Geo 識別不能性を用いた経路端点の曖昧化

浅田　真帆† 　曹　　洋†† 　吉川　正俊††

† 京都大学工学部情報学科　〒 6068501 京都府京都市左京区吉田本町 36-1
†† 京都大学情報学研究科　〒 6068501 京都府京都市左京区吉田本町 36-1
E-mail: †{asada,soyo}@db.soc.i.kyoto-u.ac.jp, ††yoshikawa@i.kyoto-u.ac.jp

あらまし　近年，スマートフォンの普及や GPS 測位機能の発達等により，人々の位置情報がかなり正確かつ容易に把握できるようになっている．このようなデータはマーケティングや都市計画など様々な場面に応用することができ，より有効に活用するためにそれらを売買する市場の整備も始まっている．しかし一方で，正確すぎる位置情報は個人を特定しかねないというプライバシリスクがあり，それを防ぐために位置情報の匿名化や曖昧化を行う必要がある．また位置情報の中でも移動経路情報は特にプライバシリスクが高いと言われるが，その匿名化・曖昧化の技術は発展途上にある．そこで本論文では経路情報の匿名化・曖昧化に重点を置き，自宅のような経路端点を特定できず，しかも経路情報としての効用性を失わない手法を提案する．この手法において，プライバシ保護の側面では，位置情報匿名化に差分プライバシの概念を応用した Geo 識別不能性 (Geo-indistinguishability) という技術を利用する．また，元の経路が最短経路であるという前提の下，出力経路は途中まで入力経路と同じ経路を辿り，Geo 識別不能性を満たす別の点へ最短経路を通って向かうという形にしており，この点で効用性を保つことも目指している．実データを用いた実験では，京都の道路ネットワーク上で様々な始点を持つ経路の曖昧化を行い，その出力経路の効用性を比較した．

キーワード　差分プライバシ

1. はじめに

近年，スマートフォンの普及や GPS 測位機能の発達等により，人々の位置情報がかなり正確かつ容易に把握できるようになっている．このようなデータは公私様々に応用することができる．位置情報の活用例としては以下のようなものが挙げられる [1]．

（1） ターゲティング

現在のユーザの位置情報からその時点で自社店舗の周辺にいるユーザに対して広告を配信したり，過去に過去にユーザが訪れた場所を基にユーザの嗜好や移動範囲を推測し，場所の推薦や住宅広告を配信したりする．Near(注1) というターゲティング広告配信技術などがこれにあたる．今日では一点の位置情報だけでなく，その系列である移動経路を収集し活用する『リアル行動ターゲティング』という手法も用いられ始めている [2]．これはある一人のユーザの移動経路を複数集めることでその行動パターンを分析し，ユーザに合わせた広告をより効果的に配信するというものである．

（2） 道路交通情報の把握

リアルタイムで人々の位置情報の時空間データを分析して交通状況を把握したり，最適な移動ルートを推薦したりする．Google Maps などで利用されている．

（3） 都市計画

都市で生活する人々の移動パターンや居住分布を分析し，都市計画に生かす．

（4） 公衆衛生

群衆の位置情報の分析を感染症の蔓延防止策に役立てる．

（5） 位置情報を用いたゲーム

『Pokemon Go』など，位置情報がなければ有効に機能しないアプリケーションがある．

以上のように位置情報が価値を持ち始める中で，より有用かつ多くの位置情報を手に入れるために，それらを売買する市場の整備が始まっている [1]．日本でも位置情報をはじめとしたパーソナル情報の流通を司る “情報銀行”の設立を目指す動きも起こっている [3]．

しかし一方で，正確すぎる位置情報は個人を特定しかねない，というプライバシリスクがある．そのため，図 1 に示すように，人々の位置情報に対するプライバシ意識は非常に高いという調査結果もある．こういったプライバシリスクを防ぐために，位置情報の匿名化や曖昧化を行う必要がある．また位置情報の中でも，移動経路情報は特にプライバシリスクが高い．例えば訪

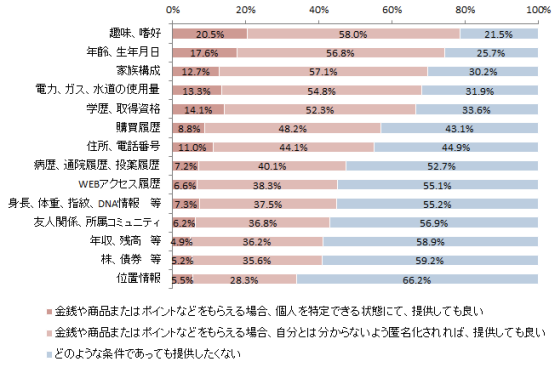

図 1　提供しても良いと思うデータの条件 (全体)（n=1,059）[4]

(注1)：https://near.co/

れた場所の位置情報が四つ分かれば 95%の確率で個人を特定することができるという研究結果 [5] や，複数の経路情報を公開することにより，よく訪れる場所や自宅の場所などが特定されてしまう可能性もある．しかしプライバシリスクの高さに反して，こういった経路データの匿名化や曖昧化の技術は発展途上にある．

そこで，本論文では経路情報の匿名化・曖昧化に重点を置き，中でも自宅のような経路の端点を特定できず，しかも経路情報としての効用性を失わない手法を考える．この手法において，プライバシ保護の側面では，位置情報匿名化に差分プライバシの概念を応用した Geo 識別不能性 (Geo-indistinguishability) という技術を利用する．また，元の経路が最短経路であるという前提の下，出力する経路は途中まで元の経路と同じ道を辿り，Geo 識別不能性を満たす別の点へ最短経路を通って向かうという形にしており，この点で効用性を保つことも目指している．

また効用性を示す指標として入力経路と出力経路の誤差を考え，経路始点やダミー端点を様々な位置にとりそれを比較する実験を行った．ここで誤差として入力経路と出力経路が囲む面積と，その二つの DTW(Dynamic Time Warping) 距離を計測した．この実験結果からダミー端点の取り方や始点の位置と効用性の相関関係について検討した．

