

観光地名なしツイートからの観光地に関する感想の抽出手法

渡邊 小百合^{1,a)} 吉野 孝^{2,b)}

受付日 2017年4月8日, 採録日 2017年10月3日

概要: 近年, ドラマやアニメの舞台への聖地巡礼等の新しい形態の旅行が出てきたことや, 外国人旅行者の増加から, 観光地に対して新しいニーズが発生してくることが考えられる. これより, 観光地側も観光客のニーズや問題点を知り, 観光地の発展につなげていく必要がある. しかし, Twitterにおいて「観光地名が入っていない観光地に関するツイート」の収集が困難であるという問題があった. そこで本研究では, 観光地名を含まないツイートからの観光地に関する感想の抽出手法について検討し, 下記を明らかにした. (1) 観光地の特徴語を含むツイートの中には, 観光地に関する感想が含まれている可能性がある. (2) 観光地名入りツイートの前後に呟かれた観光地名なしツイートの中には, 観光地に関する感想が含まれている可能性がある. (3) 観光地名入りツイートに対するリプライには, 観光地に関する感想が含まれている可能性がある. (4) 観光地名を含まない画像付きツイートには, 観光地の食べ物や催し物に関する感想が含まれている可能性がある.

キーワード: 観光情報, 情報抽出, マイクロブログ

Extraction Method of Impressions about Tourist Spots from Tweets without Tourist Spot Names

SAYURI WATANABE^{1,a)} TAKASHI YOSHINO^{2,b)}

Received: April 8, 2017, Accepted: October 3, 2017

Abstract: New needs for tourist spots may arise because the appearance of new forms of travel in Japan, in which fans visit the locations of dramas and animated programs, and the number of foreign travelers to Japan has increased in recent years. Therefore, tourist spot operators want to know the needs of tourists and any problems with the tourist spot so that they can be solved. However, insufficient tourist information was provided by Twitter. In this study, we propose an extraction method of impressions about tourist spots to apply to tweets that do not contain tourist spot names. The experiments provided the following four results: (1) Tweets containing feature words contain tourist information. (2) Tweets without tourist spot names tweeted before and after tweets containing tourist spot names contain tourist information. (3) Replies to tweets containing tourist spot names contain tourist information. (4) Tweets with images that do not contain tourist spot names contain information regarding the food and entertainment available at tourist spots.

Keywords: tourist information, information extraction, microblog

1. はじめに

観光庁による旅行・観光消費動向調査 [1] では, 2006 年から減少を続けていた日本人の旅行平均回数が, 2010 年以降からその減少が止まっている. これは, ドラマやアニメの舞台への聖地巡礼等の新しい形態の旅行が出てきたことが要因として考えられる [2]. また, 観光庁による訪日外国人旅行者数・出国日本人人数の推移 [3] では, 外国人旅行者

¹ 和歌山大学大学院システム工学研究科
Graduate School of Systems Engineering, Wakayama University, Wakayama 640-8510, Japan

² 和歌山大学システム工学部
Faculty of Systems Engineering, Wakayama University, Wakayama 640-8510, Japan

a) watanabe.sayuri@g.wakayama-u.jp

b) yoshino@sys.wakayama-u.ac.jp

が2012年から増加しており、今後も2020年に行われる東京オリンピック等により、外国人旅行者数は増加していくと考えられる。このように、観光客の観光行動が変化している一方で、ドラマ・アニメとのコラボイベントや外国語への対応といった観光地に対する新しいニーズが発生している。観光地側がそれに気づかなければ観光客は離れていってしまう。このことから、観光地側も観光客のニーズや問題点を知り、観光地の発展につなげていく必要がある。

先行研究として、Web上から観光地に関する感想を抽出し、類似性を可視化するシステムの開発を行ったが、Twitterからは観光地に関する感想がほとんど得られないという問題があった。その原因として、「観光地名が入っていない観光地に関するツイート」を考慮していなかったことが考えられる。たとえば、「偕楽園の梅の花はややがっかり感があった」のような、ツイート内に観光地名が含まれているものは収集できるが、「今年の梅の花は微妙」のような、観光地名を含んでいない観光地に関するツイートは収集できていない。

本研究では、観光地名を含まないツイートからの観光地に関する感想の抽出手法を提案する。観光地周辺の位置情報が付加されたツイートや、観光地の画像付きツイートは、観光地名が入ってなくても観光地に関する情報である可能性が高い。また、観光地名を含むツイートの前後のツイートや、観光地名を含むツイートをしたユーザーと一緒に行動したフォロワの同時刻のツイートにも観光地に関する情報が含まれている可能性が高いと考えた。これより、観光地周辺の位置情報付きツイートや画像付きツイート、観光地名入りツイートの前後のツイート、そのユーザーの人間関係等を用いて、観光地名なしツイートから観光地に関する感想を抽出できるのかを検証する。本論文では、観光地名なしツイートからの観光地に関する感想の抽出手法の概要と、実験結果と考察について述べる。

