

場所と時間の制約に対処する SNS データを用いた タイムリーなイベント提示

工藤瑠璃子[†] 丸 千尋[†] 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: †ruriko-k@ogl.is.ocha.ac.jp, oguchi@is.ocha.ac.jp, ††chihiro@ogl.is.ocha.ac.jp,
†††enomiki@jp.ibm.com, ††††{nakao,shu}@iii.u-tokyo.ac.jp, ††††sane@cc.kogakuin.ac.jp

あらまし 近年、様々な SNS が人々の間に普及し、利用者が多様な情報を発信している。そのような Web 上の情報には、ユーザの体験やローカルイベント、地域特有の情報など、特定の場所にいる人にとって有益な情報が含まれているため、SNS データから観光地や観光ルート、地域の特徴を抽出する研究が盛んである。しかし、ユーザの地理的、時間的制約を考慮し、「その時・その場」で役立つ情報を抽出している研究は少ない。本研究では、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 SNS の代表である Twitter から抽出する。そして、外国人観光者向けに提示する情報を多言語化し、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。

キーワード Twitter, SNS, 観光情報, 情報抽出

Timely Event Presentation Using SNS Data That Can Cope with Limitations of Place and Time

Chihiro MARU[†], Chihiro MARU[†], 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 Hakozaicho, 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: †ruriko-k@ogl.is.ocha.ac.jp, oguchi@is.ocha.ac.jp, ††chihiro@ogl.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

近年、様々な SNS(ソーシャル・ネットワーキング・サービス)が人々の間に普及し、一種の情報伝達手段として一翼を担っている。SNS の代表である Twitter [1] は、ツイートを読む(=情報収集)とツイートする(=情報発信)というシンプルな機能を軸に作られており、その使いやすさとモバイル端末からも投稿できるという利便性から、多くの人が気軽に情報を発信している。例えば、訪れた観光地の感想や、映画や舞台を観に行った感

想をその場でツイートしたり、お祭りやイルミネーションのようなイベントでは写真や感想、混雑具合をツイートすることも多い。実際に行った感想ではなく、これから行くイベントなどの予定をツイートするユーザもよく見られる。また、「リツイート」機能を利用してツイートを拡散し、多くの人に情報を知ってもらうことも可能であるため、イベントの告知などにも使われている。そのような目的で使用する際には、ツイートの文字数が 140 字に制限されていることから、ツイート本文には重要な

情報が簡潔にまとめられているという特徴もある。このように Twitter を代表とする SNS には、ローカルイベントや地域特有の情報など、特定の場所にいる人にとって有益な情報が含まれている。しかし、有名な観光スポットなどの情報はガイドブックや WEB サイトから取得できるが、それらの媒体に載っていないようなローカルな情報や今まさに開催されているイベントを取得するのは、膨大な情報の中から自力で探し出さなければならず、非常に困難である。特に、場所と時間の制約がある旅行者などが必要とする「その時」「その場所」に関する情報を発信している媒体は少なく、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。

また、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]。急増しているインバウンドへの情報配信手段を充実させるため、IoT デバイスの開発も進んでいる。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、様々な情報配信をする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のスマートフォンに情報を配信する「BeaconCast」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、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、バス到着地周辺のイベント情報、店舗のクーポン、リアルタイムで更新されるお得な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。

本研究では、このような IoT デバイスを基盤と考え、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 SNS の代表である 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する。

本論文の構成は以下の通りである。2. 章で関連研究について述べ、3. 章で提案システムの概要を紹介する。提案システムの一部である、イベント分類についてを 4. 章で、配信準備についてを 5. 章で説明する。最後に、6. 章で本稿をまとめる。

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本研究では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。

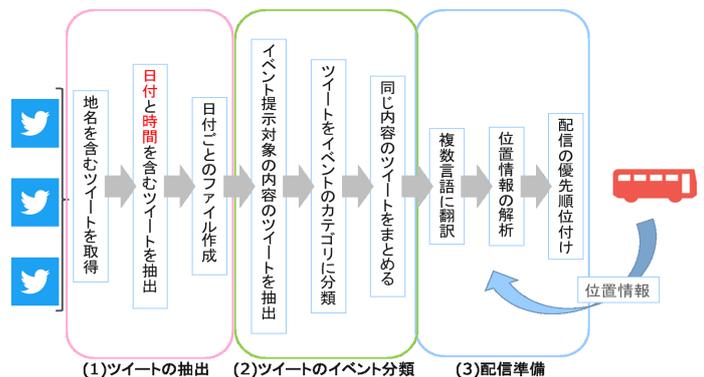


図 1 提案システムの概要

(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取得
2. ツイートの本文に日時を含むツイートを抽出
1. で取得したツイートの本文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、日付ごとのファイルにまとめる。