2. 関 連 研 究

この章では経路情報の保護に関する先行研究を紹介する．前提として経路情報の公開には二つの手法があるとする．一つは複数の経路をまとめて公開するものであり，それぞれの経路を一つのレコードとして考える．またもう一つの手法ではある一つの経路を点の系列とみて公開するものであり，この見方をする手法はまだ少ない．本論文では経路情報を後者のように見ることとする．

2.1 k-匿 名 化

k-匿名化 [6] とは近年最もよく用いられているプライバシ保護技術の一つで，同じ属性を持つデータが k 件以上存在する（k-匿名性を満たす）ようにデータを変換することで、個人が特定される確率を k 分の 1 以下に低減させるというものである．

複数の経路に対しこの手法を経路情報の保護に応用した手法で用いられる概念が，(k, δ)-匿名化というものである [7]．ここで δ は保護を行いたい範囲を表すパラメータである．この手法では，x 軸と y 軸が平面上の点の位置を表し，z 軸が時間を表す 3 次元空間を考える．このとき経路は (x, y, z) の系列集合と見なすことができる．ここで匿名化したい経路を T とし，そのそれぞれの時間における位置を中心とした半径 δ の円を考える．全ての時間における点の位置がこの円の中に含まれる経路を $k-1$ 個集めたとき，T は (k, δ)-匿名化されているといえる．

また Huo らの研究 [8] では，一定時間以上滞在した場所を “stay point” と定義し，その点を通る経路の集合を用いて k-匿名化を行う．

このように k-匿名化を利用した経路情報保護の手法は多くあるが，同時に欠点もある．データ所有者を特定しようとする “攻撃者” が与えられた経路情報以外の情報を持っていた場合プライバシ情報が漏洩してしまう可能性がありプライバシリスクが高まってしまう点 [9] や，ここで紹介した手法がすべてそうであったように，k-匿名化を行うには他のユーザの経路が複数必要であるという点がそれにあたる．次章で紹介する差分プライバシの概念を用いた手法ではこのような欠点がない．

3. 差分プライバシに基づく経路情報の保護

3.1 差分プライバシの定義

差分プライバシ [10] とは，データベース中の個人データの含まれるレコードの内容を攻撃者から保護しつつ，データベース全体に対する統計的解析を可能とする仕組みである [11]．以下にその数学的定義を述べる．

数学的定義を行うための準備としてデータベースモデルについて述べる．データベース D の要素となるレコードは属性の集合とし，属性の集合の取り得る値のインデクスを i とする．ただし，$i = 1, \ldots, m$ である．以下では D を x_i の出現回数で表現し，これをヒストグラムと呼ぶ．これで表現された二つのデータベース D，D' の距離を定義するため，$l1$ ノルムを定義する．

［定義 1］（$l1$ ノルム） m 次元ベクトル $\boldsymbol{x}$ の $l1$ ノルム $\|\boldsymbol{x}\|_1$ は次式で定義される．

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{m} |x_i|$$

これを用いると，ヒストグラムで表現された D と D' の距離は $\|D - D'\|_1$ となる．ここから差分プライバシの数理モデルを定義する．

［定義 2］（差分プライベート） $D \in \mathbb{N}^m$，$D' \in \mathbb{N}^m$ とする．メカニズム M が以下の条件を満たすとき，M は (ϵ, δ)-差分プライベートであるという．

$\forall S \subseteqq Range(M)$ 及び $\forall D, D' \in \mathbb{N}^m$ に対して，$\|D - D'\|_1 = 1$ であるとき

$$Pr[M(D) \in S] \leqq \exp(\epsilon) Pr[M(D') \in S] + \delta$$

なお，$\delta = 0$ の場合，M は ϵ-差分プライベートという．

ここで $\forall S \subseteqq Range(M)$ は，M が生成する可能性のあるデータベースの部分集合の全てを意味する．

3.2 差分プライバシを用いた経路情報の保護手法

差分プライバシに基づき経路情報を曖昧化する手法の例として，Jiang らの研究がある [12]．この研究では経路を点の系列と見なし，入力された経路の始点と終点はそのまま出力するとし，その間の点の集合を，差分プライバシを満足し，かつ元の経路上の点から自然な距離かつ角度にある点の集合に変換するという手法を取っている．また Dong らの研究手法 [13] では，移動経路推薦システムにおいて，その分析元となる道路上のユーザそれぞれの移動経路の始点と終点を差分プライバシの概念を用いて特定できないようになっている．このように差分プライバシを用いた経路情報の保護手法の研究はいくつかあるがまだ数が少なく，次に述べるケースのように，現在存在する手法の応用ができない場合もある．本論文で提案する手法の応用例として，1. で述べた “リアル行動ターゲティング” を考えて

いる．このとき扱うのは一人のデータ所有者の複数の経路情報であるが，このような場合に経路情報を曖昧化する手法は知られていない．また一般に，プライバシを保護するためには元の経路に雑音を加えることになるが，保護の度合いを大きくしようとして大きい雑音を加えると，出力経路の効用性が低下する．プライバシ保護と効用性の向上のバランスが難しい問題である．

先述したように，本研究ではある一人のデータ所有者が，自宅と外出先の間の経路情報を雑音を加えて複数公開する場合を主な応用例として考え，差分プライバシの概念を用いて経路端点を曖昧化する手法を提案する．このとき，元の経路は最短経路であると仮定し，雑音を加えて得られる経路は，端点を曖昧化した上でしかも最短経路になるようにする．

4. Geo 識別不能性

この章では，Geo 識別不能性 [14] の概念とそれを満足する手法について紹介する．

4.1 概　　要

まず始めに用語説明をする．$\mathcal{X}$ をユーザの存在しうる点の集合とし，ユーザの存在しうる点をある曖昧化機構によって曖昧化した結果得られる点の集合を $\mathcal{Z}$ とする．また，計算機構 K は，$\mathcal{X}$ 内のある点 x に対して，$\mathcal{Z}$ 上の確率密度分布を与えるとする．

