Vol.2022-ICS-205 No.1
2022/2/21

人の存在確率を考慮した位置情報プライバシ保護手法の提案

石禾 里帆[1,a)]　清 雄一[1,b)]　田原 康之[1,c)]　大須賀 昭彦[1,d)]

概要：近年，GPS 搭載のスマートフォンの普及により混雑情報を公開するシステムなどの，人々の位置情報を収集し統計を公開するシステムが広く使用されている．それらのシステムは便利である一方，位置情報という個人情報の漏えいリスクを抱えている．個人の情報を直接的に公開していなくても，統計データから推測することが可能な場合がある．個人の位置情報からは，自宅や職場や，病院へのアクセスなどから健康情報までも推測できてしまう．よって，各々の位置情報データをそのまま扱うのではなく，保護を施したうえでのデータの利用技術が必要である．必要以上のプライバシ保護を施すことは，ユーザーの利便性を下げることに繋がるため回避したい．そこで今回は差分プライバシを応用して，ある時点での人の存在確率から算出されるプライバシ保護の度合いを基準として，それぞれの位置について提案手法から適切なプライバシパラメータを導き，位置情報にノイズを加えプライバシを保護する手法を提案する．その結果客観的指標であるユーザーの利便性を測る指標 SQL(Service Quality Loss) を抑えることができた．

1. はじめに

1.1 背景

近年，GPS 搭載のスマートフォンの普及により，混雑情報を公開するシステムなどの，人々の位置情報を収集し統計を公開するシステムが広く使用されている．それらのシステムは便利である一方，個人の特定リスク，個人情報の漏えいリスクを抱えている．個人の情報を直接的に公開していなくても，統計データから推測することが可能な場合がある．個人の位置情報からは，自宅や職場や，病院へのアクセスなどから健康情報までもを推測できてしまう．よって，各々の位置情報データをそのまま扱うのではなく，保護を施したうえでのデータの利用技術の活用が必要不可欠である．

こうした位置情報データを守るプライバシ保護の保証基準として，差分プライバシを応用した Geo-I(Geo-Indistinguishability)(節 2.3) が注目を集めている．これは，位置情報について摂動法によるノイズ付加をユークリッド平面上で施す際の保証基準を示す．

差分プライバシとは，攻撃者の背景知識，攻撃手法アルゴリズムに依らず，データセットのプライバシの開示を抑えることのできる安全的指標として広く認められている．差分プライバシは基本データセットに対しての保護である．個別のデータをサーバに送るときなどに，送るデータ自体に保護を施す手法としては，差分プライバシの概念を拡張した ϵd_X-privacy という概念が知られている．Geo-I はこの ϵd_X-privacy を位置情報に応用している．

Geo-I の基準を満たし，自分の位置情報にランダムなノイズを加えて真の値を保護する摂動法の一つに PL という手法が知られている．これはラプラス分布を用いてノイズを加えている．

一般的に保護の度合いが強い (付加するノイズが大きい) とユーザーの有用性が下がる．ユーザーの有用性とプライバシ保護の度合いはトレードオフである．よって，適切なプライバシパラメータを定めることは，データの保護と有用性のどちらも担保するうえで重要である．

1.2 目的

一般的に差分プライバシ (節 2.1) を用いてプライバシ保護を行う場合，プライバシパラメータ ϵ の値は一定に定められ適用される．しかし，プライバシパラメータ ϵ の値が一定だったとしても，地図内のポイントの配置や人の存在確率などの影響によりプライバシ保護の度合いの評価である LPr(節 2.7) は等しくならない．それぞれの位置の特性が反映された結果となる．

例えば，現実の地図の中で，人がいる場所，いない場所というのは地図の中である程度定まっている．例えば橋のない川の上に人はいない．そういった地図の特性を反映すると攻撃者が真の位置を見つけられる可能性が高くなる．

[1] 電気通信大学
The University of Electro-Communications
〒 182-8585 東京都調布市調布ヶ丘 1-5-1
[a)] isawa.riho@ohsuga.lab.uec.ac.jp
[b)] seiuny@uec.ac.jp
[c)] tahara@uec.ac.jp
[d)] ohsuga@uec.ac.jp

今回は，人の存在確率を考慮したそれぞれの位置の LPr が一定以上の値になるよう，プライバシパラメータ ϵ をそれぞれの位置ごとに定め，保護する手法を提案する．つまり，それぞれの位置に適切なプライバシパラメータ ϵ を求める手法を提案する．例えば，攻撃者が真の位置を見つけやすい位置においてはプライバシパラメータ ϵ の値を小さくすることにより一定以上の保護を保証できるようにする．

