

7K-03 SNS データを用いた場所と時間に対応する情報配信のためのイベント名取得手法

工藤 瑠璃子[†] 榎 美紀^{††} 中尾 彰宏^{†††} 山本 周^{†††} 山口 実靖^{††††} 小口 正人[†]
[†]お茶の水女子大学 ^{††}IBM Research - Tokyo ^{†††} 東京大学 ^{††††} 工学院大学

1. はじめに

2020年の東京オリンピック開催が確定した2013年以降、訪日外国人は急増しており、観光情報の配信手段を充実させる必要がある[1]。しかしながら、ガイドブックや有名観光サイトに掲載されることが少ないローカルイベントや今まさに開催されているイベントの情報は、取得するのは困難である。また、地理的、時間的制約を持つ観光客が必要とするのはその場で役立つ情報であるが、そのようなスポット的な情報を発信するシステムは少ない。そこで本稿では、旅行者などの時間とともに移動していく人に有用な情報をTwitter[2]から抽出し、インバウンド対応のタイムリーな情報提示手法を提案する。さらに、構造化されていないツイート本文から直接取得するのが困難な情報をWebから補足することで、より多くの情報を提供できることを示す。

2. 提案システム

観光者などに有用な情報をタイムリーにインバウンド対応で提示するために以下のシステムを提案する。提案システムの概要を図1に示す。本システムの詳細については[3]で議論されている。

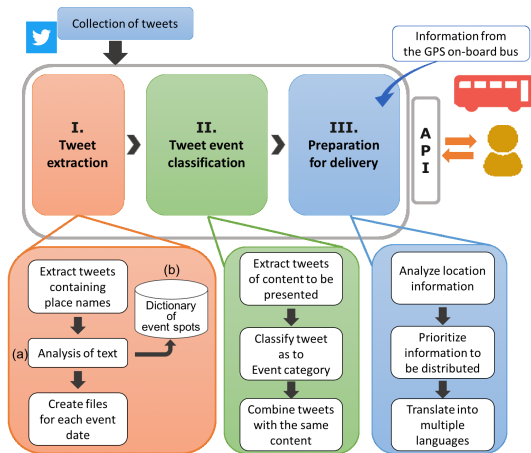


図1: 提案システムの概要

I. ツイートの抽出

1. Twitter APIのキーワード検索で地名をキーワードに設定し、地名に紐付いたツイートを収集
2. 取得したツイートの本文を解析し情報整理
 - (i) イベントの日時が記載されているツイートを抽出する。
 - (ii) 1. で設定した地名、イベントスポット辞書 (b)、正規表現を利用してイベント開催地を取得する。

Event name collection method for information distribution corresponding to place and time using SNS data

[†] Ruriko Kudo, Masato Oguchi
^{††} Miki Enoki
^{†††} Akihiro Nakao, Shu Yamamoto
^{††††} Saneatsu Yamaguchi
 Ochanomizu University ([†])
 IBM Research - Tokyo(^{††})
 Tokyo University (^{†††})
 Kogakuin University(^{††††})

- (ii) 1. で設定した地名、イベントスポット辞書 (b)、正規表現を利用してイベント開催地を取得する。

II. ツイートのイベント分類

1. 抽出されたツイートをイベントのカテゴリごとに分類

III. 配信準備

1. 不完全な文でツイートされる場合も多いため、ツイート中の必要な情報を整理
イベント開催日時、場所、イベントのカテゴリ、イベント名等を取得する。
2. 提供する情報を複数言語に翻訳
3. ユーザの位置情報を取得
4. 提供する情報の優先順位付け

3. 配信のためのイベント情報の取得

本節では、実際のツイートを用いて本システムでカテゴリ分類したツイートを対象に優先順位付けを行う部分について検証する。どのような順で情報提示すればユーザにとって有益なのかはユーザによって異なり、例えば、順位付けの条件として、イベントの開催地や人気度、開催時間、金額が考えられる。今回はその中の一例として、Distance Matrix API[4]を使用して車による移動時間の小さい順に順位付けを行う、イベント開催地に着目した実験結果について述べる。

3.1 ツイートのみによる情報取得

ユーザの位置に「新宿駅」を選択し、「音楽イベント」の情報に順位付けを行った結果の一部を図2に示す。実験の詳細については[3]に説明されている。図2に示されるように、移動時間に基づいて配信順序が決定されることがわかる。しかし、図2中の(i, ii, iii)のように、イベント名が正しく取得されないものがあった。そこで、次節にてイベント名の取得精度の向上を試みる。

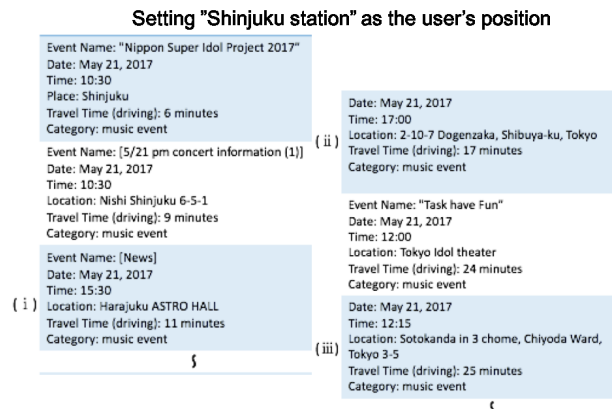


図2: 順位付け結果

3.2 Webからの情報取得

ツイートには文字数以外の制約はないため、記述の自由度が極めて高い。記述がある程度形式化された日時などの情報は、正規表現を使用して自動的に抽出することが可能であるが、「イベント名」のような形式の定まっていない記述は、自動的に抽出することが困難である。そこで、ユーザに提示する有益な情報を増やすために、外部情報を用いて情報を補足する機能を提案する。