(2) ツイートのイベント分類

1. イベント提示対象の内容のツイートを抽出
2. 抽出されたツイートをイベントのカテゴリごとに分類

(3) 配信準備

1. 提供する情報を複数言語に翻訳
2. ユーザと提供するイベントの位置情報を取得
3. 提供する情報を並び替え

4. ツイートのイベント分類評価

本章では、提案システムの一部であるツイートのイベント分類について説明する。

場所と時間を含むツイートには観光者などに有用でないツイートも多く存在する。そこで、有用でないツイートを排除するためにツイートの分類を行う。また、観光者などに有用なツイートには、様々なジャンルが混在しており、ユーザの嗜好にあった情報提供ができない。そのため、抽出した有用なツイートをカテゴリごとに分類する。

機械学習を用いた有用判定とカテゴリ分類の精度の評価実験の詳細を以下に示す。

4.1 カテゴリの設定と実験環境

本論文における「有用なツイート」とは以下のカテゴリのイベントに関するツイートとする。

音楽イベント	寄席・お笑い	舞台
映画	展覧会	ポケモン GO その他

カテゴリは、都市情報雑誌の代表格である「東京 Walker」[13]のイベントカテゴリを参考にし、予備実験で地名と日時を含むツイートの種類を分析した際に、一定数のツイートが得られると判明したカテゴリを設定した。

学習データは、2016年9月10日に取得したツイートを人手で有用か有用でないか、と各カテゴリに分類したものを使用する。実験データには、2016年12月6日～2016年12月13日に取得したツイートの中で、イベント開催日が12月13日のツイート1000件を使用した。

機械的にツイートを分類するためには、ツイート本文を数値化する前処理が必要である。MeCab[15]を用いて必要な品詞を取り出し、Bag-of-wordsにより数値化を行い、ツイートのベクトルモデルを作成する。ここで、MeCabにより取り出す品詞は「名詞、動詞、形容詞、副詞」とした。MeCabの形態素辞書には、IPADIC[16]を使用している。

評価実験の検証方法としては、ツイートを正解のツイートと不正解のツイートに人手で分け、機械学習による判定結果の正答率と、正解のツイートと機械学習による判定結果との適合率、再現率、F値を求める。

4.2 実験結果

4.2.1 Support Vector Machine

2クラスの分類を行うことができる学習機械であるSVM(Support Vector Machine)を用いて、ツイートのイベント分類を行った結果について説明する。SVMのソフトウェアにはSVM-Light[14]を使用した。

SVMによるツイートの分類は、まず有用か有用でないかの二対一分類を行う。次に有用なツイートをカテゴリごとに分類する二対他分類を行う。SVMは2クラスの分類であるため、カテゴリ分類では、あるカテゴリとその他のカテゴリ全部でSVM判別モデルを構築していき、多カテゴリに分類する。今回はカテゴリを6個設定したので、6個のSVM判別モデルを作成し判定を行った。どのカテゴリにも判別されなかったツイートはその

他のカテゴリとする。

有用か有用でないかに分ける二対一分類(1)、カテゴリごとに分ける二対他分類(2)の結果を表1,2に示す。(1)の分類では、1000件のうち200件で検証を行った。(2)の分類では、(1)で有用であると判定された482件のうち、カテゴリごとに100件で検証を行った。

表1 Performance of classification(1).

Accuracy	Precision	Recall	F値
0.8650	0.9091	0.8333	0.8857

表2 Performance of classification for each category(2).

カテゴリ	Accuracy	Precision	Recall	F値
音楽イベント	0.8700	0.8867	0.8703	0.8784
寄席・お笑い	0.9100	0.6500	0.8666	0.7428
舞台	0.8100	0.1363	1.0000	0.2399
映画	0.9300	0.4000	0.8000	0.5333
展覧会	0.9600	0.4285	1.0000	0.5999
ポケモン GO	1.0000	1.0000	1.0000	1.0000

(1),(2)の分類ともに高いAccuracyとRecallの結果が得られたが、表2のPrecisionとF値には偏りが出てしまった。

「舞台」のPrecisionの値が極めて低いのは、「舞台」以外のカテゴリのSVM判別モデルでも重みが大きくなる素性が、「舞台」のSVM判別モデルの重みの大きな素性の中にも存在していることが原因であると考えられる。以下に「舞台」と「音楽イベント」の各々のSVM判別モデルにおいて、重みの大きかった素性の一部を挙げる。