[定義 3]（Geo 識別不能性） 任意の二つの点 x, x' において以下が成り立てば，機構 K は ϵ-geo 識別不能性 (ϵ-geo-indistinguishability) を満たすという [14].

$$d_{\mathcal{P}}(K(x), K(x')) \leqq \epsilon d(x, x')$$

但し，$d(\cdot,\cdot)$ はユークリッド距離である．ここで機構 K によって計算された確率分布 $K(x)$ と $K(x')$ の距離 $d_{\mathcal{P}}(K(x), K(x'))$ について説明する．攻撃者の事前知識を，$\mathcal{X}$ 上の事前分布 π で表す．ただし，$\pi(x)$ は，場所 x に割り当てられた確率とする．π と Bayes の法則から，K によって計算された $Z(\in \mathcal{Z})$ を観測することにより，事後確率分布 $\sigma = Bayes(\pi, K, Z)$ が計算できる．ここで $\sigma(x) = \frac{K(x)(Z)\pi(x)}{\sum_{x'} K(x')(Z)\pi(x')}$ となる．これを用いて得られるのがある集合 S 上の二つの分布 σ_1 と σ_2 の距離であり，$d_{\mathcal{P}}(\sigma_1, \sigma_2) = \sup_{S \in S} |\ln \frac{\sigma_1(S)}{\sigma_2(S)}|$ と定義される．また ϵ は正の実数であり，距離単位のプライバシ保護レベルと考える．

ϵ-geo 識別不能性を満たす一つの機構が二次元ラプラス機構である．ある点 x と，それを中心とする半径 A の円に含まれるある一点 x' を考える．それぞれの点を入力した時，二次元ラプラス関数に従う雑音を加えることで出力されうる点の確率密度分布を計算することができる．このとき，2 点それぞれを中心とする二次元ラプラス確率密度分布の差が小さければ Geo 識別不能性を満たすという．ここで確率密度分布の差は $d_{\mathcal{P}}(K(x), K(x'))$ である．

5. 提 案 手 法

この章では，経路の端点が Geo 識別不能性を満足するように曖昧化された経路を出力する手法について述べる．入力する経路は最短経路であるとし，出力する経路も同様に最短経路であり，かつ元の経路に近いものとなるようにする．

5.1 手法の概要

前提として，地図上の道路ネットワークは，枝に距離を表す重みが付いた無向グラフ $G(V, E, w)$ で表現されるとする．ここで，V は節点の集合，E は枝の集合，$w : E \to \mathbb{R}^+$ は，各枝に重みを与える関数である．入力経路はこのグラフ内のノードを通る．

経路は $\boldsymbol{x} = [x_1, ..., x_n]$ のように表す．また，任意の二点 $x, y \in V$ の間の最短経路を $< x, y >$ で表す．

データ所有者は保護したい範囲を，保護したい端点 x_n を中心とした半径 r の円として指定することができる．この円を $C_r(x_n)$ のように表すものとする．またそのプライバシ保護レベル ϵ と，出力となる曖昧化された経路の端点であるダミー端点の個数 m を指定することができる．最終的に出力される経路の端点はこの複数のダミー端点からランダムに選ばれるとする．

また曖昧化された経路を出力する際の準備として，次の範囲を定義する．

[定義 4]（PTA(Possible Terminal Area)） 距離付き無向グラフ G が与えられたときに，ある点 o から別の点 q までの最短経路中に点 p が現れるとき，q を（G における）点対 (o, p) の *possible terminal* と呼ぶ．G における点対 (o, p) のすべての *possible terminal* の集合を（G における）(o, p) の *PTA (Possible Terminal Area)* と呼ぶ．特別な場合として，点対 (o, o) の PTA は G のすべての点となる．

5.1.1 手法の流れ

手法の大まかな流れは以下の通りである．

- STEP1(ダミー端点の選択)

経路 $\boldsymbol{x}$，保護範囲となる円の半径 r，保護レベル ϵ，ダミー端点の個数 m を入力．これらをもとに，$C_r(x_n)$ の中で ϵ-Geo 識別不能性を満たす点を m 個出力し，それらを $g_1, g_2, \ldots, g_m$ とする (図 2).

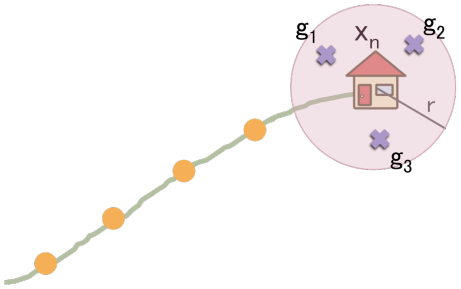

図 2　STEP1

- STEP2(x_k の選択)

経路 $\boldsymbol{x}$ 中の点のうち以下の条件を満足する点 x_k を選ぶ．

$C_r(x_n)$ 内のすべての点は点対 (x_1, x_k) の PTA に含まれる．

しかし，$C_r(x_n)$ 内の点のうち点対 (x_1, x_{k+1}) の PTA には含まれないものがある．

このような点は一意に求まることに注意されたい．概要を図 3 に示す．

- STEP3(経路の変換)