2. 関連研究

本章では、ツイート文を用いた情報抽出に関する研究、ツイートの付加情報を用いた観光に関する研究を示し、本研究の位置づけを明らかにする。

2.1 ツイート文を用いた情報抽出に関する研究

眞野らは、ユーザーがお気に入りしたツイート群を用いて、クラスタの特徴を抽出する手法を提案した[4]。同じクラスタに属するユーザーがお気に入りしたツイート群の特徴語を分析し、そのクラスタがどのような嗜好、属性を持っているのかを可視化する。邱らは、URLを含むツイートを用いてニュース記事を収集する手法を提案した[5]。ニュース記事のURLを含むツイートと、URLを含まないツイートの類似度によって、ニュース記事に関するツイートを収集する。本研究では、観光地に関する感想を含むツイートの

特徴語から、観光地名は含んでいないがその特徴語を含んでいるツイートを観光地に関する感想として収集を行う。

免田らは、Twitterに投稿された観光地に関するツイートを利用した観光地情報推薦システムを開発した[6]。観光地名を含むツイートが観光地に対して好印象か悪印象かを判断し、地図上に表示する。Ritterらは、ツイートの固有表現と日付からイベントを抽出する手法を提案した[7]。特定の時間と強く関連している固有表現を持つツイートをイベントに関連するツイートとして抽出する。どちらの研究においても、観光地名やその観光地に関する語を含むツイートのみを利用しているが、本研究では、観光地名を含むツイートだけではなく、観光地名を含まないツイートからも観光地に関する感想の抽出を行う。

2.2 ツイートの付加情報を用いた観光に関する研究

金子らは、Twitterに投稿された位置情報付き画像ツイートから視覚的なイベントの検出手法を提案した[8]。各地域において検出したキーワードを含むツイートの各画像から特徴量を抽出し、クラスタリングする。奥らは、位置情報付きツイートと位置情報付き画像を用いた観光スポット推薦システムを開発した[9]。対象とする観光地名を含んだ位置情報付きツイートによる観光地の活動領域と、位置情報付き画像による観光地の活動領域を合成して、観光地の活動領域を推定し、その領域内に含まれるツイートの特徴から観光スポットを推薦する。Leeらは、Twitterから社会的イベントを検出するための、ツイートの時間と位置情報を用いた地理的規則性の測定手法を提案した[10]。対象領域において、位置情報付きツイートの眩かれた時間、位置、ユーザーの行動から通常の地理的規則性を推定し、その規則性から外れているときをイベントとして検出する。

これらの研究においても、ツイートの位置情報や画像、眩かれた時間を用いて観光地に関する情報を抽出している。しかし、位置情報付きツイートの数はきわめて少なく、位置情報による抽出には限界がある。本研究では、位置情報付きツイートだけでなく、観光地の画像付きツイート、観光地名入りツイートの前後のツイート、そのユーザーの人間関係等を用いて、観光地名を含んでいないツイートからも観光地に関する感想の抽出を行う。

3. 観光地に関する感想の抽出手法

本章では観光地名なしツイートからの観光地に関する感想の抽出手法について述べる。図1に抽出手法の手順を示し、以下にその具体的な内容を示す。

- (1) 観光地名入りツイートと位置情報付きツイートの収集
位置情報付きツイートと位置情報なしツイートから、観光地名を含むツイートを収集する(図1(A)-a, (A)-b)。位置情報付きツイートにおいては、奥らの活動領域推定手法[9]を用いて、観光地に関することが

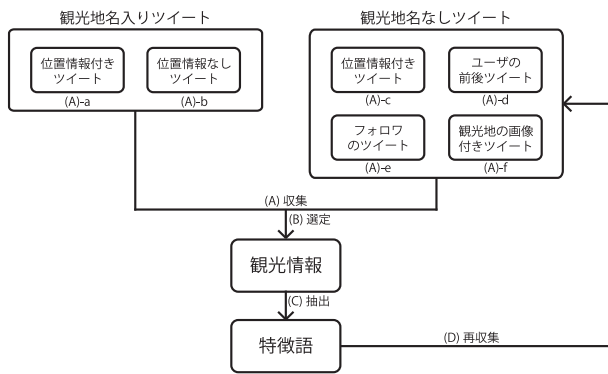


図 1 観光地に関する感想の抽出手順

Fig. 1 Extraction procedure of impressions about tourist spots.

