

位置情報とタイムスタンプの有用性を調整可能な 移動軌跡匿名化手法

非会員 千葉 智樹^{*a)} 非会員 清 雄一^{*}
非会員 田原 康之^{*} 正員 大須賀昭彦^{*}

A Method of Trajectory Anonymization with Adjustable Usefulness

Tomoki Chiba^{*a)}, Non-member, Yuichi Sei^{*}, Non-member, Yasuyuki Tahara^{*}, Non-member, Akihiko Ohsuga^{*}, Member

(2019年11月28日受付, 2020年5月8日再受付)

Movement trajectories can provide useful information for all fields. However, they may have high privacy parameters, sharing trajectory data with other operators without anonymization carries the risk of linking movement trajectories to individuals. Therefore, it is necessary to consider applying privacy protection to trajectory data. Anonymization indicators, such as k -anonymity are generally adopted for anonymization of trajectory data. Some studies modify position information to achieve anonymity. This modification sometimes produces inaccuracies in data sets. Moreover, they cannot decide on which balance to change location information and timestamp. In this study, to reduce the modification distance of the position, we propose an algorithm that allows the mismatch of time when the position information is acquired within a certain range. Further, we define indicators that represent distortions of position and time information. As a result of comparing the proposed method and the existing method, the usefulness of the proposed method is shown.

キーワード: 匿名化, 有用性, プライバシ, 位置情報, 移動軌跡

Keywords: anonymization, utility, privacy, location information, trajectory

1. はじめに

近年の測位技術の発展により, 人々の位置情報をスマートフォンや車から容易に取得することができる。人々の位置情報を時系列に取得したデータを移動軌跡という。この移動軌跡データを分析することで, 人々の移動パターンや趣味嗜好を推定することができ, マーケティングや都市開発など多くの場面で有用な情報となる。しかし, 取得したデータを加工せずに他の事業者へ提供してしまうと, 位置情報を取得された個人のプライバシーが侵害される危険性がある。そのため, 移動軌跡データを提供するにはプライバシーの保護に配慮する必要がある。

データの匿名化には, k -anonymity⁽¹⁾ や l -diversity⁽²⁾ な

どの匿名性指標が一般的に使用されている。 k -anonymity⁽³⁾ は, データベースの準識別子 (QID) を k 人以上が同一になるように一般化することで, 攻撃者から個人の特特定を防ぎ, センシティブ属性を保護している。

近年, これらの匿名性指標を満たすように移動軌跡を匿名化することを目的とした手法が多数提案されている。位置情報において k -anonymity を実現するためには, 位置かタイムスタンプの少なくとも一方に誤差を与える必要がある。ここで, 多くの既存手法は位置のみに誤差を与えてタイムスタンプは保持しているため, 匿名化処理によって位置情報が大きく変更されてしまうという課題がある。しかし, 移動軌跡のデータを扱うシナリオによっては, タイムスタンプに多少の誤差を加えてでも, 位置の精度を向上させたい場合がある。そのため, 本研究ではこの位置情報とタイムスタンプの精度のバランスを任意に調整が可能な手法を提案する。(したがって, タイムスタンプの精度のみを優先する場合は既存研究と匿名化結果が同じになる。)ただし, 位置の時系列順序は移動軌跡として重要な情報として考えられるので, 位置の時系列順序への影響は与えない範囲でタイムスタンプに誤差を与えることとする。タイム

a) Correspondence to: Tomoki Chiba. E-mail: chiba.tomoki@ohsuga.lab.uec.ac.jp

^{*} 電気通信大学大学院情報理工学研究所

〒182-8585 東京都調布市調布ヶ丘1-5-1

Graduate School of Informatics and Engineering, The University of Electro-Communications

1-5-1, Chofugaoka, Chofu, Tokyo 182-8585, Japan

スタンプが一時間単位のデータが売買されていることから、タイムスタンプの順序を保持してかつ、一時間程度で量子化をしてもデータに価値はあると考えられる⁽⁴⁾。

本論文では、既存手法において発生する位置の修正度を抑制できることを、シミュレーションによる移動データを用いた実験によって示す。また、位置情報とタイムスタンプの修正度を表す指標を提案する。これにより、データ利用者が位置と時間の精度の優先度を決定できることを示す。なお、本論文は著者らの国際会議論文⁽⁵⁾のアルゴリズムを拡張し、ダミーを新たに追加する手法をとったものである。詳細は第4章(4.5)節に記載する。

本論文の構成は次の通りである。2章で関連研究について述べる。3章では問題定義をして本論文の位置づけを示す。4章において新規アルゴリズム及びタイムスタンプの精度に関する評価指標を提案し、5章では実験及び評価を行う。最後に6章で本論文の結論をまとめる。

2. 関連研究

移動軌跡の匿名化は、位置情報の取得方法、準識別子(QID)の考慮の有無、匿名化をするデータの利用法や使用する匿名化指標などによって分類される⁽⁶⁾⁽⁷⁾。QIDとは、ユーザの性別や住所、郵便番号など、それ単体では個人が特定されることはないが、複数組み合わせることで個人が識別されてしまう可能性があるような情報である。本節では、移動データを他事業者に公開する際に発生するプライバシーリスクを低減させることを目的としている匿名化手法に焦点を絞っている。位置情報の匿名化における k -anonymityは、 k 人以上の位置情報が配置されるようにエリアを定め、そのエリアのみを要約的に表示することで匿名性を保証している。このエリアを最小化することが、位置情報匿名化手法の一般的な目的となる⁽⁸⁾⁽⁹⁾。

k -anonymityの概念をベースに、位置情報の誤差を利用して組み合わせた指標に (k, δ) -anonymity⁽¹⁰⁾という匿名性指標が存在する。この指標は位置情報に誤差があることを利用して、 k 個の位置情報がそれぞれ誤差 δ を半径とする円が重なっている部分に位置している時に、その円内にある全ての位置情報は (k, δ) -anonymityを満たすとしている。Abulら⁽¹⁰⁾は、 (k, δ) -anonymityを満たすように移動軌跡を匿名化する手法 NWA (Never Walk Alone)を提案している。この手法は、移動軌跡をタイムスタンプのデータが一致するものに分類した後、分類されたグループごとに移動軌跡をクラスタリングし、クラスター中心から半径 $\delta/2$ の3次元シリンダー状のエリアに k 個の移動軌跡が位置するように最短距離で各位置情報に誤差を与えている。なお、移動軌跡の要素1点ごとにおける修正距離には閾値が設けられており、閾値を超えるような距離を修正しなければ匿名性を維持できない軌跡は削除される。この操作を行うことにより、移動軌跡を (k, δ) -anonymityに関して匿名化している。 (k, δ) -anonymityは k -anonymityを満たさないとの指摘もあるが⁽¹¹⁾、これは最終的な出力を k