一律にプライバシパラメータ ϵ の値を下げて適用することはユーザーの利便性を著しく欠く恐れがある．したがって，それぞれの位置に適切なプライバシパラメータを定める手法を提案する．

2. 関連研究

2.1 ϵ-差分プライバシ

差分プライバシは攻撃者の事前情報によらずデータセットを守ることのできる数学的に厳格でかつ強力な基準として広く用いられている [1]．差分プライバシによる保護は，暗号化による保護ではなく，データやデータから算出された結果にノイズを加えることによる保護である．暗号化による保護でないため計算負荷が低く，導入が容易な傾向にある．$S \subseteq Range(K)$ であり，データベース D, D' が隣接しているとき，メカニズム K によって保護を行うとして，以下の式を満たしているとき ϵ-差分プライバシを満たしている．隣接とは，レコードが一か所異なることを指す．例えば，D が D' からある一つのレコードを除いたデータベースであったり，D が D' のあるレコードを何か別のレコードと入れ替えたデータベースであるとき，D と D' は隣接しているという．

$$Pr[K(D) \in S] \leq exp(\epsilon) \times Pr[K(D') \in S] \qquad (1)$$

この式は，隣接するデータベース同士による結果の見分けがつかない場合，レコードの異なる箇所を特定できないからプライバシが保護されているという意味を示している．例えば，あるレコード A 以外の情報をすべて知っている攻撃者がいたとき，データベースの結果から逆算して A についての情報を得ることが可能になりうる．しかし，隣接しているデータベース同士の見分け，つまり違いが判らない場合にはこの特定は難しくなる．

この保証に従って保護を施すことで，有用性を保ちつつプライバシを保護することができる．

2.1.1 post-processing property

post-processing property は，差分プライバシの性質の一つである．ある真の値に対して差分プライバシを適用した後，真の値を用いないで加工を行なった場合，変わらず差分プライバシが保証される [2].

2.2 単体のデータへの拡張 ϵd_X-privacy

上述の ϵ-差分プライバシの定義 (節 2.1) はデータベースに関しての保護を意味する．つまり位置情報データを格納したデータベースに対しては保証されるが，それぞれの位置情報をサーバに毎回 1 つずつ送る場合などについては保証されていない．こういった単体のデータに対しての保証について，Chatzikokolakis ら [3] はデータベース上でのみ定義されていた差分プライバシのデータへの拡張を行った．

あるドメイン Z 上の確率分布を与えるあるドメイン X 上のメカニズム $K : X \to P(Z)$ があるとする．$P(Z)$ は，Z 上の確率分布を表す．このとき，X 上の”距離”を表す dX を用いて，ある $\epsilon \in \mathbb{R}+$，メカニズム K が任意の $x, x' \in X, Z \subseteq Z$ について，以下の式を満たしているとき，ϵdX-privacy を満たしている．

$$\frac{K(x)(Z)}{K(x')(Z)} <= \epsilon dX(x, x') \qquad (2)$$

あらかじめ定められたプライバシパラメータ ϵ を用いて算出したノイズを，自分の真の位置に加えることは計算負荷の低いものである．そのため，それぞれのデバイスでも計算ができ，位置情報システムに取り入れやすいと考えられる．

2.3 Geo-I(Geo-Indistinguishability)

Geo-I(Geo-Indistinguishability) は，差分プライバシの概念を応用して位置情報データに適用させた概念である．Geo-I は摂動法の中でも特に注目されている概念である [4]. Geo-I は，ϵd_X-privacy を位置情報データに適用させている．

x と x' のユークリッド平面上での距離を $d(x, x')$ として，$\epsilon \in \mathbb{R}+$ について, メカニズム K が以下を満たすとき，パラメータ ϵ で Geo-I が保証される．

$$dP(K(x), K(x')) \leq \epsilon d(x, x') \qquad (3)$$

2.4 PL 法 (The Planar Laplace mechanism)

Geo-I を満たすデータの保護手法として PL 法が知られている [4]．ラプラス分布を用いてパラメータ ϵ からノイズを生成し，真の位置に加える手法である．

r について 4 によって算出されたノイズを代入，θ について [0,2π) で一様分布の確率からランダムに値を算出して代入し，p について [0,1) で一様分布の確率からランダムに値を算出して代入し 真の値 x に，$< rcos(\theta), rsin(\theta) >$ をノイズとして加え，そのノイズを加えた座標と，ありうる座標系の中で一番近いものを選択し，それをメカニズム適用後の値とする．関数 W はランベルトの W 関数を意味する．

