．

位置データは様々な応用が可能であり，その有用性に基づくニーズから，それらを売買する市場の整備も始まっている．しかし一方で，正確に位置情報を公開することには個人が特定されてしまうなどのプライバシリスクも伴う．そのためには，まず，位置情報プライバシ選好，すなわち“ある場所と時間において，位置情報を公開するかどうか”を指定することが必要となるが，ユーザにとって，このような選好を全ての場所と時間に関して把握し指定することは難しい．そこで本研究ではユーザの意思決定を支援するために，オンラインショッピングサイトなどで利用される商品推薦の概念を利用し，位置情報プライバシ選好を推薦するシステムを提案した．実験によって，予測の正誤，真の評価値と予測の誤差について評価を行い，提案手法がプライバシ保護を満足しながら高い有用性を担保していることを示した．

また，位置情報プライバシ保護のために，従来は自分の位置情報を送る前に，ランダムな雑音を加えることでプライバシを保護する摂動法が提案されているが，道路ネットワークを考慮していなかった．近くのレストランを探すなどの近傍検索などの位置情報サービスは，道路ネットワークを用いた方が近傍を正確に表現できることから，より良いサービスを提供可能である．そこで，このような位置情報サービスを対象とし，プライバシ保護の強さと雑音付加後の情報の有用性を両立させた摂動法の手法を開発した．

研究期間全体を通じ，各ユーザ個人の要求に応じたプライバシ保護基準でパーソナルデータを収集，解析し，公共的利益を創出するための基盤技術を開発した．

５．主な発表論文等

〔雑誌論文〕（計 2 件）

① Yang Cao, Masatoshi Yoshikawa, Yonghui Xiao, Li Xiong: "Quantifying Differential Privacy in Continuous Data Release Under Temporal Correlations," IEEE Transactions on Knowledge and Data Engineering, Volume: 31, Issue:7, pp. 1281-1295, July 2019. DOI: 10.1109/TKDE.2018.2824328

② Yang Cao, Li Xiong, Masatoshi Yoshikawa, Yonghui Xiao, Si Zhang: "ConTPL: Controlling Temporal Privacy Leakage in Differentially Private Continuous Data Release," PVLDB 11(12): pp. 2090-2093, 2018. DOI: 10.14778/3229863.3236267

〔学会発表〕（計 2 件）

① 鍋谷 優斗, 吉川 正俊: "データベース問合せ結果販売利益の分配に関する研究", 第 10 回データ工学と情報マネジメントに関するフォーラム （第 16 回日本データベース学会年次大会）, G1-6, 2018 年 3 月.

② Rachana Nget, Yang Cao, Masatoshi Yoshikawa: "How to Balance Privacy and Money through Pricing Mechanism in Personal Data Market," roceedings of the SIGIR 2017 Workshop On eCommerce co-located with the 40th International ACM SIGIR

Conference on Research and Development in Information Retrieval, August 2017.

③ Juyeong Park, Masatoshi Yoshikawa, Hiroyuki Kato: Cell-Based Provenance for Scientific Data. ACM/IEEE Joint Conference on Digital Libraries (JCDL 2017): pp. 289-290, June 2017.

④ Yang Cao, Masatoshi Yoshikawa, Yonghui Xiao, Li Xiong: "Quantifying Differential Privacy under Temporal Correlations," Proceedings of the 33rd IEEE International Conference on Data Engineering (ICDE 2017), pp. 821-832, April, 2017. DOI: 10.1109/ICDE.2017.132

〔図書〕（計　　件）

〔産業財産権〕
○出願状況（計　　件）

名称：
発明者：
権利者：
種類：
番号：
出願年：
国内外の別：

○取得状況（計　　件）

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕
ホームページ等

６．研究組織

(1)研究分担者
研究分担者氏名：浅野　泰仁
ローマ字氏名：ASANO Yasuhito
所属研究機関名：京都大学
部局名：情報学研究科
職名：特定准教授
研究者番号（8 桁）：20361157

(2)研究協力者
研究協力者氏名：曹　洋
ローマ字氏名：CAO Yang