Vol.2019-ICS-193 No.5
2019/2/27

タイムスタンプに対して柔軟な移動軌跡匿名化手法の提案

千葉 智樹[1]　清 雄一[2]　田原 康之[2]　大須賀 昭彦[2]

概要：移動軌跡はあらゆる分野で有用な情報となりうる．しかし, 移動軌跡は高いプライバシ性を有する可能性があるため, 移動軌跡のデータを加工せずに他事業者に提供すると, 移動軌跡を個人と結び付けられてしまう危険性がある．そのため, 移動軌跡データを応用するにはプライバシの保護を考慮する必要がある．移動データの匿名化には $k-$ 匿名化などの匿名化指標が一般的に採用されている．$k-$ 匿名化を移動軌跡データに適用したものに $\mathcal{NWA}(\mathcal{N}\text{ever }\mathcal{W}\text{alk }\mathcal{A}\text{lone})$ という手法がある．この手法は位置データに対してアプローチを行い, $(k,\delta)-anonymity$ という匿名性指標を提案したうえで, その匿名性を保証している．$\mathcal{NWA}$ では位置情報のみに修正を加えることで, $(k,\delta)-anonymity$ を保証している．本研究では, $\mathcal{NWA}$ の匿名性を実現させるために伴う位置の修正距離を抑えることを目的として, 位置情報を取得した時間の不一致を一定の範囲で許容するアルゴリズムを提案する．この手法では, 位置情報だけでなく位置の取得時間を表すタイムスタンプに対しても修正を加えている．この手法において発生する位置の取得時間の歪みを表す指標を提案し, この指標と従来研究で提案されている位置修正度に関する指標に重みを付与したものを有用性指標として提案している．提案手法と既存手法の有用性を比較した結果, 提案手法の有用性が示された．

1. はじめに

近年の測位技術の発達により, 人々の位置情報をスマートフォンなどから容易に取得することができる．人々の位置情報を継続して取得した時系列の移動データは, 分析することでマーケティングや都市開発などあらゆる場面で有用な情報となりうる．しかし, 取得したデータをそのまま他事業者に提供してしまうと, 個人が特定されプライバシが侵害される危険性がある．この背景から, 移動データを利用する際はプライバシの保護を考慮する必要がある．

データの匿名化には, $k-$ 匿名性 [4] や $l-$ 多様性 [5] などの指標が一般的に使用されている．$k-$ 匿名化は, 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定を防ぎ, センシティブ属性を保護している．この指標を移動データに適応させたものに, $(k,\delta)-anonymity$[1] という指標が提案されている．この指標は, 位置情報に誤差 δ があることを利用し, 互いに半径 δ の範囲内に位置している 2 つの点は一意に識別されないということを主張している．

この $(k,\delta)-anonymity$ という指標の下で移動データを匿名化しているのが, $\mathcal{NWA}(\mathcal{N}\text{ever }\mathcal{W}\text{alk }\mathcal{A}\text{lone})$[1] という手法である．この手法は, まず前処理としてタイムスタンプ数が同一, つまり位置情報の取得時間の長さが同一である軌跡を分類した後, その分けられた軌跡の中でさらに軌跡の位置座標と位置情報の誤差をもとにクラスタリングを行い, クラスターの中心を匿名シリンダーの中心軌跡としてそれぞれの軌跡の各タイムスタンプにおける座標をクラスター中心軌跡から $\delta/2$ の距離になるように修正を行っている．

本研究では, 時間の精度を一定の割合で粗くすることが可能な場面を前提に, タイムスタンプに対して柔軟な移動データ匿名化アルゴリズムを提案する．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, 提案手法において発生するタイムスタンプの修正度を表す指標を定義し, パラメータを用いて位置と時間の精度の優先度をデータ利用者が決定できることを示す．

2. 関連研究

移動軌跡の匿名化には, 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ベー

[1] 電気通信大学情報理工学部総合情報学科
Department of Informatics, Faculty of Informatics and Engineering, The University of Electro-Communications, Chohu, Tokyo 182-8585, Japan
[2] 電気通信大学大学院情報理工学研究科情報学専攻
Department of Informatics, Graduate School of Informatics and Engineering, The University of Electro-Communications, Chohu, Tokyo 182-8585, Japan

