

Swarm ツイートの分析に基づく地域混雑度可視化システム

川井 裕介[†] 熊本 忠彦[†]

[†] 千葉工業大学情報科学部情報ネットワーク学科 〒275-0016 千葉県習志野市津田沼 2-17-1

E-mail: ts1432047fa@s.chibakoudai.jp, tkumamoto@net.it-chiba.ac.jp

あらまし 本論文では、ユーザが指定した地域・時間帯における混雑度を予測し、可視化するシステムを提案する。具体的には、Twitter を介して投稿される Swarm ツイートから投稿者の位置情報（緯度・経度）を抽出し、メッシュ化された各地域におけるツイート投稿数を単位時間ごとに調べ、過去 3 週間分の平均ツイート投稿数と比べることで、当該地域がいつもに比べ、どのくらい混雑するのか（あるいは空くのか）をヒートマップを用いて Google マップ上に可視化する。このシステムを用いることで、ユーザは任意の地点の混雑状況が今後数時間においてどのように変化するかを視覚的に確認することができるようになる。

キーワード 位置情報共有サービス, Swarm, Twitter 連携, 基準地域メッシュ, ヒートマップ, Google マップ

1. はじめに

近年、利用者が増えている位置ベースソーシャルメディア（Location-based Social Media）[1] の一つに Swarm [2] がある。Swarm は、スマートフォンやタブレット端末から利用可能なモバイルアプリケーションであり、自分が今どこにいるかを友人らと共有することができる。特に、特定の場所（観光スポットや店舗・施設あるいはイベント会場など）にチェックインすることで、自分の詳細な位置情報を友人らに知らせることもできる。また、Twitter [3] と連携させると、チェックインした場所に関する情報を含むツイート^(注1) が自動的に生成され、Twitter 上に投稿されるようになる。Swarm におけるチェックイン自体はあらかじめ設定された友人ら以外には配信されないが、Twitter と連携させているユーザも数多く存在しており、そのようにして投稿されたツイート（本論文では Swarm ツイートと呼ぶ）は Twitter API [4] を用いて収集することができる。

そこで本論文では、ユーザが指定した地域・時間帯における混雑度を予測し、可視化するシステムを提案する。すなわち、Twitter 上に投稿された Swarm ツイートを収集し蓄積することで、ユーザが指定した任意の地点とその周辺における今日のツイート投稿数と過去 3 週間分の平均ツイート投稿数の時間変化を比較できるようにするとともに、その比較結果に基づいて当該地域がいつもの混雑度に比べより混雑するのかそうでないのかを予測する手法を提案する。さらに、この予測した混雑度を用いてヒートマップを描画し、Google マップに重ねることで、その地点周辺の混雑状況（いつもより混雑するのかそれとも空くのか）を可視化するシステムを提案する。

2. 関連研究

ツイートや写真データに付与されたジオタグ（緯度・経度）に

(注1)：厳密に言えば、このツイートそのものには位置情報は付与されていないが、ツイート内に含まれているデータを解析することで、チェックインした場所の緯度・経度を取得することが可能となる。

関する数値を含むタグ）を用いた研究 [5] [6] [7] は数多くなされている。例えば、谷ら [5] は、ジオタグ付きツイートを分析し、投稿ユーザらが共通に利用している交通路を抽出する手法を提案しており、桜並木に沿って移動する花見客の交通路を可視化している。また、大森ら [6] は、写真共有サイトにおいて公開されている写真に含まれているジオタグ等を用いて、海岸線や線路、高速道路のような線からなる地理的特徴を抽出する手法を提案しており、評価実験により海岸線の 64%～82%を実際の海岸線から 500m 以内に描くことができることを確認している。一方、斉藤ら [7] は、Twitter 上に投稿されたツイートとツイートに付与されたジオタグから投稿者の行動を予測する手法を提案しており、評価実験により単純統計を用いた従来手法より高い予測精度が得られることを確認している。

