地図の非対称性に起因する差分プライバシ脆弱化の軽減

平石　亮太† 　吉川　正俊†† 　曹　　洋†† 　藤田　澄男††† 　五味　秀仁†††

† 京都大学工学部情報学科　〒 606–8501 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　〒 606–8501 京都府京都市左京区吉田本町 36–1
††† ヤフー株式会社 Yahoo! JAPAN 研究所　〒 102–8282 東京都千代田区紀尾井町 1–3
E-mail: †hiraishi.ryouta.73m@st.kyoto-u.ac.jp, ††{yoshikawa,yang}@i.kyoto-u.ac.jp,
†††{sufujita,hgomi}@yahoo-corp.jp

あらまし　近年, パーソナルデータの重要性は増しており, 多くのデータを収集し利用することはマーケティングや提供するサービスの質の向上にとって必要不可欠となりつつある. 本論文では特に位置情報を差分プライバシで保護することを考えるが, 実用上, 有限の地図を考える必要があることや海などの人が位置しない場所の存在によって, 地図に非対称性が生まれ, 差分プライバシが脆弱化する問題が発生する. この問題について詳細に議論するために, グリッド上の位置情報での差分プライバシを考える. グリッド上の位置情報に対して, 差分プライバシを適用する手法を定義し, この設定の下で差分プライバシの脆弱化ついて詳細を議論する. また, この問題を軽減する手法を提案し, 実際に問題が軽減されることを実験により示す.

キーワード　位置情報の保護, 差分プライバシ, 局所差分プライバシ

1　はじめに

近年, パーソナルデータの重要性は増大し続けており, パーソナルデータを利活用することはビジネスや社会全体のために必要不可欠となってきている. しかし, 人々が提供したデータは悪用される可能性があるため, ユーザがパーソナルデータを提供する際には自身の端末でランダムな雑音を加えてから収集者に提供することが考えられる. この考えは Warner[1] による手法が元になっており, 差分プライバシ [4] を用いた概念に拡張され, 局所差分プライバシ (Local Differential Privacy, LDP)[7][8] と呼ばれている. また, データの収集者はユーザのデータ提供への動機付けのため, 提供されたデータが利用された場合には各個人に雑音の量に応じた補償を分配する. このようなモデルを導入することで, データの収集者はパーソナルデータを活用したビジネス上の意思決定や分析を行うことができる一方で, ユーザも収集者がデータを絶対に悪用しないという完全な信用をする必要なく, プライバシが保証されたデータを提供することでサービスやデータの提供に対する補償を受けることができる. これは最終的にスマートシティの構築といった社会全体への利益につながる. 本研究では特に個人の位置情報に焦点を当てる. ユーザは自身の位置情報に雑音を加えてから位置情報を提供するとし, 収集者はこれらの雑音を含むデータを利用することを考える. このようなデータは天気予報のサービスや近場の店舗を検索するサービスに利用することができる. また, 大量のデータを収集し集約することができれば, 店舗のオープンやサービスの提供エリアの拡大といったビジネス上の意思決定や, 渋滞や混雑状況などの情報を提供するアプリケーションなどで利用できると考えられ有用である.

位置情報のためのプライバシ保護基準としては Andrés ら [6] の Geo-Indistinguishability (以下 Geo-I と略記する) が広く知られている. しかし, Geo-I で用いられる 2 次元ラプラスメカニズムは Takagi ら [3] により道路情報を考慮することにより, 有用性やの保護度合いが低下することが示されている. [3] では道路ネットワークをグラフとして表現し, グラフ上の差分プライバシである Geo-Graph-Indistinguishability(GGI) を用いることでこの問題を解決したが, 海の近くなど周辺に頂点が少ない地域や地図の “端”ではプライバシ保護の度合いが下がってしまう.

本論文では, この問題について詳細に議論するために, 扱いが容易なグリッド上の位置情報を考え, 位置情報をグリッド上で切り分けられた領域単位で表す. これは, 平面上の格子点を頂点とするグラフと考えることができ, GGI の特殊な場合といえる. グリッド上の個人の位置情報を収集者に提供する際のプライバシ基準として, 差分プライバシ [4] を一般のデータに拡張した Chatzikokolakis ら [2] の $d_{\mathcal{X}}$-privacy を用いる. また, 雑音を加えるためのメカニズムとして $d_{\mathcal{X}}$-privacy における指数メカニズムのナイーブな拡張である Grid-Exponential-Mechanism($GridE$) を提案する. しかし, 長方形の地図において $GridE$ で雑音を加えた際には, Geo-I と同様に海上などの人が以内と考えられる場所に出力されてしまう可能性がある. 解決法として各領域に重みを割り当て, その重みを考慮した指数メカニズムである Weighted-Grid-Exponential-Mechanism($WGridE$) を提案する. これらの差分プライバシを保証するための手法を提案したのち, 地図の”端”付近では他の領域よりもプライバシの保護度合いが低下してしまう現象が発生することを実験により示す. 最後に, この境界脆弱性問題を解決するための手法を提案し, この手法を適用した場合にはどの手法も適用しなかった場合よりも境界脆弱性問題が改善されることを示す

2 基礎事項

この章では本研究で利用する差分プライバシに関する基礎事項を紹介する.

2.1 差分プライバシ

Dwork[4] によって提案された差分プライバシ (differential privacy) は, 数学的に厳密に定義されたプライバシ基準である. 直観的には, 差分プライバシはある一つのレコードの存在がデータベースへの問合せの答に与える影響が小さいことを表している. 同一サイズの 2 つのデータベース D_1, D_2 が 1 つのレコードのみ異なる場合に, D_1, D_2 は隣接しているという. このとき, 差分プライバシは以下のように定義される.

定義 1 (ϵ-差分プライバシ)**.** $\epsilon \in \mathbb{R}^+$ について, あるメカニズム K が任意の隣接したデータベース D_1, D_2 と $\forall S \subseteq range(K)$ に対して,

$$\Pr[K(D_1) \in S] \leqq e^{\epsilon} \Pr[K(D_2) \in S]$$

を満足するとき K は ϵ-差分プライバシを満たすという.