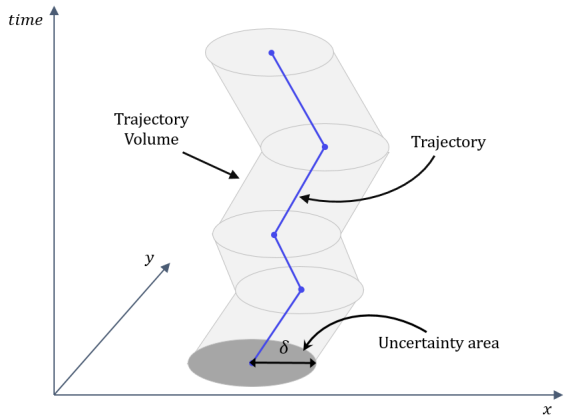


図 1　誤差を持つ移動軌跡. Abul ら [1] の作成した図を参考に作成.
Fig. 1　Uncertain trajectory based on figure Abul et al[1]

スの匿名化手法などによって分類される [3][14]. 本章では, 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.

2.1　$k-$ 匿名性

$k-$ 匿名性 [4] とは, プライバシ情報を含んだデータレコードにおける準識別子 (QID) が同一であるレコードが少なくとも k 個存在することを証明する匿名化指標である. QID が同じレコードが k 個以上存在するように処理をする (QID の一般化) ことにより, 攻撃者は $1/k$ の確立でしか標的のレコードを特定できないことになる. しかし, $k-$ 匿名化だけでは解決できない攻撃もあることから, $l-$ 多様性 [5] などの発展形も提案されている. 本研究では QID を含まない位置情報を扱う前提であるため, このような発展形は扱っていない. 位置情報の匿名化においては, 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.

2.2　移動軌跡の匿名化手法

$k-$ 匿名性の概念をベースに, 位置情報の誤差を利用して組み合わせた指標に $(k, \delta) - anonymity$[1] という手法が存在する. 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の最大誤差 δ を半径とする円内にある時に, その円内にある全ての位置情報は $(k, \delta) - anonymity$ を満たすとしている. Abul ら [1] は, 移動軌跡をクラスタリングして, $(k, \delta) - anonymity$ を満たすようにクラスター中心から半径 $\delta/2$ のシリンダー内に入るように最短距離で各軌跡の位置情報に修正を加えている. 誤差を持った移動軌跡と誤差エリアのイメージを図 1 に示す. さらに Abul らは $\mathcal{NWA}$ の時間的制約を緩和するために, $\mathcal{W}4\mathcal{M}(\mathcal{W}\text{ait for }\mathcal{M}\text{e})$[2] という新規匿名化手法を提案している. $\mathcal{NWA}$ では位置間の距離の定義を Euculid 距離としているが, $\mathcal{W}4\mathcal{M}$ 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.

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-匿名性をベースにした研究もある. Nergiz ら [7] 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.

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 [13]. この手法は前述の $\mathcal{NWA}$ 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 次元ベクトル (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.

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保護を目的とした研究もある. Primault ら [10] 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.

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 [12] 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.

3.　問題定義

本章では, 本研究において扱う問題及び, 想定するシナリオを述べる.

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, つまり位置情報を測位された位置と次に測位された位置の間の移動は一定の速度で直線に移動していることとする. また, 測位された位置情報には誤差が最大でも δ あるとし, 移動軌跡には各点を中心とする半径 δ の円が与えられる.

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は不明であるし, 移動データを分析業者に提供する際にプライバシリスクが発生する. 移動データの分析に関するシーンは特に指定していないが, 位置情報の測位時刻が加工されてはいけないような分析は想定しないこととする. タイムスタンプを考慮しないようなデータは実際に販売されており, 需要があることが分かる [8].

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. すなわち, 匿名化前のデータセットを D, 移動軌跡を $\tau = \{(id, x_i, y_i, t_i) | \forall i \in \mathbb{Z}, t_i < t_{i+1}\}$, データセットからユーザ識別子である id を抜いたデータを $V(\tau)$, としたとき, 攻撃者は $V(\tau)$ の部分集合 $A(\tau)$ $(\subset V(\tau))$ を持つものである. この定義は, Tsubasa ら [13] の定義を参考にしている.

4. 提案手法

4.1 概要

本章では, 既存手法 $\mathcal{NWA}$ において発生する位置情報の精度低下を低減させることを目的とした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を提案する. 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.

