第57回 プログラミング･シンポジウム 2016.1.8-10

移動履歴の匿名化手法の許容解導出に関する検討

疋田 敏朗[1]　山口 利恵[1]

概要：個人情報保護においてプライバシ保護は重要であると言われており，利用履歴を匿名化したとしてもその個人特有の利用履歴がある場合には本人の特定がしやすいと一般的に言われている．近年発展が著しいスマートフォンのアプリケーションやIoTデバイス，特にICカードの移動履歴についても利活用の期待が高まる一方で，その履歴情報のプライバシー保護の重要性が問題となりつつある．
本論文ではスマートフォンのアプリケーションやIoTデバイスの移動履歴に関して，個人を特定しないという匿名化に関して，最適解に近い許容解の導出方法に関する検討を行った．さらに移動履歴の匿名化について説明をしたうえで，最適解を得る方法と空間を階層的に符号化するGeoHashを用いて，移動履歴を一次元に符号化し，その上でGeoHashの階層構造を利用することで移動履歴の許容解を高速に実現する方法について検討を行っている．より良い許容解を得るために符号化の分割数を変更する方法と木構造の階層数の組み合わせを変更して従来の方法よりも精度が良い許容解を得るための方法を示し，より最適解に近い方法を模索する．

キーワード：位置情報，移動履歴，k-匿名化，R-Tree

1. はじめに

近年，GPSを始めとする測位デバイスが携帯機器の機能の一部として搭載されるようになり，さらに携帯網の発展により携帯機器が測位した位置情報をセンター上のサーバに送信し，蓄積することが現実的になっている．さらにスマートフォンとそのアプリケーションや各種IoTデバイスの普及によって，一般のユーザから大量に位置情報を収集・蓄積することができるようになっており，蓄積した位置情報を活用することで従来は困難であった新たなサービスが次々と生まれるようになっている．

また鉄道乗車券のICカードが進展し，これらの乗車券の利用履歴を用いることでIoTデバイスの移動履歴と同様に利用者の移動履歴を得ることも難しくなくなった．

これらの各種IoTデバイスやICカードから得られる移動履歴情報を活用すれば移動履歴の推定ならびに予測を高精度に行うことができると考えられる．

しかしながら，IoTデバイスやICカードから発生する移動履歴情報はその内容に位置情報やその移動履歴を含み，個人を特定しうる情報やセンシティブな情報が含まれることから，その履歴情報をそのまま利活用することはできない[*1]．

すでに，鉄道乗車に関するICカードの移動履歴が事業者間で提供されたり[1]，個人の移動履歴に

[1] 東京大学大学院情報理工学系研究科
The University of Tokyo

[*1] 正確には同意を取得しさえすれば履歴情報をそのまま利活用することは不可能ではない．

関するプライバシ保護に関する懸念が示されるなどの事例も発生している．それらの懸念に対応するためには履歴を匿名化することによって対応を行うことが考えられる．すでに携帯事業者では位置情報の履歴を匿名化してプライバシ保護を行いつつ統計を行う [2] という事業が始まっている．

交通履歴データに関しても同様に移動履歴はプライバシを保護し，個人を特定しないまま移動履歴を利用するための変換手法が望まれている．我々は都市開発においては最も必要と考えられる移動者の出発地と目的地の履歴に着目して匿名化手法を検討 [3] した．

本論文では，まず 2 章で位置情報と移動履歴の匿名化の現状に関する紹介を行う．従来の移動履歴情報の匿名化に関する研究について紹介を行う．

その上で 3 章では移動履歴の匿名化の要件と移動履歴の匿名化の最適手法について論じる．

次に 4 章では移動履歴を高速に匿名化するための方法について，矩形化を利用した階層化符号化方式における手法について説明を行う．

そして 5 章ではこれらの手法に関して，より詳細な許容解を求めるための手法について検討を行うこととする．

最後に 6 章でこれらの手法に関する議論を行う．

2. 履歴データの匿名化と位置情報と移動履歴の匿名化の現状

本章では，個人情報保護法における個人情報の定義について説明を行うともに，まず一般的な履歴データの匿名化について説明し，そのうえで位置情報や移動履歴の匿名化の現状について述べる．