人内包するエリアとして要約表示することで、各エリアに k 人以上が存在することを保証できることから、定義により k -anonymityを満たす。

さらにAbulらは NWA の時間的制約を緩和するために、 $W4M$ (Wait for Me)⁽¹²⁾という新規匿名化手法を提案している。 NWA では位置間の距離の定義をEuclid距離としているが、 $W4M$ では座標平面上の距離に加えてタイムスタンプの距離も兼ね合わせたEDR(Edited distance on Real sequences)で定義しているため、クラスタリングの際に単純な位置の距離だけでなく、いつの時点でその場所に位置していたかという時間的ズレを考慮して計算している。本研究と異なる点としては、 $W4M$ は距離の定義を変えているために、位置における距離と時間における距離を同時に考えている。そのため、任意のパラメータに対して匿名化結果およびそれに伴う情報損失度が一意に定まる。しかし、提案手法の場合は位置と時間を別々に定義し、重み α を用いて位置と時間の情報損失度に差をつけることができるため、データ分析者が任意のバランスに匿名化後のデータの精度を調整することができる。また、 $\alpha = 1.0$ の場合は、位置情報の測位時刻を全く考慮しないため、精度は少なくとも劣らないことが明らかである。

同様に背景知識として位置情報を持ち、QIDを考慮しないような移動軌跡匿名化手法として、 k -anonymityをベースにした研究もある。Nergizら⁽¹³⁾は、移動軌跡を要約表示によって一般化したのちに、ランダムに軌跡の要素である点を再構築させることでデータの有用性をできるだけ損なわないように匿名化をしている。本研究とは、ベースの匿名性指標が k -anonymityか (k, δ) -anonymityという点で異なる。

また、移動軌跡をリアルタイムで匿名化する手法も提案されている⁽¹⁴⁾。この手法は前述の NWA がリアルタイムに匿名化したい環境の場合に、データの抽象度が高くなってしまうことを指摘し、空間座標を表す2次元の座標 (x, y) を測位するたびに匿名化処理を行い、さらにデータの抽象度を抑えるために合成や分割といった動的再構成を行うことで、匿名チューブの構成を動的に組み替えている。この手法は、リアルタイムにデータを分析するような環境に特化させている。本研究との相違点はリアルタイム性の有無である。本研究ではリアルタイムに匿名化することを想定しておらず、データ保持者が分析者にデータを提供する前に匿名化を行うこととしている。そのため、本研究はリアルタイム性を要していない。

上記の匿名化手法は攻撃者モデルの背景知識は位置情報であったが、それとは別に移動パターンによる攻撃からのプライバシー保護を目的とした研究もある。Primaultら⁽¹⁵⁾は、規則的な移動をする習慣があることを主な問題点として挙げ、POI(point of interest)が攻撃者に特定されないように、移動データのトレース間の距離を一定に保つことで滞在した場所を攻撃者に把握されないようにしている。本研究では、移動軌跡を形成する各要素の関係性に関するプライバ

シリスクは考慮しておらず，各要素がそれぞれのタイムスタンプにおいて (k, δ) -anonymity を満たすように匿名化をしている。

QID を考慮した移動軌跡の匿名化について研究しているものもある。Sui ら⁽⁶⁾は QID がレコードに含まれる位置情報を対象とし，ユーザの大規模なモビリティ情報を収集してユーザ属性の多様性の低さに伴うプライバシーリスクを分析している。この研究は Wi-Fi ネットワークによって位置情報を取得することを前提においており，屋内環境における移動データに適している。本研究において想定しているデータは，位置情報を取得されるユーザの識別子，位置情報，タイムスタンプの情報を持つ時系列移動データであり，QID を含むようなデータを想定していない。

3. 問題定義

〈3・1〉 移動軌跡 移動軌跡とは，あるユーザにおける位置情報を時系列に並べたデータである。3 次元の時空間上においてはポリラインで表され，各点間における移動は一定の速度で直線に移動していることとする。匿名化前のデータセットを D ，軌跡の要素時系列順に前から i 番目の空間座標を表す 2 次元の座標を (x_i, y_i) ，時系列 i 番目のタイムスタンプを t_i とした時，移動軌跡を τ を $\tau = \{(id, x_i, y_i, t_i) | \forall i \in \mathbb{Z}, t_i < t_{i+1}\}$ と定義する。また，測位された位置情報には一様に誤差が δ あるとし，移動軌跡には各点を中心とする半径 δ の誤差エリアの円が与えられる。実際の移動データでは誤差は一様ではなく，位置情報が測位された場所によって変動することが考えられるが， NWA ではデータセット内の位置情報の中における最小の誤差を δ として定義することによって直接的に対処できると結論付けているため，本研究においても同様に最小の誤差として想定する。

〈3・2〉 位置情報の利用シナリオ 都市開発やマーケティングを目的として，人々の移動の特徴を抽出するために移動軌跡のマイニングを行うことを想定する。また，データ保持者と分析者は異なる事業者であるようなモデルを想定する。この分析機関は信頼性が不明であるとし，移動データを分析業者に提供する際にプライバシーリスクが発生する。ここで，移動データを分析業者に提供する前にデータ保持者が提案手法によって匿名化を行うことにより，プライバシーリスクを低減する。

