

SNS を活用した日本人向けおよび外国人向け観光地情報の抽出 — 文化的背景の異なる人々の興味の違いを拾えるか —

星野 祐子[†] 石井 英里子[‡] 柴田 賢太郎^{*} 大友 隆秀^{*} 山田 光穂[†]

[†] * 東海大学情報通信学部 〒108-8619 東京都港区高輪 2-3-23

[‡] 鹿児島県立短期大学 〒890-0005 鹿児島県鹿児島市下伊敷 1-52-1

E-mail: [†] {hoshino, myamada}@tokai.ac.jp, [‡] erikoishii@k-kentan.ac.jp, ^{*} {5bji2104, 8bjnm006}@mail.u-tokai.ac.jp

あらまし 観光案内の Web ページでは、日本語でも英語でも同じ内容の情報が提供されていることが多いが、文化的違いにより、興味が異なるのであれば、日本人向けと外国人向けの Web ページで異なる内容の情報提供をすべきである[1]。しかし、国内のコンテンツを作成している担当者は日本人が多く、多文化社会出身者の興味を理解しているとは言い難い。このため、日本人と外国人が興味を持っている話題を Web 発信すべき情報として推薦できれば、Web コンテンツの作成に役立つと考えている。これを実現するため、本研究では日本語と英語の Twitter データを収集し、それぞれからある特定地域に関連した話題の抽出ができるか検証を行う。

キーワード SNS, twitter, 観光地, 異文化, 話題抽出, LDA, Web ページ, 位置情報

Extracting sightseeing spot information using SNS

— Development of a new system to differentiate the interests of local tourists and foreign tourists in Japan —

Yuko Hoshino[†] Eriko Ishii[‡] Kentarou Shibata^{*} Otomo Takahide^{*} Mitsuho Yamada[†]

[†] School of Information and Telecommunication, Tokai University 〒108-8619 2-3-23 takanawa minato-ku, Tokyo

[‡] Kagoshima Prefectural College 〒890-0005 1-52-1 shimoishiki kagoshima city, Kagoshima

E-mail: [†] {hoshino, myamada}@tokai.ac.jp, [‡] erikoishii@k-kentan.ac.jp, ^{*} {5bji2104, 8bjnm006}@mail.u-tokai.ac.jp

Abstract The tourist information on websites in foreign languages is usually translated from the local language, so the same and sometimes less information is provided to foreign tourists than local tourists. In addition, there is a tendency that people from different cultural background might have different travel preference from local tourists [1]. However, this might be a burden for tourist web creators. For this reason, we develop a useful system to recommend the information not only for local tourists but also foreign tourists. In this research, we collect twitter streaming data written in Japanese from local tourists and English from foreign tourists, and then, verify that to extract topics which related one tourist area.

Keywords SNS, twitter, tourist spot, different culture, topic extraction, LDA

1. はじめに

日本ではオリンピックの開催や観光立国へ向け、インバウンド観光客を増やそうとしている。海外からの観光客の多くは地方自治体の観光局ホームページのような Web 上のデータを中心に情報収集をしている。しかし、いくつかの地方自治体の観光案内ホームページを見てみると、日本語の内容は充実しているのに対し、英語や他言語の情報量が非常に少ないところもある。これは地方の小さな自治体では Web ページの更新も数少ない役所関係者が行っているため、英語での文章

作成や web ページを作成する技術の不足、作業時間の不足が原因ではないかと考えられる。また、日本語と同等の情報量を提供している Web ページでも、その多くは日本語で提示されている内容をそのまま英語や多言語に翻訳しただけのものであり、日本人と外国人との文化的背景の違いによる興味の対象を考慮しているところはほぼない。しかし、訪日外国人に関連する統計データ[2]やオープンデータを利用した行動調査[1]、また、街頭でのインタビューでも訪日外国人と日本人とは興味対象が異なり、訪れる観光スポットにかなり

の差異があることがわかっているため、この文化的背景の違いを考慮し、日本語 Web ページと英語や他言語 Web ページでは異なる内容を提示したほうが良いのではないかと考える。

