

# ユーザの位置情報プライバシーを考慮した SNS データからのイベント情報検索

理学専攻・情報科学コース 2240640 石神京佳 (指導教員: 小口正人)

## 1 はじめに

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、位置情報を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。そこで本研究では、ユーザの位置情報をダミー位置に変換し、サーバへの問い合わせに利用する。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として、メッシュへの近似手法、差分プライバシーを位置情報匿名化に拡張した技術である Geo-Indistinguishability を用いた手法を実際の区画に対し実装し、匿名度合いや有効性の比較を行う。SNS の一種である X (旧 Twitter) からイベント情報を取得し、ユーザの位置情報プライバシーを考慮して地理的な制約条件を満たしながら、大量の SNS データからイベント情報を検索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。提案するシステムモデルを図 1 に示す。サーバはパブリックなツイート

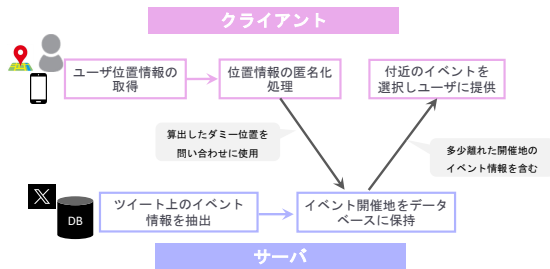


図 1: 提案システムモデル

データからのイベント情報抽出・データベースへの格納を行う。クライアント側はユーザの位置情報にプライバシー保護処理を施し、サーバへイベント情報の問い合わせを行う。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 ツイートデータからのイベント情報抽出

X(旧 Twitter) API を使用しキーワード(地名)を含むツイートを収集する。続いて、取得したツイートデータから、正規表現を用いて日付・イベント名・イベント開催地が含まれるツイートを抽出し、それらをイベント情報として重複を除きデータベースに格納する。

### 2.2 地域メッシュ情報の付与

地域メッシュは、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 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。

### 2.3 プライバシを考慮した問い合わせ

#### 2.3.1 提案手法 1 (メッシュへの近似手法)

ユーザの位置情報プライバシーを保護するために、ユーザ位置から算出した第 3 次メッシュの中心位置をダミー位置としてサーバへの問い合わせに使用する。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る。問い合わせの効率化を図るため、イベント密集度合いに応じて問い合わせ範囲を変更する。主要駅 5 箇所(新宿、渋谷、池袋、東京、品川)中の最寄りの駅との距離で以下のように場合を分け、図 2 に示した 3 種の範囲のうちから問い合わせ範囲を選択する。

500m 圏内に主要駅がある場合：

図 2 ①の範囲からイベント情報を検索(1km 四方)

1km 圏内に主要駅がある場合：

図 2 ②の範囲からイベント情報を検索(2km 四方)

1km 圏内に主要駅がない場合：

図 2 ③の範囲からイベント情報を検索(3km 四方)

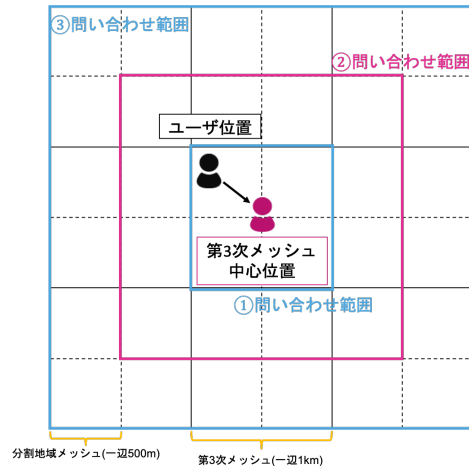


図 2: ダミー位置とユーザ位置の関係 (提案手法 1 メッシュへの近似手法)

#### 2.3.2 提案手法 2 (Geo-I を用いた位置匿名化手法)

Geo-I 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、ユーザ位置の匿名化を保護度合い  $\epsilon$  で行う。Geo-I を満たすメカニズム  $K$  として、Bordenabe [1] らが提案した、ユーザ位置が  $x \in X$  で

ある事前確率  $\pi_x$  を用いつつ、ユーザ位置  $x$  とダミー位置  $z$  のユークリッド距離  $d_Q(x, z)$  の期待値が最小になるように最適化を行う手法を採用する。携帯端末の位置情報データを集計した、単位時間あたりの分割地域メッシュ単位の滞在人口数データ [2] を用いた重みづけをユーザ位置の事前確率  $\pi_x$  とし、ユーザ位置の匿名化を保護度合い  $\epsilon$  で行う。求めた  $\epsilon$ -Geo-I を満たす確率分布に従ってダミー位置を算出し、問い合わせの位置として使用する。今回の実装では図 3 に表すように、分割地域メッシュ 25 区画に対し、滞在人口数リアルデータを事前確率  $\pi_x$  とし、ダミー位置として選択された各メッシュの中心位置の緯度経度をサーバに送信し、イベント情報の問い合わせを行う。