定義中の $\mathbb{R}^+$ は正の実数全体の集合を表す. ϵ はプライバシパラメータやプライバシバジェット (privacy budget) などと呼ばれるパラメータであり, 一般に ϵ が小さいほど隣接したデータベースをメカニズムの入力としたときに出力が同じ集合に含まれる確率の比が 1 に近づくため, 出力のみを観測したとき 2 つのデータベースを識別するのが困難になる. 結果, プライバシ保護の度合いは大きくなるがデータの有用性 ("正確さ") は低下する. ϵ-差分プライバシを満たすメカニズムとしては, 元のデータベースに問合せを適用した結果にラプラス分布から抽出したランダムな雑音を加えるラプラスメカニズム [4] や, 答の"質"を数値で表す関数に依存した確率で結果を出力する指数メカニズム [5] がよく知られている.

2.2 指数メカニズム

指数メカニズムは McSherry ら [5] によって提案された, ϵ-差分プライバシを満たすメカニズムである. ラプラスメカニズムとの大きな相違点はラプラスメカニズムは数値データにのみ適用できるメカニズムであるのに対し, 指数メカニズムは数値以外のデータにも適用できる点である. 指数メカニズムではメカニズムの出力 o の "質"を数値で表すクオリティ関数 $q : (\mathbb{D} \times O) \to \mathbb{R}$ を定義する. ここで, $\mathbb{D}$ はデータセット全体の集合を, O はメカニズムの可能な出力の集合を, $\mathbb{R}$ は実数全体の集合をそれぞれ表す. クオリティ関数の sensitivity Δq を以下のように定義する.

$$\Delta q = \max_{\forall o, D \simeq D'} |q(D, o) - q(D', o)|$$

ここで, $D \simeq D'$ は D と D' が隣接していることを表す. クオリティ関数の sensitivity はレコードが一つだけ異なるデータベースに対するクエリの結果と元のデータベースに対するクエリの結果のクオリティ関数の最大の変化量を表す. この sensitivity を用いて指数メカニズムは以下のように定義される.

定義 2 (指数メカニズム)**.** クオリティ関数 $q : (\mathbb{D} \times O) \to \mathbb{R}$ とプライバシパラメータ $\epsilon \in \mathbb{R}^+$ に対して, 指数メカニズム $\mathcal{M}_q^{\epsilon}(D)$ は $o \in O$ を $\exp(\frac{\epsilon q(D,o)}{2\Delta q})$ に比例する確率で出力する. すなわち,

$$\Pr[\mathcal{M}_q^{\epsilon}(D) = o] = \frac{\exp(\frac{\epsilon q(D,o)}{2\Delta q})}{\sum_{o' \in O} \exp(\frac{\epsilon q(D,o')}{2\Delta q})}.$$

2.3 $d_{\mathcal{X}}$-privacy

Chatzikokolakis ら [2] は, 差分プライバシの定義をデータベースだけではなく一般的なデータに対するものへと拡張した. ドメイン $\mathcal{X}$ 上のメカニズム K がドメイン $\mathcal{Z}$ 上の確率分布を与え, $d_{\mathcal{X}}$ は $\mathcal{X}$ 上の "距離"を表すとすると, $d_{\mathcal{X}}$-privacy は以下のように定義される.

定義 3. 任意の $x, x' \in \mathcal{X}, Z \subseteq \mathcal{Z}$ について以下の不等式

$$\frac{K(x)(Z)}{K(x')(Z)} \leqq \epsilon d_{\mathcal{X}}(x, x')$$

を満たすとき, K は $\epsilon d_{\mathcal{X}}$-privacy を満たすという.

$d_{\mathcal{X}}$-privacy では任意の二つのデータに対してそのデータ間の距離に応じて識別の困難さが変化し, 距離が小さいデータではメカニズムの出力の分布が近くなり. 出力を観測するだけでは実際のデータが二つのうちのいずれであるかの識別が困難であることを表現している. データベース上の差分プライバシは, $\epsilon d_{\mathcal{X}}$-privacy において $d_{\mathcal{X}}$ をハミング距離 d_h とし, $d_h = 1$ となるデータベースだけを考慮したものと等価である.

3 関連研究

この章では関連研究を紹介する.

3.1 差分プライバシの位置情報への適用

位置情報に差分プライバシを適用する研究も盛んにおこなわれている. 最も広く知られた概念は Geo-indistinguishability (Geo-I) である. これは, 前章で紹介した $d_{\mathcal{X}}$-privacy において x を各位置, $d_{\mathcal{X}}$ をユークリッド距離として定義したものと捉えることができる. Geo-I では位置情報に確率分布から抽出した雑音を加えることで, 正確な位置情報を推定することを困難にする. 2 次元ラプラス分布を用いたラプラスメカニズムは Geo-I を満たすことが示されている.

Takagi ら [3] は Geo-I が道路ネットワークを考慮できていないことに注目した. Geo-I では道路ネットワークを考慮することによって有用性の低下や, 道路ネットワークに関する情報をもつ攻撃者に対するプライバシ保護の度合いの低下が生じることを示し, 道路ネットワークを考慮したプライバシ基準として位置情報をグラフで表現した Geo-Graph-Indistinguishability(GGI) と GGI を満たすメカニズムである Graph-Exponential-Mechanism (GEM) を提案した.

3.2 局所差分プライバシ

データベース内のデータのプライバシを保護するための差分プライバシでは信頼されたサーバが個人の正確なデータを収集し集約したのちに, メカニズムを適用することでプライバシを保護していたが, サーバは実際には収集したデータを不正に利用したり, 第三者に販売したりする可能性があり必ずしも信頼できるとは限らない. そこで, ユーザがスマートフォンなどの自身の端末上で差分プライバシのメカニズムを用いてデータに雑音を加えてからサーバに送信する局所差分プライバシ (LDP)[7][8] が注目されている. LDP ではサーバはユーザの雑音が加えられたデータしか観測できず, 正確なデータを知ることはできない.

