

Perturbationを利用した参加型センシングにおける 復元精度向上のためのインセンティブ付与手法の検討

鈴木 孝男^{1,a)} 伊藤 昌毅^{1,b)} 瀬崎 薫^{1,2,c)}

概要: スマートフォンに代表される, 様々なセンサを搭載したモバイルデバイスが普及したことにより, 一般ユーザーがセンシングに参加する参加型センシングに注目が集まっている. 参加型センシングの問題として, ユーザーのプライバシーが流出するリスクがあること, 十分なユーザー確保が困難であることが挙げられる. これら問題を解決するために過去多くの研究がユーザーのプライバシー保護, ユーザーへのインセンティブ付与について取り組んできたが, これら問題を同時に扱った研究は少ない. 特に Perturbation 技術を用いたプライバシー保護を行った場合のインセンティブ付与に取り組んだ研究は筆者の知る限り存在しない. 本研究では Perturbation 技術によってプライバシー保護が施されたデータを集計した後再構築して得られる時空間データと, 既知の時空間データとの適合度を求め, 満たすべき適合度との乖離に応じてインセンティブの確保を行うことで再構築の復元精度を向上させる手法を提案する.

キーワード: 参加型センシング, プライバシー保護, インセンティブ付与, Perturbation, 時空間データ

Study on Incentive Method to Improve Accuracy of Reconstruction After Perturbation in Participatory Sensing

Abstract: The proliferation of mobile devices equipped with various sensors made a strong focus on participatory sensing. Participatory sensing has the problems of users' privacy leakage and inadequate participants. Though many researches aiming at preserving users' privacy and incentivizing users' participation were conducted, a few dealing with these problems simultaneously were done. In particular, to the best of my knowledge, there is no incentive method when users' privacy are preserved with perturbation. We propose an incentive mechanism that we calculate goodness-of-fit between a reconstructed spatio-temporal data and a known model of this data, and decide an amount of incentive according to the difference between desired goodness-of-fit and this calculated value in order to improve accuracy of reconstruction.

Keywords: Participatory Sensing, Privacy Preserving, Incentivizing, Perturbation, Spatio-temporal Data

1. はじめに

スマートフォンに代表される, 様々なセンサを搭載したモバイルデバイスが普及したことにより, 一般ユーザーがセンシングに参加する参加型センシングに注目が集まっている [1]. ユーザーがモバイルデバイスを持ち歩き, 周囲の

情報をセンシングすることで, 既存の固定型センサを用いるセンシング方法よりも, 低コストかつ高粒度なセンシングが可能となる [2]. しかし参加型センシングの問題として, プライバシーの流出, 及び必要十分な参加者の確保が挙げられる.

まずプライバシーの流出については, モバイルデバイスから送信されるセンサ情報, 例えば GPS から得られる位置情報などの単体のセンサ値, もしくは複数のセンサ値を組み合わせて, センサ値以外にも一般向けに公開されている情報を用いることでユーザーのプライバシーが流出してしまうことが分かっている [3]. プライバシー保護手法はこれまで多数提案されており, データの送信元を匿名化

¹ 東京大学 生産技術研究所
Institute of Industrial Science, The University of Tokyo

² 東京大学 空間情報科学研究センター
Center for Spatial Information Science,
The University of Tokyo

a) t-suzuki@mcl.iis.u-tokyo.ac.jp

b) mito@iis.u-tokyo.ac.jp

c) sezaki@iis.u-tokyo.ac.jp

することでデータとその送信元ユーザーとの結びつけを防ぐ手法や [4], [5], [6], データの粒度を変える手法 [7], データにノイズを乗せる Perturbation 技術を用いたプライバシー保護手法 [8], [9] 等が提案されている. Perturbation 技術を用いたプライバシー保護手法 [8], [9] は, [4], [5], [6] のようなデータの送信元を匿名化することでプライバシーを保護する手法と比べて, 匿名化に必要な複数のサーバーや, データの交換のための複数のユーザーのような, 信頼できる第三者機関 (TTP) が必要ないという点で優れている [8]. またノイズ付加後に送信するデータの値は元のデータと異なるので, データの値からプライバシーが流出するリスクが低いという点でも優れている [8].