さらに観光地を訪れた人々が Twitter に代表される SNS で情報発信し、このような Web 上の口コミを見た人がさらにその場所を訪れることも多くなっている。SNS データは利用者の多い国・地域の偏りがおおく[3]、訪日外国人客全般的な傾向を把握するには向かないが、国別に分けて傾向調査をするには向いているデータと言える[4]。この SNS の情報をもとに、現在、日本人の間で話題になっている場所、外国人(英語話者)、その他言語圏の人々の間で話題になっている場所をそれぞれ抽出することができれば、地方自治体などが運営する観光用 Web ページへ掲載すべき日本人向けスポット情報、外国人向けスポット情報として、情報推薦できるのではないかと考えている。この手法はオープンデータなど、元データが公開されるまで待つ必要がなく、常にデータ収集が可能であるため、より早く情報抽出ができ、インバウンドによる地域活性を目指す地方に有益なものとなるのではないだろうか。

このようなシステムの開発を目指し、本研究では、日本語と英語の Twitter ストリーミングデータを取集し、各データ群からある特定地域に関連した話題の抽出ができるのか、システムの面からの調査・検証と文化的面からの調査・検証を行う。合わせて、位置情報付 tweet をもとに、混雑状況の表示と混雑理由の抽出ができるかについても検証を行う。

2. 関連研究

これまで SNS などのインターネット上で発信された大量データに潜在的に存在する話題に関する研究は数多くある。我々はこれまで、より正確に話題を抽出するために、抽出する話題数の決定方法や抽出した話題の類似度、話題の変遷と消滅について研究を進めてきた[5][6][7][8]。しかし、これまでの研究は、抽出した話題をもとに、選挙や流行語大賞がどのような結果になるかなどを予測検討するなど、日本国内またはある国・地域に住む人を対象としたものであり、文化的背景の異なる人々が同じ場所で、どのような情報を発信するのか調査、検討はしてこなかった。

本間らによるジオタグ付き Tweet データによる分析では、投稿時刻と投稿位置によって、ツイートに性格付けや時空間特性を抽出し、それをもとに地域特性を求めているが、そこに何があるのか、なぜそこに関する発信が多いのかは明確にされていない[9]。相原はスマートフォンアプリによる移動体データ、SNS データを使用した観光動態調査や行動ログ解析による目的地

性指標を示す等をしているが、これも同様にその地域がどのような特性を持っているのかを示しており、その場に何があり、なぜその場所にいるのかは求められていない[4]。

文化的側面を考慮した研究としては、伊藤らによる場所依存型ガイドシステムの提案などがあるが、これは日本にいるものが外国人に対して伝えたいと思っているものをガイドとして提供するもので、訪日外国人が実際にほしいと思っている情報を提供しているとは限らない[10]。同様に福井らによる温泉地宿泊施設とインターネットに関する研究でも旅行者に対して便利なシステム、そこから得た情報をもとにした購買行動の分析などは行われているが、外国人客実際に体験したいと思うような情報が提供できているかは不明瞭である[11]。

日本の地方自治体の多くは観光局、観光協会などと呼ばれる部署を持ち、独自の Web ページを作成して、その地域の観光情報を発信しているが、これらの内容も、これまでの研究と同様に日本人観光客、外国人観光客が今知りたいと思っている内容が提供できているのか不明である。

これらの問題を解決するためには、文化的背景の異なる人々の興味がどこにあるのか、いまネット上の口コミで広がりを見せている場所やイベントを積極的に発信していく必要があるのではないかと考えている。SNS 情報に含まれている、観光局などの担当者がまだ気づいていないスポットなどを抽出し、それを担当者が精査したうえで Web ページに掲載することは、SNS だけでは信ぴょう性が高いとは言えない情報でも、より早く正確に、その場所を訪れようとしている人々に伝えることができるのではないだろうか。

また、我々はアプリケーション画面などのインターフェースについて視覚的観点からの評価も行ってきた[12]。両研究を合わせることによって、よりの確でわかりやすい情報提供をする Web ページの作成ができるのではないかと考えている。