移動軌跡は (k, δ) -anonymity を満たすように匿名化されることを想定している。本研究のゴールは，あらゆるデータ分析者に対するデータ有用性が最大となるように移動軌跡を匿名化することである。4 章で説明する提案手法の目的は，〈4・2・3〉項に示されている情報損失度の指標を最小にすることである。また，本研究ではプライバシーリスクが一定という条件下で，位置情報とタイムスタンプの精度の比率を調整することを想定している。これは，タイムスタンプを修正することによるプライバシ保護効果があることを前提としている。タイムスタンプの修正によるプライバ

シ保護効果は，次の例によって示される。例えば，位置情報の誤差 20m 程度で同じ経路を同一時刻にたどった Alice と Bob，1 時間後に同じ経路を辿った Carol の三人の移動軌跡があるとする。また，Alice と Bob が経路をたどっている時刻に，Carol は十分離れていることを想定する。この 3 つの移動軌跡を (k, δ) -anonymity の指標によって匿名化することを考えた時，時間に誤差を加える前は Alice と Bob に関して $(2, 20)$ -anonymity で匿名化することができる。ここで Carol の移動軌跡に 1 時間の誤差を与えると，Alice, Bob, Carol の三人が同一時刻に同じ経路をたどっているデータとなり， $(3, 20)$ -anonymity で匿名化することができる。 (k, δ) -anonymity の匿名性指標は k の値が大きいほど匿名性が高いという指標であるため，Carol の移動軌跡に時間の誤差を与えたことで匿名性を高めたことになる。

〈3・3〉 攻撃者モデル 本研究における攻撃者は，あるユーザの移動軌跡のうち，部分的な位置情報もしくは移動軌跡を背景知識として持っている人物とする。部分的な位置情報を持つとは，例えばある特定の時間に特定の場所で攻撃者がターゲットユーザに会っている，もしくは目撃していることを指す。攻撃者の目的は，上記の背景知識と移動軌跡データを照合することによって，その移動軌跡が誰のものであるかということとを判定することとする。

4. 提案手法

〈4・1〉 概要 本節では，タイムスタンプ修正アルゴリズム，移動速度修正アルゴリズム，およびダミー追加アルゴリズムを提案する。これらは匿名化処理によって発生する位置情報の精度悪化を低減させることを目的としている。本研究では， NWA によって位置情報を匿名化を行い，この過程で発生する位置修正度を分析している。また，タイムスタンプの修正に伴う時間の修正度を表す指標を提案し，既存指標である位置の修正度を表す指標と併せてパラメータ α を用いた全体の情報の損失度を表す指標を提案する。

本手法では 1 度匿名化を行った後で，生データと匿名化されたデータをもとにタイムスタンプ修正アルゴリズムおよび移動速度修正アルゴリズムによって生データのタイムスタンプに修正を加えている。最終的に，データ分析者に提供される全ての移動軌跡は (k, δ) -anonymity を満たす。提案システムの全体像は Fig. 1 に示す。Fig. 1 における“anonymization method”は，出力データが (k, δ) -anonymity を満たすように匿名化する処理としている。これより，提案手法の最終的な出力データはこの匿名化処理を施した後の“anonymized data”となる。したがって，提案手法における最終的な出力は，後述するパラメータ α の値に関わらず， (k, δ) -anonymity に関して匿名性，並びにプライバシーリスクが一定となる。

〈4・2〉 既存指標および提案指標 既存指標および提案指標の説明に用いる記号を Table 1 に示す。この表における τ や D といった匿名化前の記号は全て raw data を示

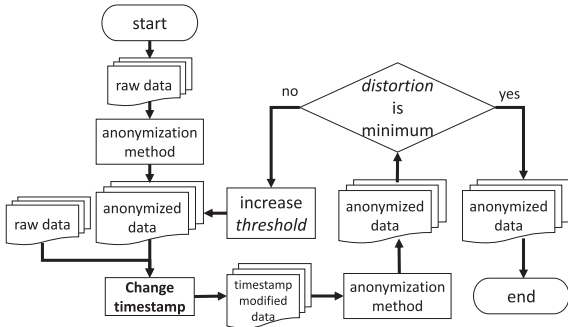


Fig. 1. Flowchart of the proposed system

Table 1. Notation

Notation	Description
τ, τ'	Trajectory of user before and after anonymization
$t_i(\tau)$	Timestamp in i^{th} position information of trajectory τ
$\tau[t]$	Position coordinates in timestamp t of trajectory τ
$ T(\tau) $	Total number of timestamps in trajectory τ
D, D'	Data set before and after anonymization

している。位置修正度は Abul ら⁽¹⁰⁾が提案している指標を本研究にも適用している。この既存指標に基づいた時間修正度を提案し、双方の指標に重み α を付与した式を、全体の有用性指標として提案する。

〈4・2・1〉 位置修正度 位置情報の有用性を示す評価指標は、Abul ら⁽¹⁰⁾の定義を基にしている。これは NWA の処理で発生する位置の修正度を示している。位置修正度を表す評価式を式 1 に示す。

$$LocD(\tau[t], \tau'[t]) = \begin{cases} Dist(\tau[t], \tau'[t]) & \text{if } \tau' \text{ is defined;} \\ \Omega & \text{otherwise;} \end{cases} \quad (1)$$