LDP を用いたデータの収集や集約に関する先行研究をいくつか紹介する. Chatzikokolakis ら [9] は全ての位置が要求されたプライバシを満たすようにグラフを構築することで Elastic metric を計算し, 人口が多い地域と少ない地域でプライバシ保護の度合いが異なってしまう問題を解決した. Chen ら [10] は LDP のもとで safe region と呼ばれる, ユーザがその領域にいることは開示してよいが, その領域内の任意の 2 点では差分プライバシが満たされていなければならない領域を考え, これを用いた Personalized Local Differential Privacy (PLDP) を定義した. また, PLDP を満たす randomized-responce に基づくメカニズム及び各領域内のユーザ数をカウントするための手法を提案した. これにより, 同じユーザの中でも場所ごとにプライバシ保護への要求の度合いが異なる状況に適応した. Gu ら [11] はデータごとにプライバシ保護への要求の度合いが異なることを考慮したプライバシ基準を定義し, ランダムリスポンスを用いて MSE(平均二乗誤差) を最小化するように最適化問題を解くこと, でこのプライバシ基準を満たすメカニズムを求めた.

4 グリッド上での位置情報のプライバシ保護

4.1 領域単位での指数メカニズム

本研究ではユーザの位置情報を収集する際のプライバシの問題を考える. このようなデータは天気予報のサービスや近場の店舗を検索するサービスといった, ユーザ個人へのサービス提供だけでなく, 多くのデータを収集し集約することでビジネスにおける意思決定や事業展開のために利用できる他, このデータを使用したサービスにも利用できる. 例えば, ビジネスの場面では新たな店舗のオープンやサービス展開の場所を, 各位置にいる人数を補助情報として決定することが考えられる. また, 集約データをリアルタイムに利用できれば渋滞予測や混雑状況に関する情報を提供するサービスが可能になる.

このようにユーザの位置情報は有用であるが, 一方でユーザの自宅や職場といったセンシティブな情報の解析や, ユーザの位置情報のトラッキングに利用される可能性があるため, プライバシへの配慮が必要となる. ユーザは位置情報をデータ収集者に提供するが, これらの組織は必ずしも信頼できるとは限らないため, ユーザは正確なデータを提供するのではなくスマートフォンなどの自身の端末上で現在位置に雑音を加えてから提供する LDP の概念を用いるものとする.

表 1: 記号の意味

記号	意味
$\mathcal{R}$	地図上の領域全体の集合
r_{ij}	i 行 j 列の領域
r, r', r_k	領域（具体的な位置を指定しない場合）
$K(r)(r')$	メカニズム K が r を入力とし r' を出力する確率
$\pi(r)$	ユーザが領域 r にいる事前確率
$d_{\mathcal{P}}(\sigma_1, \sigma_2)$	確率分布 σ_1, σ_2 間の距離
$d_{\#}(r, r')$	領域 r の中心と領域 r' の中心間の距離
w_{ij}	領域 r_{ij} に割り当てられた重み
$\Pr(O = r' \mid A = r)$	ユーザが領域 r に位置するときにメカニズムの出力（観測された領域）が r' である確率, $K(r)(r')$ とほぼ同義
$\Pr(A = r \mid O = r')$	ユーザが r' に位置することが観測されたとき, 実際は領域 r に位置する事後確率

道路ネットワークを考慮したプライバシ基準として 3.1 節で述べた, GGI があるが, このプライバシ基準を満たす GE で雑音を加えると 海の近くなど周辺に頂点が少ない地域や, 実用のために抽出した有限の地図上の “端”ではプライバシ保護の度合いが低下してしまう. 本論文ではこの問題について詳細に議論するために, データの扱いが容易なグリッド上の位置情報を考える. グリッドは対象とする地域の地図をほぼ長方形の領域に切り分けることで構成され, ユーザの位置情報はグリッド内の領域単位で表す. これは, 平面上の格子点を頂点とするグラフと考えることができ, GGI で考えている状況の特殊な場合といえる. グリッド以外で表現された位置情報への拡張は今後の研究課題とする. ただし, 地図のグリッド表現は地域メッシュのように広く使われている規格が存在するため, ユーザ, 収集者およびデータ利用者の間でのデータに対する共通認識が容易になるという実用上のメリットが存在するため, 本研究の内容も有用であると考えている.

以降, 本論文で使用する記号とその意味を表 1 に示した.

本研究で使用するグリッド上の位置情報に関しては以下のように定義する. 縦 V 個× 横 H 個のほぼ長方形の領域に分割されたグリッド状の地図を考える. $\mathcal{R}$ を領域全体の集合とするとき, i 行 j 列の領域を $r_{ij} \in \mathcal{R}$ と表し, 各ユーザは実際の位置を含む領域に位置しているとする. ただし, 具体的な位置を考慮する必要がないときには単に r, r' や r_k といった表記も使用する. このとき, 各ユーザのプライバシを保護するために 2.3 節で説明した $\epsilon d_{\mathcal{X}}$-privacy において $\mathcal{X}$ を $\mathcal{R}$, $d_{\mathcal{X}}$ を領域の中心間の距離 $d_{\#}$ と定義し, 任意の $R \subseteq \mathcal{R}$ に対して以下のような不等式が成り立つようにする.

$$\frac{K(r_1)(R)}{K(r_2)(R)} \leqq \exp(\epsilon d_{\#}(r_1, r_2)) \tag{1}$$

直観的には, この不等式は r_1 と r_2 の距離が近いときにはユーザが領域 r_1 にいるときのメカニズムの出力の確率分布と r_2 にいるときのメカニズムの出力の確率分布が似ているため, r_1 と

r_2 を識別するのが困難であることを表現している.

このプライバシ基準を満たすメカニズムとして $d_{\mathcal{X}}$-privacy における指数メカニズムをグリッドのデータに適用したメカニズムである, Grid-Exponential-Mechanism を用いる. これは以下のように定義される.

定義 4 ($GridE$(Grid-Exponential-Mechanism))**.** プライバシパラメータ $\epsilon \in \mathbb{R}^+$ が与えられたとき, Grid-Exponential-Mechanism $\mathcal{M}^{\epsilon}_{d_\#}(r)$ は $r' \in \mathcal{R}$ を $\exp(-\frac{\epsilon}{2}d_\#(r,r'))$ に比例する確率で出力する. すなわち,

$$\Pr[\mathcal{M}^{\epsilon}_{d_\#}(r) = r'] = \frac{\exp(-\frac{\epsilon}{2}d_\#(r,r'))}{\sum_{r_{i'j'}\in\mathcal{R}} w_{i'j'}\exp(-\frac{\epsilon}{2}d_\#(r,r_{i'j'}))}$$

Grid-Exponential-Mechanism ではユーザの正確な位置が出力される確率が最大であり, その領域からの距離が大きくなるにつれて出力される確率は小さくなる.