$$C^{-1}(p) = -\frac{1}{\epsilon}(W_{-1}(\frac{p-1}{\epsilon}) + 1) \qquad (4)$$

真の位置 x を x' に曖昧化する確率は，x と x' のユーク

リッド平面上での距離を $d(x, x')$ として，次式 5 のように表される．この PL 法では位置についてのデータの小数点を丸めている．その丸めの影響についても下記の式は考慮されている．

$$D_\epsilon(x)(x') = (\frac{\epsilon^2}{2\pi})e^{-\epsilon d(x,x')} \tag{5}$$

2.5 攻撃者のモデル

Shokri ら [5] は，メカニズムを用いて位置情報を曖昧化した場合の，攻撃者のモデルについて提案した．攻撃者の持つユーザーの位置の事前知識を表す確率分布を $\pi_a(r)$, ユーザーが使用しているプライバシ保護メカニズムによる位置 r を r' を曖昧化する確率を $K(r)(r')$, r' の入力に対してユーザーの真の位置は $\hat{r}$ であるかどうか判別する攻撃者の予測関数を $h(r')(\hat{r})$，r と $\hat{r}$ のユークリッド平面上での距離を $d(r, \hat{r})$ として, 攻撃者の予測関数は次の線形計画問題を解くことにより求まる．

$$\begin{aligned} \text{maximize} \quad & \sum_{r,r',\hat{r}} \pi_a(r)K(r)(r')h(r')(\hat{r})d(r,\hat{r}) \\ \text{subject to} \quad & \sum_{\hat{r}} h(r')(\hat{r}) = 1, \forall r' \\ & \sum_{\hat{r}} h(r')(\hat{r}) \geqq 0, \forall r', \hat{r} \end{aligned} \tag{6}$$

2.6 有用性 SQL(Service Quality Loss)

Shokri ら [6] は，メカニズム K を用いて曖昧化を施しプライバシを保護するにあたって，有用性の評価式として SQL(Service Quality Loss) を提案した．SQL(Service Quality Loss) とは，プライバシ保護を施した際の誤差などを定量化している．ユーザーにとっては，サービスを利用するときに，真の値に近い値で利用できるとサービスの質が良いとされる．逆に，プライバシ保護の度合いが大きくても，真の値から離れた値でサービスを利用する場合サービスの質が悪いと判断される．こういったユーザーにとっての有用性を SQL で判断する．

ユーザー u が r にいる確率を $\pi_u(r)$，ユーザーが使用しているプライバシ保護メカニズムによる位置 r を r' を曖昧化する確率を $K(r)(r')$, r と r' のユークリッド平面上での距離を $d(r, r')$ として, SQL は次の式で算出される．

$$SQL(\pi_u, K, d_q) = \sum_{r,r'} \pi_u(r)K(r)(r')d(r,r') \tag{7}$$

2.7 プライバシ保護の度合い

Shokri ら [6] は，メカニズムを用いて曖昧化を施しプライバシを保護するにあたって，プライバシ保護の度合いの評価式として LP(Location Privacy) を提案した．

ユーザー u が r にいる確率を $\pi_u(r)$，ユーザーが使用しているプライバシ保護メカニズムによる位置 r を r' を曖昧化する確率を $K(r)(r')$, r' の入力に対してユーザーの真の位置は $\hat{r}$ であるかどうか判別する攻撃者の予測関数を $h(r')(\hat{r})$, r と r' のユークリッド平面上での距離を $d(r, r')$ として, LP は次の式で算出される．

$$LP(\pi_u, K, h, d_q) = \sum_{\hat{r},r',r} \pi_u(r)K(r)(r')h(r')(\hat{r})d_q(\hat{r}, r) \tag{8}$$

今回は，高木ら [7] によって提案された，各ユーザーが自分のいる位置 r で保護メカニズムを用いた場合のプライバシ保護の度合いの定式化 LPr についても用いる．つまり，LPr は各位置によってプライバシ保護の度合いが変わることを意味している．

ユーザーの位置を r, ユーザーが使用しているプライバシ保護メカニズムによる位置 r を r' を曖昧化する確率を $K(r)(r')$, r' の入力に対してユーザーの真の位置は $\hat{r}$ であるかどうか判別する攻撃者の予測関数を $h(r')(\hat{r})$，r と r' のユークリッド平面上での距離を $d(r, r')$ として, LPr は次の式で算出される．

$$LPr(r, K, h) = \sum_{r',r} K(r)(r')h(r')(\hat{r})d_s(\hat{r}, r) \tag{9}$$

3. 提案手法

提案手法の説明で用いる記号とその意味を以下表 1 に示す.