公演	席	舞台	シアター	演劇	朗読
上演	枚	ミュージカル	劇	演出	開演

ライブ	LIVE	演奏	ワンマン	予約
料金	コンサート	ピアノ	開演	アイドル

上記に列挙した素性の中には、「開演」という素性が共通して含まれている。このように複数のカテゴリの判別モデルに共通して出現する、重みの大きな素性が存在する場合、あるツイートが複数のカテゴリにおいて正解と判別されてしまい、分類の精度が下がってしまうと考えられる。そこで、カテゴリ「舞台」に多く混在していた[音楽イベント]との二対一分類のSVM判別モデルを作成、分類し、「舞台」でないツイートを排除した。二対一分類で有用であると判定された482件のうち、二対他分類のカテゴリ分類で「舞台」に振り分けられた106件に、新しく作成したSVM判別モデルを適用した結果を表3に示す。

表3 Performance of classification(3).

カテゴリ	Accuracy	Precision	Recall	F値
舞台	0.9439	0.3809	0.9411	0.5516

表3に示されているように、Recallの値が少し下がってしまったが、PrecisionとF値の値を改善することができた。これより、カテゴリ「舞台」に多く混在していた「音楽イベント」を排除

できたことが分かる。

次に,MeCabの形態素辞書を変更し,前述の実験と同じ条件で実験を行った。

前述の実験で使用していた形態素辞書である IPADIC は,2007年3月以降更新されておらず,2007年3月以降に流行った単語などは収録されていない。実際に,カテゴリ「映画」の判別モデルにおいて重みが大い素性である「六本木」「ヒルズ」や,「TOHO」「シネマズ」といった素性は,「六本木ヒルズ」「TOHOシネマズ」と分割されるべき素性であるが,IPADICでは,このような固有表現に対応しておらず,望ましくない形態素に分割されてしまっている。また,「舞台」のツイートに多く出現する「赤坂 ACT シアター」という固有表現は,2008年3月に完成した劇場であるため,IPADICでは「赤坂」「ACT」「シアター」の3つの形態素に分割されてしまう。このように,IPADICはしかるべき機能によって作成された信頼性の高い辞書であるが,収録されている単語が長い期間更新されていないため,流行語や新語,芸能人の名前や,新製品の名前,ランドマークの名称や通称が多く含まれる SNS に投稿されるような文章の解析には不向きな点もある。そこで,LINEのData Labsで開発された辞書である,mecab-ipadic-NEologd [17]を併用することにした。

mecab-ipadic-NEologd は,多数の Web 上の言語資源から得た新語を追加することでカスタマイズした MeCab 用のシステム辞書であり,以下のような特徴があげられる。

(1) IPADIC では複数の形態素に分割されてしまう固有表現を採録

(2) 毎週 2 回以上更新

(3) Web 上の言語資源を活用して更新時に新しい固有表現を随時追加

MeCab の形態素辞書に mecab-ipadic-NEologd を併用して実験を行った結果を表 4,5 に示す。

表 4 Performance of classification.

Accuracy	Precision	Recall	F 値
0.8750	0.9175	0.8396	0.8768

表 5 Performance of classification for each category.

カテゴリ	Accuracy	Precision	Recall	F 値
音楽イベント	0.8400	0.8484	0.9032	0.8750
寄席・お笑い	0.8600	0.5357	0.9375	0.6818
舞台	0.7200	0.3500	0.8750	0.5000
映画	0.9400	0.5454	0.8571	0.6667
展覧会	0.8300	0.3703	1.0000	0.5405
ポケモン GO	1.0000	1.0000	1.0000	1.0000

有用判定, カテゴリ分類ともに, 前述の実験結果に比べ, 全体的に F 値が高い結果が得られた。特に, 表 2 では極端に低い値が出てしまっていたカテゴリ「舞台」の Precision と F 値については, 一回の分類で表 3 の値に近づけることができ,mecab-ipadic-NEologd を併用したことによる効果が見られた。この先の実験では,MeCab の辞書には mecab-ipadic-NEologd を併用することとする。

