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　３０　年　　６　月　１９　日現在

機関番号：　６２６０３
研究種目：　基盤研究(C)（一般）
研究期間：　2015～2017
課題番号：　１５Ｋ００１９５

研究課題名（和文）状態空間モデルに基づく統計的アクセス制御手法の研究

研究課題名（英文）Statistical access control based on a state-space model

研究代表者
南　和宏（Minami, Kazuhiro）

統計数理研究所・モデリング研究系・准教授

研究者番号：１０５７９４１０

交付決定額（研究期間全体）：（直接経費）　　3,600,000円

研究成果の概要（和文）：近年、様々なビックデータが利活用されており、その主な対象は我々の行動履歴に関する時系列データである。しかし行動履歴データは時空間の相関性が高く、時系列データの安全な公開には新しいアクセス制御手法が不可欠である。本研究では、アクセス制御の問題を状態空間モデルにおける状態推定の逆問題として定式化し、内部状態の機密性を保証する観測モデルの統一的設計手法を確立した。

研究成果の概要（英文）：In recent years, various kinds of big data, most of which are concerned with our behavior history, has been actively utilized for data analysis. Since such historical data on our behaviors contains sensitive information on our privacy, we need a method for controlling access to sensitive information. However, time-series data on our activities tends to have high temporal and spatial correlations, we need a new access-control method for protecting time-series data under the presence of various inference attacks. In this research, we formulated the problem of access control as the inverse problem of state estimation in the state space model, and established a unified framework for designing a proper enforcement mechanism for access control.

研究分野：　情報セキュリティ

キーワード：　アクセス制御　匿名化　状態空間モデル　隠れマルコフモデル　統計開示抑制　最適化

１．研究開始当初の背景

現在、我々の行動履歴に関する情報はデジタル化され、様々なサービス提供者に収集、分析されている。しかし我々の行動履歴は多くの個人情報を含み、その機密性の保護が重要である。特に行動履歴データを２次利用の目的で流通する場合、情報の受け手の信頼性に応じて提供する情報の範囲を適切に制限するアクセス制御が必要である。

しかし元の時系列データ間の相関性が強い場合、一部のデータを秘匿しても他の公開したデータから機密情報が推測される問題が存在する。つまり、従来のように秘匿したい情報のみ非公開にするだけではなく、統計的な機密情報の推論を考慮した新しいアクセス制御技術の確立が必要となってきた。

２．研究の目的

本研究の目的は、公開された情報から非公開の機密情報の推論するプロセスを統計的な情報空間モデルでモデル化し、時系列多次元データを安全に公開するためのアクセス制御手法を確立することである。具体的には、機密情報の推論プロセスを統計的な状態空間モデルにおける観測情報から内部状態の推定問題として定式化し、秘密の情報を保護するため、公開可能な情報の一部を敢えて非公開とする拡大的なアクセス制御を行う新規手法の確立を目指した。

３．研究の方法

研究は、1) 時系列データのアクセス制御に関する理論的なセキュリティモデルの構築、2) そのモデルにおける安全性の要件を満たすアクセス制御手法の実装、3) 実データに提案手法を適用したときにどの程度有益な情報を公開できるか、データ効用を指標とする実証的評価、の３つからなる。セキュリティモデルにおける安全性は、公開データが与えられたときの、非公開データが取りうる値の条件付き確立を隠れマルコフモデルの内部状態推定問題の枠組みで定式化した。また評価の対象とするデータ・セットは、一般に研究用に公開されている位置情報セットと公的調査データの分析において、一般的な集計表等の表データを対象とした。

４．研究成果

本研究は主に位置情報を対象とし、ユーザーが移動に関する制約、習慣をマルコフチェーンでモデル化し、隠れマルコフモデルによる安全性の定式化を行い、さらに複数ユーザーの移動履歴の匿名化データを公開する場合の安全性研究にモデルを確立した。

通常、匿名化処理を行う場合、k 個以上のレコードが同じ値をとるようにデータを一般化する k-匿名化という処理を行うことが一般的である。ただし、位置情報に対する k-匿名化の適用は、ユーザーの移動パターンに関する習慣、道路形状の制限等の原因によりうまく機能しない。例えば、容易に入手可能な道路地図の情報を用いれば、位置情報の粒度を粗くした k-匿名化データから元の位置情報軌跡が容易に復元される場合が多い。図１は二人のユーザーの移動軌跡を 2-匿名化した例を示す。二人の移動軌跡を同一に見せるため、二人の位置が同じ太線の矩形に入るように位置情報を一般化している。しかし、破線で示す道路形状が重ねると、二人の元の移動軌跡が復元されてしまう。