表 1 記号の意味

記号	意味
u	ユーザ
a	攻撃者
$\pi_u(r)$	ユーザ u が r にいる確率
$K(r)(r')$	メカニズム．位置 r を r' に曖昧化する確率
$d(x, x')$	x と x' のユークリッド距離
$SQL(\pi_u, K, d)$	曖昧化した位置と位置 u との距離 d の期待値
$LP(\pi_a, K, h, d)$	a が推測する位置と真の位置との距離 d の期待値
$LPr(r, K, h, d_p)$	位置 r でメカニズム K を用いた場合， ユーザ u の真の位置と予測関数 h を用いる 攻撃者の予測する位置との距離 d の期待値

3.1 アルゴリズム

提案手法では，LPr(式 9) というプライバシ保証の評価の値を求め，それを基準としてプライバシパラメータを変更する．

まず，ある一定の ϵ_{before} を定めると位置 r を r' に曖昧化する確率 $K_{before}(r)(r')$ が定まる．これと，x と x' のユークリッド距離 $d(x, x')$, ユーザー u が r にいる確率を $\pi_u(r)$ 用いると，攻撃者モデル (節 2.5) に従って 攻撃者が r' を観測したときにユーザーの位置を $\hat{r}$ と予測する確率 $h(r')(\hat{r})$ が求められる．ここでの $\pi_u(r)$ には，想定している空間の中

で人がどのように分布しているかというデータから事前情報を算出して代入する．

次に，使用する ϵ の値ががそれぞれの位置ごとに定められたメカニズムを算出する．位置 r で使用するパラメータを ϵ_r と置く．そのメカニズムによる位置 r を r' に曖昧化する確率を $K_{after}(r)(r')$ と置く．

ある与えられた位置 r に対して，求められた $h(r')(\hat{r})$ および，$K_{after}(r), d(x, x')$ を用いると，式 9 から LPr を算出できる．$K_{after}(r)$ は，算出した LPr が基準値 *base_value* を上回るように算出する．

$\epsilon_r = \epsilon_{init}$ を初期値として，算出した LPr が基準値 *base_value* を上回らなかった場合，$\epsilon_r = \epsilon_r/\alpha, (1 < \alpha)$ の処理を行う．ここで，ϵ_{init}, α はハイパーパラメータである。以上の操作によって位置 r に対して，条件を満たした ϵ_r と $K_{after}(r)$ を求める．

この与えられた位置 r に対しての操作を全ての r について行い K_{after} を算出する．

つまり，この提案手法は LPr についてプライバシ保護を保証するアルゴリズムである．プライバシパラメータの値の意義については，値から通常想像がつかない．しかし LPr は，メートル (m) という単位で算出される．これは値から保護の程度が把握できる．こういった保護の想像がつきやすいという点もこのアルゴリズムの特徴である．

3.1.1 K についての処理

K は確率を示すため，任意の位置 r について理論的に次が成り立つ．

$$\sum_{r'} k(r)(r') = 0 \tag{10}$$

式 5 に従うと，位置 r と位置 r' 間のユークリッド距離，ϵ からメカニズムによる位置 r を r' に曖昧化する確率が求まる．これを K の配列に格納できる．しかし，今回のプログラムは地図の中のある一定の区域を切り取った空間である．よって，これをそのまま配列に格納しても式 10 が満たされない．例えば，この配列が仮定している空間の一番端 (0, 0) から広がる曖昧化する確率は円形で広がっており，空間外にも広がってしまっている．

式 10 を満たすために，今回は post-processing property(節 2.1.1) という概念を用いる．もし，ある真の位置 r について仮定している空間外に曖昧化した場合，仮定している空間内からランダムに曖昧化する位置を選ぶという手法をとる．真の位置を使わず，仮定している空間内に一様分布で確率が存在するとして曖昧化した位置を定めるということである．差分プライバシを用いて一度曖昧化を行なった後は，真の値を使っていないため post-processing property を満たす．

