

大規模イベントにおける現地ツイートの分析と混雑状況可視化

山崎 裕太郎^{†1} 福原 知宏^{†1†2} 山田 剛一^{†1} 阿倍博信^{†1} 増田 英孝^{†1}

概要: 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。一方、SNS の 1 つである Twitter は、その手軽さからリアルタイム性の高いサービスとされ、イベントに参加しながらツイートする人は多い。本報告では、我々が開発に取り組んでいる、Twitter から大規模なイベントに関するツイートを収集し、その中から現地で投稿されたツイートを抽出して分析し、会場内の人々の動きを可視化するシステムについて述べる。

キーワード: Twitter, イベント可視化, 現地ツイート抽出, 群衆分析

Analyzing and Visualizing Attendees in Large-scale Events from Onsite Tweets

YUTARO YAMAZAKI^{†1} TOMOHIRO FUKUHARA^{†1†2} KOICHI YAMADA^{†1}
HIRONOBU ABE^{†1} HIDETAKA MASUDA^{†1}

Abstract: In specific events or facilities, many people gather in a short period of time. Especially at events that sell goods, matrices may occur at specific booths for purchasing popular items. However, it is difficult to comprehensively grasp where and how many people are gathering. On the other hand, Twitter, which is one of SNS, is regarded as a service with high real-time property from its handiness, many people tweet while participating in the event. In this paper, we describe a system that collects tweets about large events from Twitter, extracts and analyzes the onsite tweets from them, analyzes it, visualizes locations of people in the venue.

Keywords: Twitter, Event visualization, Onsite tweet extraction, Crowd analysis

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 万人が参加する規模となっている^{a)}。そういった場面では、イベント参加者としては、人混みや購入できる見込みのない場所は避けたいと考え、イベント主催側としては、混雑自体を発生させないようにしたいと考える。これらの改善のためには、会場全体の混雑状況を把握する必要があるが、実際にそれを捉えることは難しい。

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。SNS の 1 つである Twitter は、その手軽さからリアルタイム性の高いサービスとされ、コミケのようなイベントでも、そこに参加しながらツイートする人は多い。

本報告では、我々が開発に取り組んでいる、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投稿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 [1]と、それを用いて「コミックマーケット」を対象に行った検証実験について述べる。

2. 関連研究

丸ら [2] の研究では、災害等の緊急時にユーザがネットワークの状況に関心を寄せることに着目し、通信障害に関する初期キーワードを検索キーワードとし、それで取得したツイートを地名名詞ごとにまとめ、そこで実際に通信障害が発生しているかどうかを分類し、緊急地震速報からの情報と合わせることで、復旧のための地域ごとの重要度を決定している。

石野らの研究 [3]は、移動元、移動先、移動手段、経由地点などの情報を、与えられた文に含まれる単語の前後関係を手がかりとして機械学習で抽出して、各ユーザの災害発生時の移動経路を集約して地図上に表示するシステムを提案している。

^{†1} 東京電機大学
Tokyo Denki University
^{†2} マルティス株式会社
MULTISOUP CO., LTD.

a) <http://www.comiket.co.jp/info-a/C92/C92AfterReport.html>,

Zhang らの研究 [4]は、地名辞典 GeoNames を用いて文章内の地名の言及となる単語をツイートから抽出し、地名辞典の持っている人口、国の別名、そして国コードなどの地理学的属性と、さらに共起する地名、文脈特徴、地名辞典からの候補と抽出された地名の類似性を加えた情報を用いて場所を推定している。

以上のことから、SNS の 1 つである Twitter は、その投稿内容から場所を特定するソーシャルセンサとしての働きがあり、テキストベースによる分析でそれが可能であることが示されている。この性質に着目した上で、前述の研究では主に災害関連が対象にされていたが、本研究では特定の範囲で催されるイベントに特化し、Twitter のツイート内の文章を、場所を特定する辞書を用いて分析することによって、人々の動きを捉える。