元データ

ユーザー	時刻 t_1	時刻 t_2
田中	6	4
鈴木	10	12

一般化処理

2匿名化データ

仮名	時刻 t_1	時刻 t_2
A	{6, 10}	{4, 8, 12}
B	{6, 10}	{4, 8, 12}

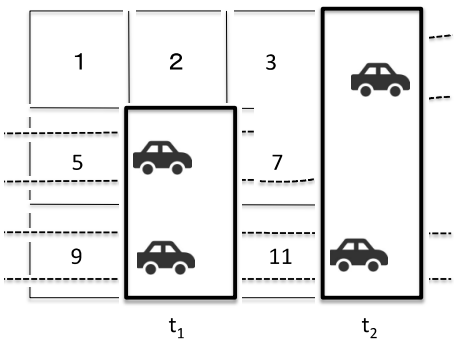


図 1.道路情報を用いた推論

また一般に匿名加工のアルゴリズムは匿名加工による情報損失を最小化することを目的とする最小化原理に基づいて設計されることが多い。そのような原理をしることで、一般化された位置領域のどの部分にユーザーが位置するかを推測することもできる。

このような位置の問題点を統一的に扱うため、本研究では位置情報の時系列データを確率的な状態遷移と捉えるマルコフ過程でモデル化した。つまり各ユーザーの位置情報データは遷移行列から確率的に生成される時系列データと捉える。このデータは一般にユーザーの機密情報を含むため、適切な匿名化加工が必要になる。我々はこの匿名加工を隠れマルコフモデルにおける記号出力行列としてモデル化した。この行列が元データにどのように確率的に変換されるかを記述する行列である。各行が元データの値、各列は変換後に取りうる値に相当する。そして、要素 (i,j)は、i 行目の元の値が j 番目の値に変換される確率を格納する。

この２つを図２に示す隠れマルコフモデ

ルで統合し、匿名化データの生成プロセスをモデル化することで、匿名化データの安全性を公開可能な観測情報を与えられたときに、非公開の内部状態を推定する条件付き確率として定式した。

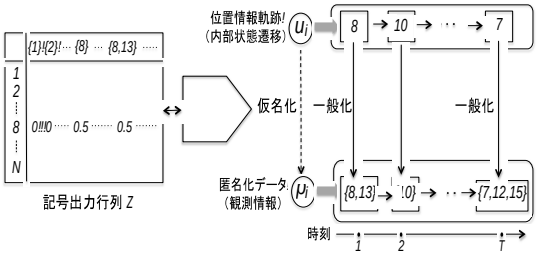
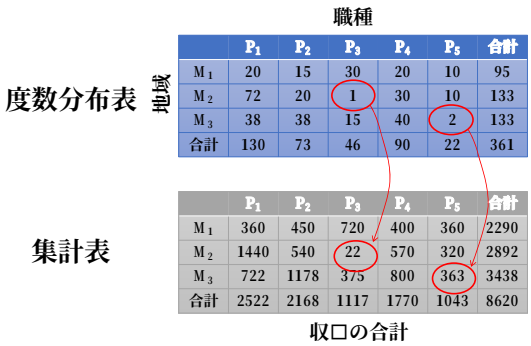


図 2. 隠れマルコフモデルによる匿名化処理のモデル化

さらに位置情報の匿名化データの安全性を検証するツールを既存のビタビアルゴリズムを拡張することで実装した。今後の課題としては、安全な匿名化加工を設計するための設計手法の確立が挙げられる。

本研究は、さらに公的統計の分野で一般的な表データに関するプライバシー保護の研究を行った。表データの場合、セル単位で集計する調査客体の数が少ないと機密情報が漏洩するリスクが高まる。例えば、図 3 に示す度数分布表と集計表では、セル (M_2, P_3) の調査客体数は 1 なので、集計表のセル値 (22) からその客体の収入が分かる。また客体数 2 のセル (M_3, P_5) の場合、2 つの調査客体の一方は、集計表のセル値から自身の収入を差し引くことでもう一つの客体の収入が算出できてしまう。



	P1	P2	P3	P4	P5	合計
M1	20	15	30	20	10	95
M2	72	20	1	30	10	133
M3	38	38	15	40	2	133
合計	130	73	46	90	22	361

	P1	P2	P3	P4	P5	合計
M1	360	450	720	400	360	2290
M2	1440	540	22	570	320	2892
M3	722	1178	375	800	363	3438
合計	2522	2168	1117	1770	1043	8620

図 3. 表セルからの機密情報漏洩

したがって、あるしきい値より小さい度数のセルは情報漏えいリスクが高いと判断し、そのセル値を秘匿する必要がある。ただし、表データは、行計、列計に関する線形の関係式を内包し、秘匿したセル値の復元が可能なため、追加で他のセル値を秘匿する 2 次秘匿処理が必要となる。

2 次秘匿するセルを決定するには、秘匿セル値に十分な不確定性を確保しつつ情報損失（例えば、秘匿セル数）を最小化する最適化問題を解く必要がある。この 2 次秘匿処理を自動化するツールを *R* 言語で開発した（図 4）。