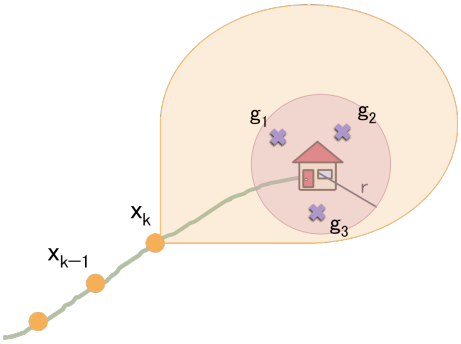

図 3　STEP2

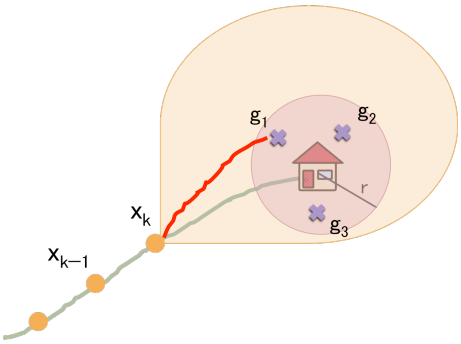

図 4　STEP3

STEP1 で求めた m 個のダミー端点の中から一つをランダムに選択し，それを g_l とする．STEP2 で求めた点 x_k から g_l に至る最短経路 $< x_k, g_l >$ を求める．元の経路のうち始点から点 x_k までの経路と $< x_k, g_l >$ を繋げた経路 $\boldsymbol{x'}$ を，曖昧化された経路として出力する (図 4)．

5.2　具体的な手法

経路 $\boldsymbol{x}$ 上の連続する 2 点 x_{k-1}，x_k をそれぞれ始点，中継点としたときに最短経路の終点となりうる点の集合 $V_{[x_{k-1},x_k]}$ を 5.2.1 に述べる Goal-select で求め，$V_{[x_{k-1},x_k]}$ が端点 x_n を中心とした半径 r の円を被覆するような点 x_k を発見する．また，5.2.2 に示す Geo-select により，プライバシ保護を行いたい範囲の中で Geo 識別不能性を満足する点 g_l を求める．元の経路の始点から x_k までの経路と，x_k から g_l までの最短経路 $< x_k, g_l >$ を繋げたものを曖昧化された経路として出力する．

- 入力データ
  - グラフ $G(V, E, w)$

地図上の道路ネットワークを表す．ノードとエッジの繋がり，ノード間のコスト，ノードの地図上における緯度経度を情報として持つ．

  - 経路 $\boldsymbol{x} = [x_1, ..., x_n]$

$\boldsymbol{x}[i]$ は $\boldsymbol{x}$ の i 番目の点，即ち x_i であるとする．

  - 半径 r

x_n が曖昧化された点 x'_l は，x_n を中心とした半径 r の円の中にあるとする．

  - 保護の度合い ϵ

x_n を中心とした半径 r の円の中におけるプライバシ保護レベル．これが小さくなればなる程プライバシ保護の度合いは小さくなる．

- 出力データ

以下に示す処理を行い曖昧化された経路 $\boldsymbol{x'} = [x_1, ...x_k, x'_{k+1}, ..., x'_l]$ を出力データとする．

- 処理

入力データを出力データに変換する際の処理の流れは以下の通りである．

  - STEP1

k を $n-1$ とする．

5.2.1 に示す Goal-select の入力において，始点 x_s を $\boldsymbol{x}[k-1]$，中継点 x_m を $\boldsymbol{x}[k]$ とし，処理を行う．このとき出力される集合 $V_{[x_{k-1},x_k]}$ が，x_n を中心とする半径 r の円に含まれるノードの集合を被覆するかどうかを調べる．被覆する場合は STEP2 へ，しない場合は k の値を 1 減らし，STEP1 の処理を再度行う．

  - STEP2

5.2.2 に示す Geo-select を呼び出し，x_n を中心とした半径 r の円の中で ϵ-Geo 識別不能性を満たすある m 個の点を計算し，その中から一つをランダムに選択し，ダミー端点 x'_m とする．$\boldsymbol{x}[k]$ から x_n' に至る最短経路 $< x_k, x'_l >$ を $\boldsymbol{y} = [x_k, x'_{k+1}, ...x'_l]$ とする．

  - STEP3

$\boldsymbol{x}$ の k 番目から n 番目までの要素 $[x_k, ..., x_n]$ を $\boldsymbol{y}$ に置き換えた経路 $\boldsymbol{x'} = [x_1, ...x_k, x'_{k+1}, ..., x'_l]$ を作成し，これを曖昧化された経路として出力する．

5.2.1　Goal-select

始点と中継点を入力した際に，その最短経路の終点となりうる点の集合を求める範囲を求める．

- 入力データ
  - グラフ $G(V, E, w)$

始点 x_s，中継点 x_m ともに，このグラフ上のノードであるとする．

  - 始点 x_s
  - 中継点 x_m

- 出力データ

最短経路の終点となりうる点の集合 $V_{[x_s,x_m]}$ を出力とする．

- 処理
  - STEP1

Dijkstra 法を用いて，x_s を始点とした時にグラフ上のそれぞれのノードに至る最短経路を求める．

  - STEP2

STEP1 で求めた最短経路の集合の中から x_m を通るものを探索し，見つかった場合にはその経路の x_m 以降のノードを $V_{[x_s,x_m]}$ に追加する．

  - STEP3

すべての最短経路に対して STEP2 の処理が終われば，$V_{[x_s,x_m]}$ を出力する．

5.2.2　Geo-select

4. において定義した Geo 識別不能性を満足する点を選択する．

- 入力データ
  - グラフ $G(V, E, w)$

それぞれのノードは地図上の緯度・経度の情報を持つ．

– 点 x

グラフ上の保護したいある一点.

– 半径 r

– 保護の度合い ϵ

– 出力する点の個数 m

– パラメータ u, δ_θ

Geo 識別不能性を満たす点を出力するメカニズムの精度を調整するパラメータ.

- 出力データ

点 x を中心とした半径 r の円の中で ϵ-Geo 識別不能性を満たすある m 個の点 $g_1, g_2, \ldots, g_m$ を出力する.

- 処理

– STEP1

点 x とグラフ上のそれぞれのノードのユークリッド距離を計算し，x を中心とした半径 r の円の中にあるグラフ上の点の集合 R を求める.

– STEP2

ある範囲 r_{max} の中で *epsilon*-geo 識別不能性を満足するためには，以下の定理を満たす必要がある.