まず，ある位置 r について $K(r)$ を収めている配列には式 5 に従って確率を収める．その後次の式 11 で求められる値を全ての位置に格納する．すると，式 10 を満たす．

$$\frac{(1 - \sum_{r'} K(r)(r'))}{n \times n} \tag{11}$$

3.2 実際の実装について

提案手法では，初めに定めた一定の ϵ から関数 $K_{before}(r)(r')$ や $h(r')(\hat{r})$ などの関数を求める．しかし，これは通常のプログラミング言語では定義が難しい．また，これらを適用する地図の範囲は非常に限定されている．

そこで本研究では，$100(n-1) \times 100(n-1)\mathrm{m}^2$ の仮想的空間に関数が対応していることにした．$(n \in \mathbb{N})$ 空間は 100m × 100m で分割されており，つまり $n \times n$ 箇所のマス目がある．

そして，K や h などの関数は配列で表すことにした．K, h に関しては二次元配列になる．有効な位置は $n \times n$ 箇所存在するため，$(n \times n) \times (n \times n)$ の大きさの 2 次元配列となる．

3.3 実際の位置情報サービスへ適用について

それぞれの位置ごとに算出されたプライバシパラメータ ϵ を用いて LP 法 (節 2.4) を施しプライバシを保護する．

4. 実験

仮想的な設定で 2 種，シミュレーションによって生成した人の位置データの 2 種，合計 4 種のデータから人の存在確率などを導いて実験を行う．

4.1 仮想的な空間

$n = 100, 100 \times 100$ マスの空間を仮定する．そして，1 マスのサイズは 100 m × 100 m とする．そして，人の存在確率が一様分布となっている空間および人の存在確率が一部の位置に集中している空間の 2 種類について検討する．

まず一様分布の空間について検討する．それぞれのマスに均等に確率が割り当てられるので一つのマスの確率は 1/10000 となる．今回この空間のことを unifrom と称することとする．

人の存在確率が一部の位置に集中している空間について検討する．(0, 0) から (0, 10000), (10000, 0), (10000, 10000) に空間が広がっているとする．この中で，(8000, 8000), (8000, 8900), (8900, 8000), (8900, 8900) に囲まれた 100 マスのみに人のいる存在確率が集中しているとする．1 マスの確率は 1/100 となる．今回この空間のことを ununifrom と称することとする．

4.1.1 シミュレーションによって生成した人の位置データを用いた空間

Siafu tool[8] というシミュレーションツールによって生成したデータを使用した．Siafu tool は OSS であり，地図の中で，典型的な人の行動のモデルを用いて人の行動についてのデータを得ることができる．

8.4 km ×8.4 km の会社やレストラン，公園などが配置された空間に，1,000 人から 10,000 人のユーザーが行動する設定となっている．このシミュレーションのデータについては Yuichi ら [9] の先行研究を参考に作成した．

$n = 100, 100 \times 100$ マスの空間を仮定する．そして，1 マスのサイズは 100 m × 100 m とする．シミュレーションによって作成された人の位置データセットのうち，人が同じ場所に密集している率が高い場合と低い場合の 2 種類のデータについて実験を行った．空間を 100 × 100 マスに区切り，シミュレーションデータから人の存在の割合を算出した．その結果 0 でない箇所が 452 である空間 (人が同じ場所に密集している率が高い場合) と 2471 である空間 (人が同じ場所に密集している率が低い場合) を比較することにした．今回は，人が同じ場所に密集している率が高い場合のデータを「密データ」と称し，人が同じ場所に密集している率が低い場合のデータを「疎データ」と称することとする．

4.1.1.1 人が同じ場所に密集している率が高い場合

シミュレーションデータによる分布を次の図 1 に示した．

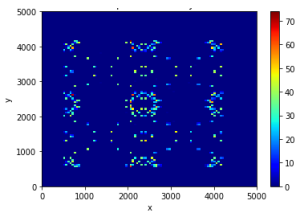

図 1 密データにおける人の分布

4.1.1.2 人が同じ場所に密集している率が低い場合

シミュレーションデータによる分布を次の図 2 に示した．

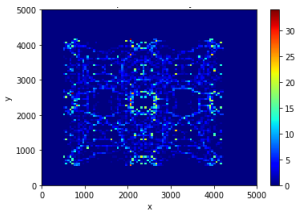

図 2 疎データにおける人の分布

4.2 パラメータ設定について

今回の実装では，$\alpha = 5.0$ とした．ϵ_{init} については 0.1 と 0.5 の 2 種類について試行した．*base_value* については 150, 300, 600, 900, 1200 について試した．ある一点から両隣上下は 100m, 斜め上下は $100\sqrt{2}$ であることから，このようなパラメータ設定にした．ϵ が一律の場合については，ϵ が 0.5，0.1，0.02，0.004 の場合について実験を行なった．