本研究は匿名化処理することで生じてしまう情報の損失を抑えることに焦点を絞っているため, 移動データの匿名化アルゴリズムについては既存手法 $\mathcal{NWA}$ で担保されていることとする. 実際には, $\mathcal{NWA}$ によるクラスタリング手法は厳密には $(k, \delta) - anonymity$ を満たしていないという指摘もあるが [11], 本研究においては特別言及しない. 本手法では 1 度 $\mathcal{NWA}$ による匿名化を行った後で, 元のデータと $\mathcal{NWA}$ を実行した後の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元のデータセットのタイムスタンプに修正を加え, 再度 $\mathcal{NWA}$ を実行してその出力を最終的な匿名化後の出力として提供されることを想定している. 提案システムの全体像は図 2 に示す.

4.2 既存指標及び提案手法

既存指標および提案指標の説明に用いる記号を表 1 に示す. 位置修正度は Abul ら [1] 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.

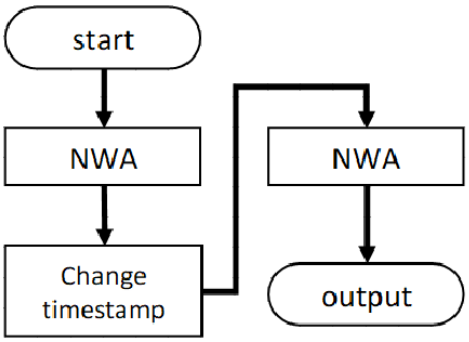


図 2 提案システムの全体像
Fig. 2 Flowchart of the proposed system

表 1 Notation

記号	意味
τ, τ'	匿名化前, 後におけるユーザの移動軌跡
t_i	i 番目の位置情報におけるタイムスタンプ
$\tau[i]$	軌跡 τ のタイムスタンプ t_i における位置座標
T_τ	軌跡 τ における一連の位置情報測位時間の長さ
D, D'	匿名化前, 後のデータセット

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 [1] の定義を使っている. これは $\mathcal{NWA}$ 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.

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

この評価式は, 軌跡の各タイムスタンプに対しての修正度を示している. 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, 削除されてしまった場合はペナルティ 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 つの軌跡で生み出す位置情報の歪みは以下の式 2 で定義される.

$$LD(\tau, \tau') = \sum_{t \in T_\tau} LD(\tau[t], \tau'[t]) \tag{2}$$

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式 3 で定義される.

$$LD(D, D') = \sum_{\tau \in D} LD(\tau, \tau') \tag{3}$$

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, 節 4.2.1 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式 4 に示す.

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

式 4 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. $TimeDist$ 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いくつ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, 式 1 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 $TimeDist$ 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式 5 で定義される.

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

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.

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

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.

$$Distortion = \alpha \cdot LD + (1 - \alpha) \cdot TD \quad (0 \leq \alpha \leq 1) \quad (7)$$

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.

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, 既存手法である $\mathcal{NWA}$ を用いて移動データを匿名化する際に生じる位置情報の修正を低減させるため, 位置情報の測位時刻を表すタイムスタンプを必要に応じて削除している.

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 $\mathcal{NWA}$ 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 Algorithm1 に示す.

Algorithm 1 Search all trajectory that needs to change timestamp

```
Input: D, D', δ, threshold
for all D do
    for i = 0 to T_τ do
        γ ← Dist(τ[i], τ'[i])
        if γ > threshold then
            τ.flag ← true
        else if τ' is undefined then
            τ.flag ← true
        end if
    end for
end for
for all D do
    if τ.flag ⟺ true then
        τ ← ChangeTime(D, threshold, δ)
    end if
end for
```

Algorithm1 における $threshold$ とは, 位置が近いとする距離の定義である. これは, 後述する $\mathcal{NWA}$ 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, 参照される軌跡がタイムスタンプ修正対象になっている場合, 参照する意味が薄れてしまうために, 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, 仮に $threshold$ が 0 の場合でも, 参照される軌跡は $\mathcal{NWA}$ 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, 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, 式 7 の値が最小になるように調整している.

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に分類された結果

位置の修正距離が大きくなってしまう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こともである.

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, 第 5 章で記述する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う可能性がある.

タイムスタンプの修正の際に参照される軌跡は, $\mathcal{NWA}$ 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 $\mathcal{NWA}$ 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 $\mathcal{NWA}$ 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.