3. 大規模イベントと提案システムの概要

3.1 コミックマーケットとは

コミックマーケットとは、同人誌即売会イベントのことであり、通常年に 2 回、8 月と 12 月に開催されるため、参加者からはそれぞれ「夏コミ」、「冬コミ」と呼ばれる。サークル参加者は、同人誌を中心に様々なジャンルのコンテンツを出展し、それを一般参加者が買い求める。サークル参加者はカタログにサークルとして登録され、スペース番号という一意に振られた文字列が与えられる。登録されたサークル情報は、紙媒体の他、会員制の Web サイトで閲覧することができる。Web 上で閲覧できるカタログ(Web カタログ)^{b)}からは、サークル名、執筆者名(出展者名)、登録ジャンル、概要文の他、niconico^{c)}、pixiv^{d)}、Twitter のアカウントの登録情報を確認することができる。

3.2 本研究における現地ツイート

本研究では、ツイートを現地ツイートであるか否かの 2 種類に分類して取り扱う。まず、現地ツイートとは、イベントが開催期間中に、会場内でリアルタイムに投稿されたツイートのことを指し、コミケを例にとると、ブースに到着したこと、現在の待機列の様子、ブースの残り在庫の速報などが挙げられる。一方、現地ツイートではないものは、まとめサイトと呼ばれる、世間の主にインターネット上での出来事をまとめた記事を投稿するサイトで、イベントの様子がまとめられた記事の共有を目的としたツイートが挙げられる。現地ツイートの実例を表 1 に示す。ブースに到着したことを示すツイートでは、「(目的のブース)に到着」といった文が見られる。待機列の様子を示すツイートは、表 1 の例の場合、ブースの待機列が東京ビッグサイトの東 4 ホールから東 6 ホールまでを折返していること、また別

のブースでは本来東 4 ホールに属するが東 7 ホールまで列が伸びていることが示されている。残り在庫状況を知らせるツイートでは、頒布物が残りわずかであることや、購入数に制限がかけられたことを知らせるものが見られ、これは例えば 1 人 1 品限りの場合「1 限」と表現される。一方、現地ツイートでないものは、インターネット上の情報をまとめた記事を扱う「まとめサイト」と呼ばれるサイトの記事の共有機能によるツイートや、イベント会場にいない人が現地ツイートを見たときに示す反応といったものが見られる。

表 1 ツイートの例

分類	ツイート例
ブースに到着	上海アリス待機列到着
待機列の様子	apricot+ の待機列が東 4 折返し東 6 折返しの最後尾東 5 とか 比村乳業、待機列最後尾は東 7 を越えそうな勢い#C93
残り在庫状況	杏仁豆腐先生のあんかけどおふ、すべて 1 限に変更#C92
記事共有	【C92 叶姉妹まとめ】案の定地獄の待機列！！しかし本人がいる東 7 ホールはファビュラスな香りに包まれ、コミケとは思えない清涼感に
外部の反応	TL にめっちゃ夏コミのコスプレ写真があがってる・・・行きたかった・・・ワイは絶賛仕事シゴ・・・

3.3 システムコンセプト

本研究の目的は、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、独自に構築した辞書を用いて分析し、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し、イベント会場内の人々の動きを混雑状況として示すことで、イベント参加者には行動プランの判断材料を、イベント主催側には混雑が発生しづらい会場設計の判断材料を提供することである。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、後述する2つの実験でツイートを収集した結果、共に全体に対する割合が5%未満なため、データとして利用するには不十分である。

本目的を果たすためのシステムは、収集した雑多なツイート群から現地ツイートを取り出し、それらが投稿された場所を推定して、可視化データの生成を行い、そのデータを地図上にマーカーとしてプロットする。プロットされた可視化データを閲覧するシステムのイメージ図を図 1 に示す。地図表示部内にプロットされたマーカーが各場所でのツイートの数を表しており、ツイート数に応じて、ヒートマップのように色に変化を持たせている。また、各マーカーをクリックすると、その場所の詳細情報が場所詳細表示部に、そこでのツイートの一覧がツイート一覧表示部に表示

b) <https://webcatalog.circle.ms>
c) <http://www.nicovideo.jp>

d) <https://www.pixiv.net>

される。また、システム画面の右側にある時間帯操作部で、表示するデータの時間帯の変更ができる。

4 システムの実装

4.1 システム概要

実装した提案システムの画面の様子を図 2 に示す。本システムの利用には、対象期間中に収集したツイートから、構築した辞書を用いてデータを検出し、その中から現地ツイートに分類されたものを地図にプロットしたものを用いる。