4.2.2 ランダムフォレスト

本節では,分類器にランダムフォレストを用いてツイートのイベント分類を行った結果について述べる。

ランダムフォレストとは,決定木を弱学習器とする集団学習アルゴリズムである。2クラスの分類を行う SVM は,各クラスごとに学習を行う必要があり,本研究では有用判定を行った後に,6個の SVM を作成してカテゴリ分類を行っていたが,ランダムフォレストは多クラスの分類を行うことが可能であるため,有用判定を行わず,「有用でない」「音楽イベント」「寄席,お笑い」「舞台」「映画」「展覧会」「ポケモン GO」の7クラスの分類を一度に行うことにした。

分類は,Pythonの機械学習ライブラリである scikit-learn [18]を用いて行った。

学習データ,実験データ,検証方法の条件は SVM を用いた実験の条件と変えていない。検証件数は,各カテゴリごとに実験データ 1000 件のうち 400 件で行った。

ランダムフォレストによるカテゴリ分類の結果を表 6 に示す。

表 6 Performance of classification for each category.

カテゴリ	Accuracy	Precision	Recall	F 値
音楽イベント	0.8950	0.8750	0.7583	0.8125
寄席・お笑い	0.9550	0.7948	0.7561	0.7750
舞台	0.9875	0.5000	0.4000	0.4444
映画	0.9925	1.0000	0.7272	0.8421
展覧会	0.9875	0.8666	0.8125	0.8387
ポケモン GO	1.0000	1.0000	1.0000	1.0000

表 6 に示されている通り,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。カテゴリ「舞台」においても,SVM の場合と同等の結果を得ることができた。

SVM とランダムフォレストのどちらの分類器を用いて分類を行った場合でも,最終的には高い精度が得られることが分かったが,両者を比較すると,ランダムフォレストの方が F 値が全体的に高い結果が得られ,精度がより優れていた。この精度は,学習データを増やすこと,高頻度すぎる単語,低頻度すぎる単語を削除する等の手法によって,更に高めることができると考える。

また,処理速度についても,SVM のように総当たりでカテゴリ分類を行う必要のないランダムフォレストの方が,処理にかかる時間が少なく優れていた。

本研究では,情報をリアルタイムに処理してユーザに提示したいため,ツイートのイベント分類にはランダムフォレストを適用することにする。

5. 配信準備

本章では,システムの一部である配信準備について,情報の多言語化と配信の優先順位付けに分けて説明する。

5.1 情報の多言語化

インバウンドに対応するため,ツイート本文を多言語に翻訳する必要がある。しかし,ツイート本文は不完全な日本語で書かれていることも多く,ツイート本文を原文のまま翻訳すると正しく翻訳されないことが多い。そのため,ツイート本文中の情報

を予め整理し、必要な情報のみを提供する。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」[11] を使用して翻訳する。翻訳する言語は、訪日外国人の国籍別の割合 [12] を基に、英語、中国語 (簡体字、繁体字)、韓国語を選んだ。情報を整理する前処理を行わなかった英語翻訳の例を図 2 に示す。ツイートの情報を整理する前処理を施した例を図 3 に示す。図 3 を英語、中国語 (簡体字、繁体字)、韓国語で翻訳した例を図 4,5,6,7 に示す。

<紙わざ大賞26> 12/13(火)-17(土)10:00-19:00(最終日18:00まで)
東京交通会館 東京都千代田区有楽町2-10-1 私制作の紙わざ大賞 歴代入
選作品 今年26回 <南天雄鷄像> 25回 <飛翔-白鷹> 24回 <ヴェ...
https://t.co/zhblqq79ZH

<Paper's Awards 26> 12/13 (Tue) 17 (soil) 10: paper of 00 to
19:00 (until the final day 18:00) 2-10-1 I make Tokyo Traffic Hall
in Chiyoda Ward, Tokyo, Yurakucho Awards past winners of this
year and 26 times southern rooster image> 25 times flying-
Shirataka> 24 times ヴェ... https://t.co/zhblqq79ZH

図 2 Result of translation(full text).

イベント名: <紙わざ大賞26>
日付: 12/13
時間: 10:00
場所: 千代田区
カテゴリ: 展覧会

図 3 Result of organizing data.

Event Name: <Paper Techniques Grand Prix 26>
Date: 13/12
Time: 10 00
Location: Chiyoda Ward
Category: exhibition

図 4 Result of translation(英語).

事件名称:<紙的雑技大賞26>
日期:12/13
时间:10:00
地点:千代田区
类别:展览会

図 5 Result of translation(簡体字).

事件名稱:<紙的雜技大獎26>
日期:12/13
時間:10:00
地點:千代田區
類別:展覽會

図 6 Result of translation(繁体字).

이벤트 명:<종이 화대상 26>
일시: 12/13
시간: 10:00
장소: 지요다구
분류: 전람회

図 7 Result of translation(韓国語).