ツイートされる領域を推定し、その範囲内のツイートも収集する (図 1(A)-c)。

(2) 観光地名入りツイートの前後のツイートの収集

観光地名を含むツイート (図 1(A)-a, (A)-b) の前後 3 時間以内のツイートを収集する (図 1(A)-d)。1カ所の観光地に観光目的で訪れる場合、3 時間あれば十分に観光できると考えたことと、観光地に関するツイートをすると同時に、1 ツイートごとの間が長く空くのではなく、数時間の間に続けて複数回ツイートするのではないかと考えた。これより、本研究では、収集する対象を前後 3 時間以内としている。しかし、観光地によって滞在時間が異なる可能性があるため、今後収集する時間の幅を検討する必要がある。

(3) フォロウのツイートの収集

観光地名を含むツイート (図 1(A)-a, (A)-b) をしたユーザーのフォロウのツイートを収集する (図 1(A)-e)。観光地名入りツイートをしたユーザーのフォロウの中に、そのとき一緒に行動していた人がいる可能性があるため、ユーザーのフォロウのツイートから、観光地名入りツイート付近のツイートを収集する。このとき、(2) と同じく、観光地名入りツイート (図 1(A)-a, (A)-b) の前後 3 時間以内に呟かれたフォロウのツイートを対象としている。

(4) 観光地の画像付きツイートの収集

観光地の画像付きツイートを収集する (図 1(A)-f)。観光地の判断には、ランドマーク検出ができる Google Cloud Vision API^{*1}を用いる。

(5) 特徴語の抽出

(1)~(4) において収集したツイート (図 1(A)-a~f) の中から観光地に関する感想を選定し (図 1(B))、収集された観光地に関する感想を含むツイートの特徴語を抽出する (図 1(C))。本研究において、観光地に関する感想を含むツイートとは、観光地に対する印象

や不満等が含まれているものとしている。形態素解析システム JUMAN^{*2}を用いてツイート (図 1(A)-a~f) の分かち書きを行い、特徴語の抽出には tf-idf を用いる。抽出対象の品詞は、観光地の特徴を表す語となりうると考えられる名詞、形容詞、動詞としている。本研究では、tf-idf 値が高い上位 20 件を、再収集を行う特徴語の対象とする。このとき、地域に依存する単語ほど観光地の特徴語となりうると考えられるため、tf-idf に新たな重み付けをする必要があるが、今は検討中である。

(6) 特徴語を含むツイートの収集

観光地名なしツイート (図 1(A)-c~f) から、(5) で抽出した特徴語を含むツイートを収集する (図 1(D))。特徴語を含む位置情報付きツイート (図 1(A)-c) や、特徴語入りツイートの前後ツイート (図 1(A)-d)、特徴語入りツイートをしたユーザーのフォロウのツイート (図 1(A)-e)、画像付きツイート (図 1(A)-f) を収集対象とし、すでに収集したツイートは除外する。

(7) 観光地に関する感想の再収集

(5) と (6) を繰り返すことにより、観光地名が入っていないツイート (図 1(A)-c~f) から観光地に関する感想を再収集する。

4. 実験

4.1 検証項目

本実験では、観光地名なしツイートから観光地に関する感想が抽出可能か検証するため、以下の 2 つの実験を行った。

(1) 位置情報と特徴語を用いた抽出

(2) 前後ツイート、フォロウのツイート、画像付きツイートをを用いた抽出

前後ツイートやフォロウのツイート、画像付きツイートは短期間のツイートでも傾向が分かる可能性があるが、位置情報付きツイートは数が少なく、特徴語は時期によって変化する可能性がある。これより、位置情報付きツイートと特徴語については、長期間のツイートをを用いる必要があるため、今回は位置情報と特徴語による抽出と、前後ツイート、フォロウのツイート、画像付きツイートによる抽出で分けて行っている。どちらの実験においても、観光地の感想かどうかは人手で判断しており、判断基準は、ツイートが「観光地に対する感想であるか」としている。たとえば、「伏見稲荷大社の雰囲気が好き」は観光地に対する感想であるが、「伏見稲荷大社に来てる人の着物が素敵」は着物に対する感想のため、観光地の感想ではない。

*1 <https://cloud.google.com/vision/>

*2 <http://nlp.ist.i.kyoto-u.ac.jp/index.php?cmd=read&page=JUMAN>

4.2 検証概要

4.2.1 位置情報と特徴語を用いた抽出

本実験では、トリップアドバイザーによる日本観光ランキング*3において1位である伏見稲荷大社についての感想を抽出する。本実験で使用するツイートは、2016年1月1日から2016年6月30日の間に呟かれた約1億3,000万ツイートである。これらのツイートは、Twitter Streaming APIを用いてTwitterからリアルタイムに収集したものであるが、期間中に収集の停止や二重収集といった不具合が生じたため、一部不完全である。使用するツイートのうち、位置情報付きツイートは約79万ツイートであった。本実験では、「伏見稲荷」を含むツイートの特徴語を抽出し、その特徴語を含むツイートの特徴語の抽出までを行う。

4.2.2 前後ツイート、フォロワのツイート、画像付きツイートをを用いた抽出

本実験では、トリップアドバイザー*4において、9地域それぞれで1位の観光地と、トリップアドバイザーによる日本観光ランキング*3において1位である伏見稲荷大社の、合計10カ所の観光地についての感想を抽出する。表1に本実験で収集する対象の観光地を示す。日本観光ランキングとは、日本すべての観光地によるランキングであるため、伏見稲荷大社は「日本全体」としている。本実験では、2016年11月23日の0時0分から23時59分の間に呟かれたツイートの中から、Twitter*5の公式のキーワード検索を用いて観光地名を含むツイートを検索し、その前後のツイートとフォロワのツイートを調べる。キーワード検索を利用している理由としては、Twitter Streaming APIによる収集は全ツイート中の1%のみであり、観光地名入りツイートの前後のツイートが収集されない可能性があるためである。実験対象とした期間の設定理由は、11月23日は祝日であり、家族や友人達と観光地に訪れる人が多いと考えたためである。画像付きツイートについては、観光地名

表1 収集する対象の観光地

Table 1 Collection of target tourist spots.

