

モバイル端末からのイベント情報検索における Geo-indistinguishability を用いたユーザ位置匿名化手法の検討

石神 京佳[†] 榎 美紀^{††} 小口 正人[†]

[†]お茶の水女子大学

^{††}日本アイ・ビー・エム株式会社

1 はじめに

近年、ソーシャルネットワーキングサービス（以下 SNS とする）の普及に伴い、SNS 上にはローカルイベントや開催中のイベントをはじめ、大小様々な規模のイベントに関する情報が投稿されるようになった。それらの膨大なイベント情報をサーバ側に収集し、ユーザがモバイル端末の位置情報を提供して、現在地付近のイベント情報を得るようなサービスは近年多く存在するが、位置情報などの個人情報をもそのままサーバに送信して情報分析に使用されることはプライバシー上の懸念がある。本稿では人流データ [1] を利用し、差分プライバシーを位置情報匿名化に拡張した技術 Geo-Indistinguishability（以下 Geo-I とする）[2] を満たすようなユーザ位置のダミー位置を算出し、サーバへの問い合わせに使用することで、よりプライバシー保護の指標に則った度合いで、ユーザの位置情報を曖昧化する手法を提案する。

2 提案システムの概要

本研究ではユーザの位置情報プライバシーを保護しつつ、膨大な SNS 上のイベント情報から、ユーザに適したイベント情報を推薦するシステムの構築を目指す。提案するシステムの概要を図 1 に示す。

サーバ側は先行研究である工藤ら [3] が提案する手法を参考に、パブリックなツイートデータからのイベント情報抽出・分

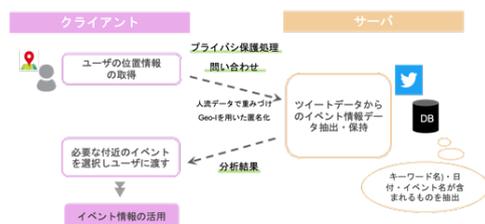


図 1: 提案システムの概要

析処理を行う。クライアント側はユーザの位置情報プライバシー保護処理をしてからサーバへイベント情報を問い合わせる。クライアントは問い合わせ結果で得られたイベント情報から、ユーザが必要としている範囲や個数などの制約を満たすイベント情報を抽出しユーザへのイベント情報推薦を行う。

2.1 地域メッシュ情報の付与

地域メッシュ [4] は、統計に利用するために緯度・経度に基づいて地域を隙間のない網の目のような区画としたものである。イベント開催地にスポット名または住所が保存されているイベント情報に対し、Geolocation API [5] を用いて開催地の緯度経度を取得し、地域メッシュコードを算出し保持しておく。本研究では、現在日本で用いられている、昭和 48 年 7 月 12 日行政管理庁告示第 143 号に基づく「標準地域メッシュ・コード」のうち一辺の長さ約 1km の第 3 次メッシュ、一辺の長さ約 500m の分割地域メッシュを使用する。

2.2 Geo-I を用いた位置情報匿名化

Geo-I は、Andrés ら [2] によって提案された、差分プライバシーにおけるデータを位置情報、ハミング距離をユークリッド距離に置き換え、位置情報データの保護に応用したプライバシー保護基準である。X をユーザの位置の集合、Z をメカニズム K によって曖昧化した結果得られる位置の集合としたとき、任意の二つの点 $x, x' \in X$ において以下が成り立てば、位置 $z \in Z$ を確率 k_{xz} で出力するメカニズム K は ϵ -Geo-I を満たすという。

$$k_{xz} \leq e^{\epsilon d_X(x, x')} k_{x'z}$$

[6] で提案した地域メッシュを用いたイベント情報問い合わせ手法では、第 3 次メッシュの中心位置をダミーの位置としてサーバへの問い合わせに使用した。この処理を行うことで、サーバに付与されるユーザの位置情報は第 3 次メッシュコードに限られ、約 1km 四方の区画に曖昧化されていたが、ユーザ位置に関わらずプライバシー保護度合いが固定的であり、攻撃者が周辺の人流や道路ネットワークを考慮した予測を行うと予測精度が高くなってしまいう課題があった。そこで本提案手法では、ユーザの位置情報を地域メッシュと人流データを利用し、Geo-I を満たすようなダミー位置を算出し、サーバへの問い合わせに使用することで、ユーザ位置の匿名化を保護度合い ϵ で行う。Geo-I を満たすメカニズム K は、ラプラスメカニズムをはじめ

A: Study of User Location Anonymization Method Using Geo-indistinguishability for Event Information Search from Mobile Devices

[†]Kyoka Ishigami ^{††}Miki Enoki [†]Masato Oguchi

[†]Ochanomizu University

^{††}IBM Japan, Ltd.