さらに、ジオタグ等を用いて混雑度を可視化するという研究 [8] もある。例えば、菅野ら [8] は、Twitter 上に投稿されたジオタグ付きのツイートを解析することで、災害発生時に人々が密集している場所を抽出するとともに、その密集地から避難場所への避難行動を推測することで、新たな密集地の発生場所を予測し、可視化するシステムを提案している。

いずれの研究もジオタグの有用性を示しているが、その一方で、ジオタグが付与されたツイートの少なさが問題点として挙げられている。本研究で対象とした Swarm ツイートは、その性質上、観光スポットや人気ショップ、イベント会場などの人が集まる場所では投稿数が増える傾向にあることから、そのような場所での人の流れの把握あるいは誘導に有用であると考えられる。

3. 提案システムの構成

図 1 にシステム全体の構成を示す。以下、この構成に従って各処理モジュールについて説明する。なお、提案システムの Swarm ツイート分析モジュールと位置情報取得モジュールは、オフライン処理であり、常に実行されている。地域混雑度予測モジュールと可視化モジュールは、オンライン処理であり、

ユーザからの要求（混雑度の予測を知りたい地点・時間帯に関する情報の入力）に応じて動作する。

3.1 Swarm ツイート分析モジュール

Swarm ツイート分析モジュールは、Twitter [3] を介して投稿される Swarm ツイートを Twitter Streaming API [4] を用いてリアルタイムに収集し、データベースに格納する。Swarm ツイートには、投稿ユーザがチェックインした場所に関する情報として、ベニュー名（施設名・地名など）や所在地情報（都道府県名や市区町村名など）、施設 TwitterID（当該ベニューの Twitter アカウント名）、短縮 URL（Swarm サイトにおける当該ベニューのチェックインページの在りか）が含まれているが、これらのデータが投稿日時とともにデータベースに格納される。

3.2 位置情報取得モジュール

位置情報取得モジュールは、新規に登録された各 Swarm ツイートに含まれている短縮 URL を一定時間（現在は 3 秒）ごとにデータベースから取り出し、その URL を辿ることで、その Swarm ツイートが投稿された位置（緯度・経度）に関する情報を取得する。しかしながら、緯度・経度に関する情報は粒度が非常に細かく、データスパースとなるため、そのままでは利用しづらい。そこで本研究では、この緯度・経度情報を基準地域メッシュコード [9] [10] に変換したうえで、緯度・経度情報とともにそれぞれの Swarm ツイートと関連付け、データベースに格納することにする。

なお、地域メッシュ [9] [10] とは、緯度・経度に基づいて地域をメッシュ状に区分したものであり、昭和 48 年 7 月 12 日行政管理庁告示第 143 号に基づいて定義されている。この地域メッシュには、第 1 次地域区画、第 2 次地域区画（統合地域メッシュ）、基準地域メッシュ（第 3 次地域区画）といった階層構造があり、メッシュのサイズが異なっている。本研究で採用した基準地域メッシュは、一辺の長さが約 1km^(注2) となっており、それぞれのメッシュには「5339-46-00」といったコードが付与されている。ちなみに、この例では、「5339」が第 1 次地域区画コードであり、東京都と神奈川県を大部分を含む、一辺の長さが約 80km のメッシュであることを示している。「5339-46」は、第 2 次地域区画コードであり、千代田区から葛飾区までの一片の長さが約 10km のメッシュであることを示している。第 2 次地域区画は、第 1 次地域区画を 64 分割（縦横それぞれ 8 等分）したものであるため、「5339」という第 1 次地域区画には「5339-00」から「5339-63」までが存在する。「5339-46-00」が基準地域メッシュコードであり、霞が関・日比谷公園付近の一辺の長さが約 1km のメッシュであることを示している。基準地域メッシュは、第 2 次地域区画を 100 分割（縦横それぞれ 10 等分）したものであるため、「5339-46」という第 2 次地域区画には「5339-46-00」から「5339-46-99」までが存在する。

