

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手法

石神 京佳 (指導教員：小口 正人)

1 はじめに

近年、ソーシャルネットワークサービス（以下 SNS とする）の普及に伴い、SNS 上には固定的なメディアに掲載されていないような様々な有益な情報が存在するようになった。そこで我々は、SNS 上の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ていることに注目した。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研究されている。しかし、問い合わせユーザにとって、位置情報をはじめとする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ー上の懸念がある。

本研究では、SNS の一種である Twitter[1] 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する処理を施すことで、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

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。

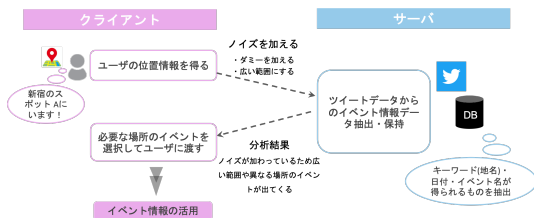


図 1: 提案システムの概要

サーバ側は同研究室工藤ら [2] が提案する手法を参考にし、パブリックなツイートデータからのイベント情報抽出・分析処理を行う。クライアント側はユーザのプライバシーに関わるような情報の保護処理をしてからサーバへイベント情報の問い合わせをする。クライアントは問い合わせ結果で得られた、ノイズが加わった位置もしくは広い範囲で開催されるイベントから、ユーザが必要としている範囲のイベントを抽出する。

2.2 イベント情報抽出

Twitter API[3] 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とイベント名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データベースに格納する。

2.3 地域メッシュ情報の付与

本研究では、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¹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」から第 3 次メッシュ（基準地域メッシュ）、2 分の 1 地域メッシュ（分割地域メッシュ）を使用する。

2.4 プライバシーを考慮した問い合わせ

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 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する。本研究では、ユーザ側に現在地から短距離順の上位 k 個以上のイベント情報を得たいというような条件があると仮定し、1 回目の問い合わせで得られたイベント数が k 個より少なかった場合範囲を広げて 2 回目の問い合わせを行う。このように、得られたイベント数に応じて範囲を段階的に広げる問い合わせを繰り返す。図 2 に、問い合わせで用いたダミー位置と問い合わせ範囲の広げ方の概要を示す。

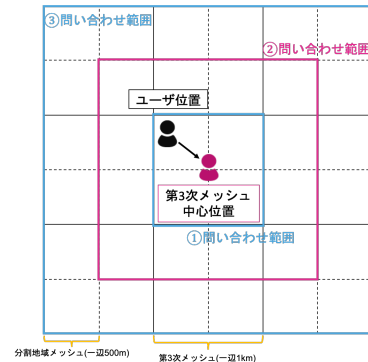


図 2: ダミー位置と問い合わせ範囲の広げ方

3 地域メッシュを用いたイベント問い合わせ

3.1 実験データ

地域メッシュを用いたイベントの問合せの実験を行うために、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を取得した。表 1 に設定したキーワード（地名）

¹東京 23 区内のアミューズメント施設、ミュージアム、ショッピング施設、エンターテインメント施設、温泉、劇場、ホールを東京ウォーカー [6] とナビタイム [7] から抜粋しイベントスポット辞書を作成する。ツイートにイベントスポット辞書内の要素が存在する場合は、その要素をイベント開催地として保持

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。

表 1: キーワードと取得したイベント数の関係

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

山手線沿線のターミナル駅である新宿や渋谷をキーワードとしたイベント数が比較的多いことが分かる。

3.2 問い合わせ結果

例として、都内の駅やスポット 6 種類のユーザ位置を設定し、それぞれに対し 3 次メッシュの中心位置から範囲を広げる問い合わせを行い、問い合わせ範囲内で開催されるイベント情報を取得した。図 2 に示す①1 回目問い合わせ範囲,②2 回目問い合わせ範囲,③3 回目問い合わせ範囲で段階的に問い合わせ範囲を広げた。計 3 回の問い合わせで得られたユーザ位置とイベント取得数の関係を表 2 に示す。

表 2: 3 次メッシュ中心位置から範囲を広げた問い合わせ結果

ユーザ位置	問い合わせで得られたイベント数(個)		
	1 回目問い合わせ	2 回目問い合わせ	3 回目問い合わせ
新宿三丁目駅	44	144	180
東京都庁	2	122	178
東京芸術劇場	29	39	45
東京ドーム	1	1	10
六本木ヒルズ	17	19	30
八王子駅	7	7	7

3.3 考察

ユーザ位置によって、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せず、必要な問い合わせ回数が多くなる事が分かる。東京ドームなど、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様々で無く、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。例えば、池袋駅付近のイベント開催地の分布は以下の図 3 のようになっていて、池袋駅付近に集中して分布していることが見て取れる。

しかし、初めから広い範囲で問い合わせをすると、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、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 k 個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ふ。よって、段階的に範囲を広げる問い合わせは有効であると考えられる。

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

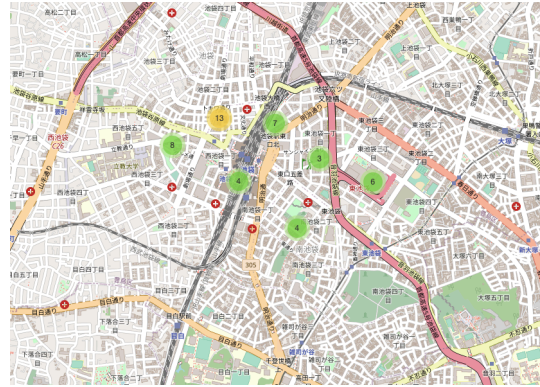


図 3: 池袋駅付近で取得したイベント開催地の分布(周辺で開催されるイベントをクラスタ化し件数表示)

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。ユーザ位置の 3 次メッシュ中心位置をサーバへの問い合わせで使用することで、サーバに付与されるユーザ位置情報は 3 次メッシュ情報に限定されることを保証した。攻撃者からユーザの位置情報の予測はされにくくなるが、問い合わせ位置をずらしたことによってデータの有効性が多少欠損してしまうことが課題として挙げられる。また今回得られたイベント情報数が問い合わせ回数に応じてあまり増加しないものがあつたため、問い合わせ回数とクライアントの計算量のトレードオフの関係を考慮し、より効率的な範囲の広げ方を検討していきたい。またユーザが移動しながら都度問い合わせる場合、攻撃者の予測の精度は高くなってしまふため、ユーザの移動経路を保護できるようなモデルも構築していきたい。

参考文献

- [1] Twitter: <http://twitter.com/>
- [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>
- [6] Walker+東京都のおでかけスポット一覧: https://www.walkerplus.com/spot_list/ar0313/
- [7] NAVITIME: <https://www.navitime.co.jp/category/>