3W-03

参加型センシングにおけるユーザーの位置情報特定を防ぐためのプライバシー保護手法の提案

鈴木　涼　　早川　知道　　伊藤　孝行
名古屋工業大学大学院 工学研究科

1　はじめに

近年，多数のセンサを搭載しているスマートフォンが普及したことにより，参加型センシングが注目されている．参加型センシングとは，複数のユーザーがセンサデバイスから得られるセンサ情報を共有し有効活用する，クラウドソーシングの一種である．一般ユーザーの持つスマートフォンをセンサデバイスとして使用することで，安価に参加型センシングを行うことができる．

しかし，参加型センシングが広まる中でプライバシーの問題が発生している．参加型センシングによって収集するセンサ情報には，プライバシー情報が含まれている．ユーザーがプライバシー侵害の不安を感じ，センサ情報の提供を控えてしまう恐れがある．本研究では参加型センシングに対して適用できるプライバシー保護処理の手法を提案する．

本稿の構成を次に示す．2 章でプライバシー保護の既存手法について述べる．3 章でプライバシー保護の提案手法について述べる．4 章で評価実験，および考察について述べ，最後にまとめとする．

2　プライバシー保護の既存手法

プライバシー保護の既存手法として、Randomized Response[1] がある．Randomized Response では，あらかじめ提示されたいくつかの選択肢の中から一つの選択肢を選び回答するセンシングに対して適用できるプライバシー保護手法である．

Randomized Response におけるセンサデータのプライバシー保護処理の手順は，次の通りである．

1. センサデバイスがセンサデータを取得する
2. センサデバイスがセンサデータに対しプライバシー保護処理を行い，データをサーバへ送信する
3. サーバは，センサデバイスから受け取ったデータを元にセンサデータの統計情報の再構築を行う

手順 1 において，センサデバイスは選択肢の総数 α の中から一つの選択肢を観測する．手順 2 において，センサデバイスは確率 p で得られたセンサデータを真のままサーバへ送信する．また，得られたセンサデータとは異なる $\alpha-1$ 個の選択肢をそれぞれ確率 $\frac{1-p}{\alpha-1}$ で選択し，サーバへ送信する．センサデバイスが収集したデータ分布を **A**，サーバが受け取るプライバシー保護処理を施した後のデータ分布を **Y** として，手順 2 のプライバシー保護処理は式 (1)，および式 (2) に示す通りに表すことができる．

$$\mathbf{Y} = \mathbf{AM} \tag{1}$$

$$\mathbf{M} = \begin{pmatrix} p & \frac{1-p}{\alpha-1} & \cdots & \frac{1-p}{\alpha-1} \\ \frac{1-p}{\alpha-1} & p & \frac{1-p}{\alpha-1} & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1-p}{\alpha-1} & \cdots & \cdots & p \end{pmatrix} \tag{2}$$

手順 3 において，サーバは式 (3) によって，センサデバイスが収集したデータ分布 **A** を得ることができる．

$$\mathbf{A} = \mathbf{YM}^{-1} \tag{3}$$

3　プライバシー保護の提案手法

Randomized Response は，センシング対象の選択肢の数が多い場合，得られたセンサデバイスが収集したデータ分布の正確な再構築が困難となる．また，Randomized Response は，データ分布の正確な再構築を行うために大量の通知データを必要とする．提案手法では Randomized Response を改良し，1 回のセンサデバイスによるセンサデータ取得につき，複数個の選択肢の通知をサーバに対し行う．複数個の選択肢の通知をサーバに対し行うことで，元のデータの分布に基づいて通知データ数を増加させることができ，センサデバイスが収集したデータ分布をより高い精度で得ることができる．

