

Twitter を用いた 開催中のソーシャルイベントの状況把握に関する研究

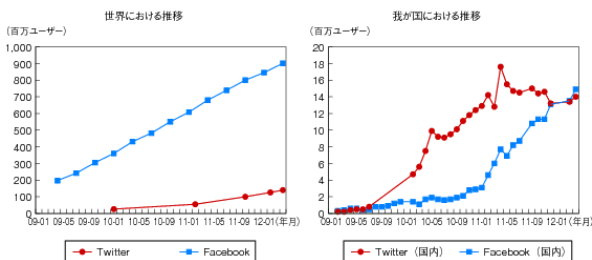
渡辺 大貴† 相場 亮‡

芝浦工業大学大学院 理工学研究科 電気電子情報工学専攻†

芝浦工業大学 システム理工学部 電子情報システム学科‡

1. はじめに

近年, Twitter[1]は情報収集のマイクロブログとして普及しており(図 1), 膨大な量のツイートデータをリアルタイムに情報を取得することができる. Twitter はソーシャルイベント(以下イベント)の情報を収集する手段としても有力であり, その情報を用いて行動する参加者も実際に存在する.



※ 各社公表のアクティブユーザー数を集計。 ※ アクティブユーザー数を集計。ネットレイティングス社公表資料、各社公表資料及び総務省資料により作成。

図 1 Twitter の利用者の変移[2]

本研究は, 開催中のイベントのツイートを分析することで, 混雑の状況を把握することを目的とする. 混雑とは, 入場待機列の状況やブースの混み具合, ブースの商品の完売情報のことである. イベントに関するタグでツイートを取得し, ツイート内容やリツイート情報を分析し, 任意のブースの混雑の情報を把握する.

2. 関連研究

村山ら[3]は, Twitter(論文ではマイクロブログ)を用いて交通機関で起きた事故などの情報を分類することで必要な情報を選別している.

坂本ら[4]は, Twitter におけるバースト(ある時間に記事が集中する状態)の断続性に着目することで, 過去のバーストの情報をを用いて新たなバーストを表わすようなキーワードを発見する手法を提案している.

本研究では, 不定期に起こる駅の混雑ではなく, 開催期間の限られているイベントに着目し, ブースの混雑情報のバーストをツイート情報から取得する.

3. 手法

混雑状況抽出の処理の流れを図 2 に示す. なお, 各構成要素は以下の通りである.

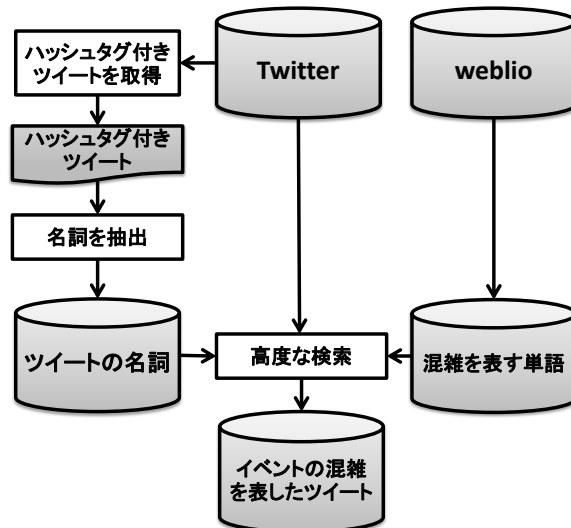


図 2 処理の流れ

3.1. ハッシュタグ付きツイートの取得

ツイートの取得には, 動的に取得する方法と静的に取得する方法があるが, 静的な取得方法は他方より API の制限が多いため, 本研究ではハッシュタグ(以下タグ)による動的な取得方法を実施する.

タグ(図 3)とは, 発言内に「#xxx」と入れて投稿すると, その記号付きの発言を検索することで, その話題のみのツイートを閲覧できる機能である.



図 3 ハッシュタグ[5]

イベントには, イベント公式もしくは参加者が設定したタグが参加者に浸透している. 本研究では各イベントのタグを検索することでそれに付随しているツイート情報を取得する. 取得する日時は, 対象のイベントの開始日 0 時から終了日 24 時までとする.