4.2 地理情報に基づいたプライバシ保護

Takagi ら [3] によると Geo-I を満たすメカニズムで雑音を加えた後のデータが, 川や海といった人がいるとは考えにくい位置を表していた場合, データの有用性やプライバシ保護の度合いが低下する. この問題に対処するために本論文ではこの問題に対処するために, 各領域 r_{ij} に重み w_{ij} を割り当てることを考える. w_{ij} はその領域が表現する実際の領域に川や海などのユーザが存在しているとは考えにくい部分の割合に基づいて決定され, 0 以上 1 以下の値をとる. 例えば, 街中など領域内の全ての位置でユーザの存在が妥当であると考えられるとき, その領域 r_{ij} の重みは $w_{ij} = 1$ とする. 一方, 海など領域内の全ての位置でユーザの存在する可能性が低いと考えられるとき, その領域 r_{ij} の重みは $w_{ij} = 0$ とする. この重みの概念を用いて GridE を拡張した Weighted-Grid-Exponential-Mechanism($WGridE$) を以下のように定義する.

定義 5 (Weighted-Grid-Exponential-Mechanism($WGridE$))**.** プライバシパラメータ $\epsilon \in \mathbb{R}^+$ が与えられたとき, Weighted-Grid-Exponential-Mechanism $\mathcal{WM}^{\epsilon}_{d_\#}(r)$ は重み w_{ij} $(0 \leqq w_{ij} \leqq 1)$ が設定された領域 $r_{ij} \in \mathcal{R}$ を $w_{ij}\exp(-\frac{\epsilon}{2}d_\#(r,r_{ij}))$ に比例する確率で出力する. すなわち,

$$\Pr[\mathcal{M}^{\epsilon}_{d_\#}(r) = r_{ij}] = \alpha(r)w_{ij}\exp(-\frac{\epsilon}{2}d_\#(r,r_{ij}))$$

ここで, $\alpha(r)$ は正規化項であり,

$$\alpha(r) = \frac{1}{\sum_{r_{i'j'}\in\mathcal{R}} w_{i'j'}\exp\left(-\frac{\epsilon}{2}d_\#(r,r_{i'j'})\right)}$$

とする. ただし, $\forall r_{i'j'} \in \mathcal{R}, 0 \leqq w_{i'j'} \leqq 1$ かつ $\exists r_{i'j'} \in \mathcal{R}, 0 < w_{i'j'}$ が成り立つとし, この定義内では $\frac{0}{0} = 1$ であるとする.

直観的には領域 r_{ij} に設定される重み w_{ij} が小さいほどその領域が出力される確率は小さくなり, $w_{ij} = 0$ ならば, その領域が出力されることはない. $GridE$ とは異なり $WGridE$ では重みの値によってはユーザの正確な位置がそのまま出力となる確率が最大であるとは限らない.

以下に $WGridE$ が式 (1) を満たすことを示す.

命題 1. $WGridE$ は $d_\#$ を距離尺度とする $d_{\mathcal{X}}$-privacy を満たす.

Proof. 任意の 2 つの領域 $r_1, r_2 \in \mathcal{R}$ と任意の $R \subseteq \mathcal{R}$ で以下が成り立つことを示せばよい.

$$\frac{K(r_1)(R)}{K(r_2)(R)} \leqq \epsilon d_\#(r_1, r_2)$$

つまり, 任意の 2 個の領域 $r_1, r_2 \in \mathcal{R}$ について

$$\frac{\Pr[\mathcal{WM}^{\epsilon}_{d_\#}(r_1) = r_{ij}]}{\Pr[\mathcal{WM}^{\epsilon}_{d_\#}(r_2) = r_{ij}]} = \frac{\alpha(r_1)w_{ij}\exp(-\frac{\epsilon}{2}d_\#(r_1,r_{ij}))}{\alpha(r_2)w_{ij}\exp\left(-\frac{\epsilon}{2}d_\#(r_2,r_{ij})\right)}$$

の最大値を調べればよい. まず $w_{ij} = 0$ のとき, 定義 5 より

$$\frac{\alpha(r_1)w_{ij}\exp(-\frac{\epsilon}{2}d_\#(r_1,r_{ij}))}{\alpha(r_2)w_{ij}\exp\left(-\frac{\epsilon}{2}d_\#(r_2,r_{ij})\right)} = \frac{0}{0} = 1 < e^{\epsilon}$$

となる. 次に $w_{ij} \neq 0$ のとき

$$\begin{aligned}&\frac{\alpha(r_1)w_{ij}\exp(-\frac{\epsilon}{2}d_\#(r_1,r_{ij}))}{\alpha(r_2)w_{ij}\exp\left(-\frac{\epsilon}{2}d_\#(r_2,r_{ij})\right)}\\ =&\frac{\alpha(r_1)}{\alpha(r_2)}\exp\left(\frac{\epsilon}{2}(d_\#(r_2,r_{ij}) - d_\#(r_1,r_{ij}))\right) \qquad (2)\end{aligned}$$

このとき, 式 (2) が最大になる場合を考えればよい. それは $d_\#(r_2, r_{i,j}) - d_\#(r_1, r_{ij})$ が最大値をとるときである. グリッド状の地図内の領域を考えているので任意の領域を経由可能である. 三角不等式より, $\forall r_{ij} \in \mathcal{R}, d_\#(r_2,r_{ij}) - d_\#(r_1,r_{ij}) \leqq d_\#(r_1,r_2)$ が成り立つので,

$$\frac{\Pr[\mathcal{WM}^{\epsilon}_{d_\#}(r_1) = r_{ij}]}{\Pr[\mathcal{WM}^{\epsilon}_{d_\#}(r_2) = r_{ij}]} \leqq \frac{\alpha(r_1)}{\alpha(r_2)}\exp\left(\frac{\epsilon}{2}d_\#(r_1,r_2)\right)$$