全体の処理の流れを図 3 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。

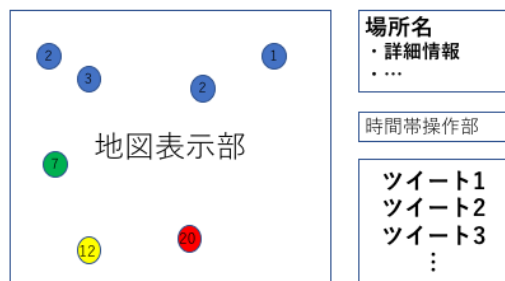


図 1 システムイメージ図



図 2 システム画面

4.2 データ処理

4.2.1 ツイートの収集

Twitter の Streaming API(filter) と、定期的な検索 API の併用によってツイートを取得する。なお、Retweet は収集の対象外とする。検索キーワードは、列の様子を述べる文に用いられる語を対象にする。対象にする語の例を挙げると、形成されている列への言及でほぼ必ず存在する「列」、また、目的の列のある場所にたどり着いたことを言及する「到着」などがある。

4.2.2 場所単語辞書の構築

イベントの各ブースなどのデータを抽出して、その位置など場所を特定する情報を場所情報とし、場所情報を持った単語を場所単語として、それらの組を辞書に登録する。場所単語はイベント内のブース名など、場所情報はブースの配置番号が対象となる。場所単語と場所情報となるものは、コミケを対象にする場合、場所単語はサークルや出展者の名前、場所情報はサークルの位置を定めるスペース番号となる。その例を表 2 に示す。

- ① ツイートの収集
- ② 場所単語辞書の構築
- ③ ツイートの抽出
- ④ 場所単語辞書の拡張
- ⑤ 現地ツイートのフィルタリング
- ⑥ 地図へのプロット

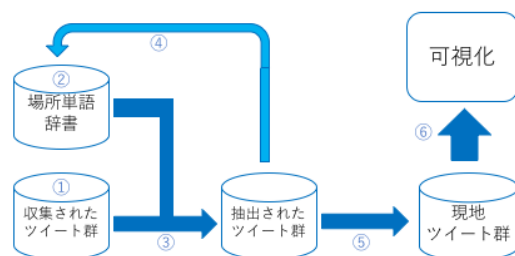


図 3 システムの処理フロー

4.2.3 ツイートの抽出

場所単語辞書をユーザ辞書として加え、mecab-ipadic-NEologd をシステム辞書として設定した MeCab を用いて形態素解析を行う。mecab-ipadic-NEologd とは、MeCab の標準のシステム辞書では対応していない固有表現や新語を定期的に自動で収録しているユーザ辞書である。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。表 3 内のツイート文に引かれた下線のように、1 つのツイートから複数の場所単語が検出された場合、それぞれ別で保存される。Bot クライアントによるツイートは除外する。また、位置情報つきツイート、Foursquare によるツイートについても、3.3 節で述べたように、そのまま利用して可視化するには件数が十分でないこと、そして本システムの処理の対象外であることから除外する。

表 2 場所単語と場所情報の例

場所単語	場所情報
ファビュラス叶組	東あ 26a
プレシャス M 組	東あ 26b
irodori	東ほ 02a
比村乳業	東 A 28a

4.2.4 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、表記ゆれや他の呼称といった、既に登録されている場所単語の中に対応する語を新たに場所単語として登録する。表記ゆれ、他の呼称について挙げると、略称、英語表記のカタカナ読み、対象の代表作などから愛称や俗称などがある。コミケでの例を表 4 に示す。例えば「ファビュラス叶組」、「プレシヤス M 組」は、芸能人の叶姉妹が出展しているサークルであるため、「叶姉妹」を追加している。この処理を終えたら、再度ツイートの抽出を行う。

4.2.5 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。ベースラインは、予備実験の結果、対象文書での TFIDF 上位 500 語を対象に、追加する素性は以下の 8 次元とする。

- URL の有無
- Twitter 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（【速報】など）
- 投稿時間

また、フィルタリングを行う前に、「Twitter for...(iOS または Android)」などの代表的な携帯端末でのクライアントを抜粋し、それらからのツイート以外は除外した。