3.3 地域混雑度予測モジュール

地域混雑度予測モジュールは、ユーザの要求に応じて、データベースから必要なデータを取り出し、任意の地域・任意の時

間帯における混雑度を予測する。

まず、地域混雑度予測モジュールは、ユーザが指定した地点を含む基準地域メッシュとその周辺にある基準地域メッシュのそれぞれにおいて、1 時間あたりの Swarm ツイート投稿数を過去 2 時間分、データベースから取得する。すなわち、現在の時刻が H 時 M 分だとすると、 $H-2$ 時台になされた投稿数 N_{H-2} と $H-1$ 時台になされた投稿数 N_{H-1} をデータベースから取得する。さらに、同じ曜日・同じ時間帯における過去 3 週間分の投稿数もデータベースから取得する。すなわち、先週の同じ曜日（すなわち 7 日前）の $H-2$ 時台と $H-1$ 時台になされた投稿数と先々週の同じ曜日（すなわち 14 日前）の $H-2$ 時台と $H-1$ 時台になされた投稿数、および 3 週間前の同じ曜日（すなわち 21 日前）の $H-2$ 時台と $H-1$ 時台になされた投稿数を取得し、それぞれの時間帯における平均ツイート投稿数 AN_{H-2} と AN_{H-1} を求める。本研究では、この平均ツイート投稿数を当該時間帯における当該地域のいつもの混雑度を表す指標として用いる。

次に、過去 2 時間分のツイート投稿数と過去 3 週間分の平均ツイート投稿数を比べることで、今後数時間における混雑度がいつもの混雑度に比べて、より混雑するのかそれとも空くのかを予測する。ここで、参考のために、混雑度予測手法の設計コンセプトを図 2 に示す。

まず、それぞれの時間帯における過去 2 時間分のツイート投稿数と過去 3 週間分の平均ツイート投稿数との差分を次式により求める。

$$\Delta_{H-2} = N_{H-2} - AN_{H-2} \quad (1)$$

$$\Delta_{H-1} = N_{H-1} - AN_{H-1} \quad (2)$$

次に、XY 平面座標系において点 $(H-2, \Delta_{H-2})$ と点 $(H-1, \Delta_{H-1})$ を通る直線の式（図 2 では赤色と橙色の線）を求める。この式を $y = ax + b$ とすると、時刻 t （但し、 $t > H$ ）における y の値は $at + b$ で求められ、この y の値がいつもの混雑度からどのくらい混雑しているのか（あるいは空いているのか）を表す指標となる。本論文では、この y の値を混雑度として扱う。なお、 y の値が正であれば、いつもより混雑するという予測を意味し、負であれば、いつもより空くという予測を意味する。また、その絶対値の大小によりどのくらい混雑するのかあるいは空くのかを表されている。

3.4 可視化モジュール

可視化モジュールは、地域混雑度予測モジュールにおいて算出された各基準地域メッシュの混雑度をもとに、Google Maps API [11] を用いてヒートマップを描画し、Google マップに重ね提示することで、ユーザが指定した時間帯における当該地域の混雑状況を視覚的に捉えることができるようにする。Google マップの操作自体には、Google マップの機能をそのまま利用しており、拡大・縮小・ドラッグ（表示範囲の移動）等により表示させる地域や範囲を変えることができる。また、予測する時間帯はスライダーにより選択することが可能となっている。