［定理 1］ $r_{max} < {}^{u}/_{\delta_\theta}$，$q = {}^{u}/_{r_{max}\delta_\theta}$ であるとする．この時 ϵ, ϵ' は以下の式を満足する.

$$\epsilon' + \frac{1}{u} \ln \frac{q + 2e^{\epsilon' u}}{q - 2e^{\epsilon' u}} \leqq \epsilon$$

この時，メカニズム $K_{\epsilon'}$ によって出力される点は r_{max} の範囲の中で ϵ'-Geo 識別不能性を満たす．つまり，$d(x_0, x), d(x'_0, x) \leqq r_{max}$ であれば次式が成り立つ.

$$K_{\epsilon'}(x_0)(x) \leqq e^{\epsilon d(x_0, x'_o)} K_{\epsilon'}(x'_0)(x)$$

ϵ を $r_{max} = diam(r)$ とした時に定理 5.2.2 を満たす最大の値 ϵ' に変換する．ここで $diam(r)$ は x を中心とした半径 r の円の中のそれぞれのノードの最大距離である.

– STEP3

x を極座標 (r, θ) に変換する．ここで出力されうる点を x_0 とすると，r は x と x_0 の距離，θ は直線 xx_0 と x 軸のなす角である.

– STEP4

x を中心とする半径 r の円内の任意の点が出力される確率を p とすると，雑音 r' は

$$r' = C_{\epsilon'}^{-1}(p) = -\frac{1}{\epsilon'}(W_{-1}(\frac{p-1}{\epsilon'}) + 1)$$

と計算できる．ここで，W_{-1} は Lambert 関数である.

また，雑音 θ' は乱数となる.

ここで計算した (r', θ') を (r, θ) に加え，緯度・経度の形に変換した点 x' を求める.

– STEP5

STEP1 で求めた集合 R に含まれる点の中で x' に最も近いものを g_1 とする.

– STEP6

STEP1〜STEP5 の処理を m 回行い $g_1, g_2, \ldots, g_m$ を求め，これを出力とする.

6. 実 験 方 法

6.1 実験の概要

経路端点を Geo 識別不能性を用いて曖昧化することにより，プライバシ保護は既に保障されている．したがって，端点を曖昧化することにより出力された経路の効用性を本実験で確かめたい．効用性を測るにあたっての指標を入力経路と出力経路の誤差とし，2 種類の値を用いた．一つは入力経路と出力経路が囲む範囲の面積 $A(m^2)$ としてそれを計測した．ここで，面積の計算には地図蔵 (注1) を利用している．またもう一つの効用性の評価指標として，次の値を定義する.

［定義 5］（RPD (Relative Path Distance)） 始点が同じある二つの経路 $\boldsymbol{x} = [x_1, x_2, \ldots, x_l], \boldsymbol{y} = [y_1, y_2, \ldots, y_m]$ $(x_1 = y_1)$ を考える．$\boldsymbol{x}$ 上の各点と始点の間の距離の，$\boldsymbol{x}$ 全体の長さに対する割合 $\boldsymbol{r_x} = [r_1, r_2, \ldots, r_l]$ を求める．$\boldsymbol{y}$ 上で $\boldsymbol{y}$ 全体の長さに対して $\boldsymbol{r}$ の割合をとる点の系列 $\boldsymbol{y_x} = [y_{x1}, \ldots, y_{xl}]$ を求め，これを $\boldsymbol{x}$ に対する $\boldsymbol{y}$ の *relative path* と呼ぶ．$\boldsymbol{x}$ と $\boldsymbol{y_x}$ の対応する点の距離の総和を $\boldsymbol{x}$ に対する $\boldsymbol{y}$ の *RPD (Relative Path Distance)* と呼ぶ.

この値において入力経路を $\boldsymbol{x}$，出力経路を $\boldsymbol{y}$ としたときの RPD を $D(km)$ とし，これを計測した．ここで，入力経路の始点を s，ダミー端点を d としたときの誤差をそれぞれ $A_{s,d}$，$D_{s,d}$ とする．誤差が小さくなればなるほど，効用性は大きくなると考える.

本実験では経路端点 o を図 5 の中心にあたる四条烏丸に固定し，出力されるダミー端点を 3 個として 2 種類の実験を行った．一つは端点からの方角が異なる 4 点を始点とする四つの経路を入力として，異なるダミー端点を持つ経路をそれぞれ三つずつ出力し，合計 12 個の曖昧化した経路とそれぞれの元の経路との誤差を比較した．これにより入力経路の始点とダミー端点の位置関係が誤差にどのような影響を与えるのかを検証した．またもう一つの実験では，端点と始点の距離，$C_r(o)$ の半径 r を変化させた時の誤差を比較し，この二つのパラメータの相関関係を調べた.

ここで，入力経路の始点と r を固定した時に最大・最小の A または D を生成するダミー端点をそれぞれ d_{max}，d_{min} とする.

6.2 データセット

本論文は道路ネットワーク上の経路曖昧化手法の提案を行っている．したがって公開されている道路ネットワークとして，Open Street Map Japan(注1) より，図 5 に示す範囲の京都市の道路データを取得した．またこのデータに対し，非連結な頂点を省く等の前処理を行っている．この処理を行った結果，頂点数は 3742 個となった.

6.3 経路曖昧化の一例

実際の経路に本論文の提案手法の処理を行った結果出力された経路の例を図 6 に示す.