4.2.6 地図へのプロット

処理されたツイートを最終的な現地ツイート群とし、時間帯ごと、場所ごとにデータを参照し、ツイート数を集計したものを、イベントの配置図などを元に得たデータを元に、各場所に座標を与えて地図にプロットする。場所ごとの集計については、個別の場所ごと、大分類の場所ごと、コミケの場合、東京ビッグサイトの東展示場のホールごとの 2 種類の分類で集計する。

5. 実験

5.1 実験目的

本研究で行う実験は、実際に構築した提案システムを用いることで、ツイートから人々の動きを可視化することができるかを、また、現地ツイートのフィルタリングの精度を検証することを目的に行った。

5.2 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。

表 3 ツイート抽出の例

場所情報	ツイート ID	ツイート文 (下線は検出された場所単語)
東 A28b	00001	パッと見た感じ、しらびさん、 <u>ワダアルコさん</u> 、 <u>ちょこ庵さん</u> 、佐々木少年さん、lack さん、サテーさん、TAa さんの列が長い
東 A05b	00001	パッと見た感じ、しらびさん、 <u>ワダアルコさん</u> 、 <u>ちょこ庵さん</u> 、佐々木少年さん、lack さん、サテーさん、TAa さんの列が長い
東 A61a	00002	<u>irodori2</u> 限！#C93

表 4 表記ゆれ、他呼称の例

サークル	表記ゆれ、 他呼称	分類
ファビュラス叶組、 プレシヤス M 組	叶姉妹	俗称
比村乳業	たわわ、比村サ ーキット	俗称
irodori	たつき監督、け ものフレンズ、 けもフレ	俗称、及びその略称
Dragon Kitchen	ドラゴンキッ チン	英語のカタカナ読み

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。

現地ツイートフィルタリングの学習データは、抽出したツイート約 12,000 件中から人手分類した約 1,200 件を用いた。10 分割交差検定によるフィルタリングの分類性能評価は、ベースラインのみの場合と素性を追加した場合とで、それぞれ表 5 の通りとなった。

表 5 現地ツイート分類性能評価

評価値	ベースラインのみ	素性追加
精度	0.820	0.841
再現率	0.854	0.883
F 値	0.843	0.863

また、追加実験として、2017 年 12 月 29 日から 12 月 31

日にかけて、C92 と同じ会場で行われた「コミックマーケット 93」(C93) を対象に、C92 とほぼ同様の条件で行った。また、この実験では、実際に会場に赴き、発生している行列の規模の調査を行った。

5.3 結果

C92 での実験結果を図 4、図 5、C93 での実験結果を図 6、図 7 に示す。

それぞれ 1 日目の 10 時から 11 時までと、15 時から 16 時までのデータが示されている。10 時から 11 時まででは、開場してすぐの時間帯のため、各人が目的のブースで賑わっていることがわかる一方で、15 時から 16 時まででは、イベント終了の時刻が近づき、ツイート数が減少している。この時間帯では、その日の感想を述べたツイートが見られた。

また、それぞれの実験での各処理を行った時点でのツイートの件数は表 6 の通りとなった。なお、C92 では各日 9 時から 16 時を対象に、C93 では 3 日間 72 時間全てを対象にし、それに加えて検索の実行頻度を上げたため母数としての総ツイート数が C92 から増加している。また、ツイートの抽出以降の件数は、複数の場所単語が検出された同一のツイートを重複して計上した。

6. 考察

C93 での実験を例に、全体のマーカの数推移を図 8 に示す。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った。このことから、時間の経過につれて、購入を終えた人、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サークルは、「ファビュラス叶組」「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。

イベント終了後の感想ツイートは、その日に訪れたサークルやそこで購入した頒布物の内容を自身のイベント参加を終えてからまとめてツイートされるものがあり、今回の実験では、フィルタリングで取り除かれる対象としていたが、そこに人々が動いたことを示す材料となる可能性が考えられる。

可視化できなかったものとして、場所単語を検出することができなかったサークル、ツイート自体が少ないサークルが挙げられる。前者については、表 7 で挙げるように、場所単語の表記ゆれや、その他呼称が存在し、場所単語辞書構築の際に特徴語の抽出からだけでは対応できなかった