また, 現段階において参照される軌跡は, 修正対象の軌跡よりもタイムスタンプ数が同一もしくは小さい軌跡のみに限定している. これは, もし参照される軌跡のタイムスタンプ数のほうが修正対象の軌跡よりも大きかった場合, コピーしようとするとタイムスタンプ数が足りないため, ダミーを追加する必要が出てしまうからである. このアプローチをとっている研究は存在するが [2], 匿名エリア内にランダムで追加しているため, ダミーであることが推測されてしまう可能性がある. この可能性がある場合, $(k, \delta) - anonymity$ を満たさないような状況が発生するため, 本手法ではダミー追加を行っていない. タイムスタンプ修正アルゴリズムを Algorithm2 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, $Corrdis$ とは $\mathcal{NWA}$ によって修正された位置の座標と匿名化前のデータの座標との距離, つまり $\mathcal{NWA}$ 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

前節で説明したタイムスタンプ修正アルゴリズムのみでは, $\mathcal{NWA}$ におけるタイムスタンプの制約を緩和したものの, 時間の修正度に対する位置修正度の低減率は 1,2 割程度であった. そこで, 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 Algorithm3 に示す.

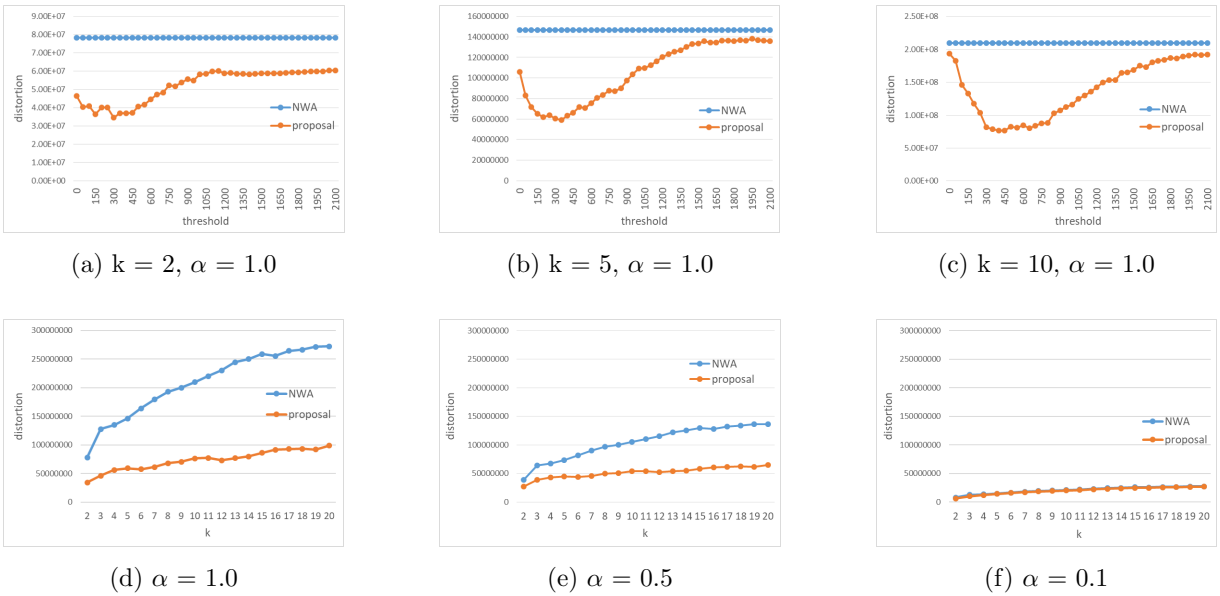
移動速度修正アルゴリズムは, タイムスタンプ修正アルゴリズムにおける条件に加え, もしタイムスタンプをスタートから順番に数えていき, 同一の順番でもし 2 点間の距離が閾値を超えてしまっている場合, タイムスタンプ修正対象の軌跡において, 次のタイムスタンプの点との距離を計算させている. もしこの時点で 2 点間の距離が閾値以下であった場合, スキップした点を削除対象とし, 2 点間の距離

Algorithm 2 Change Timestamp

Input: τ_k, D, $threshold$, δ
for all $\tau \in D$ **do**
  if $|Timestamp(\tau_k)| \iff |Timestamp(\tau_i)|$ **then**
    if $Dist(\tau_k, \tau_i) \leq \delta \cap for\,all\,timestamp\,Corrdis(\tau_i) \leq threshold$ **then**
      for all $timestamp$ **do**
        $time(\tau_k) \leftarrow time(\tau_i)$
      end for
    end if
    $diss \leftarrow |Starttime(\tau_k) - Starttime(\tau_i)|$
    $disg \leftarrow |Goaltime(\tau_k) - Goaltime(\tau_i)|$
  else if $|\tau_k(t)| > |\tau_i|(t) \cap diss \leq disg$ **then**
    if $Dist(\tau_k, \tau_i) \leq \delta \cap Corrdis(\tau_i) \leq threshold$ **then**
      copy all timestamps of τ_i from the beginning to the end
      delete surplus points of τ_k
    end if
  else if $|\tau_k(t)| > |\tau_i|(t) \cap diss > disg$ **then**
    if $Dist(\tau_k, \tau_i) \leq \delta \cap Corrdis(\tau_i) \leq threshold$ **then**
      copy all timestamps of τ_i from the end to the beginning
      delete surplus points of τ_k
    end if
  end if
end for