3. 調査手法

本研究での Twitter を使用した話題抽出の調査・検証手順を以下に示す。

3.1. データ収集

まず、Twitter の Geolocation を用いて、ある地域を対象とした日本語と英語の tweet データを収集する。今回の調査では、多くの外国人観光客が訪れ、日本人も多い場所である東京都中央区、千代田区、港区の都心 3 区を中心にデータ収集を行った。収集期間は 2018 年 7 月初旬～11 月初旬である。ただし、一部設備点検

による停電のため、抜けている日付があり、収集日数は計 113 日間である。

また、1 日あたりの平均収集件数は日本語ツイートが約 28,000 件、英語ツイートが約 1,000 件となっている。

3.2. トピック抽出

次に収集したデータ群に対して、LDA 法[13]によるトピック抽出を行った。まず、日本語 tweet データよりトピック抽出を行った。抽出トピック数は先行研究[5][6]をもとに日本語 80 に設定した。トピック抽出をするにあたり、形態素解析が必要となるが、この処理には MeCab[14]を使用している。また、MeCab 標準辞書に加えて、新語を登録した mecab-ipadic-NEologd[15]、および、今回の研究対象地域である都心 3 区に六本木ヒルズや東京タワーなど関連するキーワードを集めた独自の辞書を作成し、標準辞書と合わせて使用している。この辞書の追加により、モデル地域に関するワードをより抜き出しやすくする。また、我々の先行研究[8][16]のように、このモデル地域に関する辞書に含まれる単語のトピック割合を増加させるようにしている。表 1 に日本語データから抽出したトピックの一部、表 2 に英語データから抽出したトピックの一部とそのトピックに含まれるワードの一部を示す。今回の処理では全 1 トピック当たり 15 ワードで抽出している。

表 1 都心 3 区から抽出した日本語トピックの例
(7 月 4 日)

topic0	港区	品川駅	世界	KEIKYU
	京急	赤坂	毎日	つけ麺
topic1	SHIBUYA	子供	まーす	新橋駅
	セブンイレブン	SEJ	ホント	アプリ
topic2	東京駅	渋谷区	千代田区	渋谷駅
	バスタ新宿	東海道新幹線	基本	歌詞
topic3	SHINJUKU	W／	いいね	新宿区
	出して	押上	目的	東京タワー
topic4	言葉	キレ	お互い	BEER
	只今	見た目	動き	用事
topic5	周り	本人	三丁目	悩み
	NEKO	判らん	パリ	ロンドン
topic6	千代田区	秋葉原駅	秋葉原	ワイ
	全く	つくばエクスプレス	駅前	イヤホン

表 2 の英語データから抽出したトピックには一部日本語ワードが含まれてしまっている。これは日本語ワードをコピーするなどし、英文の中に埋め込む形でツ

イートしているものがあることが原因ではないかと考えている。また、形態素解析に英語専用形態素エンジンではなく、日本語同様に MeCab を使用していることも原因のひとつではないかと考えている。

表 2 都心 3 区から抽出した英語トピックの例
(7 月 4 日)

topic0	ROPPONGI	POSTED	PHOTO	TATTOOS
	EVERYONES	UNDER	TRASH	ARRIVAL
topic1	TRYNA	DEMON	MARIETARC	OLDER
	SIGNIFICANTLY	ZILLI	BELT	LRT
topic2	JULY	MAD	DAYSSSS	STARBUCKS COFFEE
	SHINAGAWA	WEAREONEEXO	JIHI	PUPS
topic3	YOUR	DAYS	LORILNOEL	TOES
	TRYING	PERSON	OUT	HER
topic4	るか	JOHNTOGRAPHY	TENTOGRAPHY	
	帝出	KーYO	ノーフューチャー	WAITING
topic5	DON	THAT	GUY	GINZA
	WITH	TOKYO	FORGET	VOTE

3.3. 抽出トピックの比較(単日データの場合)

次に日付ごとに抽出された日本語と英語のトピックの比較を行う。前述の表 1、表 2 は 2018 年 7 月 4 日単日のデータ群から抽出したトピックである。この 2 つの表を比較すると、日本語データでは「京急品川駅高架橋下にある品達（ラーメン街）」に関するトピックと推測されるものが抽出されている。英語データでは「六本木にあるタトゥー店」に関する話題と推測されるトピックが抽出されている。

