

時空間におけるクラスタリングを用いた 軌跡情報の k -匿名化法

正木 彰伍^{1,a)} 長谷川 聡¹ 千田 浩司¹

概要: 本稿では、時刻・緯度・経度で表される位置情報の時系列データである軌跡情報の新たな k -匿名化法を提案する。一般に、 k -匿名化された軌跡情報は、常に人通りが非常に多い軌跡を持つユーザーのみを含む傾向がある。そのようなデータを分析しても、新たな知見は得られにくいと考えられる。そこで、時刻-緯度-経度の三次元時空間における位置情報のクラスタリングを利用した k -匿名化法を提案する。この手法は、時空間で近接した位置情報を統一化し、時刻・緯度・経度を効果的に加工することで、人通りが比較的少ない領域の軌跡も k -匿名化データに含まれることを可能にする。このことを公開データを用いた実験により示す。加えて、 k -匿名化された軌跡情報の誤差評価の方法についても提案し、実際に評価を行う。簡易な手法と比較して、提案手法が作成する匿名化データは、より小さい距離誤差を持つことがわかった。

キーワード: プライバシー保護, k -匿名化, 軌跡情報

A k -anonymization method for trajectory data with clustering in space-time

SHOGO MASAKI^{1,a)} SATOSHI HASEGAWA¹ KOJI CHIDA¹

1. 導入

1.1 背景

昨今、個人にまつわるデータであるパーソナルデータの二次利用ニーズが高まっている。二次利用とは、パーソナルデータを取得した事業者からデータの提供を受けた別の事業者が、分析・利活用することである。二次利用におけるリスクとして、情報主体となるユーザーのプライバシー侵害が挙げられる。これを受けて、 k -匿名化 [1], Pk -匿名化 [2], [3], 差分プライバシー [4] といった技術によるプライバシー保護が研究されている。

本稿では、パーソナルデータの中でも時刻・緯度・経度で表される位置情報の時系列データである軌跡情報に着目する。スマートフォン等の GPS 搭載の携帯端末やテレマ

ティクス用の車載端末の普及により、大量の軌跡情報が事業者に蓄積されるようになってきている。分析結果の活用用途として、出店計画立案や位置情報に応じたレコメンド等幅広く想定されることから、軌跡情報は二次利用ニーズが特に高いパーソナルデータの 1 つである。

一方で、軌跡情報は機微性が特に高いパーソナルデータであり、二次利用においては適切にプライバシーを保護する必要がある。その理由として、軌跡情報の特徴に基づくリスクを以下に 3 つ挙げる。

- (1) 個々人がユニークな軌跡を持ちやすく、個人特定につながりやすい。
- (2) 自宅住所や勤務先など、個人に強く紐づく場所が軌跡から推定できる可能性がある。
- (3) 複数事業者が、ある個人のユニークな軌跡情報を取得できる可能性があり、二次利用において軌跡が照合のキーとなって、属性推定が可能となりうる。

軌跡情報のプライバシー保護には様々な研究 [5], [6], [7],

¹ 〒180-8585 東京都武蔵野市 3-9-11 日本電信電話株式会社 NTT セキュアプラットフォーム研究所

NTT Secure Platform Laboratories, NTT Corporation, 3-9-11 Midori-cho, Musashino-shi, Tokyo, 180-8585, Japan

^{a)} masaki.shogo@lab.ntt.co.jp

[8], [9], [10]があるが, 上記リスクの対策には, k -匿名化が有効な手段の1つである。

ここで k -匿名化とは, k -匿名性という匿名性指標を満たすようにデータを加工する匿名化手法である。 k -匿名性とは, 同じ準識別子(それ単体では個人を直接特定することはできないが, 他の属性と組合せることで, 個人を特定できる可能性がある属性)の組合せを持つユーザーが, 1つのデータ中に少なくとも k 個存在するという匿名性指標である[1]。つまり, 準識別子の組合せを基に個人を特定しようとしても k 人未満に絞り込むことができない, ということを保証する。したがって, 上記リスク, 特に(1), (3)の対策には有効な手段と言える。ただし, リスク(2)については, k -匿名性だけでなく, リスクのある位置情報の削除や, l -多様性[11]なども考慮すべき場合があると考えられる。

1.2 動機

軌跡情報の簡易な k -匿名化法として, 緯度-経度平面上のメッシュにマッピングすることで緯度・経度を統一化し, 入力した k 値を満たさないユーザーを削除するという方法が考えられる(3.1節を参照)。しかし, このような手法により作成された k -匿名化データは, 人通りが常時非常に多い道を通ったユーザーのみによって構成される傾向がある(図10参照)。これはマッピングされたメッシュが等しくても, 時刻がわずかでも異なれば, 異なる位置情報と判断されてしまうためである。そのため, 例えば人流分析を想定すれば, 分析結果は, 「街の中心部は人通りが多い」といった当たり前のものとなってしまう, データ分析の本来の目的である新たな知見を得ることが難しいと考えられ課題となっている。

