

SNS データを用いた場所と時間を考慮する観光情報配信システム

理学専攻・情報科学コース 1740654 工藤 瑠璃子 (指導教員: 小口 正人)

1 はじめに

2020年の東京オリンピック開催の決定や格安航空会社の増加, ビザの緩和により, 日本を訪れる外国人観光客の数は急増している [1]. 日本政府は2020年までに訪日外国人数4000万人を目標に掲げており, 観光庁をはじめとする多くの機関が訪日外国人旅行者の受入環境の整備・充実を総合的に推進している. さらに観光客増加に向けて, 快適に観光を満喫できる環境整備を進める必要があり, それぞれの観光客の嗜好に合わせた多様な情報を発信すること, 観光客にとって使いやすい形式で情報を提供することが重要になっている.

既存の外人向けの情報配信としては, ガイドブックによる王道ルートの提示や, 外国人観光客向けのWebサイトが挙げられるが, ガイドブックによる王道ルートなどの提示は恒常的に有名な観光地に関する情報が多く, 多様な情報は取得しにくいのが現状である. 一方で, SNSには情報が存在するが, ホームページのような固定的なメディアからは殆んど取れない観光情報やイベント情報も多く存在する. 特に, SNSの代表であるTwitter[2]には, ホームページのような固定的なメディアには載っていない情報や, ローカルイベントや地域特有の情報など, 大小様々な有益な情報が含まれている. ターゲットの対象となる趣味・嗜好を持つ人が多くないイベントは, 大々的なホームページなども作られず情報を探し出す事は難しいが, 特定のジャンルに興味を持つ人に向けては, そのようなニッチな情報が非常に価値ある可能性が高い. しかし, これまでそのような所をターゲットとした検索・推薦システムは存在しなかったため, ニッチな情報を取得しようとする際には膨大な情報量の中から自力で探さなければならず, 容易に有用な情報を取得するのは困難であった. 多様化する観光客のニーズに応えるためには, バリエーションに富んだ情報の発信が必要であるため, Twitter上の情報を利用した検索・推薦システムは有効である.

しかしながら, Tweetのような自然言語でかかれた非構造データから, 情報提示に必要な, 日時, 場所, イベント名の情報を綺麗に抜き出すことは困難であるため, 本研究では, この課題に取り組み, Twitterから広く大小様々なイベントを取得する. そして, Twitterから抽出した大小様々な多様な情報, 場所と時間を考慮した「その場」「その時」に観光者などが利用出来る, スポット的な有用性の高い情報を, インバウンド対応でタイムリーに配信するシステムを提案する.

2 提案システム

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 [6] で議論されている.

1. ツイートの抽出

Twitter API[3] 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. イベントに関するツイートを取得

イベント情報を含むツイートかSVMを用いて判

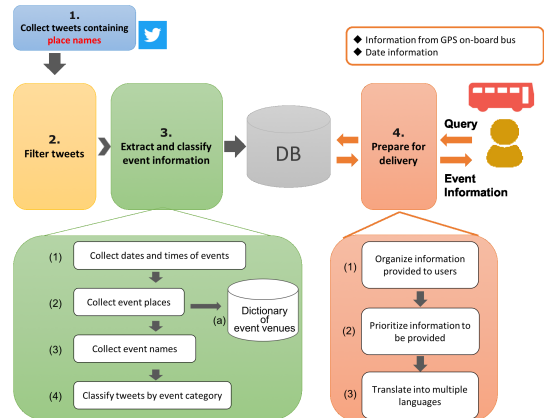


図1: 提案システムの概要

別しフィルタリング

3. イベント情報の抽出とカテゴリ分類

- (1). ツイート本文を解析し, イベントの開催日時を取得
ローカルイベントには, 1日だけ開催されるイベントと, 数日から数ヶ月などの長期間にわたって開催されるイベントがある. そこで, 提案システムでは, 1日だけ開催されるイベントと長期間にわたって開催されるイベントを区別し, 別の処理を行うことで, 正確なイベント開催日を取得することにした. イベント開催日は複数の正規表現を使用して取得する. 長期間にわたって開催されるイベントは, “○/○~○/○”, “○/○から○/○まで”, “○年○月○日-○日” のような記載を取得し, イベントの開始日と終了日を特定する. 1日だけ開催されるイベントの場合は, 日付の記載を使用して取得し開始日と終了日に同じ日付を保持する. 時間についても正規表現を使用して, 時間に関する表記が含まれているかどうかを判別する. 上述の処理を行い, 日付と時間がツイート本文に含まれているもののみを抽出する.
- (2). ツイート本文を解析し, イベントの開催場所を取得
キーワード検索で設定した地名は, 山手線の駅名, 東京23区の地名等の粒度が非常に大きな地名であるため, これらをイベント開催地として使用するのは望ましくない. そこで, 東京23区内のアミューズメント施設, ミュージアム, ショッピング施設, エンターテインメント施設, 温泉, 劇場, ホールを東京ウォーカー [4] とナビタイム [5] から抜粋しイベントスポット辞書を作成した. 辞書に登録されたスポットの記載がツイートにある場合は, そのスポットをイベント開催地としてJSONファイルに保持する. また, イベント開催地の住所がテキストに記載されている場合は, 正規表現を使用して住所を取得し, 保持する. 上述のどちらも記載されていない場合は, キーワード検索で設定した地名を保持する.

