

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

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 〒 112-8610 東京都文京区大塚 2-1-1

^{††} 日本アイ・ビー・エム株式会社 〒 103-8510 東京都中央区日本橋箱崎町 19-21

^{†††} 東京大学 〒 113-8654 東京都文京区本郷 7-3-1

^{††††} 工学院大学 〒 163-8677 東京都新宿区西新宿 1-24-2

E-mail: [†]{g1320514,oguchi}@is.ocha.ac.jp, ^{††}enomiki@jp.ibm.com, ^{†††}{nakao,shu}@iii.u-tokyo.ac.jp,
^{††††}sane@cc.kogakuin.ac.jp

あらまし 2020 年の東京オリンピック開催決定を受け、訪日外国人旅行者は堅調に増加を続けている。2020 年の政府目標である訪日外国人旅行者数 4000 万人に到達するためには、情報配信手段を今以上に充実させる必要がある。特に、場所と時間の制約がある旅行者などが必要とする「その場」「その時」に関する情報を発信している媒体は少ないため、本稿では観光に来た土地勘の無い旅行者などに有益な、その場で役立つスポット的な情報をインバウンド対応でタイムリーに配信可能にするシステムを提案する。

キーワード Twitter, SNS, 観光情報, 情報抽出

Timely Information Distribution to Foreign Tourists Visiting Japan Using SNS Data Considering Place and Time

Ruriko KUDO[†], Miki ENOKI^{††}, Akihiro NAKANO^{†††}, Shu YAMAMOTO^{†††}, Saneyasu
YAMAGUCHI^{††††}, and Masato OGUCHI[†]

[†] Ochanomizu University 2-1-1 Otsuka, Bunkyo-ku, Tokyo, 112-8610, Japan

^{††} IBM Research - Tokyo 19-21 Hakozakicho, Nihonbashi, Chuo-ku, Tokyo, 103-8510, Japan

^{†††} University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8654, Japan

^{††††} Kogakuin University 1-24-2 Nishi-shinjuku, Shinjuku-ku, Tokyo, 163-8677, Japan

E-mail: [†]{g1320514,oguchi}@is.ocha.ac.jp, ^{††}enomiki@jp.ibm.com, ^{†††}{nakao,shu}@iii.u-tokyo.ac.jp,
^{††††}sane@cc.kogakuin.ac.jp

1. はじめに

東京オリンピック開催決定前の 2012 年の訪日外国人数は 8,358,105 人であったのに対し、2013 年は 10,363,904 人、2014 年は 13,413,467 人、さらに 2015 年は 19,737,409 人に上り、東京オリンピック開催が確定した 2013 年以降、訪日外国人数は大幅に増加している [2]。それに伴い観光情報の配信手段を充実させる必要があるが、観光客等を対象とした情報配信に関して以下のような問題があげられる。まず、恒常的に人気のある観光スポットに関する情報は、ガイドブックや有名観光サイトから取

得可能であるが、ローカルイベントや今まさに開催されているイベントに関する情報は、ガイドブック等には掲載されていない。そのため、そのような情報を取得するには膨大な情報量の中から自力で探さなければならず、有用な情報を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しているのはその場で役立つ情報であるが、「その場」「その時」に関する情報を発信しているシステムは少ない。その場で役立つスポット的な情報が配信可能になれば、観光に来た土地勘の無い旅行者などに有益だと考える。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 [1] 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、Twitter のみを情報源とするのではなく、外部情報を利用することで、より多くの有益な情報をユーザーに提示可能にする。