3.2.1 外部情報を利用した情報取得手法

外部情報を用いた情報補足機能の概要を図3に示す。

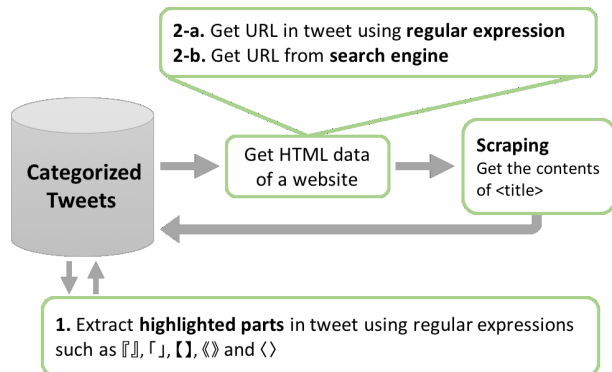


図3: 外部情報を用いた情報補足機能

正規表現を用いて強調された部分をツイートから抽出するという前節までに用いていた手法(以下、手法1)に加え、外部情報を利用する手法として、スクレイピングを行い、提示するイベントのWebサイトのHTMLから、titleタグの中身を取得する手法を利用する。スクレイピングを用いる手法においては、ツイートのWebサイトのURLが記載されている場合は、正規表現でURLを抽出しHTMLを取得する(以下、手法2-a)。URLの記載がない場合は、ツイートを検索エンジンにかけた結果の上位数件のWebサイトのHTMLデータを取得する(以下、手法2-b)。

3.2.2 外部情報を利用した情報取得の結果

手法1,2-a,2-bを適用した結果のユーザに有用な情報の取得率を評価する。実験データは2017年11月1日~2017年11月17日に収集した中で、イベント開催日が2017年11月18日(土)のツイートである。正解と判定する基準は以下の2通りを設定した。

1. イベント名が取得結果に含まれている
2. イベント名、開催場所、出演者の名前のいずれかが取得結果に含まれている

各々の正解基準に対する結果を表1,2に示す。正解とする基準が厳しい基準1において0.66、正解とする基準を一段階広げた基準2においては、0.80という取得率が得られた。

表1: 基準1を用いた結果

手法	対象件数	正解数	取得率
全手法を適用	979	643	0.66
手法1	604	456	0.75
手法2-a	77	45	0.58
手法2-b	404	142	0.35

表2: 基準2を用いた結果

手法	対象件数	正解数	取得率
全手法を適用	979	784	0.80
手法1	604	473	0.78
手法2-a	77	51	0.66
手法2-b	397	260	0.65

それぞれの手法を順に適用した結果、最終的にユーザに提示可能な有用な情報は図4になる。図4に示されているように、ツイートそのものから取得可能な情報は図4の青い斜線部のみなのに対し、外部から情報を補足することで、有用な情報が赤い斜線部の分だけ増すことが分かる。

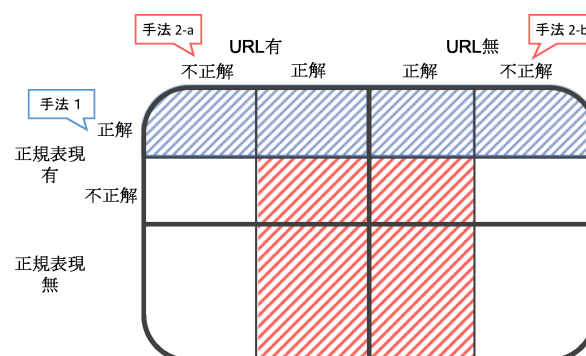


図4: 取得可能な情報の全体像

4. おわりに

場所と時間を考慮したその場その時に旅行者などが利用出来るイベント情報の配信システムを提案した。実際にツイートをを用いて検証し、より正確なイベント情報を提供するには、Webから情報を補足することが有効であると分かった。

5. 謝辞

本研究は一部、JST CREST JPMJCR1503の支援を受けたものである。

参考文献

- [1] 日本政府観光局「年別訪日外客数、出国人数の推移」<http://www.jnto.go.jp/jpn/statistics/marketingdata-outbound.pdf>
- [2] Twitter, <http://twitter.com/>
- [3] Ruriko Kudo, Miki Enoki, Nakao Akihiro, Shu Yamamoto, Saneyasu Yamaguchi, and Masato Oguchi, "Real-Time Event Search Corresponding to Place and Time using Social Stream", The 3rd IEEE International Conference on Big Data Intelligence and Computing(DataCom2017),pp.1047-1053,2017.
- [4] Google Maps API, <https://developers.google.com/maps/documentation/distance-matrix/?hl=ja>