が成り立つ. 次に $\frac{\alpha(r_1)}{\alpha(r_2)} \leqq \exp(\frac{\epsilon}{2}d_\#(r_1,r_2))$, すなわち

$$\begin{aligned}\sum_{r_{i'j'}\in\mathcal{R}} w_{i'j'}\exp\left(-\frac{\epsilon}{2}d_\#(r_1,r_{i'j'})\right)\exp\left(\frac{\epsilon}{2}d_\#(r_1,r_2)\right)\\ -\sum_{r_{i'j'}\in\mathcal{R}} w_{i'j'}\exp\left(-\frac{\epsilon}{2}d_\#(r_2,r_{i'j'})\right) \geqq 0\end{aligned}$$

であることを示す. 三角不等式より,

$$\begin{aligned}&\sum_{r_{i'j'}\in\mathcal{R}} w_{i'j'}\exp\left(-\frac{\epsilon}{2}d_\#(r_1,r_{i'j'})\right)\exp\left(\frac{\epsilon}{2}d_\#(r_1,r_2)\right)\\ &-\sum_{r_{i'j'}\in\mathcal{R}} w_{i'j'}\exp\left(-\frac{\epsilon}{2}d_\#(r_2,r_{i'j'})\right)\\ =&\sum_{r_{i'j'}\in\mathcal{R}} w_{i'j'}\exp\left(\frac{\epsilon}{2}(d_\#(r_1,r_2) - d_\#(r_1,r_{i'j'}))\right)\\ &-\sum_{r_{i'j'}\in\mathcal{R}} w_{i'j'}\exp\left(-\frac{\epsilon}{2}d_\#(r_2,r_{i'j'})\right)\\ \geqq&\sum_{r_{i'j'}\in\mathcal{R}} w_{i'j'}\exp\left(-\frac{\epsilon}{2}d_\#(r_2,r_{i'j'})\right)\\ &-\sum_{r_{i'j'}\in\mathcal{R}} w_{i'j'}\exp\left(-\frac{\epsilon}{2}d_\#(r_2,r_{i'j'})\right) = 0\end{aligned}$$

となるので,

$$\frac{\Pr[\mathcal{WM}^{\epsilon}_{d_\#}(r_1) = r_{ij}]}{\Pr[\mathcal{WM}^{\epsilon}_{d_\#}(r_2) = r_{ij}]} \leqq \exp(\epsilon d_\#(r_1,r_2))$$

が成り立ち, $WGridE$ は式 (1) を満たすことが示された. □

5 境界脆弱性問題

5.1 *GridE* および *WGridE* における問題点

この章では, 4.1 節でも少し触れた地図の非対称性から生じる式 (1) で表現される差分プライバシの脆弱化について述べる. 以後, 頻繁に使用するためユーザの位置に関する事後確率を以下の記法で表現する.

$$\begin{aligned}&\Pr(ActualLocation = r | Observation = r')\\ =&\Pr(A = r|O = r') = \frac{K(r)(r')\pi(r)}{\sum_{r''\in\mathcal{R}} \exp(K(r'')(r'))\pi(r'')} \quad (3)\end{aligned}$$

二つ目の等号ではベイズの定理を用いた. $ActualLocation$ はユーザの実際の位置, $Observation$ は雑音を加えた位置である. また, $\pi(r)$ はユーザの事前分布であり, 例えば人口密度などの情報を利用することができるが, 本論文では簡単のためユーザの分布は一様であると仮定する. この時, 式 (3) は以下のようになる.

$$\Pr(A = r|O = r') = \frac{K(r)(r')}{\sum_{r''\in\mathcal{R}} \exp(K(r'')(r'))}$$

この定義を踏まえ, GridE や WGridE において発生する問題について述べる. この問題は, 地球上に海が存在することや実用上ある有限の領域を考える必要があることで生じる地図の非対称性に起因する問題であり, 同じプライバシパラメータの値 ϵ でも, 地図の “端”に近い領域にいるユーザは他の領域にいるユーザよりもプライバシの保護度合いが低下してしまうというものである. そのことを示すために以下のような実験を行った. 縦約 115.6m, 横 141.5m（これは東京における 8 分の 1 地域メッシュの縦と横の真の長さに相当する）の領域が行方向 15 個, 列方向 15 個あるような長方形の地図を考える. 図 1(a) は各領域において $GridE$ を適用した際の入力と同じ領域が出力となる確率の最大値 $\max_{r\in\mathcal{R}} \Pr(O = r|A = r)$ と最小値 $\min_{r\in\mathcal{R}} \Pr(O = r|A = r)$ を ϵ を $\epsilon = 0.01, 0.02, \cdots, 0.09$ を設定して算出した結果である. また, 図 1(b) は各領域において $GridE$ を適用した際の各領域が出力となったときに, その領域がユーザの正確な位置であるという事後確率を事前確率が一様分布であると仮定したの最大値 $\max_{r\in\mathcal{R}} \Pr(A = r|O = r)$ と最小値 $\min_{r\in\mathcal{R}} \Pr(A = r|O = r)$ を $\epsilon = 0.01, 0.02, \cdots, 0.09$ に設定して算出した結果である. 図 1(a), 図 1(b) ともに橙色の線が最大値を, 青色の線が最小値を表しており, 最大値を取る領域は長方形の地図の四隅の領域であった. 図 1(a) と図 1(b) ではほとんどの ϵ で領域によって, “入力と同じ確率が出力される確率”と “事後確率”が異なっている. 実際, 図 1(a) では $\epsilon = 0.02$ のとき最大値と最小値には約 0.22 の差がある. 図 1(b) では $\epsilon = 0.02$ 付近では約 0.3 の差がある. また, 地図の “端”以外のほとんどの領域では事後確率は最小値に近い値をとるこ. これらの結果は, 地図の “端”の近くでは $GridE$ の入力と出力が同じとなる確率が高くなることにより, 観測された領域がそのままユーザの実際の位置である事後確率が高くなり, この観点からプライバシ保護の度合いが低下していることを示している.