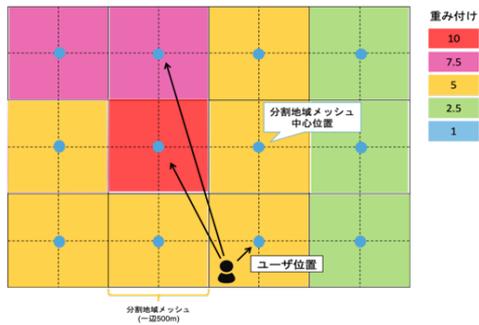


図 2: Geo-I を用いた位置情報匿名化の概要

様々なものが提案されているが、今回の実験では Bordenabe [7] が提案している事前確率 π_x を用いたメカニズム K を参考に実装を行う。分割地域メッシュ単位の人流データを用いた重み付けを π_x とし、ユーザ位置の匿名化を保護度合い ϵ で行う。求めた ϵ -Geo-I を満たす確率分布を満たすダミー位置を問い合わせの位置として使用する。今回の実験では図 2 に表すように、人流数が多い区画から少ない区画に 10 から 1 の重み付けを行ない、 ϵ の値と確率分布の関係を調査した。

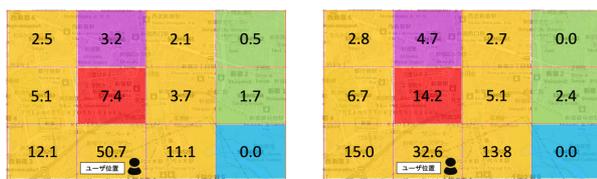
3 提案手法の評価

3.1 実験

例として新宿駅付近 500m 四方の分割地域メッシュ 12 区画に対し人流データを用いて重み付けを行い、問い合わせの位置として使用される ϵ -Geo-I を満たすダミー位置として各区画が選ばれる確率を調査する。ユーザ位置を固定しプライバシー保護度合い $\epsilon = 3.0$ としたときの結果 (図 3(a)) と、 $\epsilon = 2.5$ としたときの結果 (図 3(b)) を示す。各メッシュ区画が選ばれる確率を数値で記載している。

3.2 考察

ϵ はノイズ付きの位置から元の位置に関して漏洩する情報量を制御するパラメータであり、その値が小さいほどプライバシー保護度合いは高くなるが、ノイズは大きくなり、よりユーザ位置とは無関係な区画が選ばれる。図 3(a) $\epsilon = 3.0$ では元のユーザ位置がそのままダミー位置として選ばれる確率が 50.7% であるのに対し、図 3(b) $\epsilon = 2.5$ においては 32.6% となっていて低い。 ϵ が大きい値のときはよりユーザ位置からの距離が近い区画が選ばれ、 ϵ が小さい値のときはより人流数が多い区画が選ばれる。このような ϵ の値とユーザ位置における人流数の大



(a) $\epsilon = 3.0$ (b) $\epsilon = 2.5$

図 3: 各メッシュが選ばれる確率 (%)

小の傾向を利用し、適切な ϵ の値を調整することで、より人のいるところ、またはユーザ位置から近い場所を問い合わせの位置として選んで欲しいというユーザの要件に柔軟に対応することが可能となる。

4 まとめと今後の課題

ユーザ位置周辺という地理的な制約条件を満たすような、SNS データからのイベント情報の検索をする際の、ユーザの位置情報プライバシーを保護するためのデータ処理方法について検討した。地域メッシュと人流データを利用し Geo-I を満たすようなダミー位置を算出する処理を行った。 ϵ の値を調整することで、より人のいるところまたはユーザ位置から近い場所を問い合わせの位置として選んで欲しいというユーザの要件に柔軟に対応でき、従来の手法よりプライバシー保護の指標に則った度合いでプライバシー保護を行うことが可能となる。今回の実験では限定的な区画で実験を行い、年間の人流データを使用した。今後はより広い連続する範囲での実装、時間別や年齢別の人流データを使用し、より細かなユーザの要件に柔軟に対応し、適切な ϵ の値の設定を行いたい。

5 謝辞

本研究は一部、JST CREST JPMJCR22M2 の支援を受けたものである。

文献

- [1] Ministry of land, infrastructure, transport and tourism national open data on human flow. <https://www.geospatial.jp/ckan/dataset/mlit-1km-fromto>.
- [2] Miguel E. Andrés, Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Geo-indistinguishability: Differential privacy for location-based systems. *Proceedings of the 2013 ACM SIGSAC conference on Computer and Communications Security (CCS'13)*, pages 901–914, 2013.
- [3] Ruriko Kudo, Miki Enoki, Akihiro Nakao, Shu Yamamoto, Saneyasu Yamaguchi, and Masato Oguchi. Real-time event search corresponding to place and time using social stream. *the 3rd IEEE International Conference on Big Data Intelligence and Computing (DataCom2017)*, pages 1047–1053, 2017.
- [4] Ministry of internal affairs and communications information bureau regional mesh statistics overview. https://www.stat.go.jp/data/mesh/m_tuite.html.
- [5] Google maps geolocation api. <https://developers.google.com/maps/documentation/geolocation/overview>.
- [6] Kyoka Ishigami, Miki Enoki, and Masato Oguchi. Event information search method from sns data considering privacy of user's location information. pages 240–245, 2022.
- [7] Nicolás E Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Optimal geo-indistinguishable mechanisms for location privacy, 2014.