地域	観光地名
北海道	JRタワー (北海道札幌市)
東北	瑞鳳殿 (宮城県仙台市)
関東	横浜みなとみらい21 (神奈川県横浜市)
中部	熱田神宮 (愛知県名古屋市)
近畿	道頓堀 (大阪府大阪市)
中国	原爆ドーム (広島県広島市)
四国	松山城 (愛媛県松山市)
九州	大濠公園 (福岡県福岡市)
沖縄	首里城 (沖縄県那覇市)
日本全体	伏見稲荷大社 (京都府京都市)

*3 <https://www.tripadvisor.jp/Attractions-g294232-Activities-Japan.html>

*4 <https://www.tripadvisor.jp/>

*5 <https://twitter.com/?lang=ja>

名を含まない画像付きツイートは膨大にあるため、本実験では、観光地名を含むツイートの前後において観光地名を含まない画像付きツイートに観光地に関する感想があるかを調べる。これより、今回収集するツイートの画像は観光地に関する画像である可能性が高いと考えられるため、本実験では写真も人手で評価している。

5. 実験結果と考察

5.1 位置情報と特徴語を用いた抽出の実験結果と考察

(1) 位置情報付きツイートからの観光地に関する感想

表2に「伏見稲荷」を含むツイート数と感想の数を示す。位置情報付きツイートは図1(A)-a、位置情報なしツイートは図1(A)-bにあたる。位置情報付きツイート内の感想は20件であったのに対し、位置情報なしツイート内の感想は180件と、9倍の差があった。この結果から、位置情報なしツイート内には観光地に関する感想が多く含まれているので、位置情報なしツイートから感想を抽出する必要があることが分かる。

表3に収集した位置情報付きツイートの数と感想の数を示す。観光地名入りツイートは図1(A)-a、観光地名なしツイートは図1(A)-cにあたる。感想を含んでいる観光地名なしツイートは感想を含んでいる観光地名入りツイートの約0.8倍収集された。その例を表4に示す。「めっちゃ登ったのにまだ半分くらいやった」と「タイムスリップしたかのよう」のどちらのツイートにおいても、観光地名を含んでいないため、何についてのツイートか分からないが、伏見稲荷大社付近の

表2 「伏見稲荷」を含むツイート数と感想数

Table 2 Number of tweets containing “Fushimi Inari” and the amount of impressions.

ツイートの種類	収集したツイート数	感想数
位置情報付きツイート	7,314	20
位置情報なしツイート	1,605	180

表3 位置情報付きツイートからの観光地に関する感想数

Table 3 Amount of impressions about tourist spots in tweets with position information.

ツイートの種類	収集したツイート数	感想数
観光地名入りツイート	34	11
観光地名なしツイート	7,280	9

表4 位置情報付きツイートからの感想例

Table 4 Examples of impressions in tweets with position information.

ツイートの種類	ツイート内容
位置情報付きツイート	めっちゃ登ったのにまだ半分くらいやったww お腹すいた、さむい!
	タイムスリップしたかのよう

位置情報が付いていることにより、伏見稲荷大社についてのツイートではないかと推測することができる。この結果から、観光地付近の位置情報付きツイートには観光地に関する感想が含まれている可能性があることが分かる。さらに、位置情報付きツイートをを用いた収集によって、通常の検索で収集される感想数の約1.8倍収集されており、関連研究でも述べられているように、位置情報付きツイートをを用いた感想の収集は有効であると考えられる。

(2) 特徴語を含むツイートからの観光地に関する感想

表5に「伏見稲荷」を含むツイートの特徴語の例を示す。「京都」や「行って」のような、場所に関する単語のtf-idf値が高い傾向にあった。また、「千本鳥居」や「おもかる石」というような、伏見稲荷大社特有のものも特徴語として抽出された。その場所特有なものを特徴語として用いることで、より対象とした観光地に関係のあるツイートが収集できると考えたため、「千本鳥居」と「おもかる石」を含むツイートの再収集を行った。表6に「千本鳥居」と「おもかる石」を含むツイート数と感想の数を示し、表7に特徴語を含むツイートからの感想例をそれぞれ示す。この収集は図1(D)にあたる。観光地名は含まれていないが、特徴語を含んでいる「昨日初めて某所の千本鳥居を見たのですが、

表5 「伏見稲荷」を含むツイートの特徴語の例

Table 5 Examples of feature words of tweets containing “Fushimi Inari.”