2.1 個人情報保護法による個人の特定と識別属性

2015 年 12 月現在における我が国の個人情報保護法 [*2] においては，個人情報とは個人が特定できるような情報のほかに，『(他の情報と容易に照合することができ、それにより特定の個人を識別することができることとなるものを含む。)』という形で他の情報と照合することで個人が特定できる情報もまた個人情報であるとされている．ここで識別とはそれが誰だかわからないが特有の 1 名に分離できるということであり，特定とはそれが固有の 1 名を示すこととされる，

ここであるデータ T が存在した場合にそのリストの全項目を QID として $k-$ 匿名化を行ったデータ T' については，1 つの情報について少なくとも 2 つ以上の列が該当することから一意に特定ができないことが知られている．すなわちデータ T の全項目を QID としてに $k-$ 匿名化による変換を行うことができればその情報について個人を特定することはできないとすることができる．

2.2 匿名化の研究と移動履歴に関する匿名化の現状

あるデータ T が存在した時に，そのデータ T から個人が特定できないようにする変換を行い，データ T' を生成する作業を匿名化を行うと定義することにする．

個人を直接的かつ一意的に識別する属性，たとえば氏名 [*3]，個人番号 [*4] などを示し，これを **個体識別属性** と呼ぶ

個人を一意に識別できないとしても複数の属性を組み合わせると個人を一意的に識別できるものもある．たとえば性別，生年月日，住所などが該当する．これらの属性を**疑似識別属性** (Quasi Identifier，以下 **QID**) と呼ぶ．

たとえば履歴データとして表 1 の様な購買データが存在したとする．この時に個体識別属性は会員番号と氏名であり，擬似識別属性は住所と年齢となる，購買品は一般的な考え方ではその他の属性として位置づけ，本稿ではそのように扱うこととする．

このデータ T の匿名化を行うとする． 例えば氏

[*2] 2015 年 9 月に改正法が成立しており 2 年以内に施行されることなっている，改正法では個人を識別できる符号情報についても追加で個人情報となることとなったが現行法の個人情報の範囲を拡大するものではないとされている

[*3] 厳密には氏名だけでは同姓同名の個人が複数存在する可能性があるが，社会通念では個体識別属性とみなされている

[*4] 各個人に一意に割り当てられている番号，例えば日本でいえばマイナンバー，米国でいえばソーシャルセキュリティーナンバー．

表 1 履歴データの例

会員番号	氏名	住所	年齢	購買品
11111	東京 太郎	文京区本郷	35	模型
22222	横浜 花子	文京区小石川	32	自動車
33333	春日 坪根	文京区春日	37	フィギュア
44444	多摩 珠子	府中市住吉町	29	洗顔料
55555	電機 研究	府中市日新町	22	マスカラ
66666	黒鴉 国府	府中市宮西町	28	口紅
77777	谷津 夏見	船橋市若松	47	雑誌
88888	神崎 干潟	船橋市海神町	45	文庫
99999	千葉 梨花	船橋市湊町	40	雑誌

表 2 履歴データの例 (QID による k– 匿名化)

会員番号	氏名	住所	年齢	購買品
		文京区	30 代	模型
		文京区	30 代	自動車
		文京区	30 代	フィギュア
		府中市	20 代	洗顔料
		府中市	20 代	マスカラ
		府中市	20 代	口紅
		船橋市	40 代	雑誌
		船橋市	40 代	文庫
		船橋市	40 代	雑誌

名情報のみを削除し，会員番号のみを利用する方法は一般的には仮名化と呼ばれる．仮名化を行っても個体識別属性が残っていると一意に識別できるため，仮名化は厳密な意味での匿名化ではない [4].

表 2 はこのデータ T に対して，一般的によく知られている匿名化である k– 匿名化を適用したものである．k– 匿名化は Sweeney らによって提唱 [5] された概念で同一の擬似識別属性に対して，最低でも $n \geq k$ のデータが存在するように，擬似識別属性を曖昧化する表 2 の例の場合は，住所の末尾を落として市区町村にまとめ，年齢を細かい年齢から年代に変更することで k を満たすようにしている．また，k– 匿名化の情報では，匿名化として不十分として，データの種別を定量的に計る手法 l-diversity [6] や データの全体の割合傾向を計る手法 t-clossness といった手法も提案されている [7].