表 1: Event Table (一部省略)

Event Name	Place	Start date	End date
ウルトラマンフェスティバル 2018	サンシャインシティ	2018/07/20	2018/07/31
「自由が丘スイーツフェスタ 2018 Retro Fes.」	目黒区自由が丘...	2018/05/06	2018/05/06
第十一回 銀座柳まつり	銀座	2018/05/05	2018/05/05
あしたのショー展	東京ソラマチ	2018/04/28	2018/05/06

(3). イベント名を取得

ツイートのような非構造化データにおいて、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、本研究では、外部情報を用いた情報補足手法を使用している [7]。

(4). ランダムフォレストを用いてカテゴリ分類

カテゴリ分類を行い抽出したツイートを分類する。複数の手法でイベント分類評価の比較を行った結果、分類精度、処理速度ともにランダムフォレストが優れていることがわかっている [6]。

表 1 にイベント情報のデータベースの例を示す。

4. 配信準備

(1). ユーザに提供する情報を整理

(2). 提供する情報の優先順位付け

実際に運行されているリムジンバスと観光バスのルートに基づいて、ユーザの移動方向を考慮した順位付けを行う。順位付けの詳細については、[7] で議論されている。

(3). 提供する情報を複数言語に翻訳

インバウンドに対応するため、複数言語に翻訳する。

3 評価

本節では、従来の手法である Web サイトからの情報取得に比べ、本手法を用いることでどれくらいのイベントが取得できるか検証を行う。

従来手法には、有名イベント情報サイトである、「Walker+」を用いた。評価は、2018 年 7 月 14 日から 2018 年 7 月 18 日の 5 日間に、「渋谷」、「新宿」、「池袋」、「お台場」の 4 箇所で開催されたイベントを対象として行った。実験データには 2018 年 6 月 1 日から 2018 年 7 月 18 日に収集したツイートをを用いている。

従来手法である Web サイトだけをを用いたイベント情報の取得結果と本手法を合わせて用いた場合の取得結果の比較を図 2 に示す。

グラフから分かる通り、提案手法を用いることで取得可能なイベント情報はどの日付、どの地域においても増加する。「Odaiba」では、他の地域に比べて増加率は低い結果となったが、これは、お台場には有名観光スポットが多く、そのような場所で開催されるイベントは、そのスポットの公式アカウントからツイートされることが多いため場所に関する情報の記載がないこと、そのイベントの Web ページが作成されていて Web ページに誘導することで詳細な情報を提供する場合が多いことが原因である。このような場合に関しては、Walker+などのサイトで頻出するスポットを学習し、ツイート

をクロールする際のキーワードに含めること、ツイート中に記載されている URL 先の Web ページも解析することで提案システムにおいても対応できる。

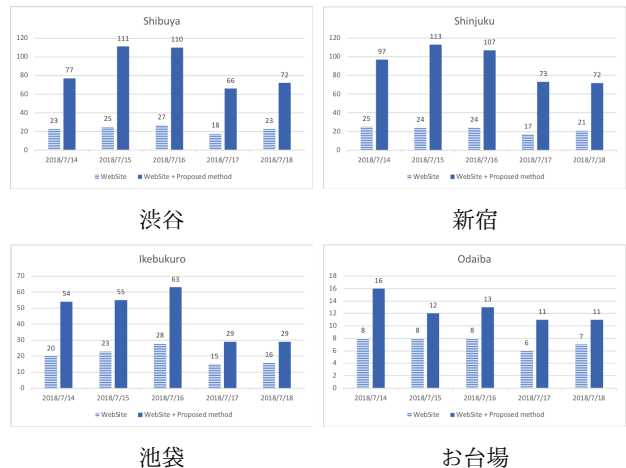


図 2: Number of collecting

4 おわりに

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 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した。また、既存の情報配信手段との比較による評価を行った。

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, https://www.jnto.go.jp/jpn/statistics/since2003_tourists.pdf.
- [2] "Twitter," <http://twitter.com/>
- [3] "Twitter Search API," <https://dev.twitter.com/rest/public/search>
- [4] "Walker+ 東京都のおでかけスポット一覧," http://www.walkerplus.com/spot_list/ar0313/
- [5] "NAVITIME," <https://www.navitime.co.jp/category/>
- [6] 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.
- [7] 工藤 瑠璃子, 榎 美紀, 中尾 彰宏, 山本周, 山口 実靖, 小口 正人, "場所と時間を考慮した SNS データを用いる訪日外国人観光客へのタイムリーな情報配信", 第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM2018), B1-3.