次に必要十分な参加ユーザーの確保については, センシングに参加する労力だけでなく, センシング時のバッテリー消費やセンシングによるプライバシーの流出などがユーザーの参加を促す障害となり, センシングに必要なユーザー数を確保するためにインセンティブ付与を行う必要がある [10]. 参加型センシングへの参加を促すインセンティブ付与に関する研究はこれまで多数なされてきており, オークション理論を用いたインセンティブ付与手法や [10], [11], データの質や量, インセンティブの制限など各種パラメータを考慮した手法 [12], [13], ユーザー同士で競わせるゲーミフィケーションを用いた手法 [14] などが提案されている.

このように参加型センシングにおけるプライバシー保護手法やインセンティブ付与手法はこれまで多くの研究がなされてきたが, これらを同時に取り扱った研究は数が少なく, [5] のようにデータの送信元を匿名化した場合のインセンティブ付与手法が知られている程度である. TTP が不要である点, データ自体からプライバシーが流出するリスクが低いという点で匿名化によるプライバシー保護手法よりも優れている Perturbation 技術は, データ数が復元精度に直結するため, 所望の復元精度を満たすためのデータ数を確保するためにインセンティブ付与が必要となるが, このようなインセンティブ付与手法は筆者の知る限り存在せず, Perturbation 技術を用いたプライバシー保護手法を実際に適用することが困難であると考えられる. そこで本研究では Perturbation 技術を用いた場合に所望の復元精度を満たすためのインセンティブ付与手法の実現を目指す.

本稿では 2 章において Perturbation 技術について述べ, Perturbation 技術に対するインセンティブ付与を考える際に解決すべき問題について 3 章で述べる. さらに 4 章にて 3 章で挙げた問題の 1 つを解決するために測定対象のモデルを利用するという方法を提案する. 最後に 5 章でまとめと展望を述べる.

2. Perturbation 技術を用いたプライバシー保護手法

本項ではプライバシー保護手法, 特に Perturbation 技術を用いた手法について関連研究を紹介する. センサデータに対してノイズを乗せることでユーザーのプライバシーを保護する試みは多数なされてきたが, データに乗せるノイズによっては元のデータが推定されてしまうことを問題とし, R.K.Ganti らはセンシングデータに相関のあるノイズを付加することによって, PCA やフィルタリング攻撃によるプライバシー流出を防ぐ PoolView というシステムを提案した [8]. 例として体重を管理するアプリケーションを取り上げ, 体重変化のモデル式を元にしてノイズを作成し, 体重データに付加することによって実際に PCA やフィルタリング攻撃によってデータが推定されないことを示した.

他の Perturbation 技術に Negative Surveys がある. Negative Surveys とは調査方法の 1 つであり, 複数選択肢がある質問に対して回答者は常に真でない選択肢を当確率で選んで回答し, 集計者は得られた回答を再構築することで全体的なデータ分布を得る手法のことである. この Negative Surveys はセンサデータを離散化してカテゴリデータとすることで参加型センシングのプライバシー保護手法として適応できるが, 1 次元データ, つまりセンシングできる対象が 1 つのみとなってしまう, 例えば気温情報と湿度情報といったような複数次元のデータに適応することができない. そこで M. M. Groat らは Negative Surveys を多次元に拡張した Multidimensional Negative Surveys を提案した [9]. 送信側はいずれの次元においても真の値を送らないことでプライバシーを保護する.

3. 解決すべき問題

本項では Perturbation 技術に対するインセンティブ付与を考える際に解決すべき問題について述べる.