また、提案システムは以下のような IoT デバイスを基盤と考えて構築する。東京大学では 2015 年にネットワーク仮想化に対応したアクセスポイントを用いて、様々な情報配信サービスを行う実証実験を実施している [3]。この実証実験では、Wi-Fi の無線ビーコンに情報をのせユーザーのスマートフォンに情報を配信する BeaconCast 技術を実装したアクセスポイントを、空港と都内を結ぶリムジンバスに設置しており、インターネットにアクセスできない場合のメッセージ配信を実現している。リムジンバスの乗客は、専用アプリケーション「LimoCast」をインストールすることで、車内に配置されたアクセスポイントを利用して、通常の無線 LAN アクセスによるインターネット接続を楽しむと同時に、リアルタイムに更新されるバス到着地周辺の情報を受け取ることが可能になる。これは、通信契約をしていないインバウンドへの情報配信手段としても有用である。また、本研究の提案内容は、現在地を把握できインターネットに接続できる端末を持つ特定のユーザーに対し情報を配信できる環境であるなら利用する事ができるため、汎用性の高い手法となっている。

2. 関連研究

Twitter から観光情報や地域特徴を抽出する手法は、前田ら [4]、長谷川ら [5]、中川ら [6]、佐伯ら [7] によっても提案されている。前田らの手法では、Twitter 位置情報を用いて、人の高頻度滞在地と低頻度移動地を求め、移動件数と距離から地域の魅力を算出し、その魅力が固有性によるものなのか、利便性によるものなのかを区別することで観光地を抽出している。長谷川らは、Twitter 上に投稿されたコンテンツの中から、地域の特徴を表す特徴語を抽出して地域特徴語辞書の構築を行い、構築された辞書を利用して Twitter からユーザーの観光体験を検索する手法を提案している。中川らは、ツイートされた時刻に着目して観光ルートを推薦する手法の問題点を改善し、観光スポット間の距離を考慮した観光ルートを推薦するための手法を提案している。また、佐伯らは、外国人のツイートに出現する地名に注目し外国人向けの観光情報抽出の手法を提案している。これらの研究では、蓄積された情報を解析するというスタティックな情報提供になっている。本研究ではダイナミックに、リアルタイムに情報を収集して配信することを目指している点で、これらの研究とは異なる。

タイミングを考慮した情報推薦手法の研究として、向井ら [9]、矢部ら [8] があげられる。向井らは、時系列におけるツイートの急激な増加 (バースト) が発生した時、そのタイミングで何らかのイベントが起きており、ユーザーの興奮度合いが上昇し寛容な気分になっている可能性が高いという考えに基づき商品情報の推薦を行う手法を提案している。矢部らは、ユーザーの地理的・時間的な状況の変化に応じて、意味解釈を伴う情報選択に基づく

情報配信システムを提案している。配信情報とユーザーの関係性を計量するためのベクトル計量空間を設定し、ユーザーの地理的状況と配信情報の特徴量を動的に計量することに成功している。

3. 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 [11] で議論されている。

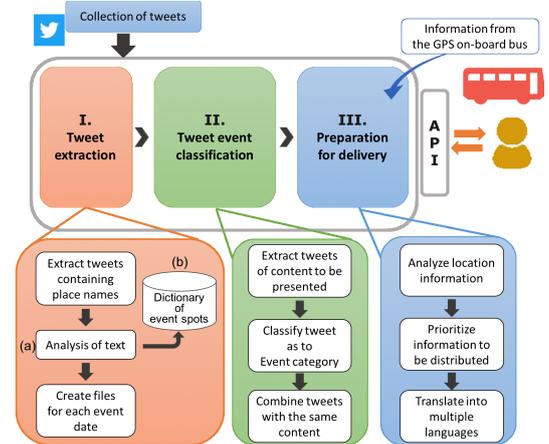


図 1 提案システムの概要

1. ツイートの抽出

1. Twitter API [10] 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. 取得したツイートを解析し情報整理

取得した JSON ファイルからツイート本文、アカウント名、ID、RT 数、公式アカウントかどうか、フォロワー数を抜粋し、JSON ファイルにまとめ直す。さらに、イベント開催日時、開催場所を、ツイート本文を解析することにより取得する。

[イベントの開催日時の取得]

キーワード検索で取得したツイートのテキストに日付と時間が記載されているかを、正規表現を使用して判定する。日付と時間が記載されているツイートを抽出し、イベント開催日ごとのファイルにまとめる。

[イベント開催地の取得]