特徴語	tf-idf 値
京都	0.479
千本鳥居	0.422
行って	0.217
写真	0.159
おもかる石	0.080

表6 2種類の特徴語をそれぞれ含むツイート数と感想数

Table 6 Number of tweets and impressions containing each of two types of feature words.

特徴語	収集したツイート数	感想数
千本鳥居	104	27
おもかる石	10	3

表7 特徴語を含むツイートからの感想例

Table 7 Examples of impressions in tweets containing feature words.

ツイートの種類	ツイート内容
「千本鳥居」を含むツイート	昨日初めて某所の千本鳥居を見たのですが、圧巻でした。天気も良くて、最高のロケーション。
「おもかる石」を含むツイート	おもかる石が軽かったことにびっくりしてる。重い暗示かかりすぎていたかしら 笑

圧巻でした」や、「おもかる石が軽かったことにびっくりしてる」のような、伏見稲荷大社に対する印象や感想のツイートが新たに30ツイート収集できた。この結果から、特徴語を含むツイートには観光地に関する感想が含まれている可能性があることが分かる。

表8に「千本鳥居」を含むツイートと「おもかる石」を含むツイートの特徴語の例を示す。どちらの結果においても、「鳥居」や「重い」のような、再収集に用いた語句を表す単語が抽出された。しかし、「伏見稲荷」を含むツイートの特徴語とは違い、「鳥居」や「重い」といった単語は他の観光地に関するツイートや、観光地とは関係ないツイートにも含まれている可能性がある。これより、「千本鳥居」と「おもかる石」を含むツイートの特徴語を用いた収集では、伏見稲荷大社に関する感想だけではなく、他の観光地に関する感想も出てくる可能性がある。この要因として、収集対象の観光地数が少ないことが考えられるため、今後収集対象とする観光地数を増やして再実験する必要がある。さらに、特徴語によって収集したツイートの中に他の観光地の情報も混ざっている場合、どの語を対象とした観光地の特徴語とするのか検討する必要がある。また、特徴語を含む感想数は「伏見稲荷」を含む感想数の約0.2倍収集された。このことから、1回の特徴語による再収集で通常の検索で収集される感想数の約1.2倍収集できたので、特徴語を含むツイートをを用いた感想の収集は有効であると考えられる。しかし、再収集を繰り返すことで、今回の実験結果のように、観光地の特定が困難な特徴語が多く抽出される可能性があるため、特徴語を用いた収集については今後さらに検討していく必要がある。

(3) 考察

本実験においては、位置情報付きツイートと特徴語を用いた観光地に関する感想の抽出が有効であるかの検証を行い、以下のことが分かった。

1. 観光地付近の位置情報付きツイートには感想が含まれている可能性がある。
2. 特徴語を含むツイートには感想が含まれている可

表8 「千本鳥居」を含むツイートと「おもかる石」を含むツイートの特徴語の例

Table 8 Examples of feature words of tweets containing “Senbonotorii” and tweets containing “Omokaruishi.”

ツイートの種類	抽出された特徴語	tf-idf 値
「千本鳥居」を含むツイート	鳥居	0.290
	しんどい	0.193
	圧巻	0.193
「おもかる石」を含むツイート	重い	0.632
	暗示	0.316
	びっくり	0.316

表 9 収集されたツイート数と感想数および感想の増加率

Table 9 Number of collected tweets, the amount of impressions, and the amount of increase of impressions.

観光地名	収集したツイート数 ^{※1}	観光地名入りツイートの感想数	観光地名なしツイートの感想数 ^{※2※3}	増加率
JR タワー	965(34)	3	3(1)	1.0
瑞鳳殿	3,032(99)	21	8(2)	0.4
横浜みなとみらい 21	6,675(215)	4	2(0)	0.5
熱田神宮	9,770(321)	30	7(1)	0.2
道頓堀	37,706(1,147)	31	7(2)	0.2
原爆ドーム	8,790(286)	19	3(2)	0.2
松山城	2,128(71)	7	2(0)	0.3
大濠公園	2,222(75)	7	4(1)	0.6
首里城	4,398(145)	13	7(2)	0.5
伏見稲荷大社	23,750(788)	96	64(19)	0.7
合計	99,436(3,181)	230	134(30)	0.6

※1 ()内は観光地名入りツイートから収集されたツイート数

※2 観光地名入りツイートの前後3時間以内に呟かれたツイートの中から収集

※3 ()内は観光地名なしツイートの感想のうち、リプライに含まれる感想数

表 10 観光地名入りツイートの前後ツイートからの感想例

Table 10 Examples of impressions in tweets before and after tweets containing tourist spot names.