Algorithm 3 Change Moving Speed

Input: τ_k, τi, $threshold$, δ
for all $timestamp \in \tau_i$ **do**
  for $m = 1$ to $|Timestamp(\tau_i)| - |Timestamp(\tau_k)|$ **do**
    if $Corrdis(\tau_i) \leq threshold \cap Dist(\tau_i[time], \tau_k[time + m])$ **then**
      delete points $\tau_k[time]$ to $\tau_k[time + m]$
      break
    end if
  end for
end for



(a) k = 2, $\alpha = 1.0$　(b) k = 5, $\alpha = 1.0$　(c) k = 10, $\alpha = 1.0$

(d) $\alpha = 1.0$　(e) $\alpha = 0.5$　(f) $\alpha = 0.1$

図 3　閾値および k に関する distortion

Fig. 3　Distortion on each threshold and each privacy parameter

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している.

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, 第 4.2.3 項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 $\mathcal{NWA}$ と提案手法を比較し, その精度を検証する. 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.

5.2 データセット

移動体シミュレータである Siafu[6] を用いて移動データを作成した. 移動体の移動範囲は 4.2 km × 4.2 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 $\delta = 200$ m としている. NTT ドコモにおける位置情報の誤差は, 50m 未満, 300m 未満, 300m 以上, の 3 段階で表されており, 200m の誤差は, 比較的正確な位置情報と定義されている距離にあたる [9].

5.3 実験結果

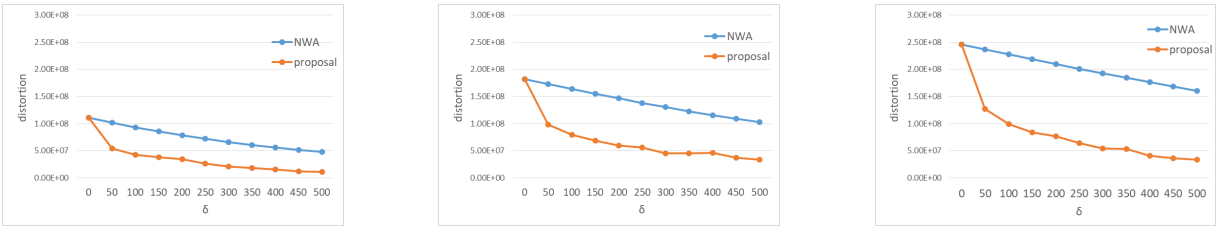
実験結果を図 3 に示す. 図 3[a] から [c] までは, 各閾値に関する distortion の結果であり, 図 3[d] から [f] までは, 各プライバシパラメータ k に関する distortion の結果である. また, 位置情報の誤差 δ に関する distortion の結果を図 4 に示す.

5.4 考察

閾値に関する distortion の推移の結果から, 第 4 章で述べた通り, 閾値に関して比例したグラフにはならないことが分かる. また, やや単峰性な形状が見られるが, 厳密には滑らかな曲線は描いていない. この曲線が滑らかであった場合, 黄金分割探索により distortion 最小値探索アルゴリズムを大幅に効率化することが可能だが, 的確に最小値を取る方法には現状全探索に近いアルゴリズムを取っている. しかし, 厳密に最小値を取ることを目的とせず, 最小値に近いもの, あるいは最小値の推測として値を求めるのであれば, さらに効率化を図ることは可能である. また, 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 $\mathcal{NWA}$ よりも優位な結果を示した. 3(d) 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. 位置情報の誤差 $\delta = 0$ 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 $\mathcal{NWA}$ 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図 4 から読み取ることができる. k が 10 より先まで増加した場合でも, 同様な傾向が表れることが考えられる.

既存手法である $\mathcal{W}4\mathcal{M}$ との比較であるが, 異なる点としては, 既存手法は距離の定義を変えているために, 位置にお



(a) k = 2, $\alpha = 1.0$　(b) k = 5, $\alpha = 1.0$　(c) k = 10, $\alpha = 1.0$

図 4　δ に関する distortion

Fig. 4　Distortion on each δ

ける距離と時間における距離を同時に考えており, 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が, 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, 位置情報の精度においては上であることが明らかである.