本システムでは、山手線の駅名、東京 23 区の地名をキーワードに設定しツイートを取得している。イベント開催地として、キーワード検索で設定した地名を使用することは、地名の粒度が非常に大きいため望ましくない。そのため、本システムでは、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 [13] とナビタイム [14] から抜粋しイベントスポット辞書を作成する。ツイートにイベントスポット辞書内の要素が存在する場合は、その要素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に対応するため、キーワード検索で設定した地名も JSON ファイルに保持する。

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次に、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 [11]。ランダムフォレストによるカテゴリ分類の結果を表 1 に示す。カテゴリ「舞台」以外では、F 値が 0.8 前後になり、高い精度で分類が行われている。

表 1 Performance of classification for each category.

Category	Accuracy	Precision	Recall	F-measure
Music Event	0.8950	0.8750	0.7583	0.8125
Vaudeville/Comedy	0.9550	0.7948	0.7561	0.7750
Theater	0.9875	0.5000	0.4000	0.4444
Movie	0.9925	1.0000	0.7272	0.8421
Art	0.9875	0.8666	0.8125	0.8387

III. 配信準備

1. ユーザに提供する情報を整理

不完全な文でツイートされる場合も多いため、イベント名、開催日時、カテゴリ等のユーザに提供する情報を取得し、整理する。

2. 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、整理した情報を情報通信研究機構が開発した自動翻訳エンジン「みんなの自動翻訳@TexTra」[12] を使用し、複数言語に翻訳する。

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

4. 情報の配信スケジュールリング

本章では、実際のツイートを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。

4.1 移動時間に基づいた情報配信

本節では、Distance Matrix API [15] 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。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 [11] に説明されている。図 2 に示されているように、移動時間に基づいて配信順序が決定されていることがわかる。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見られる。

この課題については、5. 章にて外部情報を利用したイベント情報の取得精度の向上を試みる。

4.2 バスのルートに基づいた情報配信

前節で述べた移動時間による順位付けでは、単にユーザの位置から移動時間が小さい順に並べ替えているだけになっており、ユーザの移動方向と逆方向で開催されているイベントも提示されることになる。本システムが想定しているユーザは、バスの

Setting "Shinjuku station" as the user's position

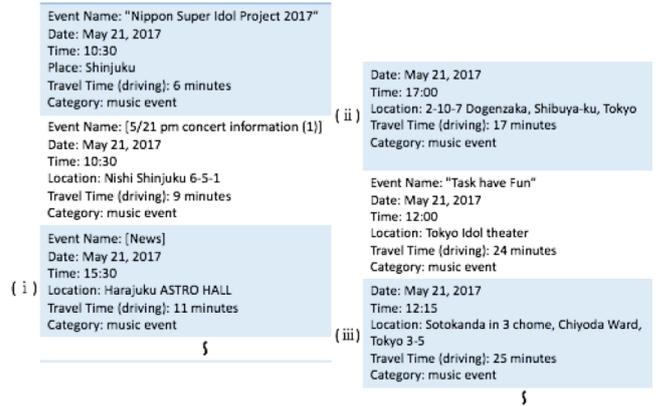


図 2 順位付け結果

乗客であるため、バスの移動方向に逆らった方向のイベント情報が提示されるのは、ユーザにとって不要な情報も提示されることになってしまい好ましくない。そのため、本節では、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い、実際のデータを用いて実験を行なった結果について説明する。

バスのルートに沿って移動方向先のイベント情報のみを提示するために、本研究では、配信するイベント情報を登録するコレクション (以下、テーブル 1) とバスの運行ルート情報を登録するコレクション (以下、テーブル 2) をデータベースに作成した。データベースにはドキュメント指向データベースである MongoDB を用いた [18]。