この評価式は、移動軌跡の各要素における位置情報の修正度を示している。(1) 式の $Dist$ 関数は、二点間のユークリッド距離を表している。例えば、 $\tau[t] = (x_1, y_1), \tau'[t] = (x_1', y_1')$ とした時、 $Dist(\tau[t], \tau'[t]) = \sqrt{(x_1 - x_1')^2 + (y_1 - y_1')^2}$ となる。匿名化処理後も位置情報が削除されていない場合は、匿名化処理によって修正された距離を、削除されている場合はペナルティ Ω として、そのデータセットにおける最大修正距離に置き換えている。この評価式が出力する値が大きければ大きいほど、情報の損失度が上昇、つまり分析される際のデータの有用性が低下していることになる。1 つの軌跡が生み出す位置情報の歪みは以下の (2) 式で定義される。

$$LocD(\tau, \tau') = \sum_{t \in T} LocD(\tau[t], \tau'[t]) \quad (2)$$

データセット D が全体で生み出す位置情報の歪みは以下の (3) 式で定義される。

$$LocD(D, D') = \sum_{\tau \in D} LocD(\tau, \tau') \quad (3)$$

〈4・2・2〉 時間修正度 ユーザの位置情報の測位時刻

を表すタイムスタンプの修正度を表す指標は、〈4・2・1〉項で説明した既存指標を参考に考案した。時間修正度を表す評価式を (4) 式に示す。

$$TimeD(\tau[t], \tau'[t]) = \begin{cases} TimeLag(\tau[t], \tau'[t]) \cdot period & \text{if } \tau' \text{ is defined;} \\ \Psi & \text{otherwise;} \end{cases} \quad (4)$$

(4) 式は、軌跡の各タイムスタンプに対する時間の修正度を表している。 $TimeLag$ とは、処理の前後において測位された位置情報のデータにおけるタイムスタンプが何個ずれたかを表しており、 $period$ は位置情報の測位時刻の周期である。また、(1) 式の場合と同様に処理によって点が削除された場合、その点における $TimeLag$ はペナルティ Ψ として、その軌跡の処理前、つまり元データにおける総タイムスタンプ数 $\times period$ で定義される。1 つの軌跡が生み出す位置測位時刻の歪みは以下の (5) 式で定義される。

$$TimeD(\tau, \tau') = \sum_{t \in T} TimeD(\tau[t], \tau'[t]) \quad (5)$$

データセット D が全体で生み出す位置測位時刻の歪みは式 6 で定義される。

$$TimeD(D, D') = \sum_{\tau \in D} TimeD(\tau, \tau') \quad (6)$$

〈4・2・3〉 有用性指標 データ分析における有用性指標は、プライバシーを保護したうえでデータの修正度を最小限に抑えることが求められる。本研究では、データ分析において位置の精度を重要視するか、もしくは時間の精度を重要視するかはターゲットを絞っておらず、分析者が任意に選択できるようにするため、前述した位置修正度と時間修正度の評価式に重みを付与した式を有用性指標としている。本研究における有用性指標を式 7 に示す。

$$Distortion = \alpha \cdot LocD(D, D') + (1 - \alpha) \cdot TimeD(D, D') \quad (0 \leq \alpha \leq 1) \quad (7)$$

この式の値が小さければ小さいほど、有用性が高いという意味になる。また、パラメータ α は位置情報の精度を重視する度合いであり、0 から 1 までの数においてこの値が大きいほど位置情報の精度を重要視する度合いが高まる。

この評価式には位置の次元と時間の次元が混在しているが、分析者が具体的に α を指定するのではなく、「位置次元もしくは時間次元での誤差をもう一方の次元に合わせて定義する」ことで、任意の分析者に最適な α が一意に定まる。例として、時間の単位が h 、座標間の距離の単位が km のデータセットに対し、ある分析者が「1 時間の誤差は位置における $2 km$ の誤差と同程度の悪影響を及ぼす」と定義することを考える。この時、 $Distortion = \alpha LocD + (1 - \alpha) TimeD$ における $TimeD$ はもともと時間の誤差だが、2 倍することで位置の誤差となる。ここで、2 倍した $TimeD$ と元の $LocD$

を足し合わせると全体が距離の誤差になり，その誤差を最小化するように匿名化を行うことで，分析者任意のバランスにおいて有用性低下を抑制した匿名化を行うことができる。 α の値は， $TimeD$ の値が $LocD$ の2倍となるように設定すればよいので，上記定義において $\alpha = 1/3$ と一意に定めることができる。ここでは， $Distortion$ の値そのものに意味があるのではなく， $Distortion$ 式の値が最小値をとるように匿名化することに意味がある。

〈4・3〉 タイムスタンプ修正アルゴリズム タイムスタンプ修正アルゴリズムは，分析者任意の有用性を最大化するため，位置情報の測位時刻を表すタイムスタンプを(7)式の値が小さくなるように追加，あるいは削除している。初期段階として，タイムスタンプを修正する軌跡をデータセット内から探索する。タイムスタンプを修正する条件は，匿名化前後のデータセットを比較したときに軌跡の少なくとも1つのポイントにおける位置の修正距離が閾値を超えた場合，もしくは軌跡が NWA によって削除されていることとしている。探索アルゴリズムをAlgorithm 1に示す。基本的な流れとして，このアルゴリズムはタイムスタンプ修正対象となった移動軌跡に対し，指定範囲のエリア内でタイムスタンプが異なり，かつ匿名化処理による位置情報の修正距離が小さいような軌跡を探索することから始まる。この指定範囲とは，Algorithm 1中にも出現する $threshold$ である。この値は位置が近いとされる距離の定義であり，ここでは参照される軌跡の探索範囲という意味を持つ。探索により見つかった軌跡は，タイムスタンプを参照される移動軌跡として定義される。参照される軌跡のタイムスタンプを修正対象の軌跡にコピーすることで，修正対象の移動軌跡の位置修正度は，参照される軌跡と同等に抑えることができる。ここで参照される軌跡となる条件は，修正対象の軌跡とタイムスタンプが一致せず，軌跡の各要素に対して修正対象の移動軌跡との位置座標の距離がそれぞれ $threshold$ 以下であり，かつタイムスタンプ修正対象となっていないものとして設定されている。