(注1)：http://japonyol.net/editor/calculate.html

(注1)：https://openstreetmap.jp/

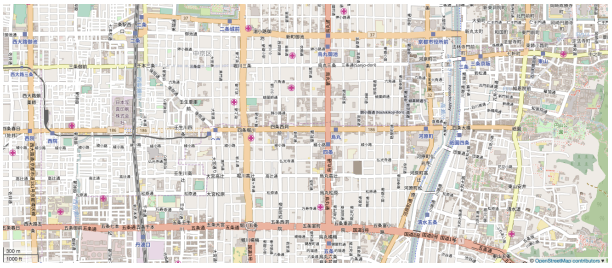

図 5　Open Street Map Japan から取得した範囲

図 6　元の経路 (緑) と曖昧化を行った経路 (オレンジ) ($r = 1.0(km)$, $\epsilon = 0.01$)

6.4　経路始点とダミー端点の位置関係の差異による比較

6.4.1　実験の概要

本論文で提案する手法ではダミー端点の候補となる点は複数個出力される．したがってどの点をダミー端点とするかにより誤差は異なってくる．本実験では，端点から見て北東，北西，南西，南東に位置する 4 点 s_1, s_2, s_3, s_4 を入力経路の始点とし，ダミー端点を $r = 0.3km$ とした時に $C_r(o)$ に含まれる 3 点 d_1, d_2, d_3 とし，それぞれの組み合わせに対して出力される A_{s_i,d_j}, D_{s_i,d_j} $(i = 1, 2, 3, 4, j = 1, 2, 3)$ の値を比較した．始点，端点，ダミー端点の地図上の位置を図 7 に示す．

図 7　実験に用いた経路群の始点，端点，ダミー端点

6.4.2　結　　果

誤差の計測結果を表 8，9 に示す．図中の黄色い線で囲まれているものが最小の誤差をとる全体の結果として，A，D ともに，始点からの距離が近いダミー端点の方が誤差が小さくなる傾向にあることが確認された．しかし始点 s_3 の結果のように，出力される経路によってはそれが成り立たない例もあるため，距離以外の影響を与える因子について検討することを今後の課題としたい．

6.5　半径 r と経路端点から始点までの距離の差異による比較

6.5.1　実験の概要

本論文で提案する手法では，$C_r(o)$ の半径 r を指定することができる．これが大きくなればなるほど，より端点 o から遠い位置にある点がダミー端点として出力されるようになる．実際の点から遠い位置にある点を端点とすることにより，誤差が大きくなり出力経路の効用性が小さくなる怖れがある．

また一方で，誤差が効用性に与える影響は入力経路の長さによって異なる．たとえばある経路 $\boldsymbol{x} = [x_1, \ldots, x_n]$ とその部分経路 $\boldsymbol{x_{i,n}} = [x_i, \ldots, x_n]$ を考えたとき，$\boldsymbol{x}$ に対する出力経路は $\boldsymbol{x'} = [x_1, ...x_k, x'_{k+1}, ..., x'_l]$ となり，$\boldsymbol{x_{i,n}}$ に対する出力経路は $i \leqq k$ のときを考えると，$\boldsymbol{x'_{i,n}} = [x_i, ...x_k, x'_{k+1}, ..., x'_l]$ となる．このとき誤差は $\boldsymbol{x}$，$\boldsymbol{x_{i,n}}$ に対して同じ値をとるが，$\boldsymbol{x}$ と $\boldsymbol{x'}$ の異なる点の占める割合は $\boldsymbol{x}$ と $\boldsymbol{x'_{i,n}}$ のそれよりも大きくなることは明らかである．したがって誤差について入力経路の長さによる相対評価を行う必要があり，その値を $A_c = \frac{A}{|\boldsymbol{x}|}$，$D_c = \frac{D}{|\boldsymbol{x}|}$ とする．ここで，$|\boldsymbol{x}|$ は経路 $\boldsymbol{x}$ の長さとする．

以上より本実験では，$C_r(o)$ の半径 r と経路端点から始点までの距離という 2 種類のパラメータを変化させた場合の A と A_c，D と D_c を比較する．半径 r を 0.3km，1.0km，2.0km と変化させた時，それぞれダミー端点 $d_{11}, \ldots, d_{13}$，$d_{21}, \ldots, d_{23}$，$d_{31}, \ldots, d_{33}$ を出力した．また経路始点は端点からの距離の異なる 3 点 s_1, s_5, s_6 を用いた．それぞれの点の地図上の位置を図 10 に示す．

6.5.2　結　　果

- 絶対評価

得られた実験結果を 3 種類に分けて示す．d_{min}，d_{max} に対する A の値と入力経路の始点と r を固定した時の三つのダミー端点を用いた出力経路の A の平均値をそれぞれ図 11，12，13 に示す．また D に関するそれぞれの値を図 14，15，16 に示す．

この結果から，図 11，14 のように誤差が小さくなるダミー端点を選択できた場合は，r が大きくなればなるほど誤差も大きくなることが確認された．また，図 12 の点 s_5 の A の値から，ダミー端点の選択の仕方によっては誤差が極端に大きくなってしまう場合もある．

またこの結果からは入力経路の始点と端点の距離と誤差の相関性は確認できなかったが，6.5.1 で述べたようにこの距離が効用性に与える影響は大きいと考えられる．次でその検証を行う．

- 相対評価

得られた実験結果を 3 種類に分けて示す．d_{min}，d_{max} に対する A の値と入力経路の始点と r を固定した時の三つのダミー端点を用いた出力経路の A の平均値をそれぞれ図 17，18，19 に示す．また D に関するそれぞれの値を図 20，21，22 に示す．

この結果から，入力経路が長い場合は誤差が効用性に与える影響が小さいことがわかる．したがってそのような場合には r を大きくしてダミー端点を経路端点から離れた場所にとってプライバシ保護の度合いを大きくしても，効用性の損失が少なくて済むという可能性が示された．