(注2)：緯度の間隔でいえば 30 秒、経度の間隔でいえば 45 秒に相当している

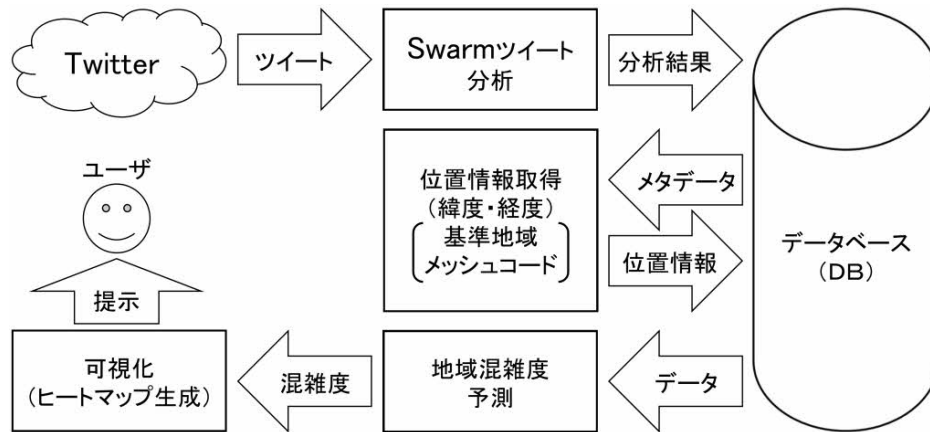


図 1 提案システムの構成

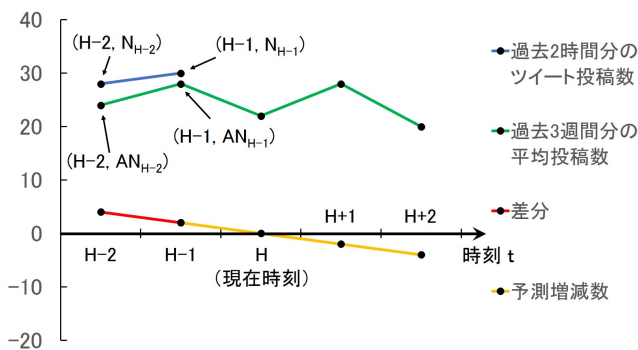


図 2 地域混雑度予測手法の設計コンセプト

4. 実行例

提案システムは、PHP と Javascript を用いて実装された Web アプリケーションであり、ブラウザ上で動作する。以下、各モジュールの実行例を示す。

4.1 Swarm ツイート分析モジュール

Swarm ツイート分析モジュールは、まず、Twitter Streaming API [4] を用いて日本のユーザを対象とする公開ストリームをリアルタイムに収集し、Swarm ツイート生成時に用いられる特徴的な文字列である「I'm at」を含むツイートを Swarm ツイートとして抽出する。ここで、実際に抽出した Swarm ツイートの例を表 1 に示す。なお、実際には計算機資源の制限により、ツイート内に「東京」あるいは「Tokyo」という語を含むツイートだけを対象としている。

次に、Swarm ツイート分析モジュールは、Swarm ツイートからベニュー名や施設 TwitterID、所在地情報、短縮 URL と投稿日時を抽出し、データベースに格納する。ちなみに、表 1 に示した Swarm ツイートからは、ベニュー名、施設 TwitterID、所在地情報、短縮 URL としてそれぞれ「東京ビッグサイト (東京国際展示場)」、「t_bigsight」、「江東区, 東京都」、「<https://t.co/XXXXXX>」^(注3) が抽出される。なお、施設 TwitterID は付与されていない場合もあり、その場合は空データと

(注3)：短縮 URL を辿ると、その Swarm ツイートの投稿者に関する情報を得ることができるため、短縮 URL の一部を伏字 (XXXXXX) にした。

表 1 Swarm ツイートの例

I'm at 東京ビッグサイト (東京国際展示場) - @t_bigsight in 江東区, 東京都 https://t.co/XXXXXX

表 2 短縮 URL のリンク先から取得可能なメタ要素の例

<meta content="35.629766394080306" property="place: location:latitude" / >
<meta content="139.79408383369443" property="place: location:longitude" / >