テーブル 1 は、提案システムにおいてカテゴリ分類された実際のツイートを管理するためのものであり、イベント名、イベント開催地などの前章で整理された情報が登録される。また、緯度・経度に基づいて地域をほぼ同じ大きさの網の目に分けたものである地域メッシュを識別するための地域メッシュコードを、イベント開催地の緯度・経度から取得し、Mesh ID として登録する。緯度・経度は、Google Maps Geocoding API [19] を使用して、前章の [イベント開催地の取得] において取得したイベント開催地をもとに取得する。今回、地域メッシュコードには、図 3 に示すような、緯度差が 30 秒、経度差は 45 秒で、一辺の長さが約 1 キロメートルである第 3 次メッシュを使用した。

テーブル 2 には、停留所や各停留所までの移動時間の情報が登録される。こちらのデータベースにもバスの停留所の緯度・経度から地域メッシュコードを取得し、Mesh ID として登録しておく。停留所が含まれる地域メッシュコード (以下、Main Mesh ID) だけでなく、そのメッシュを中心として東西南北に隣り合っている地域メッシュコード (以下、Sub Mesh ID) もセットで登録しておくことで、停留所から各方向に 2 キロメートルまでの範囲をカバーできるようにする。

配信するイベント情報を登録するテーブル 1 および、バスの運行ルート情報を登録するテーブル 2 の例を表 2,3 に示す。テーブル 2 の Time は始発停留所からのバスの移動時間である。

バスの移動方向を考慮した順位付けは、以下の手順に従う。

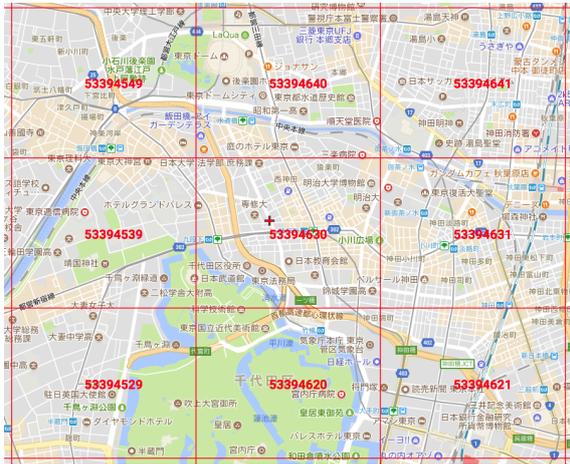


図3 第3次メッシュの例

表2 テーブル1 (一部省略)

Event Name	Place	Latitude	Mesh ID
アラジン	電通四季劇場 [海]	35.6647751	53393690
泥棒役者	六本木ヒルズ	35.6604638	53393598
大仙市ふるさとフェア	有楽町	35.6732638	53394600
アクション 2017 お台場	パレットタウン	35.6259917	53393652

表3 テーブル2 (六本木・お台場コース, 一部省略)

Bus Stop	Main Mesh ID	Sub Mesh ID (North)	Time (min.)
六本木ヒルズ	53393598	53394508	47
東京タワー	53393599	53394509	58
ダイバーシティ東京	53393642	53393652	75
東京レポート駅	53393652	53393662	81

- (1) ユーザが乗車しているバスの位置情報を取得
- (2) 取得した位置情報を利用し、テーブル2から直近で通過した停留所のレコード A を取得
- (3) 取得したレコード A よりも Time の値が大きいレコードを取得 (バスの移動方向先の停留所が含まれるレコードを取得)
- (4) 取得した各レコードに登録されている Main Mesh ID と Sub Mesh ID を取得
- (5) テーブル1から、(4)で取得した Main Mesh ID と Mesh ID が一致するレコードを取得
- (6) 取得したレコードをイベント開催時間でソート
- (7) 各レコードからユーザに提示する開催時間や開催場所などの情報を取得し整理する
- (8) バスの各停留所から1キロメートル以内のイベントとしてユーザに提示する
- (9) テーブル1から、(4)で取得した Sub Mesh ID と Mesh ID が一致するレコードを取得
- (10) レコードをイベント開催時間でソート
- (11) 各レコードからユーザに提示する開催時間や開催場所などの情報を取得し整理する
- (12) バスの各停留所から各方面に2キロメートル以内のイベントとしてユーザに提示する