3.1 復元精度の推定

所望の復元精度を満たすためにインセンティブ付与を行うには, 現在の復元精度を知る必要があるが, Perturbation 技術を用いた場合, 元々の正しいデータの分布が分からないため, 復元処理を行った後得られる分布が実際のデータ分布をどの程度復元できているか分からない. 関連研究で述べた PoolView の場合, 復元精度を元のデータ分布と復元後のデータ分布と比較することで求めているが, 現実の場面では元のデータ分布は知り得ない. さらに Multidimensional Negative Surveys において復元精度の指標は元のデータの分布を必要としないが, 全カテゴリで 1 つの評価指標となっているため, 例えばある場所の環境情報についてなど, カテゴリ毎の復元精度が分からず, 復元精度の指標として

不十分であると考えられる。

3.2 復元精度と必要データ数の対応付け

現在の復元精度が推定できた場合、その復元精度と、所望の復元精度を満たすために必要なデータ数の対応付けをする必要がある。これは用いる Perturbation 技術によって変わると考えられる。例えば、PoolView の例では数百というデータ数で十分復元できる一方で、多次元データに対してプライバシー保護する Negative Surveys ではデータのカテゴリ数や次元数に依存するが、数万、数十万オーダーのデータが必要となると考えられる。

3.3 インセンティブ付与手法の提案

必要なデータ数がわかった場合にどのようなインセンティブ付与手法を用いてそのデータ数を確保するかという問題である。用いる Perturbation 技術によっては、数万、数十万というオーダーのデータが必要となるので、効率的にこれらデータ数を確保する手法が必要となると考えられる。

4. 提案と解析

3.1 節で述べた問題の解決法として、測定対象のモデルを利用するという方法を提案する。Perturbation 技術を用いた場合、元のデータ分布を知ることにはできないが、多くの参加型センシングで測定されている空間情報は時間的、空間的な相関をもつあるモデルに従うことも多いと考えられる。このような場合、あらかじめ測定対象のモデルが既知であれば、そのモデルと再構築後のデータを比較することで Perturbation 技術を用いた場合でも復元精度を推定できると考えられる。

そこで 3.1 節で述べた問題の解決に先立ち、実際の参加型センシングのアプリケーションで測定対象となっている時空間データがどのようなモデルに従っているかを解析した。対象として一酸化炭素を取り上げる。一酸化炭素は大気汚染の原因の 1 つであり、人々の健康に影響を与える。このことから都市空間の大気汚染をセンシングする必要があるとして、実際の参加型センシングにおいても測定対象としているアプリケーションが存在している [15], [16]。今回はアメリカ海洋大気局 (NOAA) の Earth System Research Laboratory (ESRL) が公開しているアメリカ大陸の各地点でセンシングした一酸化炭素データを解析した。解析対象として選んだ一酸化炭素データの測定地点、測定地点名、及び測定高度の詳細を図 1 に示す。解析したデータの測定時期は 2010 年の 1 月から 12 月であり、各月、各場所のデータ数を表 1 に示した。1 日のデータ数の平均は 76 個である。またデータの概要を示すために、各地点の年間のデータの確率密度曲線を図 2 に示した。

参加型センシングのアプリケーションを想定する場合



図 1 一酸化炭素の測定地点、測定高度

Fig. 1 The locations and height above ground level of carbon monoxide measurement