ただし，著者らの研究 [4] によるとその他の属性もそれ自体を履歴情報として考えると擬似識別属性として考えられるため，厳密に匿名化を検討する際にはその他属性自体も擬似識別として取り扱いを行う必要がある．

次に匿名化の位置情報への拡張について述べる．位置情報について k– 匿名化を行った例 [8] は 2003 年に Gruteser らによって報告されている．この例では地点をグリッドごとに区切り，それぞれの地点情報をもとに k– 匿名化が行われている．Gkoulalas-Divanis らによるまとめ [9] によれば，k– 匿名化の手法は一般的に今あるデータを中心とした区切り方と地形情報を活用したグリッドベースの区切り方の 2 種類に分けることができると主張している．

k– 匿名化の他の匿名化手法としてはノイズを混入するという手法が挙げられる．実際の位置情報の他に複数のダミーの位置情報を挿入させることでデータ自体の匿名性を担保する手法 [10] [11] が有名である．またダミーデータの混入手法についてはより高度な手法が提案されている，Niu ら提案 [12] によればダミーデータの配置場所を統計的に検討することで，ダミーユーザの現実的な配置が可能になり，より強固な配置が可能になるとされる．

移動履歴に関してもダミーデータを加えて匿名化するという手法が提案 [13] されている．この手法はランダムにダミーデータを加えた移動履歴情報を生成することで，リアルユーザのデータを秘匿化する．しかしながらダミーを利用する方法では受領した位置情報にダミー情報がかなりの確率で紛れ込むため位置情報の利用者側から見るとデータが使いにくいという問題が発生する．例えば実際の情報に 4 倍のダミーデータを混入した場合，位置情報を的中させることが出来る確率を 20%近くに低下させることができるが，利用者から見ると 1/5 でしか正確なデータが存在しないということになる．これは特にビッグデータ処理を前提とした場合にデータ自体の信頼性がなくなること言うことを意味しているため，データの利用目的によってはこの手法は使えない．

また移動履歴をグリッド化して k– 匿名化する方法はいくつか提案されている山口 [14] の手法では単一のグリッドで k– 匿名化を実施するという

手法が提案されており，著者ら [3] は可変グリッドを利用した単体移動履歴の匿名化を提案している．

3. 移動履歴の匿名化の要件と最適化

移動履歴匿名化の目的は IoT デバイスやスマートフォンなどの移動履歴を大量に取得したデータから，個人を特定せず，元データに対してノイズの混入を行わずに，匿名性のあるデータに変換・生成することである．

3.1 移動履歴の匿名化のために必要な要件

本研究では個人が特定できない移動履歴を生成することを目的としている．昨今，Foursquare や Facebook などの SNS への投稿は位置情報を付加することが可能であったり，地点情報を付加してチェックインすることができる．そのため別の手段で収集された履歴情報や位置情報とデータ処理履歴情報を照合することで個人を特定することも難しくはなくなりつつあり，個々の履歴情報に関して，他の情報を用いた場合でも個人を一意特定しうる状態ではないことが必要ということになる．

移動履歴から個人が特定されるケースを列挙すると以下のようなケースが考えられる．

(1) **その情報自体が個人を特定できる情報を含む場合**
(2) **位置情報自体が自宅などを指し示す場合**
(3) **位置情報と時刻の組み合わせにより個人が特定される場合**
(4) **履歴が固有のために個人が特定される場合**

これらのケースに対応する匿名変換について検討を行う．1 についてはその情報自体を削除する．つまり**氏名・固有 ID などは削除**すればよい．

2 については自宅などの地理情報については**位置情報の広域化**で対応する．すなわち情報が指し示す範囲を拡大することで自宅という情報が判明しないようにする．つまり文京区本郷 7-3-1 という住所を文京区本郷 7 のようなエリアに拡大することで位置情報が固有である可能性が残るが，位置情報自体が特性を持つことはなくなる．

3 のように位置情報と時刻の組み合わせで外部から観察することで個人が特定可能である場合について検討する．これは位置情報と時刻の組み合わせが固有であることが原因なので，**位置情報と時刻の組み合わせが固有でない**ようにすればよい．