1.3 本稿の成果

本稿では, 上記課題を解決するべく, 時刻-緯度-経度の三次元時空間における, 位置情報のクラスタリングを用いた新たな k -匿名化法を提案する。この手法は, 時空間で近接した位置情報の統一化により, 緯度・経度だけでなく, 時刻も含め効果的に加工することで, 人通りが比較的少ない領域における軌跡も k -匿名化データに含まれることを可能にする。このことを公開データを用いた実験により示す。さらに, k -匿名化された軌跡情報の誤差評価の方法についても提案し, 実際に評価を行う。簡易な手法と比較して, 提案手法によって作成された匿名化データは距離誤差を小さくする効能もあることがわかった。

2. 実験に用いるデータ

本稿の実験では, 株式会社ナイトレイによって無料公開されている「疑似人流データ」[12]を用いる*1。これは

*1 クレジット:株式会社ナイトレイ, 東京大学 柴崎・関本研究室,

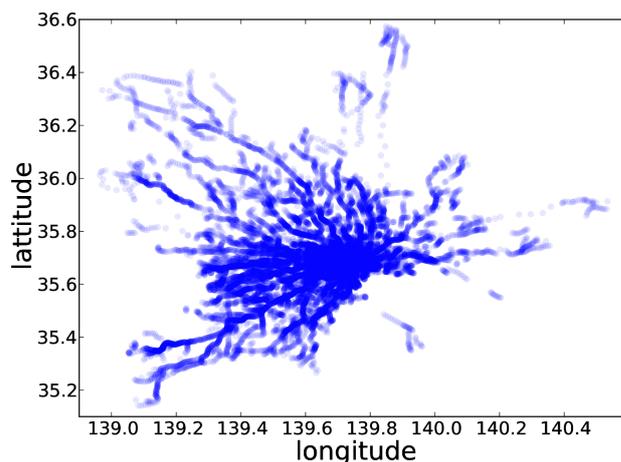


図1 疑似人流データのうち, 東京近郊2013年7月1日8時から9時までの位置情報を例示。

表1 2013年7月1日の東京近郊のデータから, 本稿で対象とする属性(ユーザーID, 時刻, 緯度, 経度)を抜き出したものの一部。

| ユーザーID | 時刻 | 緯度 | 経度 |
|--------|----------|-------------|--------------|
| 3 | 08:00:18 | 35.68367342 | 139.47106963 |
| 3 | 08:05:18 | 35.68302667 | 139.47132593 |
| 4 | 08:00:00 | 35.72072954 | 139.55554464 |
| 4 | 08:05:01 | 35.72141241 | 139.55770943 |
| 4 | 08:10:01 | 35.72167027 | 139.56001344 |
| 4 | 08:15:01 | 35.72219684 | 139.56222533 |

SNSベースの地域解析結果を元に, 移動経路の補完処理, 統計処理, 推定処理, ランダム化処理により作成された疑似的な軌跡情報である。

データに含まれる属性は, ユーザーID, 性別(推定値), 日付・時刻, 緯度, 経度, 滞在者カテゴリ(大分類), 滞在者カテゴリ(小分類), 状態(滞在 or 移動), 滞在者カテゴリIDの9つである。東京近郊については, 最小約5分刻みで24時間分のデータが, 2013年の6日分公開されている。ただし, ユーザーによって位置情報の数は異なる。

図1に, 東京近郊の7月1日8時から9時までの位置情報を例示する。色が濃い部分ほど, 位置情報が密集している。通勤や通学のために, 郊外から都心への人の流れが生じていることが見て取れる。

本稿では, 最もシンプルな軌跡情報として, ユーザーID, 時刻, 緯度, 経度の4属性のみを対象とし, 実験に用いる。表1に, 2013年7月1日の東京近郊のデータから対象属性を抜き出したものの一部を示す。

3. 本稿が提案する軌跡情報の k -匿名化法

3.1 比較用の簡易手法

本稿で提案する k -匿名化法について述べる前に, 比較の

マイクロジオデータ研究会, 人の流れプロジェクト, 東京大学空間情報科学研究センター。

ために簡易な k -匿名化法を導入する。この手法は、メッシュへのマッピングによる緯度・経度の統一化と、ユーザー削除により、 k -匿名化を行う。具体的な手順は、以下の通りである。

- (1) 緯度・経度を、最大値と最小値の間で等間隔にそれぞれ N_{lat} , N_{lon} 個に区切り、領域を $N_{\text{lat}} \times N_{\text{lon}}$ 個のメッシュに分ける。
- (2) 全ての位置情報について、緯度・経度を、対応するメッシュの中心値に置き換える。
- (3) 軌跡が等しいユーザーの数を測り、入力した k 値を満たさないユーザーを削除することで、 k -匿名性を満たすデータを作成する。

この手法のパラメーターは、 N_{lat} , N_{lon} の2つである。

3.2 提案手法

前述の簡易手法では、時刻が未加工のため、結果的に、常時人通りの非常に多い中心部の軌跡情報を持つユーザーのみを含む匿名化データが作成されてしまう。

この問題を解決するには、緯度・経度のみだけでなく、時刻も加工する必要がある。本稿では、時刻-緯度-経度の三次元時空間で近接する位置情報をグルーピングし、あるグループに属する位置情報の時刻・緯度・経度を統一するアプローチを取る。例えば、「14時59分に、A駅南口にいた」「15時01分に、A駅北口にいた」という位置情報であれば、どちらも「15時00分に、A駅にいた」という風に加工する。そのために、時刻-緯度-経度の三次元時空間で、位置情報のクラスタリングを測り、所属するクラスターの中心値に置き換えるという方法を取る。