Algorithm 1 Search all trajectory that needs to change timestamp

```

1: Input:  $D, D', \delta, threshold$ 
2: for all  $\tau \in D$  do
3:   for  $i = 0$  to  $|T(\tau)| - 1$  do
4:      $\gamma \leftarrow Dist(\tau[i], \tau'[i])$ 
5:     if  $\gamma > threshold$  then
6:        $\tau.flag \leftarrow true$ 
7:     else if  $\tau'$  is undefined then
8:        $\tau.flag \leftarrow true$ 
9:     end if
10:   end for
11: end for
12: for all  $\tau \in D$  do
13:   if  $\tau.flag \iff true$  then
14:      $\tau \leftarrow ChangeTime(D, threshold, \delta)$ 
15:   end if
16: end for

```

$threshold$ は，後述する NWA による位置修正距離の許容範囲，及び参照される軌跡の選択条件として使われている。この値が小さいほど位置修正距離の許容範囲が狭くなり，タイムスタンプ修正フラグが立つ軌跡が増加する。しかし，参照しようとした軌跡がタイムスタンプ修正対象になっている場合，タイムスタンプ修正対象になっていない軌跡しか参照しないように条件を設定しているため，参照が見送られる。そのため，仮に $threshold$ が0の場合でも，参照される軌跡は NWA による位置修正がなされていない軌跡しか参照できないため，ほとんどタイムスタンプは修正されない。つまり，タイムスタンプの修正度は閾値に比例しない。 $threshold$ の取りうる範囲は0からデータセットのエリア内における最大距離の半分と定義している。提案手法では，この $threshold$ を動かしながら，(7)式の値が最小になるように処理している。

探索アルゴリズムでタイムスタンプ修正フラグがたてられた軌跡は，タイムスタンプの修正がなされる。タイムスタンプ修正アルゴリズムでは，フラグがたてられた軌跡とタイムスタンプ数またはスタート時刻が違う軌跡のなかで位置が近い軌跡を探索し，条件に合う軌跡が存在した場合においてのみ，その軌跡のタイムスタンプを修正フラグの立った軌跡にコピーしている。

タイムスタンプ修正アルゴリズムの目的は，軌跡同士の位置が近いのにもかかわらずタイムスタンプが異なることだけの理由で同一クラスターに包含されず，タイムスタンプが同一である他の遠い軌跡のクラスターに分類された結果位置の修正距離が大きくなってしまような軌跡を対象に，タイムスタンプの相違という制約を取り払う意味もある。

ここにおいて，2つの軌跡のタイムスタンプ総数，つまりその軌跡の測位継続時間における長さが異なる場合，修正する軌跡に点を追加または削除しなければならないという状況が発生する。そこで，前節で示した時間の修正度を表す指標に基づき，修正度がなるべく小さくなるよう場合分けを行った。この場合分けを行う意味は，タイムスタンプの修正による時間の精度の悪化を最小限に抑えることである。場合分けをせずにタイムスタンプの修正を行った場合，必要以上にタイムスタンプを修正し，時間の修正度が高くなってしまふ可能性がある。

タイムスタンプの修正の際に参照される軌跡は， NWA での修正距離が閾値以下でかつ各タイムスタンプにいて2点間の距離が δ 以下であるという条件で，軌跡の識別子であるid順に若いものから選択される。参照される軌跡の NWA での修正距離が閾値を超えていた場合，その軌跡もまた，タイムスタンプ修正アルゴリズムにかけられる対象であり，結局タイムスタンプが異なることにより NWA において同じクラスター内に包含されず，タイムスタンプを修正した意味がなくなってしまう。

タイムスタンプ修正アルゴリズムをAlgorithm 2に示す。ここで， $Dist$ とは与えられた2点の座標間における距離を表し， $Moddist$ とは NWA によって修正された位置の座

Algorithm 2 Change Timestamp

```

1: Input:  $\tau_k, D, threshold, \delta$ 
2: for all  $\tau \in D$  do
3:   if  $|T(\tau_k)| = |T(\tau)|$  then
4:     if  $\forall Dist(\tau_k, \tau) \leq \delta \wedge \forall Moddist(\tau[l], \tau) \leq threshold$  then
5:       for  $l = 0$  to  $|T(\tau_k)| - 1$  do
6:          $t_l(\tau_k) \leftarrow t_l(\tau)$ 
7:       end for
8:     end if
9:      $diss \leftarrow |t_0(\tau_k) - t_0(\tau)|$ 
10:     $disg \leftarrow |t_{|T(\tau_k)|-1}(\tau_k) - t_{|T(\tau)|-1}(\tau)|$ 
11:    else if  $|T(\tau_k)| > |T(\tau)| \wedge diss \leq disg$  then
12:      if  $\forall Dist(\tau_k, \tau) \leq \delta \wedge \forall Moddist(\tau[l], \tau) \leq threshold$  then
13:        copy all timestamps of  $\tau$  to  $\tau_k$  from the beginning to the end
14:        delete surplus points of  $\tau_k$ 
15:      end if
16:    else if  $|T(\tau_k)| > |T(\tau)| \wedge diss > disg$  then
17:      if  $\forall Dist(\tau_k, \tau) \leq \delta \wedge \forall Moddist(\tau[l], \tau) \leq threshold$  then
18:        copy all timestamps of  $\tau$  to  $\tau_k$  from the end to the beginning
19:        delete surplus points of  $\tau_k$ 
20:      end if
21:    end if
22:  end for

```

Algorithm 3 Change Moving Speed

```

1: Input:  $\tau_k, \tau_i, threshold, \delta$ 
2: for  $l = 0$  to  $|T(\tau_i)| - 1$  do
3:   for  $m = 1$  to  $|T(\tau_k)| - |T(\tau_i)|$  do
4:     if  $Moddis(\tau_i[l], \tau_k[l+m]) \leq threshold \wedge Dist(\tau_i[l], \tau_k[l+m])$  then
5:       delete points  $\tau_k[l]$  to  $\tau_k[l+m]$ 
6:       break
7:     end if
8:   end for
9: end for