図4 六本木・お台場コース

バスのルートの一例として、成田空港と都内を結ぶリムジンバスである「Airport Limousine」[16]の池袋方面の路線と、都内を巡回する観光バスである「SKY BUS TOKYO」[17]の六本木・お台場コースを用いて実験を行った。六本木・お台場コースのコース図を図4に示す。図中のAからJの停留所は以下の通りである。

- A. 丸の内三菱ビル
- B. ホテルニューオータニ
- C. 東京ミッドタウン
- D. 六本木ヒルズ
- E. 東京タワー
- F. フジテレビ ダイバーシティ東京 プラザ
- G. 東京レポート駅 ヴィーナスフォート前
- H. 東京ベイ有明ワシントンホテル (東京ビックサイト)
- I. 豊洲 (ららぽーと豊洲)
- J. 築地・銀座 (築地市場・歌舞伎座)

ユーザが乗車している六本木・お台場コースのバスがC. 東京ミッドタウンを出発したと仮定した場合の実験結果を示す。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)のツイートを用了。実験結果としては、停留所DからJ周辺で開催されるイベントの情報が、各停留所ごとに表示される。次の停留所であるD. 六本木ヒルズにおけるイベント情報は図5のようになった。

図5に示されているように、停留所D. 六本木ヒルズ周辺のイベント情報が停留所からの距離とイベント開催時間に応じて、簡潔に表示されていることが分かる。

5. 配信のためのイベント情報の取得

本章では、前章の図2から分かるように、ツイートからのみの情報抽出では正確な情報取得が困難な場合に対応するための手法について検討する。

5.1 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

- Within 1 km -
Time : 10 : 00 Location : Mori Museum Category : Exhibition
Yuyu Kawakami : Solo Event 2017 - Mysterious Grand Club TYPE - A - Time : 11 : 30 Location : Roppongi Categories : music events
[The First stage of the burglar actors] Time : 11 : 55 Location : Roppongi Categories : movies
⋮
"Thief Actors" Time : 15 : 20 Location : Roppongi Hills Categories : movies
"I want to become an actor," Time : 17 : 15 Location : Roppongi Categories : movies
⋮
- Within 2 km (north) -
-[Mone Kamishiraishi LIVETOUR 2017 "and". Time : 17 : 00 Location : Akasaka BLITZ Categories : music events
⋮

図5 バスの移動方向先の停留所周辺のイベント情報

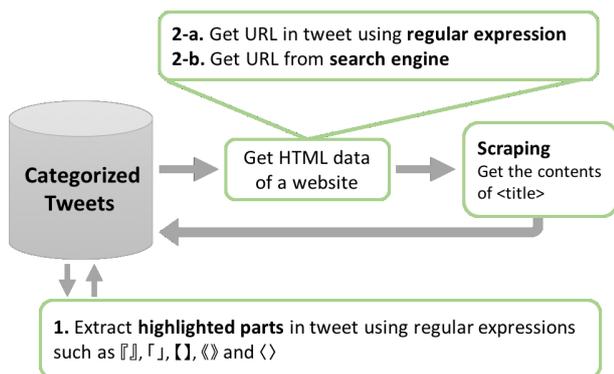


図6 外部情報を用いた情報補足機能

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。

5.1.1 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6に示す。

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

表4 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5 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

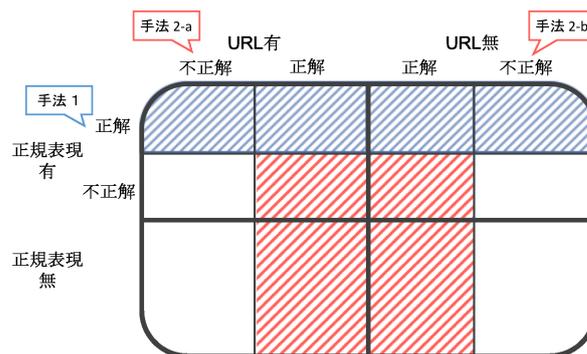


図7 取得可能な情報の全体像

法2-b)。

本システムでは、まず、手法1が適用可能なツイートに対し手法1を適用し、正解とする情報が取得できたか判定を行う。手法1において正しく情報が取得できなかった場合、手法2-aまたは手法2-bのうち、適用可能な手法で再度、情報取得を試みる。このように順に手法を適用した結果、最終的に有用な情報がどのくらい取得できるかを次節にて評価する。