提案手法の具体的な手順は、以下の通りである。

- (1) 時刻 t , 緯度 lat , 経度 lon は、次元(単位)が異なるため、単純に三次元空間を定義することができないが、以下のようにそれぞれを(最大値 - 最小値)で無次元化することで対処する。

$$t \leftarrow t / (t_{\text{max}} - t_{\text{min}}) \quad (1)$$

$$\text{lat} \leftarrow \text{lat} / (\text{lat}_{\text{max}} - \text{lat}_{\text{min}}) \quad (2)$$

$$\text{lon} \leftarrow \text{lon} / (\text{lon}_{\text{max}} - \text{lon}_{\text{min}}) \quad (3)$$

時刻については、0時0分など基準となる時刻からの時間を秒単位に変換するなどして、この操作を行う。

- (2) 無次元化した時刻-緯度-経度の三次元時空間で、位置情報のクラスタリングを測る。本稿では、 k -平均法 [13] を採用し、全ての位置情報を N_{cls} 個のクラスターに分類する。
- (3) 全ての位置情報について、時刻・緯度・経度を所属するクラスターの中心値に置き換える。この時、前述の(最大値 - 最小値)をかけることで、次元を戻す。
- (4) ある1ユーザーについて、位置情報が複数連続して同じものになった場合、1つのみ残し、あとは削除する。

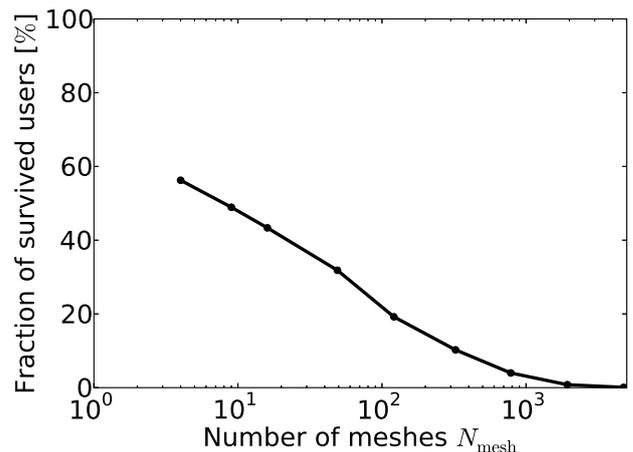


図2 簡易手法により作成された匿名化データに含まれるユーザー数をメッシュ数 N_{mesh} の関数として表示。横軸は、対数軸となっている。

- (5) 軌跡が等しいユーザーの数を測り、入力した k 値を満たさないユーザーを削除することで、 k -匿名性を満たすデータを作成する。

この手法のパラメーターは、 N_{cls} である。

4. 実験結果

本稿では、公開されている疑似人流データのうち、2013年7月1日8時から9時の東京近郊のデータに対する結果を示す。この時間帯に1つ以上の位置情報があるユーザー数は4,118人、位置情報数は33,455個ある(一人あたり8.12個の位置情報)。匿名性パラメーター k については、 $k = 3$ で固定する。

4.1 簡易手法

まず、簡易手法の結果をみていく。簡易手法のパラメーターは、 $N_{\text{lat}} = N_{\text{lon}}$ として、2から1,000の範囲で対数上で等間隔に15個取る。以降、メッシュ数 $N_{\text{mesh}} = N_{\text{lat}} \times N_{\text{lon}}$ を簡易手法のパラメーターとして用いる。

4.1.1 匿名化データのユーザー数

まず、簡易手法によって作成された匿名化データに含まれるユーザー数についてみる。図2は、メッシュ数 N_{mesh} の関数として示している*2。ただし、ユーザー数は元のユーザー数の割合で表している。メッシュ数 N_{mesh} が小さいほど、緯度・経度が粗く統一化され、軌跡のユニーク性が失われるため、削除されずに匿名化データに残るユーザー数は多くなり、 N_{mesh} 増加につれて、ユーザー数は減少する。なお、 $N_{\text{mesh}} = 4,761 (= 69^2)$ では、ユーザー数が3となり、それ以上のメッシュ数では $k = 3$ を満たすユーザーは存在しなかった。

*2 図2の横軸は、対数軸となっている。本稿ではこの図以外でも対数軸を用いたグラフを示すので注意されたい。

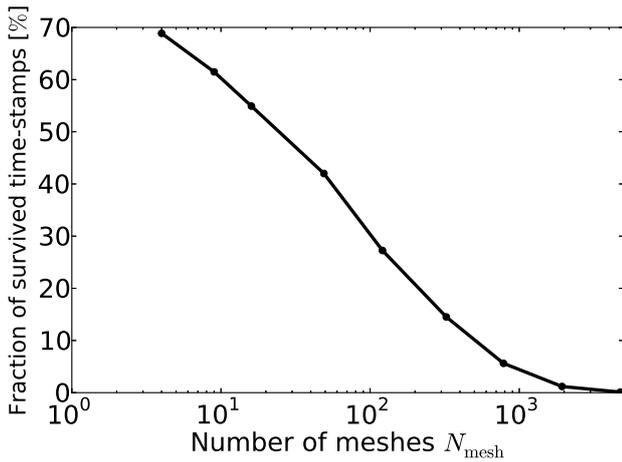


図 3 簡易手法により作成された匿名化データに含まれる位置情報の数をメッシュ数 N_{mesh} の関数として表示.