5. 評価実験

本章では SQL(Service Quality Loss) を用いて，式 7 から評価値を算出してユーザーにとっての有用性を評価する．前章で述べたように，仮想的空間 2 種類および，シミュレーションによって生成した人の位置データを用いた空間 2 種類に対してプログラムを実行し値を算出して評価を行った．

5.1 仮想的な空間

仮想的な空間 uniform と ununiform について，評価のための結果を算出した．

まず，全ての位置について同じ ϵ について SQL を算出した結果についてグラフにまとめた．uniform については表 2, ununiform については表 3 としてまとめた．

表 2 uniform データにおいて
全位置に同一の ϵ を適用した場合の SQL 算出結果

ϵ	SQL / m
0.5	659.9
0.1	2673.0
0.02	4952.9
0.004	5209.3

表 3 ununiform データにおいて
全位置に同一の ϵ を適用した場合の SQL 算出結果

ϵ	SQL / m
0.5	385.5
0.1	2808.9
0.02	5536.9
0.004	5851.2

5.1.1 unifrom 結果

提案手法に基づき，uniform データについて，$\epsilon_{init} = 0.5$ として提案手法に基づきプライバシ保護を施した場合の SQL を計算した結果，次の表 4 のようになった．$\epsilon_{init} = 0.1$ とした場合については表 5 のようになった．

MIN LPr と MAX LPr はすべての位置でそれぞれの LPr を計算したとき，すべての位置の中で最小の LPr と最大の LPr を算出して記載した．つまり，提案手法では MIN LPr が *base_value* を上回っている必要がある．

5.1.2 ununiform 結果

提案手法に基づき，ununiform データについて，$\epsilon_{init} = 0.5$ として提案手法に基づきプライバシ保護を施した場合の SQL を計算した結果，次の表 6 のようになった．$\epsilon_{init} = 0.1$ とした場合については表 7 のようになった．

表 4 unifrom データにおける SQL，最小 LPr，最大 LPr の算出結果 (ϵ_{init} = 0.5)

base_value	SQL / m	MIN LPr / m	MAX LPr / m
150	659.902	375.021	5143.26
300	659.902	375.021	5143.26
600	2142.38	600.855	5143.26
900	2332.26	902.399	5143.26
1200	2431.46	1240.12	5143.26

表 5 uniform データにおける SQL，最小 LPr，最大 LPr の算出結果 (ϵ_{init} = 0.1)

base_value	SQL / m	MIN LPr / m	MAX LPr / m
150	659.902	327.571	5662.88
300	659.902	327.572	5662.88
600	1478.58	601.507	5662.88
900	1895.29	904.208	5662.88
1200	2106.69	1210.79	5662.88

表 6 ununiform データにおける SQL，最小 LPr，最大 LPr の算出結果 (ϵ_{init} = 0.5)

base_value	SQL / m	MIN LPr / m	MAX LPr / m
150	385.5	201.24	10685
300	2563.5	310.18	10685
600	2808.9	601.26	10685
900	2808.9	902.22	10685
1200	2808.9	1200.37	10685

表 7 ununiform データにおける SQL，最小 LPr，最大 LPr の算出結果 (ϵ_{init} = 0.1)

base_value	SQL / m	MIN LPr / m	MAX LPr / m
150	795.199	38.237	11999.9
300	2122.90	38.237	11999.9
600	5731.12	38.237	11999.9
900	5856.93	38.237	11999.9
1200	5856.93	38.237	11999.9

5.2 シミュレーションデータ

シミュレーションから得たデータである密データと疎データについて，評価のための SQL の値を算出した．

5.2.1 ϵ の調節を行わない従来手法

全ての位置について同じ ϵ について SQL を算出した結果について表にまとめた．uniform については表 8, ununiform については表 9 としてまとめた．

表 8 密データにおいて，全位置に同一の ϵ を適用した場合の SQL 算出結果

ϵ	SQL / m
0.5	380.287
0.1	2193.35
0.02	4496.00
0.004	4727.52

表 9 疎データにおいて，全位置に同一の ϵ を適用した場合の SQL 算出結果

ϵ	SQL / m
0.5	379.656
0.1	2093.99
0.02	4333.87
0.004	4552.14

5.2.2 密データ

密データに対して，提案手法に基づき ϵ_{init} = 0.5 としてプライバシ保護を施した場合に SQL を計算した結果，次の表 10 のようになった．ϵ_{init} = 0.1 とした場合については表 11 のようになった．