表 3、表 4 は 7 月 5 日単日のデータ群から抽出したトピック群である。この日は日本全国的に大雨による被害が出ていた日であるため、日本語データにはこれに関連したと思われるワードがいくつか含まれている。また、翌日から開催されていたあるゲームのイベントと推測されるものが、関連する施設（スカイツリー）や地名（秋葉原）を含んだトピックとして抽出されている。表の例には含まれていないが、「スペインのチーズケーキ専門店のオープン」に関する話題と推測されるものも抽出された。英語データでは「BKAKICY」という原宿などに店舗を構えるファッションブランドに関するトピックと推測されるものが抽出されている。

表3 都心3区から抽出した日本語トピックの例
(7月5日)

topic0	大雨	CHIYODA	時点	ゲーム
	注意報	KURE	タイプ	番組
topic1	秋葉原	目黒区	URASIMA	千代田区
	笑顔	方々	大会	スカイツリー
topic2	気分	アイドル	西野	水分
	玄関	SUMMER	コンテンツ	内科
topic3	ライブ	久々	まーす	身体
	バスターミナル	神奈川県	真夏	ほんま
topic4	映画	無責任	最低	毎日
	証拠	今日も一日	回数	ポスト
topic5	100%	方面	空間	MAID
	半端ない	ツイ	教室	グループ

表4 都心3区から抽出した英語トピックの例
(7月5日)

topic0	KELET	BLAKICHY	THINK	YALL
	LOMALIA GUEST	MORNIN	ARIAKE	MAKE
topic1	BACK	TAKE	STOP	UNDERSTAND
	MORROW	RIGHT	SUPRATOKYO	お台場
topic2	BALTHIER	ROPPONGI	VIDEO	POPPER
	UEKI	BEFOR	KINDA	LOWKEY
topic3	AND	INCHES	USING	GOOD-BYE
	ZOOM	CJN	WONDERFUL	MONICAKIM
topic4	FACEBOOK	NEVER	NIGHT	FORGET
	SIDE	CRYPTO	SMILE	DEFINITELY
topic5	BEFORE	YES	ROBOTS	ALIVE
	WELL	PHOTO	AND	TOKYO

日本語データと英語データを比較すると、日本語データから抽出される話題にはイベントや飲食に関するものが多く、英語データから抽出される話題にはファッションなどに関連するものが多い。また、日本語データから上位に抽出されたトピックと類似するものは英語データの中に存在せず、英語データから上位に抽出されたトピックに類似するものが日本語データから抽出されたトピックに存在することもなかった。

3.4. 抽出トピックの比較(混成データの場合)

次に前日に抽出されたトピック群のうち、上位に含まれるものを混成したデータ群、具体的には7月4日に抽出されたトピックの上位10件を7月5日のツイートデータ群に混ぜたものを用意し、そこから抽出をおこなう。さらに、7月5日に抽出されたトピックの上位10件を7月6日のツイートデータ群に混ぜて、抽出を行う。これを繰り返し実施した。処理のイメージを図1に示す。