4.1.2 匿名化データの位置情報の数

図 3 は、簡易手法によって作成された匿名化データが含む位置情報の数 (表 1 のようなデータの行数) を、メッシュ数 N_{mesh} の関数として表している. 位置情報数は元の数の割合で表している. ユーザー数と同様に、 N_{mesh} の減少関数となっているが、やや急な減少をする. これは、長い軌跡 (位置情報の数が多い) ユーザーほど一般にユニーク性が強くなり、削除されやすいためだと考えられる.

4.1.3 匿名化データの誤差評価

ここで簡易手法によって作成された匿名化データの誤差を評価する. 削除によって、ユーザー数が変化しているため、「データの誤差評価」は単純ではないと言える. 本稿では、匿名化データに含まれる位置情報が、元の位置情報とどれほどずれているか、という点に着目して評価を行う. 具体的には、削除されたユーザーは無視して、元データの緯度・経度と置き換えたメッシュの中心値との地理的な距離を計測し、距離誤差とする. 図 4 に、匿名化データの全位置情報の平均距離誤差をメッシュ数 N_{mesh} の関数として示す. 標準偏差を、誤差棒として示している. 期待通り、 N_{mesh} の減少関数となっている.

4.2 提案手法

ここから、提案手法の結果を示す. 提案手法のパラメーターであるクラスター数 N_{cls} は、10 から 11,151 (\approx 位置情報の総数 / $k=3$) の範囲で、対数上で等間隔に 15 個取る.

4.2.1 提案手法手順 (4) による位置情報数の変化

まず、提案手法によって作成された匿名化データについてみる前に、提案手法の手順 (4) によって位置情報数がどれほど減少するかをみる. 図 5 は、提案手法手順 (4) を実行した後の位置情報数を元の位置情報数の割合に変換し、クラスター数 N_{cls} の関数として示している. N_{cls} が少ないほど、より多くの位置情報を統一化し、同じ位置情報が

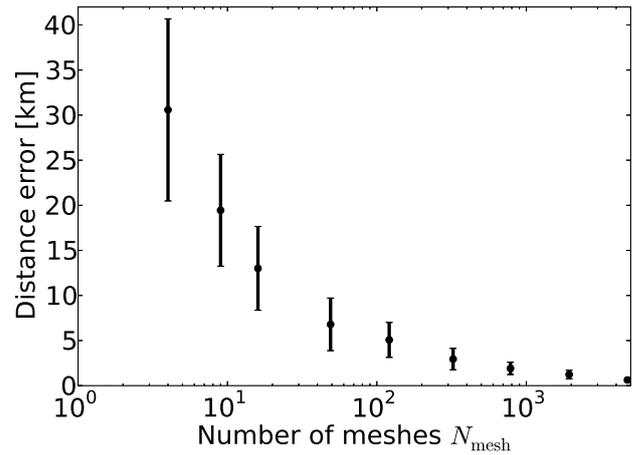


図 4 簡易手法により作成された匿名化データの距離誤差をメッシュ数 N_{mesh} の関数として表示. データ点は平均値, 誤差棒は標準偏差.

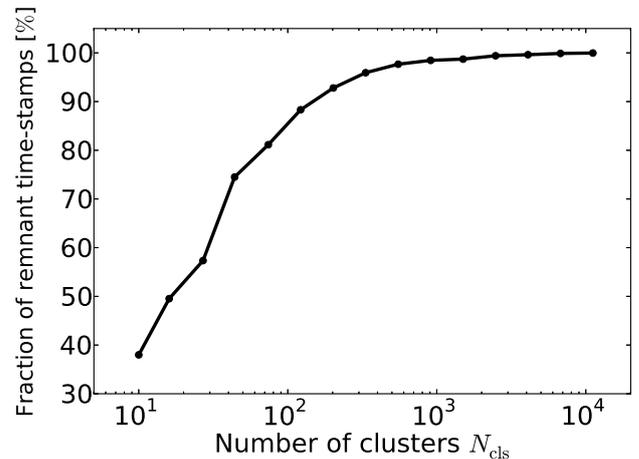


図 5 提案手法手順 (4) 実行後の位置情報数を、クラスター数 N_{cls} の関数として表示.

連続しやすいため、手順 (4) で削除する数が多くなる.

4.2.2 匿名化データのユーザー数

次に、提案手法によって作成された匿名化データについてみていく. 図 2 と同様に、匿名化データに含まれるユーザー数を、クラスター数 N_{cls} の関数として図 6 に示す. N_{cls} が小さいほど、位置情報の加工の度合いは強いいため、削除されるユーザー数は小さくなる.