表 10 密データにおける SQL，最小 LPr，最大 LPr の算出結果 (ϵ_{init} = 0.5)

base_value	SQL	MIN LPr / m	MAX LPr / m
150	399.703	150.341	5363.88
300	1549.44	300.065	5363.88
600	2192.87	600.162	5363.88
900	2193.35	900.152	5363.88
1200	2193.35	1200.38	5363.88

表 11 密データにおける SQL，最小 LPr，最大 LPr の算出結果 (ϵ_{init} = 0.1)

base_value	SQL	MIN LPr / m	MAX LPr / m
150	380.287	150.607	6521.97
300	667.314	300.508	6521.97
600	1784.54	600.044	6521.97
900	2185.64	900.430	6521.97
1200	2193.35	1201.63	6521.97

5.2.3 疎データ

疎データについて，提案手法に基づき ϵ_{init} = 0.5 としてプライバシ保護を施した場合に SQL を計算した結果，次の表 12 のようになった．ϵ_{init} = 0.1 とした場合については表 13 のようになった．

表 12 疎データにおける SQL，最小 LPr，最大 LPr の算出結果 (ϵ_{init} = 0.5)

base_value	SQL	MIN LPr / m	MAX LPr / m
150	379.656	183.387	5861.67
300	753.016	300.075	5861.67
600	2093.79	600.712	5861.67
900	2093.99	900.184	5861.67
1200	2093.99	1200.27	5861.67

表 13 疎データにおける SQL，
最小 LPr，最大 LPr の算出結果 (ϵ_{init} = 0.1)

base_value	SQL	MIN LPr / m	MAX LPr / m
150	379.656	279.065	6732.24
300	414.746	300.069	6732.24
600	1148.84	600.020	6732.24
900	1733.13	900.165	6732.24
1200	2026.25	1200.00	6732.24

6. 考察

6.1 ϵ を変化させなかった場合の SQL の変化

表 2，3，8，9 について，プライバシパラメータ ϵ を場所ごとに変化させなかった場合，特に ϵ を 0.5 から 0.1 に変化させた場合著しく SQL が上がっている．ϵ を安易に下げることはユーザーの利便性を著しく欠く可能性があることを示している．

6.2 提案手法を用いた結果について

ununiform データで ϵ_{init} = 0.1 である場合以外，提案手法に基づくプログラムの実行によるすべての結果で，MIN LPr が *base_value* を上回る結果となった．上回っている場合には LPr について保証ができている．

ununiform データで ϵ_{init} = 0.1 である場合以外，*base_value* が 1200 までのすべての結果において，全てに同じプライバシパラメータ ϵ を適用した場合の ϵ = 0.1 の SQL を下回るか，整数部分が同じである結果となった．これは，提案手法によって，SQL を抑えながら必要な場所ではプライバシパラメータの値を下げることにより保護を強化するという目標を達成した結果であると言える．

unifrom データの *base_value* が 150 と 300 で変わっていない．これは，basis value を 150 とした場合でも，全体で一番小さい LPr の値が既に 300 を上回っているため，*base_value* を 300 にした時と 150 にした時でそれぞれの場所に割り当てられたプライバシパラメータに違いがないことに由来する．

ununiform データに提案手法を ϵ_{init} = 0.1 で適用している場合，SQL も著しく上がっている．それぞれの場所で LPr がどのような値になっているか確認したところ，(85, 85) が一番小さかった．空間内でこの場所にいる確率が一番高そうだと推測されうることからこのような結果になっていると考えられる．この空間は人工的に 100 × 100 の空間の中の 10 × 10 にしか人がいないという設定になっており，10 × 10 の空間だけに着目して計算をした場合，距離と ϵ の大きさの関係のバランスが不適切になっている可能性がある．