Algorithm 1 重み削減手法

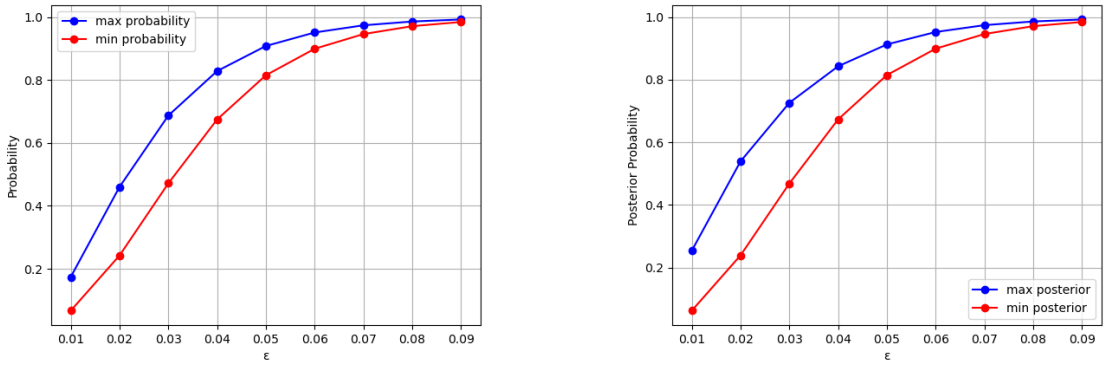
Input: 全領域のリスト $\mathcal{R}$, 重みのリスト W, 目的関数 f, 制約関数 c, 選択基準 S, ステップ数 $step$
Output: 最適化された重みの集合 W

```
obj_pre ⇐ f(R, W)
while True do
    W_0 ⇐ W
    R' ⇐ S に基づいて R 内の同じ優先順位の領域の集合を優先度順に格納したリスト
    for r_set in R' do
        W' ⇐ W
        for r in r_set do
            W'[r] ⇐ W[r] − step
        end for
        obj_post ⇐ f(R, W'), cons ⇐ c(R, W')
        if obj_post − obj_pre < 0 and cons then
            W ⇐ W'
            obj_pre ⇐ obj_post
            break
        end if
    end for
    if W = W_0 then
        break
    end if
end while
return W
```

この実験ではグリッドの地図の四隅の領域と, 確率が最小となる領域でのプライバシ保護の度合いの違いを示したが, 一般には $WGridE$ で重みが 1 未満であるような領域が近隣に存在するとプライバシが低下することが示される. 例えば, 実験で示したように便宜的に範囲を区切って取り出した地図の “端”に近い領域や, 沿岸部や離島などに相当する領域で発生する. この問題は地図内の領域と地図外の領域の “境界”や海と陸の “境界”などでプライバシの保護度合いが低下することから, 総称して “境界脆弱性問題”と呼ぶことにする. 本質的には, 境界脆弱性問題は “端”や重みが小さい領域の近くでは, 差分プライバシで互いに識別不能にするための近隣の領域が少なく, その領域との距離が大きくなるような領域が相対的に多く存在するため, 指数メカニズムにおいてその領域自身に大きな確率が割り当てられてしまうことから生じると考えられる. 次節では以上のような問題を軽減するための手法を提案する.

5.2 境界脆弱性問題を軽減するための手法

この節では前述の境界脆弱性問題を軽減するための手法を提案する. ただし, 計算量や問題の複雑さの観点から, 最適な解を解析的に求めることは困難であるので最適な解を近似するための貪欲アルゴリズムを考える. これは 4.2 節で導入した, 海などを表現するために各領域に割り当てた重みを逐次的に減少させていく方法であり, 重み削減手法と呼ぶことにする. 領域の重みを減少させることによって, その領域が出力される確率が減少することによって, 領域間でのプライバシ保護の度合いに



(a) 同じ領域が出力される確率　　**(b)** 事後確率

図 1: 同じ領域が出力される確率と事後確率の最大値, 最小値

関する不平等性が緩和されることを期待する.

逐次的に領域の重みを減少させていき, 最適な重みの近似解を得るためのアルゴリズムを Algorithm1 に示す. このアルゴリズムは, 全領域のリスト $\mathcal{R}$, 各領域の重みのリスト W, 目的関数 f, 制約関数 c, 5 行目で重みを減少させる領域を選択する順番を決定する基準 S と, 1 回の更新での重みの減少量 $step$ を入力とし, 各領域の最適な重みの近似値を出力する. このアルゴリズムでは S での選択基準で同順位の領域は同時に重みを減少させるため, 3 行目で S に基づいて $\mathcal{R}$ を同じ順位となる領域の集合のリストに変換し, 優先度の順に並び替えている. 1 行目で更新前の $\mathcal{R}, W$ を用いて目的関数の初期値を算出する. 5 行目から 16 行目では S で同じ順位となった領域の重みを $step$ 分まとめて減少させ (7 から 9 行目), その $\mathcal{R}, W'$ で目的関数の値の計算と制約を満たすかの判定を行い, (10 行目), 目的関数が改善し, かつ制約を満足するならば更新を実行し 2 行目に戻る. 目的関数が改善しなければ変更は破棄し次の順位の領域について調べる. 17 行目から 19 行目では, もし, すべての領域について重みを減少させたときに目的関数が改善しなければアルゴリズムを停止させ, 21 行目で最終的な重みを出力する.

5.3 実　　験

この節では 5.2 節で述べた手法により実際に境界脆弱性問題が改善することを示す.

5.3.1 有　用　性

実際にアルゴリズムを適用する前に, 実験でも用いるデータの有用性の指標について述べる. 一般に, データのプライバシ保護の強さと有用性にはトレードオフの関係があるため, 重み削減手法を適用しプライバシ保護の度合いを高めた結果, 適用前と比較した有用性は低下することが予想される. 有用性の指標はそのデータの利用方法によって異なるが, ここでは最も単旬で汎用的に使用できる SQL (Serevie Quality Loss) という指標を用いる. SQL は以下のように定義される.

$$SQL(\pi, K, d_{\#}) = \sum_{r,r' \in \mathcal{R}} \pi(r) K(r)(r') d_{\#}(r, r')$$

本論文では π は一様分布を仮定しているので, $\forall r \in \mathcal{R}, \pi(r) = \frac{1}{|\mathcal{R}|}$ である. SQL はユーザが存在する領域とそれに雑音を加えた後の領域との距離の期待値として解釈され, 小さいほど近い領域が出力される確率が大きくなるため有用性は高くなる. また, ある一つの領域 r におけるデータの有用性として

$$SQL(\pi, K, d_{\#}, r) = \sum_{r' \in \mathcal{R}} \pi(r) K(r)(r') d_{\#}(r, r')$$

も考える. これはある一つの領域にいるユーザが受けるサービスや, ある領域のデータを利用したい組織から見たデータの有用性を表現しており, r にいるユーザのデータに雑音を加えた後の領域と r の距離の期待値として解釈される.