4.2.3 匿名化データの位置情報数

図 7 に、匿名化データに含まれる位置情報の数を、クラスター数 N_{cls} の関数として示す (図 3 と同様). 図 6 で示したユーザー数と異なり、 N_{cls} の単調減少関数とはなっていない. $N_{\text{cls}} \lesssim 100$ では増加関数、 $100 \lesssim N_{\text{cls}}$ では減少関数となっている. なお、 $N_{\text{cls}} \approx 100$ でピーク値約 70% を持つ.

このような振る舞いを持つ理由は、以下のように理解できると考えられる. $N_{\text{cls}} \lesssim 100$ では、 N_{cls} の増加とともに

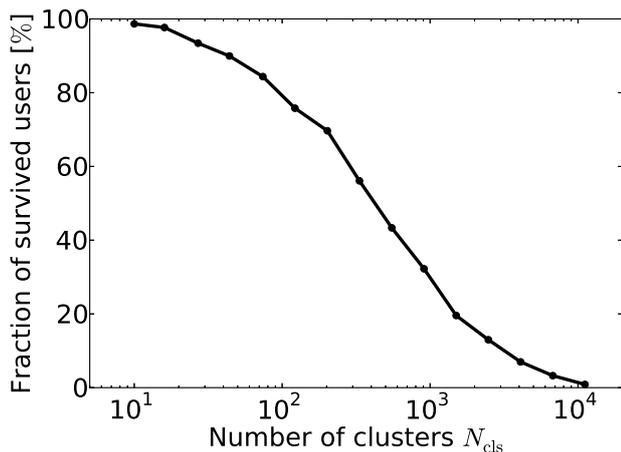


図 6 提案手法によって作成された匿名化データに含まれるユーザー数を、クラスター数 N_{cls} の関数として表示、横軸は対数軸となっている。

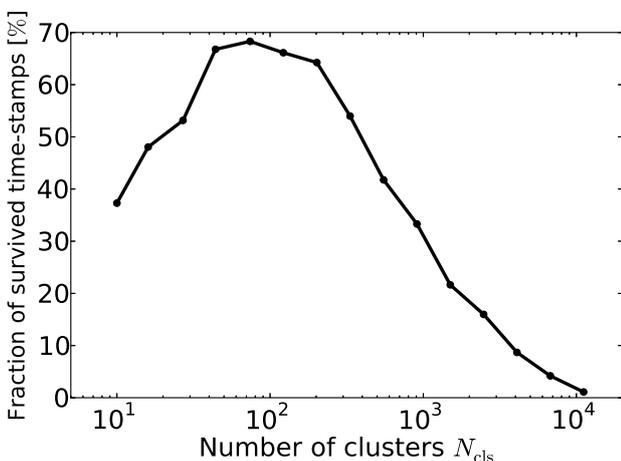


図 7 提案手法によって作成された匿名化データに含まれる位置情報数を、クラスター数 N_{cls} の関数として表示。

に、提案手法手順 (4) で削除されずに残る位置情報数が急激に増え、一方で提案手法手順 (5) で削除されずに残るユーザー数が比較的ゆるやかに減少するため、結果的に匿名化データに含まれる位置情報数が増えていく。一方、 $100 \lesssim N_{\text{cls}}$ では、 N_{cls} の増加とともに、提案手法手順 (4) で削除されずに残る位置情報数が緩やかに増加していくが、提案手法手順 (5) で削除されずに残るユーザー数が急激に減少するため、結果的に匿名化データに含まれる位置情報数が減っていく。

4.2.4 匿名化データの誤差評価

次に匿名化データの誤差を評価する。提案手法は、ユーザー・位置情報ともに削除するため、匿名化データと元データの比較はやはり単純な問題ではない。提案手法についても、4.1.3 節と同様の観点から評価を行う。具体的には、匿名化データと元データの間で、同一ユーザーの時刻が最も近い位置情報の時刻の差異と地理的距離を測る。それぞれ