これより、図 2 のように原文のままでは正しく翻訳できず、複数箇所でも誤った翻訳結果になる場合でも、図 3 のような前処理を施すことで必要な情報を簡潔に正しく提示できることが示された。

5.2 配信の順位付け

場所と日付と時間の条件に適合する情報を、ユーザにとって有益な順で提示するために、情報の順位付けを行う。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、イベントの開催地に着目した順位付けを行った。まず、本システムでは空港と都内を結ぶリムジンバスの乗客をユーザに想定しているため、ユーザの位置情報や目的地が取得可能である。Google Maps API [19] から、ユーザの位置や目的地とイベント開催地の経度、緯度を取得し、この 2 地点の距離を算出し、距離の小さい順に提示する。

ユーザの位置、目的地に「品川駅」を選択し、カテゴリ「展覧会」の順に順位付けを行った結果を図 8 示す。図 8 の結果を地図上に可視化した結果を図 9 に示す。

①	日付: 12/13 時間: 11:00 場所: 目黒 カテゴリ: 展覧会	⑤	日付: 12/13 時間: 12:00 場所: 原宿 カテゴリ: 展覧会
②	イベント名: 「宮沢賢治 十二色の絵本展」 日付: 12/13 時間: 12:00 場所: 恵比寿 カテゴリ: 展覧会	⑥	イベント名: <紙わざ大賞26> 日付: 12/13 時間: 10:00 場所: 千代田区 カテゴリ: 展覧会
③	イベント名: 「キネティックアート展」 日付: 12/13 時間: 11時30分 場所: 渋谷 カテゴリ: 展覧会	⑦	イベント名: 「第4期研究生修了発表展」 日付: 12/13 時間: 11:30 場所: 新宿 カテゴリ: 展覧会
④	イベント名: 「宮沢賢治 十二色の絵本展」 日付: 12/13 時間: 19:00 場所: 渋谷 カテゴリ: 展覧会	⑧	イベント名: 「東京四大学連合合同写真展」 日付: 12/13 時間: 11時 場所: 池袋 カテゴリ: 展覧会

図 8 Result of prioritized information.



図 9 Result of visualization on the map.

図8, 図9より, イベント開催地の近い順に情報提示の順位付けが正しく行われていることが示された。

この結果は, 経度, 緯度から算出した直線距離が小さい順に並び替えを行った結果であり, 位置情報に着目した順位付けでは, 交通手段を指定した場合の移動時間など, よりユーザが利用しやすいような条件での順位付けが必要であると考えている。

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, ユーザの嗜好に合った情報提示を視野に入れ, ツイートのイベント分類を行った。ツイートのイベント分類では, SVM とランダムフォレストで比較を行い, 多カテゴリに分類できるランダムフォレストが, 分類精度, 処理速度どちらにおいても優れていることが分かった。また, 今後さらに増えていくと考えられるインバウンドへ向け多言語で配信を実現するために, 情報を整理し, 高い精度での翻訳を可能にした。そして, ユーザにとって有益な順で情報を提示するために, 提示する情報の順位付けを行った。

今後の課題としては, 分類の精度を高めることと, 位置に着目した配信の順位付けにおいては, 地名による大まかな位置ではなくランドマークや住所等による, より正確な位置情報に基づいた実装を行うことがあげられる。また, 現在のシステムでは, 地名と日付と時間の3点をツイートの本文に含むもののみを対象としており, 解析の対象となるツイートの種類が少ない。今後はこの3点の情報の一部が欠けている場合に情報を補う手法を考え, 解析するツイートの種類を増やし, より多様な情報を提示できるようにする。

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, 清木 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] "みんなの自動翻訳@ TexTra" <https://mt-auto-minhon-mlt.ucri.jgn-x.jp>
- [12] 日本政府観光局 「国籍/月別 訪日外客数 (2003年～2016年)」: <http://www.jnto.go.jp/jpn/statistics/since2003-tourists.pdf>
- [13] "Walker+" <http://www.walkerplus.com>
- [14] "SVM light" <http://svmlight.joachims.org>
- [15] "MeCab," <http://mecab.sourceforge.net/>
- [16] "IPADIC," <https://ja.osdn.net/projects/ipadic/>
- [17] "mecab-ipadic-NEologd," <https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md>
- [18] "scikit-learn Machine Learning in Python," <http://scikit-learn.org/stable/index.html>
- [19] "Google Maps API," <https://developers.google.com/maps/?hl=ja>