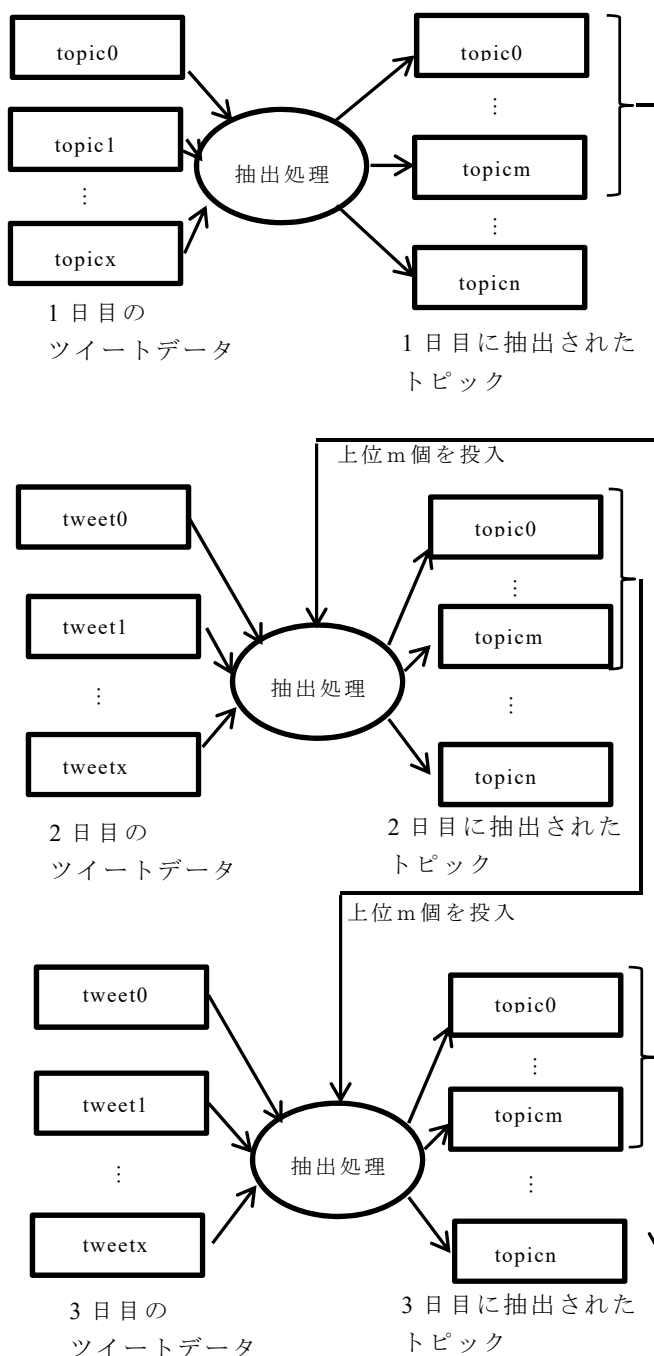


図1 混成データからの抽出イメージ

これは Twitter の場合, ある瞬間だけ 1 つの話題に関するツイートが多くなる (バズる) 現象があるため, 単日だけのデータでは観光客が興味を示すようなものが見落とされる可能性があると考え, 前日の抽出でトピック確率の高かったものを混成し, 抽出を試みた。

表 5 に日本語の 7 月 4 日 (7 月 3 日トピックの一部を混ぜた) データから抽出したトピック例を示す。

表 5 前日分を混ぜたデータから抽出した
日本語トピックの例 (7 月 4 日)

topic0	INFO	TOKYOM ETRO	港区	神宮
	ハワイ	レインボー ブリッジ	ちゃん	西新宿駅
topic1	台東区	上野駅	中目黒駅	ビキニの日
	ブランド	SAYA	海外	新宿三丁 目駅
topic2	朝ごはん	都合	全く	NOT
	RAMEN	HIME	印象	FEVER
topic3	人生	ご飯	リア	出て
	おっさん	2階	日本橋三 越	三越前
topic4	東京駅	千代田区	大阪	大手町駅
	考え	カード	寝起き	千代田線
topic5	中央区	T?KY?	KEIODEN TETSU	INFO
	CH??	銀座	セクシー	このまま

表 6 前日分を混ぜたデータから抽出した
日本語トピックの例 (7 月 5 日)

topic0	写真	休み	有楽町駅	一人
	帰り	新御茶ノ水 駅	千代田区	17時
topic1	港区	品川駅	墨田区	京急
	相手	モノレール 浜松町駅	7 月 7 日	KEIKY U
topic2	特別警 報	大雨	気象	本文
	ライン	余裕	兵庫県	岡山県
topic3	バスラ	12時	皆様	注意報
	土砂災 害	KURE	広島県呉 市	時点
topic4	情報	品川	バイト	ネット
	市ヶ谷	安達	太良	手前
topic5	広島	仲間	京浜東北	岡山
	5番	具合	一週間	錦糸町

同様に表 6 に 7 月 5 日分, 表 7 に 7 月 6 日分のデータから抽出したトピック例を示す。
前述の表 1, 表 3 のような単日データのものと混成データを比べると, 収集日だけではなく, その前から続

