7N-04

# ユーザの位置情報プライバシを考慮したSNSデータからのイベント情報検索手法についての検討

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，問合せユーザの位置情報やSNSデータなどを利用することで，ユーザに対し特定の場所や時間で開催されるイベントを推薦するシステムは既に多く研究されている．しかし，ユーザにとって，個人情報を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上の懸念があり，なかでも位置情報保護に関する研究[1]が盛んに行われている．本研究では，SNSの一種である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を考慮して地理的な制約条件を満たしながら，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

### 2.1. 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．クライアント側ではユーザのプライバシに関わるような情報の保護処理を，サーバ側では工藤ら[2]が提案する手法を参考にし，SNSデータからの情報抽出・分析処理を行う．

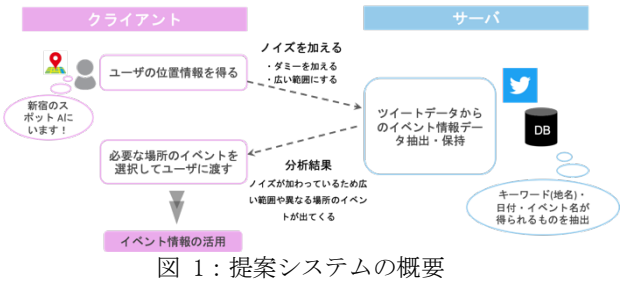


図 1：提案システムの概要

### 2.2. イベント情報抽出

キーワード（地名）を含むツイートの収集をTwitter API[3]を使用し行った．2021年の11/20から11/24，12/10から12/14のそれぞれ5日間，計10種類のキーワードを含むツイート取得した．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とイベント名が含まれるツイートを抽出した．表1に設定したキーワード(地名)と抽出できたイベント情報数(イベント名の重複を含む)の関係を示す．

### 2.3. 地域メッシュ情報の付与

ユーザの位置情報プライバシを保護するために，地域メッシュ[4]情報を用いてユーザの位置から範囲を広げたイベント情報検索を試みる．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．クライアント側もユーザ位置に対して同様の処理を行い，算出した地域メッシュコードを問い合わせに使用する．サーバはユーザ位置の地域メッシュコードと同じコード範囲内で開催されるイベント情報を返す．

本研究では，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」から以下の区画を使用する．

・**第1次地域区画(第1次メッシュ)**：緯度の間隔40分，経度の間隔1度，一辺の長さ約80km
・**第2次地域区画(第2次メッシュ)**：第1次メッシュを緯線方向及び経線方向に8等分してできる区域，一辺の長さ約10km
・**基準地域メッシュ(第3次メッシュ)**：第2次メッシュを緯線方向及び経線方向に10等分してできる区域，一辺の長さ約1km
・**分割地域メッシュ(2分の1地域メッシュ)**：第3次メッシュを緯線方向及び経線方向に2等分してできる区域，一辺の長さ約500m

Examination of event information retrieval method from SNS data considering user location information privacy
†Kyoka Ishigami
‡Miki Enoki
†Masato Oguchi



Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.

表１：キーワードと取得したツイート数の関係

| キーワード | 取得ツイート数(個) | | | | |
|---|---|---|---|---|---|
| | 11/20(土) | 11/21(日) | 11/22(月) | 11/23(火) | 11/24(水) |
| 原宿 | 16 | 16 | 11 | 18 | 11 |
| 渋谷 | 85 | 116 | 116 | 118 | 92 |
| 新宿 | 101 | 96 | 126 | 108 | 90 |
| 赤坂 | 6 | 6 | 10 | 13 | 7 |
| 代々木 | 21 | 29 | 18 | 28 | 18 |
| 池袋 | 51 | 76 | 69 | 56 | 48 |
| 六本木 | 9 | 12 | 25 | 13 | 12 |
| | 12/10(金) | 12/11(土) | 12/12(日) | 12/13(月) | 12/14(火) |
| 立川 | 6 | 16 | 8 | 10 | 8 |
| 足立 | 12 | 7 | 7 | 4 | 5 |
| 八王子 | 35 | 34 | 49 | 38 | 47 |

## 3. 地域メッシュを用いたイベント問い合わせ

### 3.1. 問い合わせ結果

例として設定した 7 種類のユーザ位置と同じ地域メッシュ圏内で開催されるイベント情報の取得数を表２に示す.

表２：地域メッシュと取得したイベント数の関係

| ユーザ位置 | 問い合わせで得られたイベント数(個) | | |
|---|---|---|---|
| | 2 次メッシュ(一辺 10km)圏内 | 3 次メッシュ(一辺 1km)圏内 | 分割地域メッシュ(一辺 500m)圏内 |
| 新宿三丁目駅 | 293 | 44 | 27 |
| 東京都庁 | 293 | 2 | 2 |
| 東京芸術劇場 | 293 | 29 | 20 |
| 東京ドーム | 16 | 1 | 1 |
| 六本木ヒルズ | 120 | 17 | 13 |
| 北千住駅 | 16 | 0 | 0 |
| 八王子駅 | 8 | 7 | 7 |

### 3.2. 考察

本研究では，2 次メッシュコード，3 次メッシュコード，分割地域メッシュコードに対し，それぞれユーザ位置の地域メッシュコードと一致しているイベント数を調査した．ユーザ位置によって，地域メッシュを用いた範囲を広げるサーバへの問い合わせをしても，得られるイベント数があまり増加していないことが分かる．東京ドームや北千住駅など，ターミナル駅から離れた位置にいた場合に顕著で，イベント開催地の分布は一様では無く，ターミナル駅付近に集中していることが理由として考えられる．よって，ユーザが参加するイベントが攻撃者に知られたくないような場合，場所によってただ範囲を広げた問い合わせではなく，ダミーのユーザ位置を加えた問い合わせを行う方がプライバシ保護の観点から効果的である可能性が高い．また，新宿三丁目駅や東京芸術劇場など，ターミナル駅付近の位置にいる場合，範囲を広げた問い合わせの結果得られるイベント数が膨大となり，ユーザが必要とする狭い範囲のイベント情報を算出するクライアント側の計算量が大きくなってしまう.

## 4. まとめと今後の課題

本研究では，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を保護するためのデータ処理方法について検討した．問い合わせで使用する地域メッシュコードの範囲を広げれば広げるほど，攻撃者からユーザ位置の予測はされにくくなるが，クライアント側の計算量が大きくなる．今後，ユーザの位置情報にノイズを加える方法として考えられる適切なダミー位置の設定と取得できるイベント情報数を調査したい.

また本研究ではユーザが移動しながら複数回問い合わせをし，攻撃者が問い合わせ結果を全て得られるような場合を考慮していない．ユーザの移動経路を攻撃者に予測されにくいようなモデルも構築していきたい.

## 参考文献

[1] 高木 駿，曹 洋，浅野 泰仁，吉川 正俊：「道路ネットワークにおける位置情報プライバシ」DEIM Forum 2019 I6-3
[2] 工藤 瑠璃子，榎 美紀，中尾 彰宏，山本 周，山口 実靖，小口 正人：「SNS データを用いた場所と時間を考慮するイベント検索手法の提案と評価」マルチメディア，分散，協調とモバイル(DICOMO2018)シンポジウム
[3]Twitter Search API, “https://dev.twitter.com/rest/public/search”
[4] 総務省情報局 地域メッシュ統計の概要, “https://www.stat.go.jp/data/mesh/m_tuite.html”
[5]Google Maps Geolocation API, “https://developers.google.com/maps/documentation/geolocation/overview



Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.