提案手法において，センサデバイスがセンサデータを取得したとき，センサデバイスは確率 p で得られたセンサデータを真のまま選択する．また，得られたセンサデータとは異なる $\alpha-1$ 個の選択肢をそれぞれ確率 $\frac{1-p}{\alpha-1}$ で選択する．前述の選択を k 回繰り返し，k 個の選択肢の組をサーバへ通知する．サーバは式 (4) によって，センサデバイスが収集したデータ分布 **A** を得ることができる．

$$\mathbf{A} = \frac{1}{k}\mathbf{YM}^{-1} \tag{4}$$

4　プライバシー保護手法の評価実験

4.1　実験概要

参加型センシングによって得られたセンシングデータに対してプライバシー保護処理を施し，本提案手法の有用性を示す．プライバシー保護の対象とするデータとして，国土交通省中部地方整備局庄内川河川事務所との受託研究によって開発している河川管理システムによって得られた見回りデータを使用する．河川管理システムでは，管理者が設定した複数地点の見回りを一般市民の方々に依頼する．見回りを行った一般市民は，見回りを行った地点をサーバへ報告する．見回りを行った地点のデータに対し，プライバシー保護手法を適用する．プライバシー保護手法の評価尺度には，Jensen-Shannon Divergence[2]，および推測可能率を使用する．真の値のデータへの復元精度の観点における

Preserving User Location Privacy for Participatory Sensing
Ryo SUZUKI　Tomomichi HAYAKAWA　Takayuki ITO
Nagoya Institute of Technology

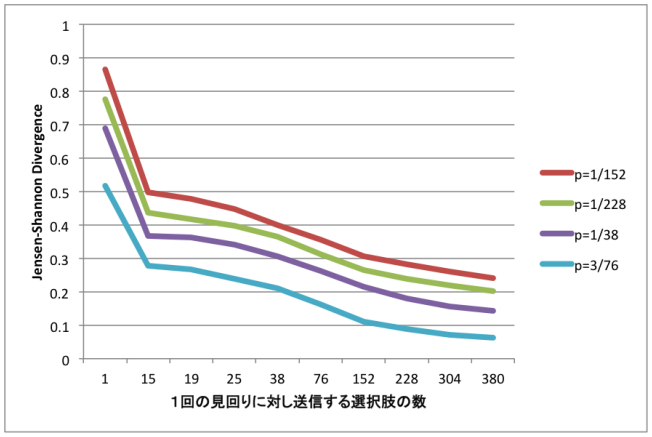

図 1: Jensen-Shannon Divergence の値

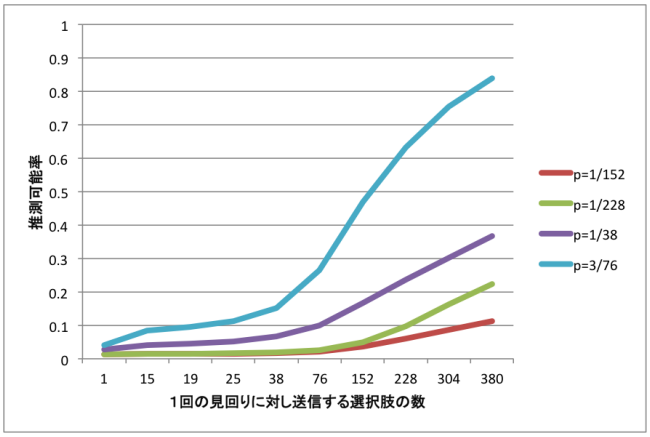

図 2: 推測可能率の値

評価尺度として，Jensen-Shannon Divergence を使用する．プライバシー保護強度の観点における評価尺度として，推測可能率を使用する．

推測可能率は，プライバシー保護処理を施したデータからプライバシー保護処理を施す前の真の値を推測できる割合である．推測可能率が小さい値をとるプライバシー保護手法は，プライバシー保護強度が高いことを表す．推測可能率は 0 以上，1 以下の値をとる．得られたセンサデータを真のまま送信する確率を p とする．センサデバイスの取り得る選択肢の総数を α とする．プライバシー保護処理を施す前の真の値と推測される選択肢の数を n とする．プライバシー保護処理を施す前の真の値と推測される選択肢の中にプライバシー保護処理を施す前の真の値が含まれる場合，推測可能率は $\frac{1}{n}$ であるとする．プライバシー保護処理を施す前の真の値と推測される選択肢の中にプライバシー保護処理を施す前の真の値が含まれない場合，推測可能率は 0 であるとする．