ツイートの種類	ツイート内容
観光地名入りツイート	福岡到着してホテルに荷物預けてぶらぶら開始。天気よければ大濠公園でまったりしようと思ってただけだな。
観光地名なしツイート	雲は多いけど薄日が差してきたのできてみた。池の周り走ってるランナーさん多し。ベンチでお弁当食べてる人も多くてまったりするには最高だなー。

能性がある。

1については、関連研究で述べられているように、位置情報付きツイートを用いることで観光地に関するツイートが得られることが分かる。2については、観光地名なしツイートにも感想が含まれていることで、「Twitterから感想があまり得られない」という問題の解決につながる。

5.2 前後ツイート、フォローのツイート、画像付きツイートをを用いた抽出の実験結果と考察

(1) 観光地名入りツイートの前後ツイートからの観光地に関する感想

表 9 に収集したツイート数と観光地名入りツイートの感想数、観光地名入りツイートの前後に呟かれた観光地名なしツイートの感想数を示す。収集したツイート数は図 1(A)-a~f、観光地名入りツイートの感想数は図 1(A)-a~b、観光地名なしツイートの感想数は図 1(A)-c~f にあたる。すべての観光地において、観光地名なしツイートから感想が収集された。その例を表 10 に示す。ある 1 人のユーザの大濠公園に関する

ツイートで、上から時系列である。「池の周り走ってるランナーさん多し。ベンチでお弁当食べてる人も多くてまったりするには最高だなー」というツイートだけでは、どの場所についての感想か分からないが、「天気よければ大濠公園でまったりしようと思ってただけだな」というツイートが前にあることによって、大濠公園に対する感想であることが分かる。この結果から、観光地名入りツイートの前後に呟かれた観光地名なしツイートの中には感想が含まれている可能性があるということが分かった。

(2) フォロワのツイートからの観光地に関する感想

表 11 に収集されたフォローのツイート数と観光地名入りツイート前後の画像付きツイート数、それぞれの感想数を示す。フォローのツイートに画像が付いている場合があるため、重複しているツイートもある。画像付きツイートはすべての観光地において含まれているが、フォローのツイートは 10 カ所中 7 カ所の観光地に含まれているという結果であった。また、本実験で収集された感想を含むフォローのツイートは、一緒に観光地に行ったのではなく、観光地に行きたいというツイートに対するリプライや、観光地へ行ったというツイートに対するリプライであった。表 9 に、各観光地のリプライに含まれている感想数、表 11 に、フォローのツイートのリプライに含まれている感想数をそれぞれ示している。リプライによる感想を含む観光地名なしツイートが本実験で収集された感想 134 件中 30 件であり、その 30 件中、フォローからのリプライが 12 件であった。その例を表 12 に示す。A さんの伏見稲荷に行くというツイートに対して、B さんが反応のツイートをする、「千本鳥居とても好きです」と A さんが観光地の印象についてリプライを送っている。そ

表 11 収集されたフォロワのツイート数と観光地名入りツイート前後の画像付きツイート数およびそれぞれの感想数

Table 11 Number of collected tweets of followers, the amount of collected tweets containing images before and after tweets containing tourist spot names, and the amount of each impressions.

観光地名	収集したフォロワのツイート数	フォロワのツイートの感想数 ※	収集した画像付きツイート数	画像付きツイートの感想数
JR タワー	771	0	11	2
瑞鳳殿	2,467	0	47	4
横浜みなとみらい 21	5,319	1(0)	67	2
熱田神宮	7,787	1(1)	49	5
道頓堀	28,373	1(1)	351	3
原爆ドーム	7,016	2(2)	40	1
松山城	1,696	0	34	1
大濠公園	1,773	1(1)	14	2
首里城	3,470	1(1)	49	4
伏見稲荷大社	18,041	6(6)	212	24
合計	76,713	13(12)	874	48

※ () 内はフォロワのツイートの感想のうち、リプライに含まれる感想数

表 12 観光地名入りツイート前後のリプライに含まれていた感想例
Table 12 Examples of impressions contained in replies before and after tweets containing tourist spot names.

ユーザ	時刻	ツイート内容 ※
A	0:05	今日は貴船と東福寺と伏見稲荷行ってくるね
B	0:06	いいな伏見
A	0:22	@B 千本鳥居とても好きです 好きです
B	0:23	@A わかる 異空間だよな. 実には行ったことない

※ 個人情報のためユーザ名は変更しているが、ツイート内容はすべて原文のまま

して、そのリプライにさらに B さんが「わかる 異空間だよな」と、観光地に対する印象を返している。これより、観光地名入りツイートに対するリプライには感想が含まれている可能性があることが分かった。本実験では観光地で一緒に行動していたフォロワのツイートから感想は得られなかったが、フォロワと一緒に観光地に行っている場合、一緒にいったフォロワの約 8 割は観光地に関するツイートにいいねやリツイートをしている傾向があった。この結果から、観光地名を含むツイートをしているユーザのすべてのフォロワを調べなくても、そのツイートにいいねやリツイートをしているユーザと一緒に行動していた可能性が高いことが分かった。その他に、ユーザと一緒に観光地へ行ったフォロワを画像にタグ付けしている場合もある。

(3) 画像付きツイートからの観光地に関する感想

表 11 に示すように、画像付きツイートはすべての観光地において、感想を含んでいるツイートがあった。画像の内容は、観光地そのものの画像が多かったが、その観光地で売られている食べ物や催し物に関する画像もあった。その例を表 13 に示す。食べ物と催し物

表 13 観光地名入りツイート前後に含まれていた観光地の食べ物や催し物に関するツイート例

Table 13 Examples describing food and entertainment available at tourist spots contained in tweets before and after tweets containing tourist spot names.