最後に 4 のように組み合わせ情報が固有であるために個人が特定できる可能性について検討する．履歴情報は同じ ID の履歴をリンクさせることで個人が特定可能なことが知られている．そこでまず移動履歴に関して，**出発地と目的地の情報のみを残し，それ以外のデータとのリンクを削除**することにする．履歴情報としては情報量を低下させることになるが都市開発のための交通量推定のデータとしてはこれで十分である．すでに 1 で固有の ID を削除してしまっているため，A さんが S 駅→ T 駅という移動を行い，その後に T 駅→ S 駅という移動を行った場合でも，それが同一ユーザの移動であるかどうかはわからない．そのため履歴の中では時刻と出発地と到着地という情報だけが残っている

次に**この時刻と出発地，到着地の組み合わせ情報から特定できない**ようにする．この場合時刻データと出発地，到着地という情報で一意特定ができない情報に変換すればよいのであるから，これと 2 の条件，すなわち位置情報を適切にエリア拡大しながら出発地と到着地の組み合わせ情報に関して，$k-$ 匿名化を行う手法があれば目的は達成される．

3.2 移動履歴の k-匿名化と最適化

次に実際の移動履歴を元に $k-$ 匿名化を行うことを考える．図 1 に出発地から目的地までの移動履歴の例を示す．この例の場合の履歴の総数は 6 である．

この移動履歴を $k = 3$ で匿名化することを考える．$k = 3$ の匿名化とは匿名化した際にどの要素も 3 以上が存在する必要があるということになる．移動履歴の場合は出発地と目的地が同じクラスタを生成し，クラスタの要素がすべて 3 以上あれば良いということになる．

図 2 に $k-$ 匿名化したクラスタの例を示す．$k-$ 匿名性を満たす匿名変換の結果は複数存在する．ここでその匿名変換に関して，$\forall x, \sum^{x} |x - x'|$ を変換誤差と定義すると変換誤差が最小になるクラ

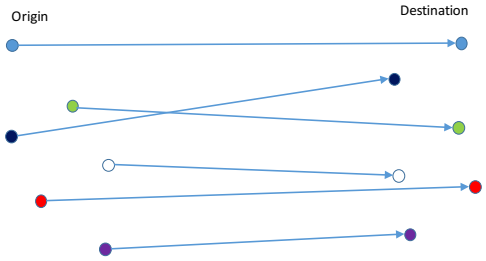

図 1 元の移動履歴

スタを探せば良いことになる．

ここで誤差を最小にする k– 匿名性変換を最適な k– 匿名化と呼ぶことにする．

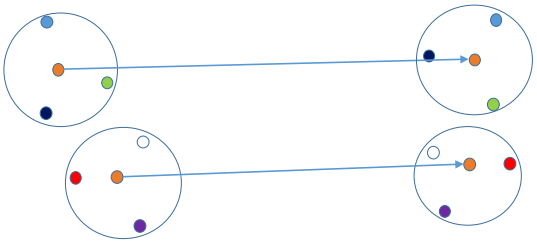

図 2 移動履歴の最適化

k– 匿名化は組合せの最適化であることからも理解できるように，一般に k– 匿名化の最適化は NP-Hard であること [15][16] が知られており，最善を尽くした場合の計算量は $O(n \log n)$ であることが知られている

移動履歴の匿名化も出発地から目的地までの移動を一つの履歴とみなすと一般の k– 匿名化と同様に NP-Hard であり，図 2 に示すような誤差を最小にする最適な k– 匿名化の計算は非常に難しい．

4. 許容解を求める方法の検討

3 章で示した要件を満たす移動履歴の最適な k– 匿名化は非常に時間がかかることから，本章では最適解ではなく，許容解を求める方法について検討を行う．

とくに移動履歴を木構造のように階層的に記述することで利用価値の高い匿名化を行う方法を提案する．まず位置情報の階層化について説明を行った上で，位置情報の階層化を移動情報に拡大する提案手法について説明を行う．

4.1 階層化表現用いた位置情報の符号化

二次元情報である位置情報を計算機で扱いやすいように，符号化する研究は以前から行われてきた．R 木 [17] は位置情報のような多次元情報 [*5] を矩形化しながら，木構造で管理することで空間インデックスとして活用できる方法として有名である．