5.1.2 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4,5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。

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7になる。図7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7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。

6. まとめと今後の課題

ガイドブックや検索エンジンなどから容易に取得できない情報、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 Twitter から抽出し、インバウンド対応のタイムリーな情報配信システムを提案した。イベントをどのような順で配信すれば、よりユーザーにとって使いやすいシステムになるのか検討するため、ユーザーの現在地からの移動時間に基づいた順位付けと、実際のバスの運行ルートに基づき、移動するユーザーのベクトル方向を考慮した順位付けを行った。また、実際にツイートを用いて検証し、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。

今後は、イベント情報の配信スケジュールについて更に検討を重ねたい。現在は、位置情報を基準に配信スケジュールを組んでいるが、ユーザーの移動にかかる時間とイベント開催時間との兼ね合い等も考慮するため、時間情報も基準に含めたスケジューリングを可能にしたい。また、イベントの人気度などの他のパラメータも取り入れた配信スケジュールの評価も行いたい。

文 献

- [1] "Twitter," <http://twitter.com/>
- [2] 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>
- [3] 訪日外国人向け O2O マーケティングも視野に iPhone・Android 両スマートフォン向け新規技術「BeaconCast」アプリを用いて 高速リムジンバス内での情報配信を行う実証実験を共同で実施 <http://www.u-tokyo.ac.jp/public/public01-261023-j.html>
- [4] 前田 高志ニコラス, 吉田 光男, 鳥海 不二夫, 大橋 弘忠, "Twitter 位置情報・テキスト情報を用いた人の移動モデルの構築と観光地推薦手法の提案", 人工知能学会 合同研究会 2015 第 9 回 データ指向構成マイニングとシミュレーション研究会
- [5] 長谷川 馨亮, 馬 強, 吉川 正俊, "Twitter からの地域特徴語辞書の構築と その観光情報検索への応用", DEIM Forum 2014 B3-4.
- [6] 中川 智也, 新妻 弘崇, 太田 学, "マイクロブログを利用した観光ルート推薦における移動効率の改善", DEIM Forum 2016 H1-3.
- [7] 佐伯 圭介, 村山 敬祐, 遠藤 雅樹, 横山 昌平, 石川 博, "外国人向け観光プロモーション支援のためのソーシャルメディア分析手法の研究", DEIM Forum 2014 B4-2.
- [8] 矢部 竜太, 倉林 修一, 清水 康, "配信情報と利用者の関係性を動的に計量する位置情報に基づく情報配信システムの提案", DEIM Forum 2012.
- [9] 向井 友宏, 黒澤 義明, 目良 和也, 竹澤 寿幸, "マイクロブログの分析に基づくユーザーの嗜好とタイミングを考慮した情報推薦手法の提案", 言語処理学会 第 17 回年次大会 発表論文集.
- [10] "Twitter Serch API," <https://dev.twitter.com/rest/public/search>
- [11] 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.
- [12] "みんなの自動翻訳@ TexTra" <https://mt-auto-minhon-mlt.ucri.jgn-x.jp>
- [13] "Walker+ 東京都のおでかけスポット一覧," http://www.walkerplus.com/spot/_list/ar0313/
- [14] "NAVITIME," <https://www.navitime.co.jp/category/>
- [15] "Google Maps API," <https://developers.google.com/maps/?hl=ja>
- [16] "Airport Limousine," <http://www.limousinebus.co.jp>
- [17] "Sky Bus Tokyo," <http://www.skybus.jp>
- [18] "MongoDB," <https://www.mongodb.com>
- [19] "Google Maps Geocoding API," <https://developers.google.com/maps/documentation/geocoding/intro?hl=ja>