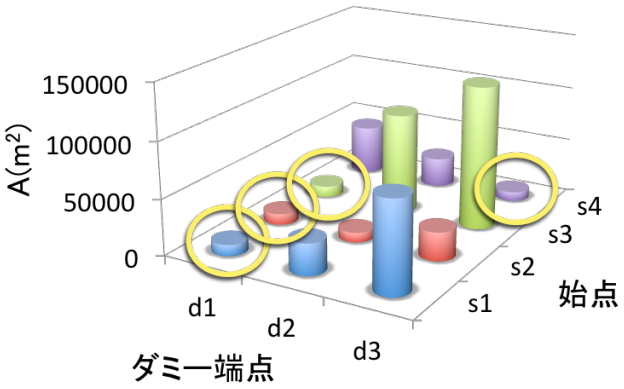

図 8　経路始点とダミー端点の位置関係が異なる場合の A の値

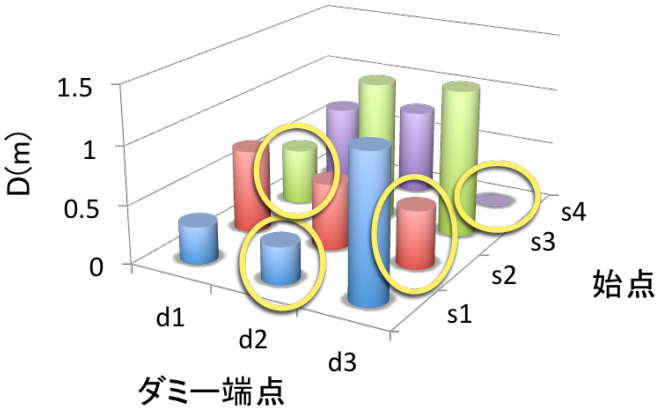

図 9　経路始点とダミー端点の位置関係が異なる場合の D の値

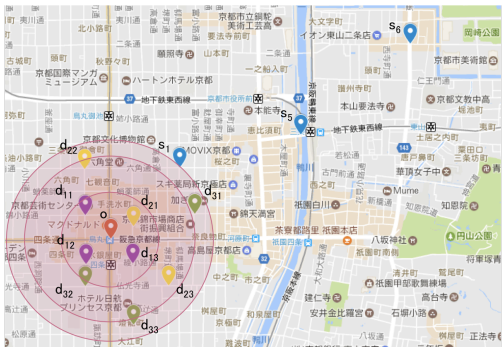

図 10　実験に用いた経路群の始点，端点，ダミー端点

7. おわりに

本研究では，プライバシ保護と経路情報としての効用性保持を両立する経路情報の曖昧化を目的として，Geo 識別不能性を用いた経路端点の曖昧化手法を提案している．提案手法では入力経路の端点 o を中心とした半径 r の円 $C_r(o)$ の中で Geo 識別不能性を満たす点をダミー端点として複数求め，その中から一点を選択してそれを端点とする経路を出力する．この手法によって求められた経路は，その端点はプライバシ保護が成り立っているだけでなく，途中までは入力経路と同じ道を辿ることから効用性の損失も少ない．

評価実験では効用性を入力経路と出力経路の誤差とみて，京都市の実際の道路ネットワークを用いた様々な経路データを使って比較を行った．変化させたパラメータの異なる 2 種類の実験を行っており，一つの実験では経路始点とダミー端点の位置関係の差異による効用性の比較を行い，ある始点に対しどのような位置にあるダミー端点を選択すればよいかを検証した．実験結果から，ダミー端点を始点から近い距離にあるものを選択した場合の方が効用性が大きくなる傾向にあることが確認された．またもう一方の実験では $C_r(o)$ の半径 r と入力経路の長さを変化させ，それぞれが効用性に与える影響を調べた．この実験結果からは，r が大きくなればなるほど絶対的な効用性は小さくなるが，経路が長い場合は誤差が効用性に与える影響は相対的に小さいことが確認された．したがって，入力経路の長さによっては r を大きくしてプライバシ保護の度合いを大きくしても効用性の損失が少なくて済む可能性が示された．

以上のように，本研究では，プライバシ保護と効用性の保持を両立する経路を得る手法を提案した．

今後，異なるデータセットを用いた実験の実施や，より実用性を考えて経路端点だけでなく中間点の曖昧化への応用の研究を行う予定である．

文　献

[1] Kanza, Yaron and Samet, Hanan, "An online marketplace for geosocial data," Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, pp. 10:1-10:4 , 2015.

[2] 横山 隆治, 楳田 良輝, " リアル行動ターゲティング," 日経 BP 社, 2015.

[3] 情報銀行, http://www.information-bank.net/index.html

[4] NTT データ研究所 : "パーソナルデータに関する一般消費者の意識調査 (2016)," http://www.keieiken.co.jp/aboutus/newsrelease/161122/supplementing01.html

[5] De Montjoye, Yves-Alexandre and Hidalgo, César A and Verleysen, Michel and Blondel, Vincent D, "Unique in the crowd: The privacy bounds of human mobility," Scientific reports, Vol.3, 2013.

[6] Sweeney, Latanya, "k-anonymity: A model for protecting privacy," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 10, No. 05, pp. 557-570, 2002.

[7] Abul, Osman and Bonchi, Francesco and Nanni, Mirco, "Never walk alone: Uncertainty for anonymity in moving objects databases," Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on, IEEE, pp. 376–385, 2008.

[8] Huo, Zheng and Meng, Xiaofeng and Hu, Haibo and Huang, Yi, "You can walk alone: trajectory privacy-preserving through significant stays protection," International Conference on Database Systems for Advanced Applications, Springer, pp. 351–366, 2012.

[9] Narayanan, Arvind and Shmatikov, Vitaly, "Robust de-anonymization of large sparse datasets," Security and Privacy, 2008. SP 2008. IEEE Symposium on, IEEE, pp. 111–125, 2008.

[10] Dwork, Cynthia, "Differential privacy," Encyclopedia of Cryptography and Security, Springer, pp. 338–340, 2011.

[11] 中川裕志, " プライバシー保護入門 : 法制度と数理的基礎," 勁草書房, 2016.