5.3.2 重み削減手法の適用結果

まずは重み削減手法を実際に適用する. 実験で使用する地図は 5.1 節と同様に縦約 115.6m, 横 141.5m の領域が, 行方向に 15 個, 列方向に 15 個あるようなグリッドとする. 重み削減手法で使用する目的関数 f_1 は以下のような関数とする.

$$f_1(\mathcal{R}, W) = \max_{r' \in \mathcal{R}} \Pr(A = r' | O = r') - \min_{r' \in \mathcal{R}} \Pr(A = r' | O = r') \tag{4}$$

直観的には, 関数 f_1 が減少するように領域の重みを減少させていくことにより領域間での事後確率の差が小さくなるため, 領域によるプライバシ保護に関する不平等性が軽減されることを期待した目的関数である. 一般には事後確率の最大値, 最小値として考慮すべきはメカニズムの出力として観測された領域のみであるとは限らないが. ユーザの事前分布が一様であるときには各領域に任意の重みを割り当てた地図に対して, 事後確率の最大値, 最小値を求める際には観測された領域のみを考えればよい.

命題 2. $\forall r \in \mathcal{R}$ に対してユーザの事前分布が一様分布である, すなわち $\pi(r) = \frac{1}{|\mathcal{R}|}$ のとき, 任意の $w_r \in [0, 1]$ を用いて $WGridE$ を適用すると,

$$\max_{r' \in \mathcal{R}} \Pr(A = r' | O = r) = \Pr(A = r | O = r) \tag{5}$$

が成り立つ.

スペースの関係上証明は省略するが, $WGridE$ の定義と三角

不等式を用いることで証明することができる．また, 今回の実験では制約関数は任意の入力 $\mathcal{R}, W$ について True を返す関数を用い, 重みの初期値は全領域で 1 であるとする．

このアルゴリズムは 5 行目で選択する領域の順番に大きく依存するが, この実験では Algorithm1 の 5 行目で選択する領域の優先順位を決定する基準 S は現時点での事後確率の大きい順とする．この選択基準は “端”に近い領域ほど重みが減少する回数が多いことから着想を得ている．

重み削減手法を目的関数 f_1 で実行した後の各領域の重みを用いて $WGridE$ を適用した際に各領域が出力されたときに, 事後確率の最大値 $\max_{r\in\mathcal{R}} \Pr(A = r|O = r)$ と最小値 $\min_{r\in\mathcal{R}} \Pr(A = r|O = r)$ を $\epsilon = 0.01, 0.02, \cdots, 0.09$ に設定して算出した結果が図 2(a) である．また, 重み削減手法適用前と適用後の SQL を算出した結果が図 2(b) である．緑色の線が手法適用後の事後確率の最大値であり, 赤色の線が適用後の事後確率の最小値である．青色の線と, 赤色の線に重なっており図 2(a) では観察できないが橙色の線は手法適用前の事後確率の最大値, 最小値である．手法適用前より適用後の方が事後確率の最大値が減少していることがわかる．特に, 適用前は 5.1 節で述べたように $\epsilon = 0.02$ で最大値と最小値で約 0.3 の差があったが, 適用後は約 0.18 に減少している．実は $\Pr(O = r|A = r)$ に関しては最大値と最小値の差は大きくなる．これは重みを減少させることによって最終的に重みが 0 に近い領域はメカニズムで出力される確率が減少することが原因である．しかし, 信頼できないサーバなどの攻撃者が実際に観測できるのはメカニズムの出力のみであるため, この確率の最大値と最小値の差が大きくなること自体にプライバシの観点から大きな問題はない．

図 2(b) は 5.3.1 で定義した, 地図全体と実験を行った全ての ϵ で重みが減少した地図の角の領域 (r_{ij} $(i = 1, j = 1)$) での SQL を算出した結果である．橙色の線と赤色の線がそれぞれ重み削減手法を適用する前の地図全体での SQL と地図の角の領域 r_{ij} での SQL を表している．また, 青色の線と緑色の線がそれぞれ手法適用後の地図全体での SQL と地図の角の領域 r_{ij} での SQL を表している．重みが減少した領域は一部であるので地図全体としてのデータの有用性の低下はわずかである．しかし, 重みを減少させた領域の SQL は手法の適用前と比較して有用性が低下しており, プライバシ保護と有用性はトレードオフの関係にあることがわかる．図 3(a) と図 3(b) は特に, 目的関数の改善量が大きかった $\epsilon = 0.02$ と $\epsilon = 0.03$ のときの手法適用後の各領域の重みを示している．全体的にもともと事後確率が大きい地図の角や端を中心に重みが減少しており, この傾向は $\epsilon = 0.01$ を除いた他の ϵ の値でも共通している．

以下, 重み削減手法について注意点を述べる．本実験ではすべての領域で重みの初期値は 1 であるとして重み削減手法をを適用したが, 重みが一般の場合にも初期値を変更するだけで適用できる．本論文のように事後確率を用いた目的関数を用いて重み削減手法を適用すると計算量が大きくなってしまう．これは, アルゴリズム内で重みの更新や領域の結合の度に全領域について事後確率を計算しなおす必要があることが原因である．ただし, 実用上は並列計算を行うことによる実行時間を削減できるうえに, 地図の変更はそれほど頻繁に行うことはないため大きな問題にはならないと考えている．

6 おわりに

本研究では, 個人がデータの収集者に対してプライバシを保護した上で領域ごとの位置情報を提供するための枠組みを提案し, その過程で地図の非対称性に起因して生じる境界脆弱性問題を軽減する手法を提案した．具体的には, グリッドに関するプライバシ保護のために $d_{\mathcal{X}}$-privacy の距離の指標として領域の中心間のユークリッド距離を用い, $d_{\mathcal{X}}$-privacy における指数メカニズムを Grid-Exponential-Mechanism という形で適用した．さらに海上などの人が存在するとは考えにくい場所に出力されてしまう問題を解決するために, 各領域に実際の地理情報を反映した重みを考慮した Weighted-Grid-Exponential-Mechanism を提案した．その上で, 実際の地理情報や地図の形状が原因で地図の “端”や重みが小さい領域の近傍でプライバシ保護が低下する境界脆弱性問題について述べ, この問題を軽減するための手法を提案し, 実際に問題が軽減することを実験により示した．