内容	時刻	ツイート内容 ※ 1 ※ 2
食べ物	23:04	今日はとりあえず首里城へ.
	23:10	内部でさんびん茶をいただけます。[さんびん茶の画像] ※ 2
催し物	12:44	大好きな場所で大切なお願いをするため神様に正式なお手紙を書いて来ました。ふと見上げたら雲が不思議な形...#熱田神宮
	13:37	お宮参りに七五三と境内は賑やか. それに加えて白無垢姿の花嫁さんにも出会えた. 幸せエネルギーも頂けて今日は良き日なり. 来て良かった. [境内の画像] ※ 2

※ 1 ツイート内容はすべて原文のまま

※ 2 [] 内はツイートに付いていた画像の内容の説明

のツイートはそれぞれ 1 人のユーザがツイートしている。食べ物に関するツイートは、「今日はとりあえず首里城へ」とツイートした後、「内部でさんびん茶をいただけます」とお茶の画像付きでツイートしているため、首里城で飲めるお茶ということが分かる。催し物に関するツイートは、熱田神宮のハッシュタグを付けてツイートした後、「お宮参りに七五三と境内は賑やか. それに加えて白無垢姿の花嫁さんにも出会えた」とツイートしているため、熱田神宮で七五三や結婚式が行われていたことが分かる。この結果から、観光地名を含まない画像付きツイートからは観光地そのものに関する感想だけではなく、観光地にある食べ物や催し物の感想が含まれている可能性があることが分かった。

(4) 考察

本実験においては、観光地名入りツイートの前後のツイート、観光地名入りツイートをしたユーザのフォローのツイート、観光地名が含まれていない画像付きツイートのそれぞれに観光地に関する感想があるかの検証を行い、以下のことが分かった。

1. 観光地名入りツイートの前後に呟かれた観光地名なしツイートの中には、感想が含まれている可能性がある。
2. 観光地名入りツイートに対するリプライには、感想が含まれている可能性がある。
3. 観光地名を含まない画像付きツイートには、観光地の食べ物や催し物の感想が含まれている可能性がある。

1と2については、観光地名なしツイートにも感想が含まれていることで、「Twitterから感想があまり得られない」という問題の解決につながる。3については、観光地に関する感想は観光地の歴史や食べ物、催し物といった、複数の種類に分けることができ、それらが観光地の主な特徴となりうる。これより、観光地の様々な感想を抽出できることで、観光地ごとの特徴が複数の指標で表せるため、類似性や問題点等の気づきの幅が広がる。また、表9の感想数の増加率において、観光地名なしツイートの感想数は観光地名入りツイートの感想数の約0.6倍である。これより、通常の検索で収集される感想数の約1.6倍収集できたので、Twitterからより多くの感想を収集する手法として有効であると考えられる。

しかし、表9において、観光地によって収集されたツイート数と感想数に差が見られる。道頓堀と横浜みなとみらい21は収集されたツイート数は多いが、感想数は少ないという結果であった。道頓堀は、「道頓堀」という名前の店があり、その店についてのツイートが半数近く含まれていたことが要因としてあげられる。これより、観光地と同一名の店等に関するツイートの判別方法を検討する必要がある。横浜みなとみらい21は、業者によるツイートが多かったことと、様々な店が含まれる場所の名前であるため、「横浜みなとみらい21」を含むツイートには感想が少ないことが要因としてあげられる。これより、観光地が複数の場所を含んだ所である場合、含まれている場所の名前でも収集を行う必要があると考えられる。また、収集ツイート数において、観光地名入りツイートが100以上の観光地は、Twitterユーザが多い若年層[11]が訪れやすい場所という共通点が見られた。原爆ドームや首里城、伏見稲荷大社は修学旅行で訪れることが多く、横浜みなとみらい21は様々な種類の店があり、デートスポットとなっている。熱田神宮は、ゲームに登場

する刀が展示されており、その刀についての話題が見られた。瑞鳳殿については、収集対象日に武将の装いをした人達によるイベントが行われていたため、100ツイート近く呟かれていたと考えられる。これより、Twitterユーザが少ない年齢層が訪れる観光地は観光地名を含むツイート自体が少ないことが考えられる。そのような観光地についてはブログ等、他のWeb上の媒体から収集することを検討する。