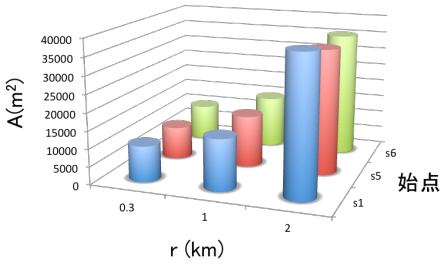

図 11　d_{min} に対する A の値

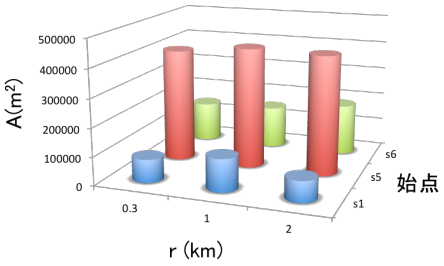

図 12　d_{max} に対する A の値

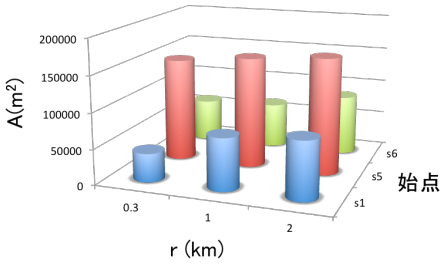

図 13　始点と r を固定した場合の A の平均値

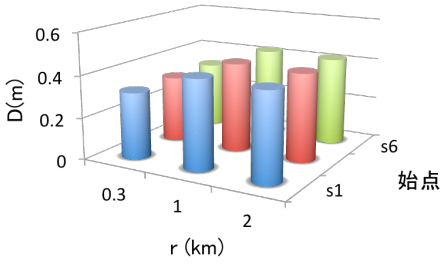

図 14　d_{min} に対する D の値

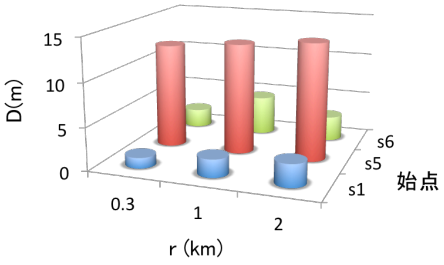

図 15　d_{max} に対する D の値

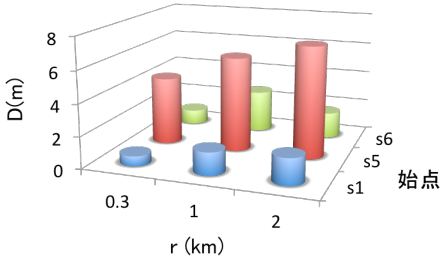

図 16　始点と r を固定した場合の D の平均値

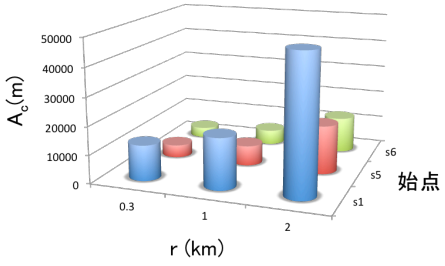

図 17　d_{min} に対する A_c の値

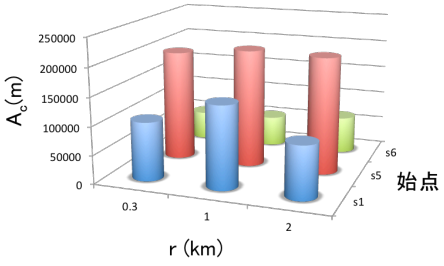

図 18　d_{max} に対する A_c の値

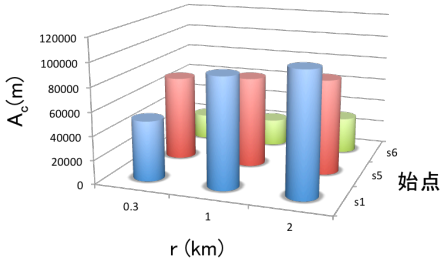

図 19　始点と r を固定した場合の A_c の平均値

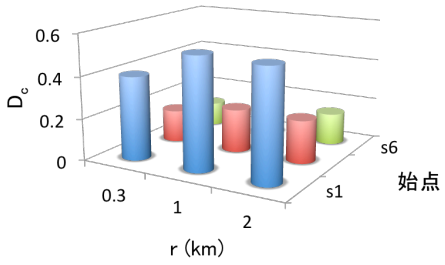

図 20　d_{min} に対する D_c の値

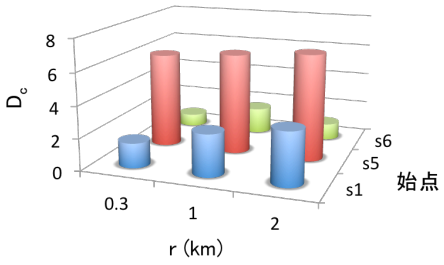

図 21　d_{max} に対する D_c の値

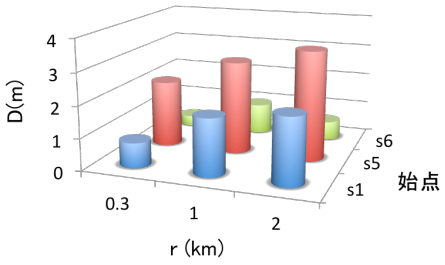

図 22　始点と r を固定した場合の D_c の平均値

[12] Jiang, Kaifeng and Shao, Dongxu and Bressan, Stéphane and Kister, Thomas and Tan, Kian-Lee, "Publishing trajectories with differential privacy guarantees," Proceedings of the 25th International Conference on Scientific and Statistical Database Management, ACM, p. 12, 2013.

[13] Dong, Roy and Krichene, Walid and Bayen, Alexandre M and Sastry, S Shankar, "Differential privacy of populations in routing games," 2015 54th IEEE Conference on Decision and Control (CDC), IEEE, p. 2798–2803, 2015.

[14] Andrés, Miguel E and Bordenabe, Nicolás E and Chatzikokolakis, Konstantinos and Palamidessi, Catuscia, "Geo-indistinguishability: Differential privacy for location-based systems," Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security, ACM, p. 901–914, 2013.

[15] Berndt, Donald J and Clifford, James, "Using dynamic time warping to find patterns in time series," KDD workshop, Vol. 10, No. 16, pp. 359–370, 1994.