プライバシー保護の対象とするデータは，地域住民，および河川管理者等が見回りを行い，サーバへ通知を行った見回りの履歴情報である．対象期間は 2015 年 9 月 29 日から 10 月 25 日までである．見回り地点数は 76 地点である．見回り履歴の件数は 1496 件であった．見回りを行った地点を真のまま送信する確率は，見回り地点数 α を考慮し，$\frac{1}{152}(\frac{1}{2\alpha})$，$\frac{1}{228}(\frac{1}{3\alpha})$，$\frac{1}{38}(\frac{2}{\alpha})$，および $\frac{3}{76}(\frac{3}{\alpha})$ の 4 種類とした．1 回の見回りに対し送信する選択肢の数は，見回り地点数 α を考慮し，1，$15(\frac{1}{5}\alpha)$，$19(\frac{1}{4}\alpha)$，$25(\frac{1}{3}\alpha)$，$38(\frac{1}{2}\alpha)$，$76(\alpha)$，$152(2\alpha)$，$228(3\alpha)$，$304(4\alpha)$，および $380(5\alpha)$ の 1 0 種類とした．

プライバシー保護の既存手法は，1 回の見回りに対し送信する選択肢の数を 1 とする手法である．本研究によるプライバシー保護の提案手法では，1 回の見回りに対し送信する選択肢の数を 15，19，25，38，76，152，228，304，および 380 とした．各パラメータ毎に 100 回の実験を行い，Jensen-Shannon Divergence，および推測可能率の値の 100 回の平均を求める．

4.2 実験結果

実験結果を図 1，および図 2 に示す．図 1 は，Jensen-Shannon Divergence の値のグラフである．図 2 は，推測可能率の値のグラフである．

1 回の見回りに対し送信する選択肢の数が高いほど，Jensen-Shannon Divergence の値は減少した．既存手法と比較して，提案手法では真の値のデータへの復元精度が向上した．1 回の見回りに対し送信する選択肢の数が高いほど，推測可能率の値は増加した．既存手法と比較して，提案手法ではプライバシー保護強度が低下した．

真の値のデータへの復元精度，およびプライバシー保護強度を考慮した最適なパラメータについて考察する．Jensen-Shannon Divergence の値，および推測可能率を，平均が 0，標準偏差が 1 となるよう正規化を行う．正規化を行った後の Jensen-Shannon Divergence の値，および推測可能率の和が最小となるパラメータが，真の値のデータへの復元精度，およびプライバシー保護強度を考慮した最適なパラメータといえる．真の値のデータへの復元精度，およびプライバシー保護強度を考慮した最適なパラメータは，見回りを行った地点を真のまま送信する確率が $\frac{1}{228}(\frac{1}{3\alpha})$，1 回の見回りに対し送信する選択肢の数が $152(2\alpha)$ のときであった．従って，真の値のデータへの復元精度，およびプライバシー保護強度の双方を考慮したとき，プライバシー保護の既存手法と比較して，本研究によるプライバシー保護の提案手法は有効である．

5 まとめ

真の値のデータへの復元精度，およびプライバシー保護強度の双方を考慮したとき，プライバシー保護の既存手法と比較して，本研究によるプライバシー保護の提案手法は有効であることを確認した．本提案手法では，汎用的なプライバシー保護手法を拡張したものであり，参加型センシングにおける特徴を取り入れることは行っていない．今後は，見回り地点間の距離等の，参加型センシング内に現れるパラメータをプライバシー保護手法に組み込むことを考えていきたい．

参考文献

[1] S. L. Warner, ”Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias”, Journal of the American Statistical Association, vol. 60, no. 309, p.63-69, 1965.

[2] J. C. Angulo, J. Antolin, S. Lopez-Rosa, and R. O. Esquivel, ”Jensen-Shannon divergence in conjugate spaces: The entropy excess of atomic systems and sets with respect to their constituents”, Elsevier Physica A, vol. 389, p899-907, 2010.