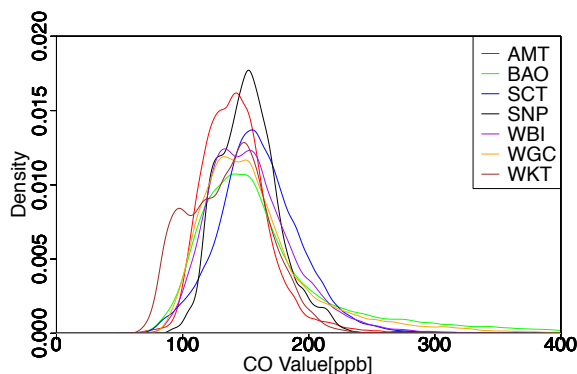


図 2 各地点における年間データの確率密度曲線

Fig. 2 The probability density curve of yearly data in all locations

に、センシングのサイクルを決めて、そのサイクルごとにデータを集計し分布を知るという状況が考えられる。このような状況のもと、センシングサイクルを 1 日として 1 日毎の一酸化炭素データの分布がどのような確率分布に従うかを調べた。その際、考えられる母集団の確率分布として、図 2 及び日毎の分布の様子から、正規分布、対数正規分布、ガンマ分布を選択した。まず 1 日のデータのサンプルに対して、上記で述べた 3 つの確率分布に当てはめる場合の最適なパラメータを最尤推定によって求め、母集団がそのパラメータの確率分布に従うとして、母集団の確率分布とサンプルの確率分布が一致するかどうかの検定をコルモゴロフ・スミルノフ検定を用いて行う。これら操作には R を使い、最尤推定は MASS パッケージの fitdistr 関数、コルモゴロフ・スミルノフ検定は ks.test 関数によって行った。地域毎に 1 年間この処理を行い、指定した母集団の確率分布に従うと判定された日数を集計し、それを 1 年間の日数で割った結果を表 2 に示す。

この結果より 3 つの分布いずれの場合に当てはめても、各地点とも 1 年のうち 70% 以上の日数が各確率分布に従うと分かる。さらにその中でも特に対数正規分布は 78.0% と、他の 2 つの分布よりも適合していることがわかった。今回検定を行ったデータは図 1 を見て分かるように測定地点が

表 1 各地点, 各月のデータ数

Table 1 The number of monthly data in all locations

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	total
AMT	2022	1878	2246	1834	2449	2398	2359	2358	2239	2278	1975	2220	26256
BAO	2479	2232	2234	2230	2179	2393	2480	2480	1995	1983	2154	2476	27315
SCT	2478	1210	2479	2400	2480	1781	2479	2457	2278	2480	2193	2480	27195
SNP	1835	2156	2429	2394	2198	147	2400	1745	1960	1999	2399	2223	23885
WBI	2480	2239	2480	2338	2480	2200	2186	2320	1959	2394	2388	2438	27902
WGC	2121	1848	1801	2053	2174	2129	2201	1993	1238	1190	2022	2047	22817
WKT	2480	2218	2496	2086	2359	2399	2473	2449	2132	2477	2400	2360	28329

表 2 各地点の 1 日のサンプルが各確率分布に従う割合

Table 2 The rate in which daily samples follow normal, log-normal, gamma distribution in all locations

	正規分布 [%]	対数正規分布 [%]	ガンマ分布 [%]
AMT	90.6(308/340)	92.4(314/340)	92.4(314/340)
BAO	56.3(198/352)	69.3(244/352)	65.3(230/352)
SCT	66.3(228/344)	71.8(247/344)	69.2(238/344)
SNP	69.3(212/306)	69.9(214/306)	70.6(216/306)
WBI	70.5(251/356)	75.8(270/356)	73.3(261/356)
WGC	72.4(257/355)	78.3(278/355)	76.9(273/355)
WKT	83.5(298/357)	87.7(313/357)	86.8(310/357)
全体	72.7(1752/2410)	78.0(1880/2410)	76.4(1842/2410)

空間的に散らばっており, また 1 年という長い期間であった。このような条件において 78.0% という高い確率で対数正規分布に従うため, 対数正規分布を一酸化炭素の一般的なモデルとしてよいと考えられる。

本章のデータ解析の結果より, 特定の測定対象を決め, その測定対象があるモデルに従う場合, そのモデルを利用し, 復元精度を推定できると考えられる。例えば参加型センシングの測定対象を一酸化炭素に決め, 復元後のデータ分布から最尤推定によってパラメータを求め, そのパラメータの対数正規分布と復元後データ分布の差を見ること, どの程度の復元精度なのかを推定できると考えられる。

5. まとめと展望

今後は測定対象が特定のモデルに従うことを前提として, 問題の解決を進めていく。問題解決の第一段階として, 4 章で解析対象とした一酸化炭素を測定する参加型センシングを想定し, 対数正規分布によって生成されたデータに対して Multidimensional Negative Surveys を施し復元するシミュレーションを行い, 予測した復元精度と, 実際の復元精度から, 予測した復元精度の正確さを調べることで, 既知のモデルを用いる提案の有用性を検証するつもりである。

参考文献

- [1] Burke, J. A., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S. and Srivastava, M. B.: Participatory sensing, *Center for Embedded Network Sensing* (2006).