```

標と匿名化前のデータの座標との距離，つまり NWA による位置の修正距離を計算する関数である。

〈4・4〉 移動速度修正アルゴリズム タイムスタンプ修正アルゴリズムの拡張として，移動軌跡の内部に当たる途中の点も削除対象とするようなアルゴリズムを考案した。移動速度修正アルゴリズムを Algorithm 3 に示す。移動速度修正アルゴリズムは，タイムスタンプ修正アルゴリズムにおける条件に加え，タイムスタンプをスタートから順番に数えていき，同一の順番でもし 2 点間の距離が閾値を超えてしまっている場合，タイムスタンプ修正対象の軌跡において，次のタイムスタンプの点との距離を計算させている。もしこの時点で 2 点間の距離が閾値以下であった場合，スキップした点を削除対象とし，2 点間の距離が閾値を超えている場合はさらにタイムスタンプをスキップして，それを距離が閾値以下の点が見つかるまで繰り返す。

〈4・5〉 ダミー追加アルゴリズム 〈4・3〉節では，タイムスタンプ数が修正対象より小さい軌跡のタイムスタンプを参照してコピーしていたが，タイムスタンプ数が修正対象よりも大きい軌跡を参照する場合もある。この場合，参照する軌跡にタイムスタンプを合わせるためには，タイムスタンプ数を増やさなければならない。この時，ダミーを追加する必要性が生じる。ここでは，位置情報のダミーを

どのように生成しているのかについて説明する。

ダミーの位置情報は基本的にランダムに生成する。ここで，ダミーの追加に伴い位置の修正度が発生してしまうため，位置情報修正度をできる限り抑えるようにダミーを生成した。本研究において匿名化手法として使用した NWA は，移動軌跡をクラスタリングし，その中心を基準として半径 δ 以内に位置するように移動軌跡の位置情報を変更しているが，ダミーはこのクラスター中心から半径 δ 以内の位置にランダムに生成した。

5. 評価実験

本章では，評価実験の内容及び評価実験に用いたデータを示し，その実験の結果を示した後に考察を述べる。

〈5・1〉 実験内容 提案手法の有用性を評価するため，〈4・2・3〉節で定義した評価指標による検証実験を行う。既存手法である NWA と提案手法を比較し，匿名化後のデータの有用性を検証する。具体的には，各閾値に対する distortion の変化を検証し，さらに各 k に対する distortion の変化を各 α に対して検証する。さらに，位置情報の誤差 δ に関する distortion の変化も検証する。

〈5・2〉 データセット 移動体シミュレータである Siaufo⁽¹⁷⁾ を用いて移動データを作成した。このツールは他の研究でも広く使われている⁽¹⁸⁾⁽¹⁹⁾。移動体の移動範囲は $4.2\text{ km} \times 4.2\text{ km}$ とし，ユーザ数 1 万人に対する 5 分毎に測位した位置情報を 4 時間分利用した。また，デフォルト値として位置情報の誤差 $\delta = 200\text{ m}$ としている。

〈5・3〉 実験結果 実験結果は Fig. 2 と 3 の通りである。Fig. 2 は各プライバシーパラメータ k と位置情報の誤差 δ についての distortion の結果を示している。

k に関する distortion の推移は，いずれのパラメータ α の値に関しても既存手法 NWA よりも優位な結果を示した。Fig. 2 (a) の結果から，タイムスタンプを修正したことにより位置修正度を 5 割以上低減させていることが確認できる。 δ に関する distortion の結果においても，提案手法の優位性を確認できた。 $\delta = 0$ の場合においては，タイムスタンプ修正アルゴリズムにおける条件により，誤差の範囲内に位置する，すなわち同一の座標に点がない場合はタイムスタンプの参照ができないため，タイムスタンプを全く修正せず，結果的には NWA と同じ出力結果になる。また， k の値が増加する毎に位置情報の修正を必要とする場面が多くなることから，distortion の値は全体的に増加しているが，提案手法はその増加の比率を低減できていることが Fig. 2 から読み取ることができる。

また，提案手法によって与えられる時間の誤差について議論するため， $\alpha = 0.5$ と 0.1 における，取得した位置情報 1 点ごとの平均修正距離 Average LocD と平均修正時間 Average TimeD の結果を Table 2 と Table 3 に示す。Table 2 の結果を見ると，各 k に対する平均時間修正度はそれぞれ 60 分程度となっている。この時間の誤差を許容できるかどうかは，厳密には分析者に依存するが，第 1 章で述べたと

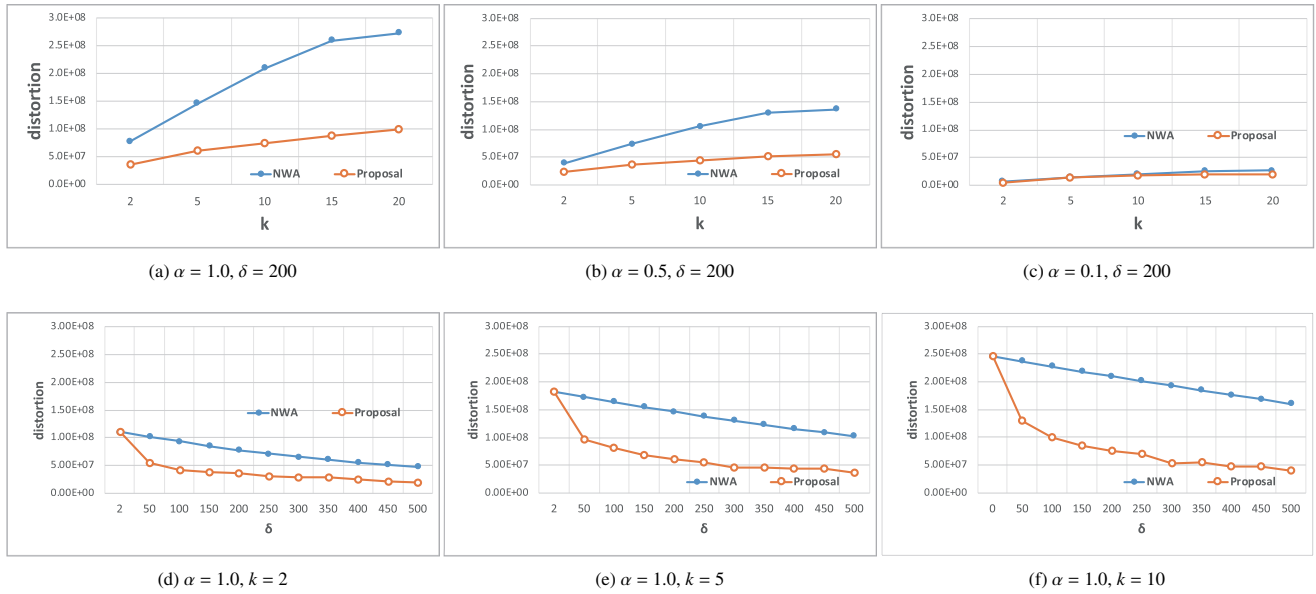
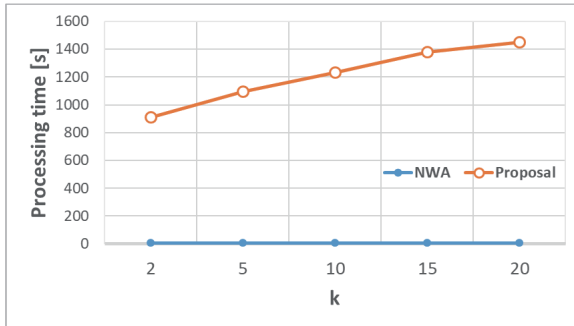