本節では位置情報を矩形情報かつ木構造で管理しつつ，計算機で扱いやすいように符号化することを考慮した手法である階層化による位置情報の符号化について説明する．Geohash[18] は位置情報を符号化するジオコーディングの一種であり，階層構造を持ちつつも位置座標を空間分割する機能を持つ．

図 3 に Geohash による階層化と領域分割の例を示す．Geohash は単一の文字列で 2 次元の空間座標を表現できる．たとえば (139.761,35.714) は xn77hmdb と表現することができる．この場合 xn77hmdb の上位エリアは xn77hmd であり，xn77hm がさらにその上位エリアとなる．

Geohash の変換方法を図 4 に示す．このように Geohash による符号化を用いることで位置情報を階層構造を持った符号化を行うことができる．

4.2 階層化表現を用いた移動履歴の匿名化

次に階層化符号を利用した移動履歴の匿名化手法を説明する．階層化符号方式で表現ができたの

[*5] 世間の R 木実装は 2 次元のみを対象としたものが数多く見られるが，本来は多次元を正しく処理できることが望ましい

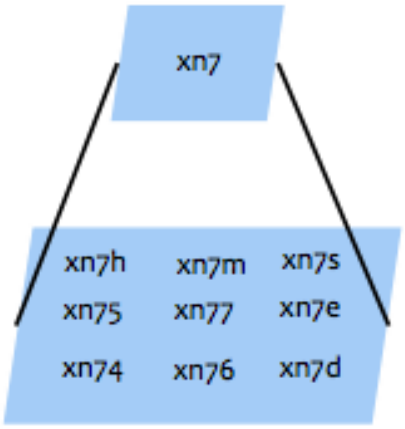

図 3 Geohash の表現と階層化

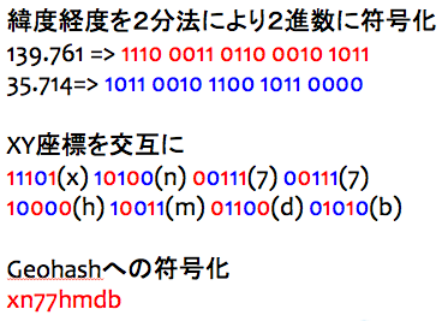

図 4 Geohash による符号化

は位置情報だけであったが，これを履歴情報に拡張することを考える．表現を簡易にするために，履歴情報自体を階層化するという手法を取ることにした．図 5 に具体的な方法を示す．本提案手法では階層化符号方式自体の出発地と目的地をそれぞれの文字単位で並べ替えることとする．こうすることで出発地と目的地を一つの文字列として表現できることになる．

提案手法の表現は元の Geohash の階層構造をそのまま維持しているので，4 文字単位でシフトすることでそのエリア構造をそのまま維持したままエリアサイズの調整ができることになる．

次に本手法を使い，$k-$ 匿名化を実施する方法について説明する．本件提案のアルゴリズムを Algorithm1 に示す．すでに単一の符号化がなされているので 4 文字単位で文字列一致を行い，該当

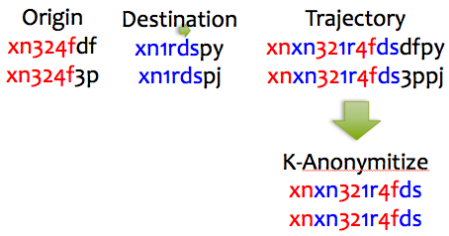

図 5 Geohash の移動履歴への拡張

の移動履歴が k 件以上存在するかどうかを探索する．k 件以上の履歴が存在した場合は該当履歴を抜き出し，残った履歴について 4 文字ごとに縮約を行いながら計算を行うことで $k-$ 匿名化を実現できる．4 文字ごとに縮約を行うのは Geohash の制限により，奇数桁と偶数桁でエリアサイズが異なるためである．Geoghash では bit 切り落としの関係上，奇数桁では正方形に近い形になり，偶数桁では経度方向に長い長方形になる．縮約サイズを同一にするために Geohash 自体を 2 文字で縮約できる 4 文字とした．

この際に同じ出発点であっても，目的地が異なる場合に関しては，異なるエリアサイズを許容する．すなわち [渋谷] から [新宿] や [六本木] への移動に関してはある程度小さなエリアで匿名化を行うが，[恵比寿] から [鷹の台] のような例であれば，[渋谷区] から [小平市] というレベルで匿名化する．