く話題も含まれるようになり, 都心 3 区に関連した話題がより多く抽出されるようになる。一方で, 何の話題なのか特定することが困難なものも増えている。

表 7 前日分を混ぜたデータから抽出した
日本語トピックの例 (7 月 6 日)

topic0	ライブ	渋谷	ステージ	女性
	渋谷区	夜中	喫煙所	11時
topic1	江東区	品川区	新木場駅	言葉
	1人	レベル	豊洲駅	ゆりかも め
topic2	東京	西日本	被害	地元
	天気	可能性	豪雨	バーレス ク
topic3	チーム	写真	豊洲PIT	スマイル
	1番	WORLDCU P	スウェーデ ン	しょう
topic4	SHIBUY A	TWITTER	早め	アカウント
	田舎	スタジオ	渋谷道玄 坂	SUKIYA
topic5	浅草駅	インスタ	現実	短冊
	MOMO	日本人	待合室	笑い

表 8 前日分を混ぜたデータから抽出した
英語トピックの例 (7 月 4 日)

topic0	ROJITHA	4TH	HAPPY	JAPAN
	21__21	TMZ	BAG	SECURE D
topic1	YOU	FORWA RD	LOOKING	THANKS
	THEMEP ARKREVI EW	PROJEC T	ONTAVYA	IPA
topic2	BEGINNI NG	FROM	END	THE
	CHARAC TERS	FATABB OT	KINOVINO ANDMOR E	FINANCI AL
topic3	ELSE	TRUMP	THE LOV E BUG	HERBIE
	クルマ	TOWER	OPERA	CITY
topic4	NICEEEEE	OOHHH	NEED	MORT
	SANSLE SAUTRE S	DAWGY	DELARGE LDN	ZIMA
topic5	CERULIS	TINIEL	BLACKPI NK	SONG
	THEME	TOWER	OPERA	FUCK
topic6	LOVING	STRATO METASH IP	STARBU CKS	FIRING
	OPPORT UNITY	APPREC I ATED	GUARD	GARO

英語データについても同様に比較を行う。表 8 に 7 月 4 日の抽出結果,の一部 表 9 に 7 月 5 日の抽出結果の一部, 表 10 に 7 月 6 日の抽出結果の一部を示す。

表 9 前日分を混ぜたデータから抽出した
英語トピックの例 (7 月 5 日)

topic0	ANIMALII FE	CUTE	JENNALY NNMEOW RI	ASS
	EAT	PLACES	TOMORR OW	HOME
topic1	LIKE	LIFE HAC K	ONE	THO
	HERE	LAB	OUT	LIVES
topic2	SQUARE	BED	DABS	MICKEYD ESSIE
	AFROJA CK	SHINAGA WA	DRA	UMEMI
topic3	THE	MILLS	MOTH	REALMA TTHEWW OLF
	ZOMBIE	MADE IN USA	WILL	MOON
topic4	THIS	REPOST	LUV	CGCARD ONA
	SHTUMM Y	TRISTAN GHILL	AGO	HURT
topic5	WEEKEN D	WOW	LALAPOR T	RETIREM ENT
	TOYOTA RACING	HARAJUK U	KYULING LEE	COMPAN Y

表 10 前日分を混ぜたデータから抽出した
英語トピックの例 (7 月 6 日)

topic0	OKISS	ZACKER HCD	CMON	STUCK
	SAY	RAPPER	XXXTENT ACION	FINE
topic1	PAPAYA MILK	LITERAL	FOOTBAL L	IMPORT ANT
	CONTASI AOUS	GOD	THANK	SONG
topic2	PEPI	SOOOO	AYAKO	REASON
	ユチョン	PROUD	HAPAEIK AIWA	WET
topic3	SCATMA NSPIKE	THAT	HAV	STILL
	TAIGA	GOOD N IGHT	WANT	DON
topic4	WITH M E	FRANCIS COBRITT T	IDOLISH7	HAPPY BIRTHDA Y
	COME	LINALOV ESIT	REMYDEL FOX	VSCOJA PAN
topic5	HERE	BUZZED	ROAM	NIHEI
	MOTION	YONGYE A	マジカルミラ イ	MARI