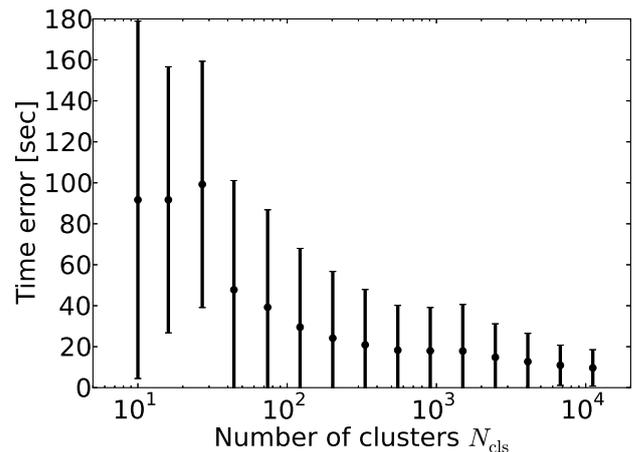


図 8 提案手法によって作成された匿名化データの位置情報の時刻誤差の平均値を、 N_{cls} の関数として表示、誤差棒は、標準偏差。

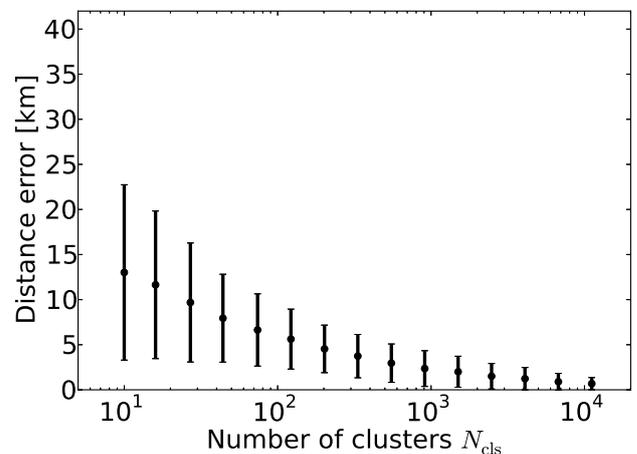


図 9 提案手法によって作成された匿名化データの位置情報の距離誤差の平均値を、 N_{cls} の関数として表示、誤差棒は、標準偏差。

を匿名化データ中の位置情報の時刻と距離の誤差とする。

4.2.4.1 時刻誤差

図 8 は、匿名化データに含まれる位置情報の時刻誤差の平均値を、 N_{cls} の関数として表している。標準偏差を誤差棒として示している。 N_{cls} が大きいほど、位置情報の加工度合いは弱いため、時刻の誤差が小さくなる。

4.2.4.2 距離誤差

図 9 は、図 4 と同様に匿名化データに含まれる位置情報の距離誤差の平均値を、 N_{cls} の関数として表している。この図でも、 N_{cls} が大きいほど誤差が小さくなるという時刻誤差と同様の傾向が見られる。

ここまでに見たように、クラスター数 N_{cls} によって、匿名化データに含まれるユーザー数、位置情報数、位置情報の時刻・距離の誤差が変化するため、匿名化データに要求される要件を満たす N_{cls} を選択することになる。

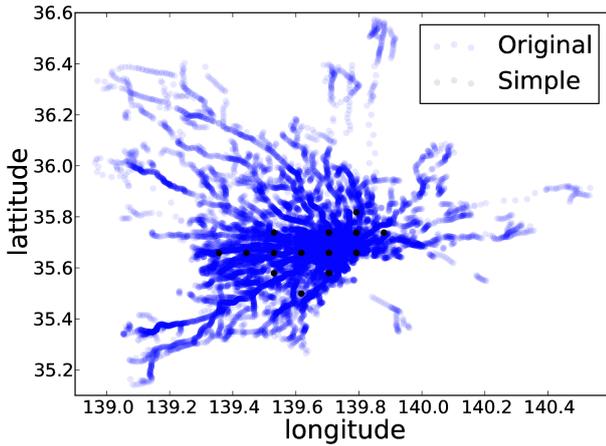


図 10 メッシュ数 $N_{\text{mesh}} = 324$ の簡易手法によって、作成された匿名化データ（平均距離誤差：2.95km）の位置情報を、元データの上に重ねて表示。

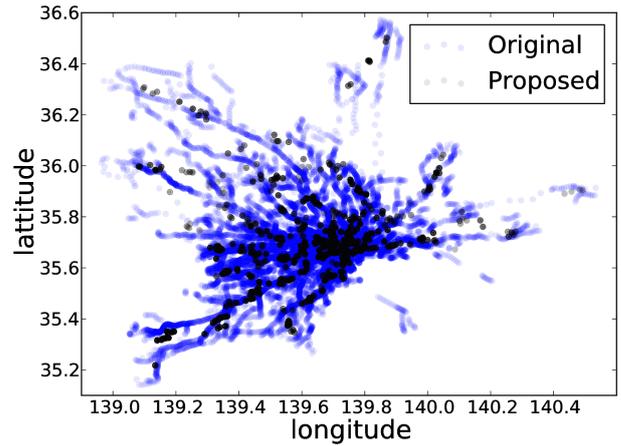


図 11 クラスタ数 $N_{\text{cls}} = 551$ の提案手法によって、作成された匿名化データ（平均距離誤差：2.96km）の位置情報を、元データの上に重ねて表示。

4.3 議論：簡易手法と提案手法の比較

最後に簡易手法と提案手法を、匿名化データの平均距離誤差を軸に比較する。平均距離誤差を軸にすることで、実質的な緯度・経度の加工レベルを物差しに、比較が可能となる。

4.3.1 提案手法は郊外の位置情報を残しているか

本稿の目的は、人通りの比較的少ない経路を通ったユーザーも匿名化データに含むことを可能にする匿名化手法の提案にある。本稿の実験に用いたデータは、東京近郊のものであるから、領域の中心部に位置する都心が人通りが非常に多く、それに比べ外縁部にある郊外は少ないと考えられる（図 1 を参照）。したがって、郊外のデータが残っているかどうか、が問題となる。図 10 にメッシュ数 $N_{\text{mesh}} = 324$ の簡易手法によって、作成された匿名化データの位置情報を、元データの上に重ねて示す。なお、この匿名化データに含まれる位置情報の平均距離誤差は 2.95km である。予想通り、都心の位置情報のみを含んでいることがわかる。図 11 は、クラスター数 $N_{\text{cls}} = 551$ の提案手法によって、作成された匿名化データの位置情報を示している。この匿名化データの平均距離誤差は 2.96km であり、図 10 で示したデータとほぼ等しい。これら 2 つの図を比較して、提案手法の方が郊外の位置情報を残していることが一目瞭然と見て取れる。