ためであること、その場所単語が一般的な単語と合致してしまい、誤検出を多く発生させてしまったことが考えられる。後者については、イベントと SNS との相性が考えられる。このことに関して、C93 での実験で現れたツイート数と、現地調査で実際に見た行列の規模を比較した。その結果、ツイートが十分に存在する場所では列が形成されていたが、その逆は必ずしも成立しないということがわかった。コミックマーケットには様々なジャンルの出展物があり、各ジャンルのファン層によって SNS の利用頻度が異なる。また、列が形成されていても、ツイートをする余裕がない或いはするほどでもない規模である場合もある。表 8 で示すように、フィルタリングされた現地ツイートの午前 9 時から午後 6 時まででの件数と実際の来場者数⁹⁾を比較すると、両者に相関があるとは言い難いものになっており、前述の通りのことが言える。これらのことから、より効果的なイベントの種類と SNS の組み合わせを検討していく。



図 4 C92 混雑状況の可視化 (1 日目 10 時台)



図 5 C92 混雑状況の可視化 (1 日目 15 時台)

e) <http://www.comiket.co.jp/info-a/C92/C92AfterReport.html>

f) <http://www.comiket.co.jp/info-a/C93/C93AfterReport.html>

7. おわりに

混雑の改善をするための判断材料を得ることを狙いに、Twitter の投稿文書から現地ツイート特定し抽出することで、イベントの会場内の人々の動きの傾向を示すことができた。本研究の手法の 1 つである場所単語辞書の実装を応用すれば、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。

今後は、ツイートの話題ごとの分類と、時間経過に伴うツイート数の変化を、ブースごとに可視化を行う予定である。



図 6 C93 混雑状況の可視化 (1 日目 10 時台)



図 7 C93 混雑状況の可視化 (1 日目 15 時台)

表 6 各処理でのツイート件数

実験対象	ツイート収集後	ツイート抽出後	現地ツイートフィルタリング後
C92	128,791	12,604	3,977
C93	677,018	66,245	20,570

表 7 対応できなかった場所単語

サークル名	分類	備考
iXima	辞書拡張未対応	「いくしま」と読む
Yan-Yam	辞書拡張未対応	「Yan Yam」とハイフンを省略されることがある
行方不明	一般単語	迷子や紛失物を指しているツイートが誤検出された
ありがとう	一般単語	感謝を示す無関係なツイートが誤検出された

表 8 ツイート数 (9 時～18 時) と来場者数

対象日	ツイート数	来場者数
C92, 1 日目	2,232	160,000
C92, 2 日目	745	150,000
C92, 3 日目	1,018	190,000
C93, 1 日目	8,658	180,000
C93, 2 日目	5,666	160,000
C93, 3 日目	7,284	210,000

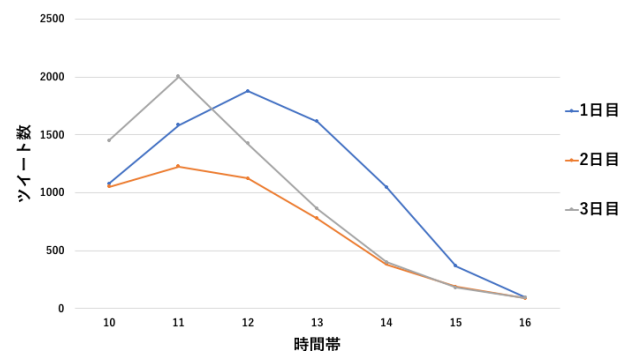


図 8 C93 でのツイート数推移

参考文献

- [1] 山崎裕太郎, 福原知宏, 山田剛一 and 増田英孝, "大規模イベントにおける現地ツイートをを用いた混雑状況可視化システムの提案", 第80 回情報処理学会全国大会, 1M-03 (2018)
- [2] 丸千尋, 榎美紀, 中尾彰宏, 山本周, 山口実靖 and 小口正人, "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, 6K-01, pp.529-530 (2016)
- [3] 石野亜耶, 小田原周平, 難波英嗣, and 竹澤寿幸. "Twitterからの被災時の行動経路の自動抽出および可視化." 言語処理学会 第 18 回年次大会: pp.907-910 (2012)
- [4] Zhang, Wei, and Judith Gelernter. "Geocoding location expressions in Twitter messages: A preference learning method." Journal of Spatial Information Science 2014.9: pp.37-70 (2014)