なる。

4.2 位置情報取得モジュール

位置情報取得モジュールは、まず、新規に登録された各 Swarm ツイートに含まれている短縮 URL をデータベースから取り出し、その短縮 URL を辿ることで、当該ベニューの Swarm サイト上のチェックインページを HTML データとして取得する。この HTML データの中にメタ要素として当該ベニューの緯度と経度が含まれているので、これを抽出する。例えば、表 1 に示した Swarm ツイートの短縮 URL を辿り、取得した HTML データには、表 2 に示したようなメタ要素が含まれているので、この Swarm ツイートが緯度「35.629766394080306」、経度「139.79408383369443」の地点において投稿されたものであることがわかる。

次に、位置情報取得モジュールは、取得した緯度・経度を基準地域メッシュコードに変換し、データベース内に格納されている当該 Swarm ツイートと関連付ける。なお、緯度・経度の基準地域メッシュコードへの変換には、Web サイト上において無償で公開されている「メッシュコード算出汎用クラス」[12] を用いている。

4.3 地域混雑度予測モジュール

地域混雑度予測モジュールは、ユーザが指定した地点を含む基準地域メッシュとその周辺にある基準地域メッシュのそれぞれにおいて、1 時間あたりの Swarm ツイート投稿数を過去 2 時間分取得するとともに、同じ曜日・同じ時間帯における過去 3 週間分のツイート投稿数も取得し、時間帯ごとに平均ツイート投稿数を算出する。そして、これらのデータを式 (1) と式 (2) に代入することで、差分値 Δ_{H-2} と Δ_{H-1} を得る。さらに、こ

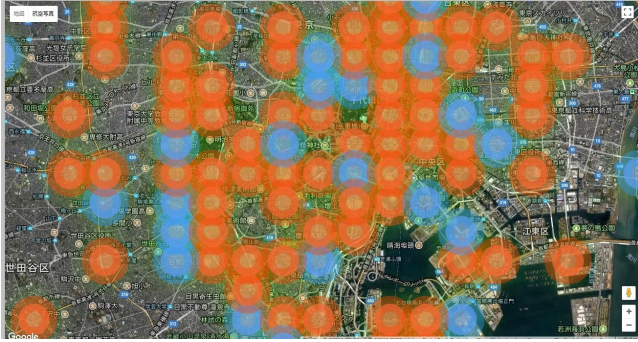


図3 Google マップの生成例（平成 30 年 1 月 9 日 11 時台）

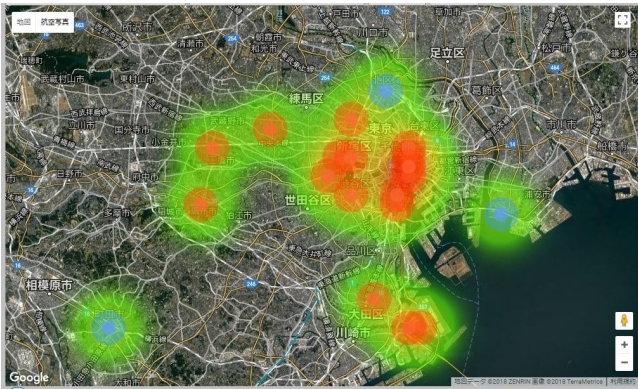


図4 Google マップの生成例（平成 30 年 1 月 9 日 7 時台）

の差分値から点 $(H-2, \Delta_{H-2})$ と点 $(H-1, \Delta_{H-1})$ を通る直線の式 $(y = ax + b)$ を求め、ユーザが指定した時刻における各基準地域メッシュの混雑度を算出する。

4.4 可視化モジュール

可視化モジュールは、地域混雑度予測モジュールにおいて算出された各基準地域メッシュにおける混雑度をもとに、Google Maps API [11] を用いてヒートマップを描画し、Google マップに重ね提示する。ここで、Google マップの生成例を図 3 に示す。