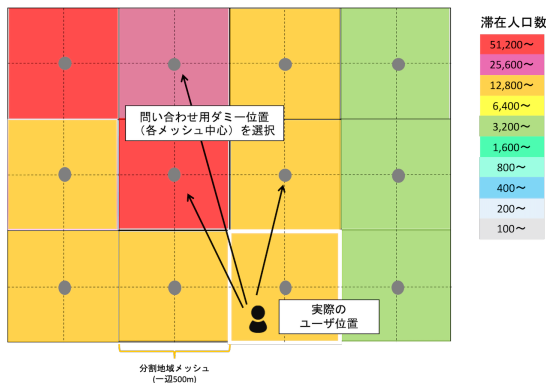


図 3: ダミー位置とユーザ位置の関係 (提案手法 2Geo-I を用いた位置匿名化手法)

### 3 提案手法 1: 問い合わせの実装

例として、都内の駅やスポット 7 種類 (新宿三丁目駅, 東京都庁, 代々木駅, 東京芸術劇場, 池袋サンシャイン, 東京ドーム, 六本木ヒルズ) のユーザ位置を設定し、それぞれに対し第 3 次メッシュの中心位置を算出しダミーの位置とした。ユーザ位置と主要駅の距離に応じて範囲を変更し、イベント情報の問い合わせを行った。表 1 にユーザ位置の最寄りの主要駅との距離, 実際に問い合わせを行った範囲, 実際に問い合わせで得られたイベント数を示す。

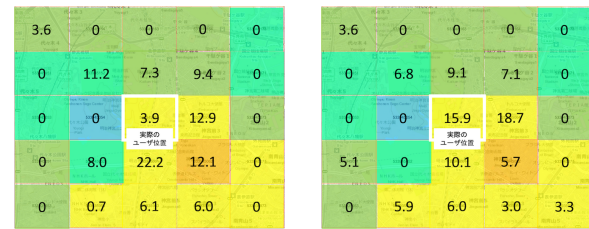
表 1: イベント情報問い合わせ結果

ユーザ位置名	最寄りの主要駅	最寄りの主要駅との距離(km)	問い合わせに使用する範囲	問い合わせで得られたイベント数
新宿三丁目駅	新宿駅	0.415	1	44
東京都庁	新宿駅	0.766	2	122
代々木駅	新宿駅	0.741	2	23
東京芸術劇場	池袋駅	0.258	1	29
池袋サンシャイン	池袋駅	0.784	2	15
東京ドーム	東京駅	3.038	3	10
六本木ヒルズ	渋谷駅	2.817	3	30

### 4 提案手法 2: 問い合わせの実装

例として、2023 年 1 月 21 日 (土) の午前 11 時から正午における滞在人口数を事前確率  $\pi_x$  とし、ユーザ位置を原宿駅周辺で固定しプライバシー保護度合い  $\epsilon = 1.5$  としたときの結果 (図 4(a)) と  $\epsilon = 2.0$  としたときの結果 (図 4(b)) を示す。区画内に各メッシュ区画が選ばれる確率を百聞率で記載し、各メッシュ区画の色は図 3 に表した、10 段階の滞

在人口数を表している。



(a)  $\epsilon = 1.5$

(b)  $\epsilon = 2.0$

図 4: ユーザ位置固定 (原宿), 各区画が選ばれる確率 (%)

## 5 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保護のためのデータ処理について検討した。提案手法 1, メッシュへの近似手法では、ユーザ位置の第 3 次メッシュ中心位置をダミー位置として設定し、サーバへの問い合わせに使用することで、サーバに付与されるユーザ位置情報は 1km 四方に曖昧化されることを保証している。問い合わせの効率化を図るため、ユーザ位置と東京都内の駅の乗降者数の多い主要駅との距離で問い合わせ範囲の場合分けを行なった。

提案手法 2, Geo-I を用いた位置情報匿名化手法では、滞在人口数のリアルデータを利用し、差分プライバシーを位置情報匿名化に拡張した技術 Geo-I を満たすようなユーザ位置のダミー位置を算出する処理を行った。単純に範囲を広げるような提案手法 1 に比べ、人口が少ない区画がダミー位置として選ばれないようにする, 同じ性年代の人が多いところをダミー位置として選ばれやすくするというようなユーザ目線のプライバシー保護の調整が可能となり, 人口分布的に異なる特徴を持つ地域に対して本手法が適応できることを確認した。また論理的なプライバシー保護指標  $\epsilon$  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った。

今回は限定的な区画で実験を行ったが, 今後はより広い連続する範囲での実装, 時間帯別などの滞在人口数データを使用し, イベント推薦に活用していきたい。適切にイベント開催地分布に応じた場合分けを行うため, 過去のイベントの発生情報利用の検討や, さらにユーザの要望に合うような情報の推薦を行うため, 他のユーザデータを利用した順位付けを考えていき, より実用的なシステムの実装を進めたい。

## 文 献

- [1] Nicolás E Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Optimal geo-indistinguishable mechanisms for location privacy, 2014.
- [2] Mobile geospatial statistics. <https://mobaku.jp>.