また、この傾向を広く確認する。人通りの比較的少ないデータを含むかどうか、は前述の東京近郊のデータの性格から、匿名化データがより広範囲にわたる位置情報があるかどうかを調べればおおそ評価できると考えられる。そこで、匿名化データに含まれる緯度・経度の最大値、最小値を測り、（緯度の最小値、経度の最小値）と（緯度の最大値、経度の最大値）の 2 点間の距離を、カバーする範囲の広さを評価する指標とする。得られた結果を、図 12 に、平均距離誤差の関数として示す。平均距離誤差の広い範囲、

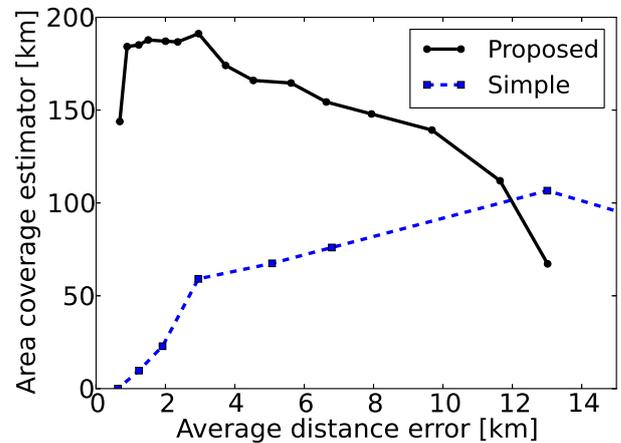


図 12 匿名化データがカバーする範囲の評価指標を、匿名化データの平均距離誤差の関数として表示。実線が提案手法、破線が簡易手法の結果。

つまり緯度・経度の加工レベルについて広い範囲で、提案手法により作成された匿名化データの方がより地理的に広範囲の位置情報を保持していることがわかる。

4.3.2 ユーザー数

次に匿名化データに含まれるユーザー数を比較する。図 13 に、匿名化データの平均距離誤差の関数として、匿名化データが含むユーザー数を示す。ユーザー数は、元データのユーザー数の割合で示している。提案手法の方が、より多くのユーザーの軌跡を保持していることがわかる。逆に見れば、ユーザー数が同程度あれば、提案手法の方がより小さい距離誤差の匿名化データを作成できることがわかる。

4.3.3 1 ユーザーあたりの位置情報数

次に、図 14 に、匿名化データの平均距離誤差の関数として、ユーザー 1 人あたりの位置情報の数を示す。提案手法により作成された匿名化データの方が、より少ない位置情報数により構成されていることがわかる。これは、提案手

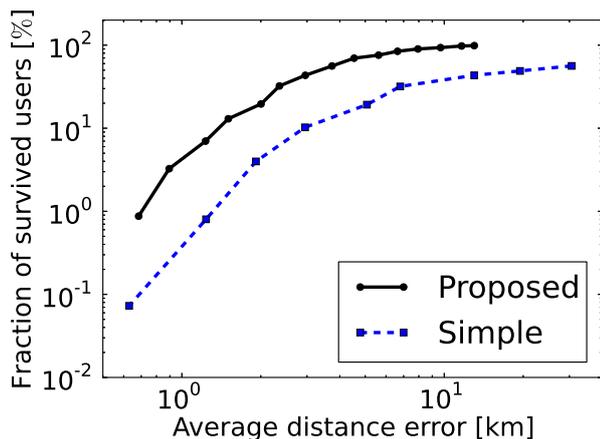


図 13 匿名化データの平均距離の誤差として、匿名化データが含むユーザー数を元の値の割合として表示。実線が提案手法、破線が簡易手法の結果。

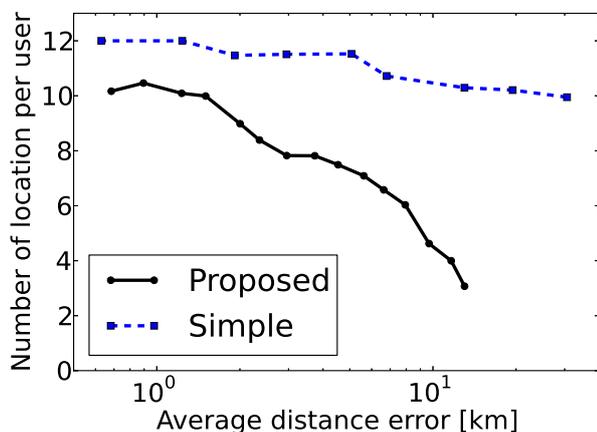


図 14 匿名化データの平均距離誤差の関数として、ユーザー 1 人あたりの位置情報の数を表示。実線が提案手法、破線が簡易手法の結果。

法が手順 (4) で位置情報を削除するためだと考えられる。

図 13 と図 14 に示した結果から、提案手法が人通りの比較的少ない領域の軌跡を持つユーザーを保持できるのは、位置情報の統一化と 1 ユーザーあたりの位置情報数の減少により軌跡のユニーク性を低下させ、匿名性を上げることで、代わりにユーザー数を増加させることができるためだと考えられる。