図 3 は、平成 30 年 1 月 9 日の 10 時台に、1 時間後である 11 時台の混雑度を予測し、可視化したものであり、表示範囲としては概ね東京 23 区全体が表示されるようになっている。図 3 において、赤い円が混雑するという予測を、青い円が空くという予測を表している。

別の Google マップの生成例として図 4 を示す。図 4 は、平成 30 年 1 月 9 日の 6 時台に、1 時間後である 7 時台の混雑度を可視化したものであり、図 3 と同様、赤い円が混雑するという予測を、青い円が空くという予測を表している。さらに、付加情報として、図 4 では Swarm ツイートの投稿状況を緑色の領域で表している。これは、Swarm ツイートの投稿状況をチェックするための機能であり、過去 2 時間におけるそれぞれの時間帯において Swarm ツイートが 1 件以上投稿された地域が緑色で表されている。言い換えれば、緑色のついていない地域では、Swarm ツイートが 0 件もしくは 1 件しか投稿されなかったため、混雑度の予測を行っていないことを示している。

5. ま と め

本論文では、ユーザが指定した地域・時間帯における混雑度がいつもの混雑度に比べ、どのくらい混雑するのか（あるいはどのくらい空くのか）を予測し、Google マップ上に可視化するシステムを提案した。このシステムでは、Twitter 上に投稿された Swarm ツイートの投稿数がいつもに比べ、多いのか少ないのかをチェックし、その時間変化を考慮することで、混雑度がどのように変化するかを予測している。さらに、その変化の度合いをヒートマップとして描画し、Google マップに重ねることで、指定した地域・時間帯における混雑度の変化予測を視覚的に捉えることができるようになっている。

今後の課題としては、表示可能な範囲を広げ、日本全国をカバーできるようにすること、実ユーザらによる大規模な性能評価実験を行い、予測精度を評価するとともに、システムの長所・短所を明らかにすること、地域混雑度の予測に際し、単純な直線回帰分析ではなく、より高度な多変量解析手法や様々な深層学習手法を導入して試みることで、Web サイトにおいて実運用することなどが挙げられる。

謝 辞

本研究の一部は、福田将治奨学寄附金により実施された。ここに記し、感謝の意を表する。

文 献

- [1] 若宮翔子, 私のブックマーク「位置ベースソーシャルメディア」, 人工知能学会誌, Vol.32, No.3, pp.448-451, 2017.
- [2] Swarm, <https://ja.swarmapp.com/>
- [3] Twitter, <https://twitter.com/>
- [4] Twitter API, <https://dev.twitter.com/>
- [5] 谷直樹, 風間一洋, 榎剛史, 吉田光男, 斉藤和巳, ジオタグ付きツイートをを用いた交通路の抽出法, 情報処理学会論文誌データベース, Vol.10, No.2, pp.31-41, 2017.
- [6] 大森雅己, 廣田雅春, 石川博, 横山昌平, ソーシャルメディア上から収集したジオタグに基づく地理的特徴の抽出と評価, 情報処理学会論文誌データベース, Vol.8, No.1, pp.1-16, 2015.
- [7] 斉藤裕樹, 高山翼, 山上慶, 戸辺義人, 鉄谷信二, マイクロブログのジオタグと発言コンテキスト解析による行動予測手法, 情報処理学会論文誌, Vol.55, No.2, pp.773-781, 2014.
- [8] 菅野真生, 遠藤雅樹, 廣田雅春, 横山昌平, 石川博, 災害時の密集地からの避難行動による混雑場所の可視化システムの提案, DEIM Forum 2015 論文集, C7-1, 2015.
- [9] Wikipedia「地域メッシュ」, <https://ja.wikipedia.org/wiki/地域メッシュ>
- [10] 総務省統計局「地域メッシュ統計について」, www.stat.go.jp/data/mesh/m_tuite.htm
- [11] Google Maps API, <https://developers.google.com/maps/>
- [12] メッシュコード算出汎用クラス, <http://white-bear.info/archives/1487>