英語データの場合も日本語データと同様に, 前から続く話題が含まれるようになる。また, 表 8 の 7 月 4 日データから topic0 として抽出されたものは「21_21 DESIGN SIGHT」というデザインのためのリサーチセンターに関するものであると推測できるが, これは単日データからは抽出されなかったものである。

3.5. 地域に関連したトピックの抽出数

指定地域から収集したツイートをもとに, 指定地域に関連する辞書を用いて, トピック抽出を行ったが, トピックは単語の集まりであるため, これらを見ただけでは何の話題であるのか推測できないものも多い。そこで, これらトピックに含まれる単語を用いて, ニュースやインターネット上の話題と一致すると推測できるものがどの程度あるのか, また, 指定地域に関連すると推測できるものがどの程度あるのか, Web 検索やトピック群を見てもらったの口頭でのアンケートで検証を行った。表 11 に 2018 年 7 月 4 日～7 月 10 日までの抽出トピックに含まれる単語を用いて, 日本語と英語別に何に関する話題であるのか推測できるトピック数を調査した結果を示す。

表 11 前日分を混ぜたデータから抽出した

	検証日	推測可能	収集地域
日本語	7 月 4～10 日	18	9
英語	7 月 4～10 日	14	4

日本語では推測可能な話題数が 18 件, そのうち都心 3 区に関連したものは 9 件, 英語では推測可能な話題数が 14 件, そのうち都心 3 区に関連したものは 4 件であった。

3.6. 抽出トピックに関する考察

ここまでの結果から, 日本語では日常生活でのニュース, その地域で開催されるイベントや飲食関連の話題が多く投稿される傾向にあり, 英語ではファッションやデザインに関する話題が多く投稿されていることがわかる。このことから, 日本語話者と英語話者の興味異なる傾向があるとわかる。

今回の対象地域である都心 3 区の観光協会の Web ページを見てみると, 日本語, 英語ともに提示されている内容はほぼ同じであり, どちらも大きな寺社仏閣, ランドマークなど一般的なガイドブックに載っている情報とさほど変わらないもので, 英語データから取得されたようなファッション関連の情報やデザイン・アート関連(美術館の紹介は載っている)の情報は非常に少なかった。

また, 抽出したトピックの中には, 都心 3 区以外の

話題，特に渋谷や新宿のような周辺地域に関する話題が比較的多くあったため，周辺地域への移動と前後してのツイートも多いと考えられる。

4. 位置情報付データによる混雑状況把握

混雑状況と混雑理由がわかれば，それを回避したい人，特に時間の限られている旅行者にとって有益ではないかと考えている。そのため，本研究では取得した Tweet データのうち，位置情報付 tweet データを用いて，混雑状況のヒートマップでの提示をし，さらにツイート内容から混雑理由も抽出，表示できるか検証をすることも目標として作業を進めている。図 2 にヒートマップの例を示す。

位置情報付ツイートは数が少ないため，これだけで何をするために，どこへ行くために，その場所にいるのか推測するには不十分であると考え，混雑理由の抽出には位置情報を含まないものも全て使用する。現在，抽出した混雑理由について，街歩きの参考なり得るか検証に入る段階である。ヒートマップの提示では，滞在者の多い都心 3 区を対象としたこともあるが，かなり広範囲な情報の提示となっているので，よりピンポイントでの情報提示をするように改善する必要がある。



図 2 ヒートマップの例（港区）

5. おわりに

今回の研究では，日本語のツイートデータと英語のツイートデータを収集し，都心 3 区の観光に関連する話題を抽出し，日本語話者と英語話者の興味の違いが出るか検証を行った。