6.3 計算時間

場所ごとに適切なプライバシパラメータ ϵ を定めるために，提案手法では初めにメカニズム K および攻撃者関数 h を計算で求めている．これは計算量が $O(n^3)$ であり，計算コストが莫大にかかってしまうという結果となった．しかし，これには解決策を提示できる．一度空間内で場所ごとに使うプライバシパラメータを定めてしまうと，その定められたパラメータを真の位置に適用することに対する計算コストはほとんどかからない．人の存在確率などの位置による性質は時間により異なるが，それは周期的なものであると考えられるため，事前に 1 回計算してパラメータを定めてしまうことによってそのあとは計算コストの負荷を回避することが可能であると考える．今回は人の存在確率について着目したが，位置が病院であるかスーパーマーケットであるかなどの位置の意味的な情報についても加味したプライバシパラメータを定めておき，真の位置に適用するなどの手法を取ることも考えられる．

6.4 ϵ を変化させることへの保証

本稿では，プライバシパラメータ ϵ を一律に用いるのではなく変化させた場合について，つまりノイズ付加量を変化させた場合についてのリスクなどを検討していない．ノイズ付加量を状況に応じて変化させるという，笹田ら [10] による類似研究が存在するが，これについてもノイズ付加量を変化させた場合についてのリスクなどを検証していない．これは今後の課題となる．

7. おわりに

7.1 まとめ

本論文では，ある時点での人の存在確率から基づくプライバシ保護の度合いを基準として位置情報にノイズを加える手法を提案した．その結果，各位置に適切なプライバシパラメータを定めることによって，客観的指標であるユーザーの利便性 SQL を抑えることができた．

7.2 今後の展望

時系列データでは，単純な差分プライバシの適用では，プライバシ保護の度合いが低下することが知られている．同一の位置について幾度も保護済みのデータを送るとその分布から真の位置が見破られる可能性がある．今回の手法でもその可能性は考えられるため，時系列データのプライバシ保護低下を防ぐ手法が今後必要となる．

また，病院などのセンシティブな情報を推測しやすい位置については，プライバシパラメータの値を下げて保護を強化する必要があると考える．位置情報や施設の意味的なものにも着目したプライバシ保護手法の提案について今後期待される．

今回の手法では位置ごとにプライバシパラメータ ϵ の値を変えている．これによる影響については評価していない．今後は，今回のようなプライバシパラメータをそれぞ

れの位置や時間で変えることによる影響を評価していく必要がある.

8. 謝辞

本研究は JSPS 科研費 JP18H03229, JP18H03340, JP18K19835, JP19H04113, JP19K12107, JP21H03496 の助成を受けたものです.

参考文献

[1] C. Dwork. differentiall privacy. In *in Automata, Languages and Programming*, pp. 1–12, 2006.

[2] Keyu Zhu, Pascal Van Hentenryck, and Ferdinando Fioretto. Bias and variance of post-processing in differential privacy. *arXiv preprint arXiv:2010.04327*, 2020.

[3] Konstantinos Chatzikokolakis, Miguel E Andrés, Nicolás Emilio Bordenabe, and Catuscia Palamidessi. Broadening the scope of differential privacy using metrics. In *International Symposium on Privacy Enhancing Technologies Symposium*, pp. 82–102. Springer, 2013.

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

[5] Reza Shokri, George Theodorakopoulos, Carmela Troncoso, Jean-Pierre Hubaux, and Jean-Yves Le Boudec. Protecting location privacy: Optimal strategy against localization attacks. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS '12, pp. 617–627, New York, NY, USA, 2012. Association for Computing Machinery.

[6] Reza Shokri, George Theodorakopoulos, Jean-Yves Le Boudec, and Jean-Pierre Hubaux. Quantifying location privacy. In *2011 IEEE symposium on security and privacy*, pp. 247–262. IEEE, 2011.

[7] 高木駿, 曹洋, 浅野泰仁, 吉川正俊. 道路ネットワークにおける位置情報プライバシ. In *DEIM Forum 2019*, 2019.

[8] Miquel Martin and Petteri Nurmi. A generic large scale simulator for ubiquitous computing. In *2006 Third Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services*, pp. 1–3. IEEE, 2006.

[9] Sei Yuichi and Ohsuga Akihiko. Location anonymization with considering errors and existence probability. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 47, No. 12, pp. 3207–3218, dec 2017.

[10] 笹田大翔, 笹田大翔, 妙中雄三, 門林雄基, FALL Doudou. ユーザの軌跡隣接性を考慮した局所差分プライバシにおけるノイズ付加量変化の検討. 信学技報, pp. 45–50. 奈良先端科学技術大学院大, サイボウズ・ラボ, 2020.

[11] Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Marco Stronati. Constructing elastic distinguishability metrics for location privacy. *arXiv preprint arXiv:1503.00756*, 2015.