

大規模イベントにおける現地ツイートを用いた 混雑状況可視化システムの提案

山崎 裕太郎[†] 福原 知宏^{†‡} 山田 剛一[†] 増田 英孝[†]

東京電機大学[†] マルティスープ株式会社[‡]

1. はじめに

特定の行事や施設等にて行われるイベントは、短期間で多くの人が集まる。特に、物品の販売を行うようなイベントでは、人気の商品を求め、特定のブースで行列が発生することがある。

しかし、具体的にどこでどの程度人が集まっているかを全体的に把握することは難しい。また、国内最大規模の同人誌即売会「コミックマーケット」(コミケ)では、3日間で約50万人が参加し[1]、会場内は大変な人混みが発生するため、詳細を捉えることは更に難しい。

一方、SNSは、その普及により様々な場面で利用されるようになり、前述のようなイベントにおいても、各地の様子等の投稿が見受けられる。丸ら[2]の研究では、災害発生時のネットワーク障害の検知に、SNSの1つであるTwitterを用いて、場所を特定するソーシャルセンサとして有効であると示されている。

本研究では、Twitterから大規模なイベントに関するツイートを収集し、その内容から現地で開催されたツイートを抽出し、会場内の人々の動きを可視化するシステムを提案する。本システムを用いることで、いつどこに人々が集まっているかを確認できる。

2. 本研究の目的

本研究の目的は、イベント参加者のツイート(現地ツイート)から、イベント会場内の混雑状況を可視化することである。会場内の人々の動きを混雑状況として示すことで、人気の少ない休憩場所などへの誘導や、混雑を避けやすいブース配置の判断材料を提供する。イベント開催期間中に投稿されたツイートを収集し、それらの文書を分析することで、各ツイートが会場内のどの場所についての内容であるかを推定する。なお、Twitterでは位置情報を付加して投稿をすることができるが、その投稿の割合は全体に対して非常に低く、データとして利用するには不十分である。

Visualization system of attendees in a large-scale event using onsite Twitter texts

[†] Yutaro Yamazaki, ^{†‡} Tomohiro Fukuhara, [†] Koichi Yamada, [†] Hidetaka Masuda

Tokyo Denki University ([†]), MULTISOUP ([‡])

3. 可視化システムの概要

可視化システムは、各種の処理を行った後のツイートを用いる。その処理の概要を図1に示す。

イベント会場内の各地点に対応する単語を最初のキーワード(場所単語)とし、文書内に場所単語を含むツイートを抽出した後、特徴語を特定し、それらを含むツイートを追加して、最後に機械学習によるフィルタリングを行い、現地ツイートを取り出す。

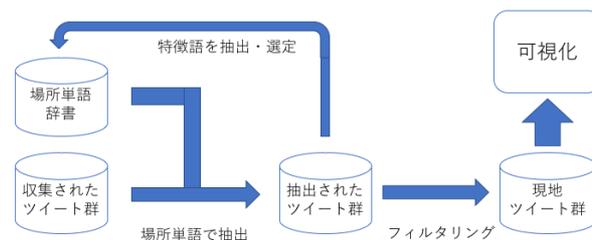


図1. システムの処理

3.1 ツイートの収集

TwitterのStreaming APIと、定期的な検索APIの実行によって収集する。なお、リツイートは収集の対象外とする。

3.2 場所単語辞書の構築

イベントに対応した場所を示す単語と、場所の位置を示す場所情報の組を辞書に登録する。

3.3 ツイートの抽出

場所単語辞書を用いて形態素解析を行う。ツイートの文書から場所単語が検出されたら、その単語から場所情報を得て、対象ツイートに与えたデータを保存する処理を行う。なお、位置情報つきツイート、Foursquareによるツイート、Botクライアントによるツイートは除外する。