このように異なるサイズでの履歴の匿名化を許容することで，従来手法のように適切なエリアサイズの検討を行わずにデータにあわせて柔軟にサイズ調整をすることができるようになった．

本手法を適用した場合の例を図 6 に示す．本手法は最適解を導くものではないが，許容解を導くことができることと必要な計算量が少ないという特徴がある．

図 2 で説明した最適解との違いは，本手法では矩形で切れる領域の位置を予め制限してあるため，履歴の形に合わせた形での切り出しができず，ある矩形領域で条件が満たされなかった場合はより大きな矩形領域として切り出す必要があること．そして矩形の大きさの調整に柔軟性がないことである．そのため図 6 では中心領域に収まらなかった履歴は大きな領域として処理せざるを得なくなっ

Algorithm 1 階層化表現を用いた匿名化

```
for all すべての履歴データ do
  Origin = 出発地を階層化表現に変換
  Dest = 目的地を階層化表現に変換
  Trip = Origin と Dest をそれぞれ 1 文字目から
    交互に記述
  Hash Key に Trip を挿入
end for
while Trip が最小文字数 (領域が最大エリア) に達する
まで do
  for all すべての Hash Key do
    if 要素の数 ≥ k then
      匿名化結果を出力し，現 Key クリア
    else
      階層化表現の末尾に文字を削除
      新 Hash Key に挿入，現 Key を削除
    end if
  end for
end while
```

ている．誤差という意味では大領域は誤差が拡大する．

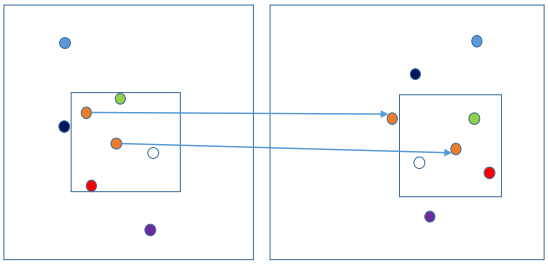

図 6 階層符号化による許容解の導入

ただし，領域や面積が必ずしも大きいことが良くないというわけではない．$k = 3$ の場合に 2 つがごく至近距離にあり，残り 1 つが離れた場合と 3 つが均等に離れた場合で領域の面積については前者のほうが大きかったとしても，誤差は後者のほうが大きい可能性はある．これは誤差の導入について単純な加法を導入したがこれが正しいかどうかという観点で議論が必要であることを示している．

5. 高速な許容解のための検討

4 章で導入した階層型符号化手法による $k-$ 匿名化の許容解の計算方法について，提案方法よりも精度が高い許容解を導く方法について検討を行う．

検討すべき点は多々あるが，以下の 2 点の方法について論じる

(1) GeoHash の符号化が Base32 で行われている問題

(2) 領域の拡大検索を出発地と目的地の両方で行っている問題

最初の問題は GeoHash の符号化方式によるものである．GeoHash はもともと位置情報を短い文字数で記述するという目的上，Base32 変換を行って 2 進数を文字変換している．2 進数表記自体が領域分割を示していることはすでに述べたとおりであり，1 文字進むごとに領域が 32 分割されることになる．

今回の手法では X,Y のサイズを同一にするために 2 文字単位の計算をしているので $32^2 = 1024$ 単位の分割になってしまうという問題がある．

対処方法としては Base32 ではなく，Base16 以下の変換による対応が挙げられる．Base16 にすることで分割を 256，Base8 で 64 に抑えることができると考えられる．

また 2^5 という奇数桁で区切ったことが 2 文字単位でのシフトを必要とした理由であるので Base16 または Base4 での変換を行うことで 1 文字シフトでも矩形領域の形状を変化させないことができるため，この問題に対しても解決することができると考えられる．この場合は Base16 で 16 単位の分割となるので 2 文字シフトよりも現実的な分割を得ることができると思われる．

次に領域の拡大検索についての方法について検討を行う．現在のアルゴリズムは出発地と目的地の木構造をそれぞれ交互に織り込んで一体とした木構造としている．そのため木構造をたどるときには出発地と目的地の双方でひとつ上の葉に移動することを要求している．