 Fig. 2. Distortion on each k and δ

 Fig. 3. Processing time with parameter $\alpha = 1.0$

 Table 2. Average *LocD* and *TimeD* on $\alpha = 0.5$

k	Average <i>LocD</i> [m]	Average <i>TimeD</i> [min]	<i>Distortion</i>
2	175.5	60.3	2.4E+07
5	303.2	61.9	3.7E+07
10	388.0	66.0	4.4E+07
15	427.4	64.9	5.1E+07
20	509.1	60.6	5.5E+07

おり、位置取得時刻を1時間単位で量子化しているデータが存在するため、少なくとも許容できる分析者は存在する。Table 3 に関しても一般的に $\alpha = 0.1$ を選ぶ分析者がどの程度の許容度なのか不明であるが、 $k = 10$ に関して $\alpha = 0.5$ に比べ、50%程度修正時間が減少しており、他のどの k に関しても $\alpha = 0.5$ の場合における修正時間よりも小さいことから、 α を変化させたことにより一方の精度を重視する効果が得られていることがわかる。

Fig. 3 は提案手法の実行パラメータ $\alpha = 1.0$ にした場合の平均実行時間の結果を表している。グラフから分かる通り、 k の増加によって実行時間も増加している。これは k が増加することで位置およびタイムスタンプの修正が必要な要素数が増加しているからである。また、*NWA* の実行時間が10秒程度に対して、提案手法は16分程度と増大して

 Table 3. Average *LocD* and *TimeD* on $\alpha = 0.1$

k	Average <i>LocD</i> [m]	Average <i>TimeD</i> [min]	<i>Distortion</i>
2	164.5	0.5	6.1E+06
5	372.3	1.6	1.4E+07
10	494.2	32.1	1.8E+07
15	447.8	47.9	2.0E+07
20	523.5	48.7	2.0E+07

いるが、実用上のシナリオにおいて提案手法を実行するのはデータ保持者が任意の事業者実際に提供する前である。ここで、匿名化処理はデータ保持者がデータを提供する前に一度だけ行うことを想定しており、リアルタイム性は必要としないため、提案手法の実行時間の長さは実用上問題ないと考えられる。しかしながら、データ数の増大に対してさらに実行時間が増えることも考えられるため、アルゴリズムの高速化は今後の課題である。

6. まとめ

本論文では、都市開発やマーケティングを目的として、人々の移動の特徴を抽出するために移動軌跡のマイニングを行うことを想定した。また、データ分析者にデータを渡す際に、生データのプライバシーリスクを低減させる処理を施す場面を想定した。既存の匿名化手法によって位置情報に誤差が発生することを示し、この誤差に関してデータ分析者がタイムスタンプとバランスを任意に取れるように匿名化を行うことを目指した。タイムスタンプを修正するというアプローチにより、タイムスタンプ修正度、および有用性指標を定義した。次に、定義した有用性指標を最小にするタイムスタンプ修正アルゴリズムを提案した。提案手法の有用性を評価するため、シミュレーションデータを用いた評価実験を行なった。結果として、既存手法における位置修正度を低減し、想定環境における匿名化後のデータ有用性を改善することに成功した。

将来課題として, **distortion** の最小値を探索するアルゴリズムの効率化によって処理時間を減らすことが挙げられる。さらに, 他の匿名化手法に適用できるように提案アルゴリズムを改善する必要がある。

謝 辞

本研究は, JSPS 科研費 JP17H04705, JP18H03229, JP18-H03340, JP18K19835, JP19H04113, JP19K12107 の助成を受けたものです。