- [2] Budde, M., El Masri, R., Riedel, T. and Beigl, M.: Enabling low-cost particulate matter measurement for participatory sensing scenarios, *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*, ACM, p. 19 (2013).
- [3] Christin, D., Reinhardt, A., Kanhere, S. S. and Hollick, M.: A survey on privacy in mobile participatory sensing applications, *Journal of Systems and Software*, Vol. 84, No. 11, pp. 1928–1946 (2011).
- [4] Wang, C.-J. and Ku, W.-S.: Anonymous sensory data collection approach for mobile participatory sensing, *Data Engineering Workshops (ICDEW)*, 2012 IEEE 28th International Conference on, IEEE, pp. 220–227 (2012).
- [5] Li, Q. and Cao, G.: Providing privacy-aware incentives for mobile sensing, *Pervasive Computing and Communications (PerCom)*, 2013 IEEE International Conference on, IEEE, pp. 76–84 (2013).
- [6] Shin, M., Cornelius, C., Peebles, D., Kapadia, A., Kotz, D. and Triandopoulos, N.: AnonySense: A system for anonymous opportunistic sensing, *Pervasive and Mobile Computing*, Vol. 7, No. 1, pp. 16–30 (2011).
- [7] Agir, B., Papaioannou, T. G., Narendula, R., Aberer, K. and Hubaux, J.-P.: User-side adaptive protection of location privacy in participatory sensing, *Geoinformatica*, Vol. 18, No. 1, pp. 165–191 (2014).
- [8] Ganti, R. K., Pham, N., Tsai, Y.-E. and Abdelzaher, T. F.: PoolView: stream privacy for grassroots participatory sensing, *Proceedings of the 6th ACM conference on Embedded network sensor systems*, ACM, pp. 281–294 (2008).
- [9] Groat, M. M., Edwards, B., Horey, J., He, W. and Forrest, S.: Enhancing privacy in participatory sensing applications with multidimensional data, *Pervasive Computing and Communications (PerCom)*, 2012 IEEE International Conference on, IEEE, pp. 144–152 (2012).
- [10] Lee, J.-S. and Hoh, B.: Dynamic pricing incentive for participatory sensing, *Pervasive and Mobile Computing*, Vol. 6, No. 6, pp. 693–708 (2010).
- [11] Koutsopoulos, I.: Optimal incentive-driven design of participatory sensing systems, *Infocom, 2013 proceedings IEEE*, IEEE, pp. 1402–1410 (2013).
- [12] Reddy, S., Estrin, D. and Srivastava, M.: Recruitment framework for participatory sensing data collections, *Pervasive Computing*, Springer, pp. 138–155 (2010).
- [13] Zhang, D., Xiong, H., Wang, L. and Chen, G.: CrowdRecruiter: selecting participants for piggyback crowdsensing under probabilistic coverage constraint, *Proceedings of the 2014 ACM International Joint Conference*

- on *Pervasive and Ubiquitous Computing*, ACM, pp. 703–714 (2014).
- [14] Kawajiri, R., Shimosaka, M. and Kahima, H.: Steered crowdsensing: incentive design towards quality-oriented place-centric crowdsensing, *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, pp. 691–701 (2014).
 - [15] Kanjo, E., Bacon, J., Roberts, D. and Landshoff, P.: MobSens: Making smart phones smarter, *Pervasive Computing, IEEE*, Vol. 8, No. 4, pp. 50–57 (2009).
 - [16] Sivaraman, V., Carrapetta, J., Hu, K. and Luxan, B. G.: Hazewatch: A participatory sensor system for monitoring air pollution in sydney, *Local Computer Networks Workshops (LCN Workshops), 2013 IEEE 38th Conference on*, IEEE, pp. 56–64 (2013).