3.2. 単語の出現頻度を抽出

取得したツイートの単語の出現頻度を調べるために, 3.1で取得したツイートデータを対象に形態素解析を行う. 形態素解析には Mecab[6]を使用する. 単語の品詞は名詞, 品詞細分類は一般か固有名詞で抽出する. それらの単語を全体一致で出現頻度の高い順に並び替え, イベントに関連する名詞の上位 10 個の名詞をツ

Detecting Situations on Twitter for Social Events

†Daiki Watanabe

†Electrical Engineering and Computer Science, Graduate School of Engineering and Science, Shibaura Institute of Technology

‡Akira Aiba

‡Electronic Information System, College of Systems Engineering and Science, Shibaura Institute of Technology

イートの名詞の集合とする。

3.3. 混雑を表す単語

本研究では、混雑を表す単語を weblio[7]の類語辞典から設定する。weblio の類語辞典で「混雑」を検索すると表 1の結果となる。

表 1 混雑の類語

意義類	類語
大勢の人などが一か所に集まること	すし詰め, 寿司詰め, ぎゅうぎゅう詰め, 満員, 混雑, ラッシュ, 超満員, 過密, 人口過多, すし詰め状態, 満杯, 寿司詰め状態, 寿司づめ状態, 人混み, パンパン, いっぱい

ここでイベントの混雑に最適な単語を考えると、「混雑」と「人混み」の2つが考えられる。本研究では、混雑状況を表す単語はこの2つの単語とする。

3.4. 高度な検索

Twitter の高度な検索[8]を用いて混雑状況を抽出する。検索方法は、混雑を表す単語の集合の任意の要素とツイートの名詞の集合の任意の要素の AND 検索を行う。日付の選択には、タグで取得した日時と同じ日時を使用する。

以上から取得された混雑状況を分析する。

4. 適用実験

● 本論文での対象

対象イベント…東京ゲームショウ 2014[9]
 イベント日時…2014年9月18日(木)~21日(日)
 (ビジネスデイ…2014年9月18日~19日
 一般公開…2014年9月20日~21日)
 取得日時…2014年9月18日0時~21日24時
 来場者数…約25万人
 (ビジネスデイ約5人, 一般公開約20人)
 取得するタグ…#tgs
 取得ツイート人数…14342人
 (ビジネスデイ7233人, 一般公開7893人)
 取得ツイート数…23087ツイート
 (ビジネスデイ10566, 一般公開12521)

● 適用実験

「#tgs」というタグを付加したツイートを取得し、そのツイート情報を形態素に分割する。その中から出現頻度の高い上位10個の名詞は以下の通りになった。ゲーム, ステージ, ショウ, ハナヤマタ, コラボ, ブース, キャラコス, SEGA, PSNOVA, 情報(降順)

次にこれらの任意の単語と混雑を表す任意の単語で AND 検索を行い、混雑状況を抽出する。日付は2014年9月18日~21日を選択する。

● 結果

適用実験を行った結果、4日間で1125ツイートを、一般公開の2日間で779ツイートを抽出することができた。4日間のツイートの中から、混雑を表すと考えられるツイートは330ツイート、一般公開の2日

間では205ツイートを確認することができた。ビジネスデイの2日間では一般公開で混雑しそうなブースを予想するツイートが多く、一般公開の2日間では実際に混雑しているブースを示唆するツイートがあった。

5. 今後の展望

現段階では、イベントの混雑状況を表すツイートの抽出が確認できたが、抽出したツイートをリアルタイムで利用するためには、以下の2点が問題となる。

● 取得方法が限られている

動的取得には、無差別にツイートを取得する API も存在するが、現在の Twitter では全体の数%のツイートしか取得することはできない。本研究ではタグ付きツイートを取得しているが、イベントのツイート全体を取得しているわけではない。全てのイベント参加者がタグを付加すれば、混雑状況をより正確に把握できることが予想される。

● API 制限が存在する

通常の検索を使うこともできるが、ツイートの取得には API 制限があり、15分に15回しかツイートを取得できない。そのため、Twitter の API 制限の緩和や動的取得の機能拡張が望まれる。

上記の問題が解決されれば、Twitter の特徴の1つである即時性を利用し、リアルタイムにイベントの混雑状況を把握できるだろう。

参考文献

[1] Twitter, <https://twitter.com/>(2015年1月9日参照)
 [2] 総務省 平成24年度情報通信白書, <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h24/html/nc123220.html>(2015年1月9日参照)
 [3] 村山敬祐, 佐伯圭介, 遠藤雅樹, 横山昌平, 石川博, “マイクロブログマイニングによるイベント時の混雑原因の特定と状況の把握”, 第6回データ工学と情報マネジメントに関するフォーラム, 2014年3月3日
 [4] 坂本翼, 廣田雅春, 横山昌平, 福田直樹, 石川博, “Twitter ストリームのバーストの断続性に着目したキーワード抽出”, 第4回データ工学と情報マネジメントに関するフォーラム, 2012年3月4日
 [5] #ハッシュタグってどうやって使うの?, [https://support.twitter.com/articles/20170159-](https://support.twitter.com/articles/20170159)(2015年1月9日参照)
 [6] MeCab, <http://mecab.sourceforge.net/>(2015年1月9日参照)
 [7] weblio, <http://thesaurus.weblio.jp/>(2015年1月9日参照)
 [8] 高度な検索, <https://twitter.com/search-advanced?lang=ja>(2015年1月9日参照)
 [9] 東京ゲームショウ2014, <http://expo.nikkeibp.co.jp/tgs/2014/>(2015年1月9日参照)