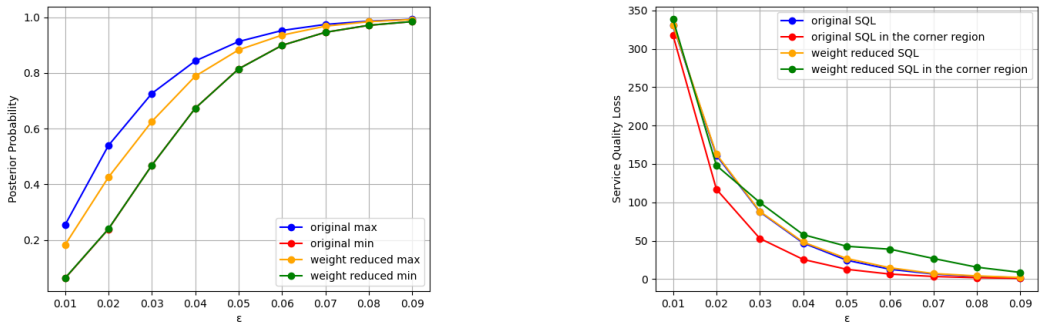
最後に, 今後の研究課題を三つ述べる．一つ目は, Takagi ら [3] の提案した道路ネットワーク上への拡張である．本論文では問題の簡単化のためにグリッド上の差分プライバシを考えたが, 道路ネットワークを考慮する方が端などの複雑な地形を詳細に表現できる．ただし, グリッドでの表現にもメリットは存在するのでデータの用途によって使い分けを行うのが良いであろう．二つ目は, プライバシ保護の度合いが個人ごとかつ位置ごとに異なると考えるのが妥当であるため, この点を考慮したプライバシ保護への拡張が考えられる．三つ目は, 単に地形だけではなく, その領域の人口のようなプライバシ保護に影響を与える情報を組み込むことであり, その上で, 実データを使用してカウンティングクエリなどの集約データを有用性の観点から議論することである．今後は以上のような課題について研究を進めていきたい．

文　献

[1] Stanley L Warner, Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

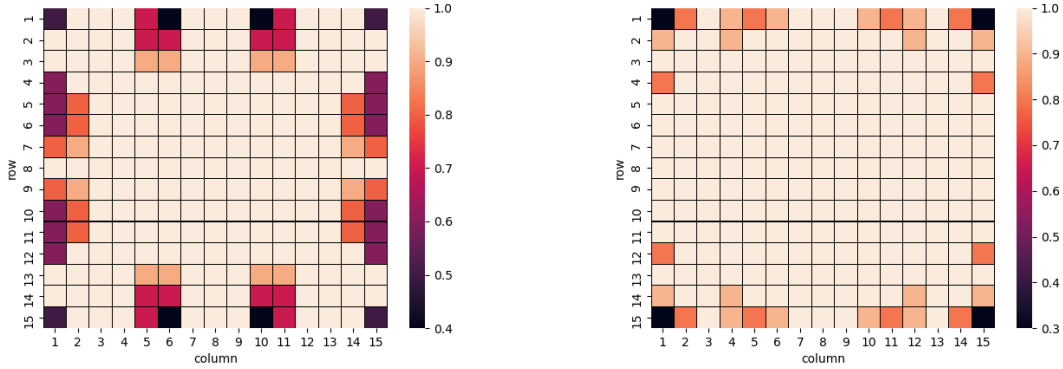
[2] Konstantinos Chatzikokolakis, Miguel E. Andrés, Nicolás Emilio Bordenabe, and Catuscia Palamidessi, Broadening the Scope of Differential Privacy Using Metrics. *Privacy Enhancing Technologies*, Vol. 7981, pp. 82–102. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[3] Shun Takagi, Yang Cao, Yasuhito Asano, and Masatoshi Yoshikawa, Geo-Graph-Indistinguishability:



(a) 事後確率　　**(b)** 地図全体と地図の角の領域の SQL

図 2: 重み削減手法の適用前後の事後確率と SQL



(a) $\epsilon = 0.02$ のときの各領域の重み　　**(b)** $\epsilon = 0.03$ のときの各領域の重み

図 3: Algorithm1 適用後の各領域の重み

Protecting Location Privacy for LBS over Road Networks. In Simon N. Foley, editor, *Data and Applications Security and Privacy XXXIII*, Vol. 11559, pp. 143–163. Springer International Publishing, Cham, 2019.

[4] Cynthia Dwork, Differential Privacy: A Survey of Results. In Manindra Agrawal, Dingzhu Du, Zhenhua Duan, and Angsheng Li, editors, *Theory and Applications of Models of Computation*, Vol. 4978, pp. 1–19. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[5] F. McSherry and K. Talwar, Mechanism Design via Differential Privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS' 07)*, pp. 94–103, October 2007.

[6] 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 - CCS ' 13*, pp. 901–914, Berlin, Germany, 2013. ACM Press.

[7] Kasiviswanathan, S. P., Lee, H. K., Nissim, K., Raskhodnikova, S. and Smith, A.: What Can We Learn Privately?, *arXiv:0803.0924 [cs]* (2010).

[8] Duchi, J. C., Jordan, M. I. andWainwright, M. J.: Local Privacy and Statistical Minimax Rates, *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pp. 429–438 (2013).

[9] Chatzikokolakis, K., Palamidessi, C. and Stronati, M.: Constructing Elastic Distinguishability Metrics for Location Privacy, *Proceedings on Privacy Enhancing Technologies*, Vol. 2015, No. 2, pp. 156–170 (2015).

[10] Chen, R., Li, H., Qin, A. K., Kasiviswanathan, S. P. and Jin, H.: Private Spatial Data Aggregation in the Local Setting, *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, Helsinki, Finland, IEEE, pp. 289–300 (2016).

[11] Gu, X., Li, M., Xiong, L. and Cao, Y.: Providing Input-Discriminative Protection for Local Differential Privacy, *arXiv:1911.01402 [cs]* (2020).