6. おわりに

6.1 本論文のまとめ

本論文では, 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, 分析業者にデータを渡す際のプライバシリスクに焦点を当てた.

その想定環境においてプライバシを保護するアプローチとして, $\mathcal{NWA}$ という手法があることを示し, その手法において位置の修正距離が発生することを示した. この位置修正距離を低減することを目的として, タイムスタンプを修正するというアプローチでタイムスタンプの修正に伴う情報の損失度を表す新たな指標を定義し, 既存の位置修正度を表す指標と併せた有用性指標を提案した. また, 提案した有用性指標を最小限に抑えるようなタイムスタンプ修正アルゴリズムを提案した.

これらの新しい指標や手法の有用性を評価するため, シミュレーションによる移動データを用いた評価実験を行った. 評価の結果, 既存手法の位置修正度の低減に成功し, 想定環境における匿名化後のデータの有用性を向上させた.

6.2 今後の課題

将来課題として, タイムスタンプの修正に伴うプライバシ保護効果の検証及び新たなプライバシ指標の考案が挙げられる. また, 閾値の変化に対する distortion 最小値の探索アルゴリズムの効率化を図り, 処理時間の高速化をすることも挙げられる. さらに, 今回はシミュレーションによる実験しか行っていないため, 実データを用いた大規模な実験をする必要がある.

謝辞　本研究は JSPS 科研費 JP16K00419, JP16K12411, JP17H04705, JP18H03229, JP18H03340, JP18K19835 の助成を受けたものです.

参考文献

[1] Abul, O., Bounchi, F. and Nanni, M.: Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases , *Proc. 24th IEEE ICDE*, pp. 376-385 2008.
[2] Abul, O., Bounchi F. and Nanni, M.: Anonymization of moving objects databases by clustering and perturbation , *Information Systems*, vol. 35, no. 8, pp. 884-910, 2010.
[3] Bounchi, F., Lakshmanan, L. and Wang, H.: Trajectory Anonymity in Publiching Personal Mobility Data , *ACM SIGKDD Explorations Newslett*, Vol. 13, No. 1 , pp. 30-42, 2011.
[4] LeFevre, K., DeWitt, D. and Ramakrishnan, R.: Mondrian Multidimensional K - Anonymity , *Proc. IEEE ICDE*, pp. 25-25, 2006.
[5] Machanavajjhala, A., Kifer, D., Gehrke, J. and Venkitasubramaniam, M.: L-diversity: Privacy beyond K-Anonymity , *ACM TKDD*, Vol. 1, No. 1, pp. 3-es, 2007.
[6] Martin, M. and Nurmi, P.: A Generic Large Scale Simulator for Ubiquitous Computing , *Proc. 3rd MobiQuitous*, IEEE, pp1-3, 2006.
[7] Nergiz, E., Atzori, M., and Saygin Y.: Towards trajectories of moving objects , *Proc. ACM GIS Workshop on Security and Privacy in GIS and LBS*, 2008.
[8] NTT docomo.: モバイル統計空間, `https://www.monaku.jp/` (2019.1.18)
[9] NTT docomo.: 測位方法, `https://www.nttdocomo.co.jp/service/search/usage/gps/` (2019.2.3)
[10] Primault, V., Mokhtar S. B., Lauradoux, C and Brunie, L.: Time Distortion Anonymization for the Publication of Mobility Data with High Utility , *Proc. IEEE Trustcom*, 2015.
[11] Trujillo-Rasua, R and Domingo-Ferrer, J.: On the Privacy Offered by (k, δ)-Anonymity , *Inf. Syst*, Vol. 38, No. 4, pp491-494, 2013.
[12] Sui, K., Zhao, Y., Liu, D., Ma, M., Xu, L., Zimu, L. and Pei, D.: Your Trajectory Privacy Can Be Breached Even If You Walk in Groups , *Proc. IEEE 24th International Symposium on Quality of Service*, 2016.
[13] Tsubasa, T and Shinya, M.: CMOA: Continuous Moving Object Anonymization, *Proceeding of the 16th International Database Engineering & Applications Symposium*, ACM, pp. 81-90, 2012.
[14] Yuichi, S and Akihiko, O.: Location Anonymization with Considering Errors and Existence Probability , *Proc. IEEE Trans. Syst., Man, Cybern., Syst.*, Vol.47, Vol.12 pp.3207-3218, 2015.