履歴 $xn771t \rightarrow xn778t$，$xn77hm \rightarrow xn778t$，$xn77hm \rightarrow xn776t$ を $k = 2$ で匿名化する例を考

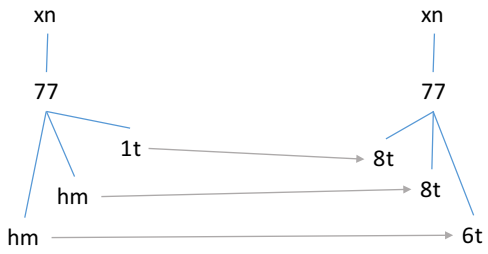

図 7 出発地と目的地の矩形領域の構造

える．図示化すると図 7 のようになる．

この場合，すべての履歴は $k = 1$ であるので第三階層で全てを匿名化することはできない．次に履歴を階層化して統合することを考える．

従来の手法で匿名化を行うと図 8 の (a) のように $xn77 \rightarrow xn77$ となる．この場合は 3 つの履歴を抜き出すことができるが情報としては出発地と目的地が $xn77$ という領域に閉じてしまい情報量が落ちている．

次に出発地または目的地のみの情報を詳細化することを考える．(b) のように $xn77hm \rightarrow xn77$ または (c) のように $xn77 \rightarrow xn778t$ とすることが考えられる．これらの場合は領域サイズは小さくなり，より詳細な移動の方法がわかったが，有効な履歴数は 2 に減ってしまう．残った履歴は必要であればより上位の階層での統合を行い，足りないようであれば削除する必要がある．

このように領域サイズを柔軟に調整をすることで従来手法より，詳細で最適手法に近い匿名化履歴を作ることができると考えられる．この場合は履歴の変換による情報損失をどのように評価するか？そして評価するための関数設計が重要となってくる．

最も簡単な手法としては木構造における階層数を出発地と目的地で加えて保持し，その合計が最大であるものを高く評価することが考えられるが，領域構造との間で差がでるという問題がある．面積であることを考慮すると階層数の自乗で加えて計算する方法というのも考えられる．領域が面積であるということを考慮するとこの方法はより現実に即していると考えられる．

また出発地と到着地のどちらを詳細に記述するべきか？という点についても評価が必要である．図 8 の場合に $xn77hm$ と $xn778t$ のどちらが重要か？ということの判断を行いやすくするために人口密集度または詳細なデータがほしい部分について一定の係数を加えて計算することなども将来検討する必要があると考えられる．

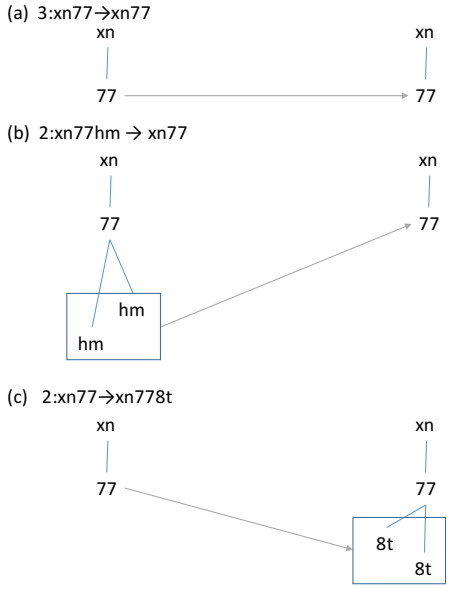

図 8 出発地と目的地の矩形領域の選択

6. まとめ

本論文ではスマートフォンのアプリケーションや IoT デバイス，特に IC カードの移動履歴に関して，個人を特定しないという匿名化に関して，最適解に近い許容解の導出方法に関する検討を行った．

本論文ではスマートフォンのアプリケーションや IoT デバイスの移動履歴に関して，個人を特定しないという匿名化に関して，最適解に近い許容解の導出方法に関する検討を行った．

さらに移動履歴の匿名化について説明をしたうえで，最適解を得る方法と空間を階層的に符号化する GeoHash を用いて，移動履歴を一次元に符号化し，その上で GeoHash の階層構造を利用することで移動履歴の許容解を高速に実現する方法について検討を行っている．より良い許容解を得るために符号化の分割数を変更する方法と木構造の階層数の組み合わせを変更して従来の方法よりも精度が良い許容解を得るための方法を示し，より最適解に近い方法を模索し，評価関数に関する今後の検討が必要であることを示した．