3.4 場所単語辞書の拡張

抽出したツイート群から特徴語を抽出し、その特徴語の表記ゆれに対応する場所単語の場所情報を与えて辞書に登録し、ツイートを追加で抽出する。

3.5 現地ツイートのフィルタリング

対象ツイート全体の名詞、形容詞、助動詞、動詞を対象にしたTFIDF値をベースラインとし、ツイートの情報から追加で素性を与え、SVMを用いて現地ツイートの特徴を学習し、フィルタリ

ングする。

3.6 地図へのプロット

処理されたツイートを最終的なツイート群とし、時間帯ごとに分け、場所ごとにツイート数を集計したものを、ヒートマップとして地図上にマーカーをプロットする。また、システム上で各地の現地ツイートを閲覧できるように処理する。

4. 実験

2017年8月11日から8月13日にかけて、東京ビッグサイトにて行われた「コミックマーケット 92」(C92)を対象に、「C92」「コミケ」「夏コミ」「列」、また「東1,東2」といった東京ビッグサイトの各ホールを表す8語を加えたものを検索キーワードにして、ツイートの収集を行った。この結果、約300,000件を収集することができた。

場所単語辞書は、C92のWebカタログに登録されている約32,000サークルのデータから構築した。サークル名、サークルの出展者を、また、連携されていればTwitterのアカウント名を場所単語に、サークルの場所を示す番号を場所情報とした。ひらがなやカタカナ、記号、英数字のみで構成され、一定文字数に満たないものは除外した。

ツイートの抽出には、場所単語辞書をユーザ辞書として、mecab-ipadic-NEologdをシステム辞書として設定したMeCabを用いた。

フィルタリングの学習では、TFIDFの上位500語に、URLの有無、公式ドメインによる画像の有無、リプライの有無、ハッシュタグの数、文字数、行数、「【(任意の語)】」の有無、投稿時間を素性として追加した。

学習データは抽出したツイート約12000件中から人手分類した約1000~1200件を用いた。フィルタリングの結果、約4,000件が現地ツイートとなった。交差検定でF値を測った結果、TFIDF上位500語のみでは約0.82、素性を追加した場合は約0.85となった。

4.1 結果

実際にデータの処理を行い、地図にプロットしたものを図2に示す。

図2では1日目の午前10時から11時までのデータが示されている。地図上の小さなマーカーは、その時間帯で現地ツイートのあったサークル、一回り大きなマーカーは、東展示場の各ホール及びそれに属するサークルの現地ツイートの数を示し、その数によってマーカーの色が変化する。

4.2 考察

マーカーの数を全体的に見ると、前半の時間帯で多く、後半の時間帯になるほど減少していつている。このことから、時間の経過につれて、購入



図2. C92 初日混雑状況の可視化
(2017年8月11日午前10時台)

を終えた人、帰宅していく人が徐々に現れていることが示されている。

また、3日間を通して特別多く現れたのが、東7ホールに属するサークル「フェアビュラス叶組」「プレシャスM組」であった。これらは芸能人の叶姉妹が出展しており、一般人に広く認知されている人物が参加したことが高い影響力を与えたことが分かる。

可視化できなかった会場内の混雑として、Twitterに投稿する時間的余裕のない短い待ち行列や、物品の発送サービスといったイベント本体と関係のない待ち行列、表記ゆれで拾いきれなかったサークルなどがある。今後、可視化可能な混雑種類の検証、効率の良い表記ゆれの取得方法を検討する。

5. おわりに

SNSの投稿文書から現地ツイートを特定し抽出することで、イベントの会場内の人々の動きを示すことができた。本研究の手法を応用すれば、他のイベント、またイベントに限らず広い範囲で用いることができる。また、今後はある場所と合わせて訪れるような場所とを関連付けて示せるようにする。

参考文献

- [1] コミックマーケット92アフターレポート:
<http://www.comiket.co.jp/info-a/C92/C92AfterReport.html>
- [2] 丸ほか, "大規模災害時におけるネットワーク制御のためのソーシャル情報を用いた障害検知システムの構築.", 情報処理学会第78回全国大会講演論文集, 529-530(2016).