単日データからも前日トピックを混成したデータからも，それぞれ違う話題が抽出された。今回

は検証できた日数が少ないため，都心 3 区にくる英語話者がファッションやデザインに興味があるという結果が続くのか，別の話題が抽出されるように変化するのか不明な点がある。今後，収集済みのデータを使用して検証を行う予定である。これをもとに Web ページへ掲載する情報の候補を探し，提示できるようにしたい。加えて，連続した日の混成データを作成する際，今回の検証では前日のトピック上位 10 件を使用した，11 件目以降にも指定地域に関連すると推測されるトピックが含まれていたため，混合するトピックの件数を増やすことも考えた。さらに，今回対象とした都心 3 区の周辺地域に関する情報も多かったため，データの収集地域をもう少し拡大することも考えたい。

また，新しい単語や流行語が次々に生まれ，特に SNS などのインターネットではこの傾向が強いので，2 つ以上の単語が並んで出現することが頻発するなどの現象があった場合，新語として辞書に自動登録する仕組みも併せて実装していきたいと考えている。

6. 謝辞

本研究は日本学術振興会 JSPS より，科研費 JP17K02129 の助成を受けたものです。

参考文献

- [1] 高松正人，大西公一郎他，外国人の視点を取り入れた 訪日観光客誘致戦略，自治体国際化フォーラム 2012 年 3 月号，pp.2-22
- [2] JNTO 日本政府観光局 統計データ，https://www.jnto.go.jp/jpn/statistics/visitor_trends/，(2019 年 1 月アクセス)
- [3] 2017 年度 ICT を活用した訪日外国人動態調査，観光庁，2018，<http://www.mlit.go.jp/kankocho/shisaku/kankochi/gps.html>(2019 年 1 月アクセス)
- [4] ビッグデータを用いた観光動態把握とその活用：動体データで訪日外客の動きをとらえる，相原 健郎，情報管理 2017 年 59 巻 11 号 p. 743-754
- [5] 藤野巖，星野祐子，LDA 法におけるトピック数の決定法およびトピックの評価法についてー Twitter ストリーミングデータを応用例としてー，第 5 回ソーシャルコンピューティングシンポジウム，2014
- [6] Making Topic Words Distribution More Accurate and Ranking Topic Significance According to the Jensen-Shannon Divergence from Background Topic, Iwao Fujino, Yuko Hoshino, ICDM2015 Proceedings, pp.185-200,2015
- [7] 小笹彰，星野祐子，「Twitter における話題の変遷追跡」，第 11 回パーソナルコンピュータ利用技術学会全国大会講演論文集，(2016)，p.64-67

- [8] 有賀章善, 星野祐子, 山田光穂, 石井英里子, 「特定分野に関するトピックの優先的抽出方法の検討」, 第12回パーソナルコンピュータ利用技術学会全国大会講演論文集, (2017), p.34-37
- [9] ジオタグつき tweet の時空間解析に基づいた地域特性抽出に関する研究, 太田 圭亮, 今井 公太郎, 本間 健太郎, 日本建築学会計画系論文集, 2017年 82 巻 731 号 p. 283-289
- [10] 日本文化理解を促進する場所依存型音声ガイドシステムの提案と効果, 伊藤 綾香, 小川 克彦, ヒューマンインタフェース学会論文誌 19 巻 (2017) 3 号, 2017 年 19 巻 3 号 p. 251-260
- [11] 温泉観光地における受給接合と情報流通の再編—群馬県草津温泉における宿泊業のインターネット利用の分析から—, 福井 一喜, 地学雑誌, 2017 年 126 巻 5 号 p. 595-615
- [12] Development of Two People's Gaze Movement Measuring System, Hideaki Takahira, Shinya Mochiduki, Mitsuho Yamada, Proceedings of IMQA2016, Technical Group of Image Media Quality, 2016
- [13] Latent Dirichlet Allocation, David M. Blei, Andrew Y. Ng, Michael I. Jordan, Journal of Machine Learning Research 3
- [14] MeCab: Yet Another Part-of-Speech and Morphological Analyzer,
<http://taku910.github.io/mecab/>
- [15] mecab-ipadic-NEologd:Neologism, dictionary for MeCab
<https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md>
- [16] A Study on Recommended Tourist Spot Information Extraction Using SNS, Yuko Hoshino, Eriko Ishii, Mitsuho Yamada, International Conference on Tourism Research (ICTR2018), Finland, 2018, pp.259-262