文 献

- (1) D. DeWitt, K. LeFevre, and R. Ramakrishnan: "Mondrian multidimensional k-anonymity", Proceedings of the 22Nd International Conference on Data Engineering, ICDE '06, pp.25-, Washington, DC, USA, IEEE Computer Society (2006)
- (2) A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramanian: "L-diversity: privacy beyond k-anonymity", 22nd International Conference on Data Engineering (ICDE'06), pp.24-24 (2006)
- (3) L. Sweeney: "Achieving k-anonymity privacy protection using generalization and suppression", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol.10, No.5, pp.571-588 (2002)
- (4) Kokusai kogyo co., ltd., Accessed on Apr. 18, 2019. <http://biz.kkc.co.jp/en/data/>
- (5) T. Chiba, Y. Sei, Y. Tahara, and A. Ohsuga: "Trajectory anonymization: Balancing usefulness about position information and timestamp", 2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS), pp.1-6 (2019)
- (6) L. Lakshmanan, F. Bonchi, and H. Wang: "Trajectory anonymity in publishing personal mobility data", *SIGKDD Explor. Newsl.*, Vol.13, No.1, pp.30-42 (2011)
- (7) Y. Sei and A. Ohsuga: "Location anonymization with considering errors and existence probability", *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol.47, No.12, pp.3207-3218 (2017)
- (8) P. Pesti, B. Bamba, L. Liu, and T. Wang: "Supporting anonymous location queries in mobile environments with privacygrid", Proceedings of the 17th International Conference on World Wide Web, WWW '08, pp.237-246, New York, NY, USA, ACM (2008)
- (9) M. F. Mokbel, C.-Y. Chow, and T. He: "Tincasper: A privacy-preserving aggregate location monitoring system in wireless sensor networks", Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, SIGMOD '08, pp.1307-1310, New York, NY, USA, ACM (2008)
- (10) F. Bonchi, O. Abul, and M. Nanni: "Never walk alone: Uncertainty for anonymity in moving objects databases", Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, ICDE '08, pp.376-385, Washington, DC, USA, IEEE Computer Society (2008)
- (11) R. Trujillo-Rasua and J. Domingo-Ferrer: "On the privacy offered by (k, δ) -anonymity", *Information Systems*, Vol.38, No.4, pp.491-494 (2013). Special section on BPM 2011 conference
- (12) F. Bonchi, O. Abul, and M. Nanni: "Anonymization of moving objects databases by clustering and perturbation", *Inf. Syst.*, Vol.35, No.8, pp.884-910 (2010)
- (13) M. Atzori, M. E. Nergiz, and Y. Saygin: "Towards trajectory anonymization: A generalization-based approach", Proceedings of the SIGSPATIAL ACM GIS 2008 International Workshop on Security and Privacy in GIS and LBS, SPRINGL '08, pp.52-61, New York, NY, USA, ACM (2008)
- (14) T. Takahashi and S. Miyakawa: "Cmoa: Continuous moving object anonymization", Proceedings of the 16th International Database Engineering & Applications Symposium, IDEAS '12, pp.81-90, New York, NY, USA, ACM (2012)
- (15) V. Primault, S. Ben Mokhtar, C. Lauradoux, and L. Brunie: "Time distortion anonymization for the publication of mobility data with high utility", 2015 IEEE Trustcom/BigDataSE/ISPA, Vol.1, pp.539-546 (2015)
- (16) K. Sui, Y. Zhao, D. Liu, M. Ma, L. Xu, L. Zimu, and D. Pei: "Your trajectory privacy can be breached even if you walk in groups", 2016 IEEE/ACM 24th International Symposium on Quality of Service (IWQoS), pp.1-6 (2016)
- (17) M. Martin and P. Nurmi: "A generic large scale simulator for ubiquitous computing", 2006 Third Annual International Conference on Mobile and Ubiquitous Systems: Networking Services, pp.1-3 (2006)
- (18) D. Cassou, J. Bruneau, C. Consel, and E. Baland: "Toward a tool-based

- development methodology for pervasive computing applications", *IEEE Transactions on Software Engineering*, Vol.38, No.6, pp.1445-1463 (2012)
- (19) M. A. R. Dantas, D. C. Nazario, I. V. B. Tromel, and J. L. Todesco: "CONTEXT MANAGEMENT: TOWARD ASSESSING QUALITY OF CONTEXT PARAMETERS IN A UBIQUITOUS AMBIENT ASSISTED LIVING ENVIRONMENT", *JISTEM - Journal of Information Systems and Technology Management*, Vol.11, pp.569-590 (2014)

千葉智樹 (非会員) 2019 年電気通信大学情報理工学部総合情報学科卒業。同年同大学大学院情報理工学研究科情報学専攻入学。移動軌跡のプライバシー保護や機械学習に対する攻撃の研究に従事。



清雄一 (非会員) 2009 年東京大学大学院情報理工学系研究科博士後期課程修了。同年 (株) 三菱総合研究所入社。2013 年より電気通信大学。現在, 同大学大学院情報理工学研究科准教授。博士 (情報理工学)。エージェント, プライバシー保護技術等の研究に従事。2016 年度土木学会水工学論文賞, 情報処理学会論文賞受賞。情報処理学会, 電子情報通信学会, 日本ソフトウェア科学会, IEEE Computer Society 各会員。



田原康之 (非会員) 1991 年東京大学大学院理学系研究科数学専攻修士課程修了。同年 (株) 東芝入社。1993~1996 年情報処理振興事業協会に出向。1996~1997 年英国 City 大学客員研究員。1997~1998 年英国 Imperial College 客員研究員。2003 年国立情報学研究所着任。2008 年より電気通信大学准教授。博士 (情報科学) (早稲田大学)。エージェント技術, およびソフトウェア工学などの研究に従事。情報処理学会, 日本ソフトウェア科学会会員。



大須賀 昭彦 (正員) 1981 年上智大学理工学部数学科卒業。同年 (株) 東芝入社。同社研究開発センター, ソフトウェア技術センター等に所属。1985~1989 年 (財) 新世代コンピュータ技術開発機構 (ICOT) 出向。2007 年より電気通信大学。現在, 同大学大学院情報理工学研究科教授。2017 年より同大学大学院情報システム学研究科研究科長併任。2012 年より国立情報学研究所客員教授兼任。工学博士 (早稲田大学)。情報処理学会フェロー。ソフトウェア工学, エージェント, 人工知能の研究に従事。1986 年度及び 2016 年度情報処理学会論文賞, 2013 年度人工知能学会研究会優秀賞, 2014 年度同学会功労賞, 2018 年度電子情報通信学会 ISS 活動功労賞 受賞。IEEE Computer Society Japan Chapter Chair, 人工知能学会理事, 日本ソフトウェア科学会理事, 同学会監事等を歴任。情報処理学会, 電子情報通信学会, 人工知能学会, 日本ソフトウェア科学会, IEEE Computer Society 各会員。