なお，本研究は東京大学空間情報科学研究センターの「人の流れプロジェクト」との共同研究である．データの整備並びに提供を行っていただいた空間情報科学研究センター各位に感謝する．

また，本稿をまとめるにあたり重要な示唆を頂いた長崎県立大学の山口先生に感謝する次第である．

参考文献

[1] ”Suica に関するデータの社外への提供についての有識者会議”：“Suica に関するデータの社外への提供について” (2014).

[2] 寺田雅之：“モバイル空間統計：携帯電話ネットワークを活用した人口推計技術とその応用 (ビッグデータ特別セッション)”, pp. 63–66 (2014).

[3] 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”, マルチメディア、分散、協調とモバイル (DICOMO2015) シンポジウム 2015 情報処理学会 (2015).

[4] 板倉陽一郎 伊藤孝一 菊池浩明 高木浩光 高橋克巳 中川裕志 疋田敏朗廣田啓一 山口利恵：“「完全な匿名化」幻想を超えて”, 暗号と情報セキュリティシンポジウム 2014 電子情報通信学会 (2014).

[5] L. Sweeney: “k-anonymity: a model for protecting privacy”, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, **10**, 5, pp. 557–570 (2002).

[6] A. Machanavajjhala, D. Kifer, J. Gehrke and M. Venkitasubramaniam: “l-diversity: Privacy beyond k-anonymity”, ACM Transactions on Knowledge Discovery from Data (TKDD), **1**, 1, p. 3 (2007).

[7] D. Rebollo-Monedero, J. Forné and J. Domingo-Ferrer: “From t-closeness to PRAM and noise addition via information theory”, Privacy in Statistical DatabasesSpringer, pp. 100–112 (2008).

[8] M. Gruteser and D. Grunwald: “Anonymous usage of location-based services through spatial and temporal cloaking”, Proceedings of the 1st International Conference on Mobile Systems, Applications and Services, MobiSys ’03, New York, NY, USA, ACM, pp. 31–42 (2003).

[9] A. Gkoulalas-Divanis, P. Kalnis and V. S. Verykios: “Providing k-anonymity in location based services”, SIGKDD Explor. Newsl., **12**, 1, pp. 3–10 (2010).

[10] H. Lu, C. S. Jensen and M. L. Yiu: “Pad: privacy-area aware, dummy-based location privacy in mobile services”, Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile AccessACM, pp. 16–23 (2008).

[11] H. Kido, Y. Yanagisawa and T. Satoh: “An anonymous communication technique using dummies for location-based services”, Pervasive Services, 2005. ICPS’05. Proceedings. International Conference onIEEE, pp. 88–97 (2005).

[12] B. Niu, Q. Li, X. Zhu, G. Cao and H. Li: “Achieving k-anonymity in privacy-aware location-based services”, Proc. IEEE INFOCOM (2014).

[13] P. Shankar, V. Ganapathy and L. Iftode: “Privately querying location-based services with sybilquery”, Proceedings of the 11th international conference on Ubiquitous computingACM, pp. 31–40 (2009).

[14] R. S. Yamaguchi, K. Hirota, K. Hamada, K. Takahashi, K. Matsuzaki, J. Sakuma and Y. Shirai: “Applicability of existing anonymization methods to large location history data in urban travel”, Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference onIEEE, pp. 997–1004 (2012).

[15] A. Meyerson and R. Williams: “On the complexity of optimal k-anonymity”, Proceedings of the Twenty-third ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS ’04, New York, NY, USA, ACM, pp. 223–228 (2004).

[16] G. Aggarwal, T. Feder, K. Kenthapadi, R. Motwani, R. Panigrahy, D. Thomas and A. Zhu: “Approximation algorithms for k-anonymity”, Journal of Privacy Technology (JOPT) (2005).

[17] N. Beckmann, H.-P. Kriegel, R. Schneider and B. Seeger: “The R*-tree: an efficient and robust access method for points and rectangles”, Vol. 19, ACM (1990).

[18] G. Niemeyer: “Geohash”, http://geohash.org/.