5. 結論

本稿では、時刻-緯度-経度の三次元時空間における位置情報のクラスタリングを利用した、軌跡情報の新たな k -匿名化法を提案した。東京近郊の擬似的な軌跡情報を用いた実験を行い、提案手法のパラメーターである時空間におけるクラスター数 N_{cls} によって、匿名化データに含まれるユーザー数、位置情報数がどのように変化するか示した。

また、匿名化データの誤差を評価する方法も提案し、時刻と距離に関する誤差評価を行った。

提案手法と比較するため、緯度-経度の二次元平面上のメッシュを利用した非常に簡易な手法も導入した。比較の結果、提案手法は確かに人通りの比較的少ない経路を通ったユーザーのデータを匿名化データに保持することを示した (4.3.1 節)。加えて、提案手法は距離誤差をより小さくすることも示した (4.3.2 節)。

クラスタリングに基づく類似手法としては、Abul らの研究がある [5]。彼らの手法は、軌跡の開始・終了時刻を揃えるように前処理が必要となる。一方、本稿の提案手法はそのような前処理は必要なく、開始・終了時刻のみならず位置情報の時刻は任意でよく、この点で優位性があると言える。

一方で、提案手法には改善の余地がある。まず、4.3.3 節でも見たとおり、提案手法は位置情報の数を減らしてしまう。実際の軌跡情報を匿名化し、分析を行うためには、削除された位置情報の推定 [6] 等により、この弱点の克服が必要となる場合もあるだろうと考えられる。その際には、本稿では取り扱わなかった性別や年齢といった位置情報以外の属性を含めた匿名化を考えることも必要となると考えられる。他にも、本稿で示した時刻・緯度・経度の無次元化の方法、クラスタリングの計測方法にも自由度があり、これらを最適化していくことは今後の課題である。

謝辞 「擬似人流データ」を公開し、無料利用を可能にくださった株式会社ナイトレイ、東京大学 柴崎・関本研究室、マイクロジオデータ研究会、人の流れプロジェクト、東京大学 空間情報科学研究センターに感謝いたします。

参考文献

- [1] L. Sweeney, “ k -Anonymity: A Model for Protecting Privacy”, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems archive, Volume 10 Issue 5, October 2002, Pages 557 - 570
- [2] 五十嵐大, 千田浩司, 高橋克巳, 「 k -匿名性の確率的指標への拡張とその適用例」, CSS, 2009.
- [3] D. Ikarashi, R. Kikuchi, K. Chida, K. Takahashi, “ k -Anonymous Microdata Release via Post Randomisation Method”, 10th International Workshop on Security, IWSEC 2015, Nara, Japan, August 26-28, 2015, Proceedings, pp 225-241
- [4] C. Dwork, “Differential privacy”, ICALP'06 Proceedings of the 33rd international conference on Automata, Languages and Programming - Volume Part II Pages 1-12
- [5] O. Abul, F. Bonchi, and M. Nanni, “Never walk alone: Uncertainty for anonymity in moving objects databases”, In Proceedings of the IEEE International Conference on Data Engineering, 2008.
- [6] M. E. Nergiz et al., “Towards trajectory anonymization: A generalization-based approach”, Transactions on Data Privacy, 2(1):47-75, 2009.
- [7] C.-Y. Chow and M. F. Mokbel, “Trajectory Privacy in

- Location-based Services and Data Publication”, ACM SIGKDD Explorations Newsletter, 2011.
- [8] R. Shigetomi-Yamaguchi et al., “Applicability of Existing Anonymization Methods to Large Location History Data in Urban Travel”, IEEE International Conference on Systems, Man, and Cybernetics, 2012.
 - [9] 寺田雅之, 竹内大二郎, 齊藤克哉, 本郷節之, 「差分プライバシー基準に基づく情報秘匿手法の一考察」, DICOMO, 2015.
 - [10] 疋田敏朗, 山口利恵, 「階層化符号表現を利用した移動履歴の匿名化手法」, DICOMO, 2015.
 - [11] A. Machanavajjhala et al., “ ℓ -diversity: Privacy beyond k -anonymity”, ACM Transactions on Knowledge Discovery from Data (TKDD), Volume 1 Issue 1, March 2007, Article No. 3
 - [12] 株式会社ナイトレイ, 東京大学 CSIS との研究活動成果として SNS 解析データを元とした「疑似人流データ」を無料公開, 入手先 (<http://nightley.jp/archives/1954>).
 - [13] J. A. Hartigan and M. A. Wong, “Algorithm AS 136: A K-Means Clustering Algorithm”, Journal of the Royal Statistical Society. Series C (Applied Statistics) Vol. 28, No. 1 (1979), pp. 100-108