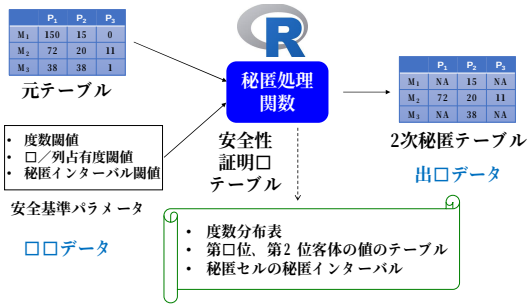


図 4. *R* 言語による秘匿処理ツール

このツールは *R* 言語の関数として提供され、*R* の分析環境で作成した表データに対して、様々な機密性ルール（(n,k)-ルール、p%ルール等）に基づく 1 次秘匿処理を実行する。さらに 2 次秘匿処理の実行時には、2 次秘匿した表データの作成に加え、その安全性を証明するための追加の表データを生成する。

現在、総務省は、国勢調査等の様々な公的調査の調査データの学術研究における 2 次利用を推進する図 5 のオンサイト利用制度を実施しているが、その中で調査参加者のプライバシー保護は不可欠であり、機密情報の漏洩を防ぐために分析結果の安全性審査を行う。本研究で開発した秘匿処理ツールは、その審査業務での活用が予定されている。

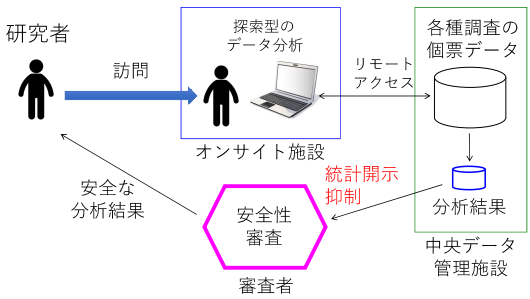


図 5.オンサイト利用による公的調査データの 2 次利用

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 2 件）

① Kazuhiro Minami and Yutaka Abe. Statistical Disclosure Control for Tabular Data in R. Romanian Statistical Review, No. 4, pp. 67–76, 2017.（査読有）

② Yoshiki Yamagata, Daisuke Murakamia, Kazuhiro Minami, Nana Arizumi, Sho Kuroda, Tomoya Tanjo, Hiroshi Maruyama. Electricity Self-Sufficient Community Clustering for Energy Resilience.

Energies, 9(7), 543, July 2016. （査読有）

〔学会発表〕（計 7 件）

① Kazuhiro Minami and Yutaka Abe. Statistical Disclosure Control for Tabular Data in R. 5th International Joint Conference New Challenges for Statistical Software - The Use of R in Official Statistics (uRos2017), November, 2017.
② Ryo Kikuchi and Kazuhiro Minami. On-site Service and Safe Output Checking in Japan. Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, September, 2017.
③ 南和宏. 菊池亮. 調査票情報のオンサイト利用における分析結果の持ち出し基準について. 経済統計学会第 61 回全国研究大会. 2017 年 9 月.
④ 南和宏. An implementation of a cell suppression algorithm for tabular data in R and its challenges, 2017 年度統計関連学会連合大会, 2017 年 9 月.
⑤ 南 和宏.集計表秘匿における差分攻撃の考察, 2017 年暗号と情報セキュリティシンポジウム.
⑥ 政野 博紀, 柴田 直樹, Gao Juntao, 南 和宏, 伊藤 実. オンデマンドバスのための乗り換えを含むリアルタイムルートスケジューリング, 第 24 回マルチメディア通信と分散処理ワークショップ .
⑦ Tomoya Tanjo, Kazuhiro Minami, and Hiroshi Maruyama. Graph Partitioning of power grids considering electricity sharing. 2nd International Conference on Environment and Renewable Energy ((ICERE), February, 2016.

〔図書〕（計 3 件）

① Kazuhiro Minami, Tomoya Tanjo, Nana Arizumi, Hiroshi Maruyama, Daisuke Murakami, Yoshiki Yamagata. Resilient Community Clustering: A Graph Theoretical Approach. In Urban Resilience -- A Transformative Approach, Chapter 7, Springer, 2016.
② Nicolas Schwind, Kazuhiro Minami, Hiroshi Maruyama, Leena Ilmola, Katsumi Inoue. Computational Framework of Resilience. In Urban Resilience -- A Transformative Approach, Chapter 12, Springer, 2016.
③ 南 和宏. 位置情報プライバシーの統計的安全性に向けて, 岩波データサイエンス, page 133–145, Vol. 4, 2016.

〔産業財産権〕

○出願状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

６．研究組織
(1)研究代表者
南 和宏（MINAMI, Kazuhiro）
統計数理研究所・モデリング研究系・准教授
研究者番号：10579410

(2)研究分担者
丹生 智也（TANJO, Tomoya）
国立情報学研究所・クラウド基盤研究開発センター・特任助教
研究者番号：40635067

(3)連携研究者
（ ）

研究者番号：

(4)研究協力者
（ ）