5.3 考察のまとめ

位置情報と特徴語を用いた抽出と、観光地名入りツイートの前後のツイート、フォローのツイート、画像付きツイートをを用いた抽出の実験の結果、以下のことが分かった。

- (1) 観光地の特徴語を含むツイートの中には、感想が含まれている可能性がある。
- (2) 観光地名入りツイートの前後に呟かれた観光地名なしツイートの中には、感想が含まれている可能性がある。
- (3) 観光地名入りツイートに対するリプライには、感想が含まれている可能性がある。
- (4) 観光地名を含まない画像付きツイートには、観光地の食べ物や催し物の感想が含まれている可能性がある。

この結果から、観光地名を含んでいないツイートにも観光地に関する感想が含まれていることが分かったとともに、ツイートの位置情報、特徴語、観光地名入りツイートの前後ツイート、フォローのツイート、画像付きツイートをを用いることで、通常の検索よりもTwitterからより多くの感想が収集できることが分かった。また、収集された感想を含むツイートは、表7や表10に示したような、単純な観光地に関する感想ツイートが多く見られた。しかし、単なる感想だけではなく、観光地に対するニーズや不満が分かるツイートも収集されていた。表14に、観光地に対するニーズや不満を含む感想例を示す。どちらのツイートも図1(A)-dから収集されたツイートである。「夜の千本鳥居はほんとなんかあそこで日本刀構えて『卍解…』とか言いたい」からは、観光地でのコスプレに対するニーズが分かり、「稲荷といえばやっぱりこのあったかい電球が一番でした・・・せめての救いで暖色系LEDにして欲しかったです」からは、観光地の電球の色が変わったことに

表 14 観光地に対するニーズや不満を含む感想例

Table 14 Examples of impressions containing needs and complaints for tourist spots.

ツイートの種類	ツイート内容
観光地に対するニーズを含むツイート	夜の千本鳥居はほんとなんかあそこで日本刀構えて『卍解…』とか言いたい
観光地に対する不満を含むツイート	稲荷といえばやっぱりこのあったかい電球が一番でした・・・せめての救いで暖色系LEDにして欲しかったです

対する不満が分かる。これより、本手法を用いることで観光地側が観光客のニーズや不満といった、観光地の改善や発展につながる情報を得ることができると考えられる。

今後の課題としては、観光地と一緒に写っていない食べ物や催し物の画像からはランドマークを特定できないため、その対応についての検討や、マイナーな観光地についても感想が収集できるかの検証を行うことがあげられる。さらに、今回は手法をすべて手作業で行ったが、それらの自動化についても検討する必要がある。

6. おわりに

本論文では、観光地名なしツイートからの観光地に関する感想の抽出手法の提案と、その実験について述べた。実験の結果、以下の4点を明らかにした。

- (1) 観光地の特徴語を含むツイートの中には、感想が含まれている可能性がある。
- (2) 観光地名入りツイートの前後に眩かれた観光地名なしツイートの中には、感想が含まれている可能性がある。
- (3) 観光地名入りツイートに対するリプライには、感想が含まれている可能性がある。
- (4) 観光地名を含まない画像付きツイートには、観光地の食べ物や催し物の感想が含まれている可能性がある。

この結果、観光地名なしツイートを用いることで、通常の検索よりもTwitterからより多くの感想が収集できることが分かった。また、収集された感想を含むツイートには、観光地に対するニーズや不満が分かるツイートも含まれていたため、本手法を用いることで観光地の改善や発展につながる情報を得ることができると考えられる。今後の課題として、マイナーな観光地についても感想が収集できるかの検証や、抽出手法の自動化があげられる。

参考文献

- [1] 国土交通省観光庁：旅行平均回数「旅行・観光産業の経済効果に関する調査研究」(2014年版)，国土交通省観光庁（オンライン），入手先 (<http://www.mlit.go.jp/common/001136064.pdf>) (参照 2017-04-05)。
- [2] 堀内和哉，小山友介：アニメ聖地巡礼を通じた街興しに関する調査研究，第19回進化経済学会北海道大会，pp.1-56 (2015)。
- [3] 国土交通省観光庁：訪日外国人旅行者数・出国日本人人数の推移，国土交通省観光庁（オンライン），入手先 (http://www.mlit.go.jp/kankocho/siryou/toukei/in_out.html) (参照 2017-04-05)。
- [4] 眞野裕也，青山俊弘：ミニプログユーザの記事嗜好を用いたクラスタ発見，*Journal of JACT*, Vol.15, No.3, pp.43-46 (2010)。
- [5] 邱起仁，樋山淳雄：ニュース記事に関連するTwitterの投稿の収集手法の提案，情報処理学会研究報告，Vol.2013-DBS-158, No.22, pp.1-6 (2013)。
- [6] 免田哲矢，Kryssanov, V.V., 林勇吾，小川均：Twitterを用いたリアルタイム情報収集による観光地情報推薦システム，情報処理学会第73回全国大会，第4分冊，pp.647-648 (2011)。
- [7] Ritter, A., Mausam, Etzioni, O. and Clark, S.: Open Domain Event Extraction from Twitter, *KDD '12, Proc. 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.1104-1112 (2012)。
- [8] 金子昴夢，柳井啓司：Twitterからのジオタグ画像収集による視覚的イベント検出，電子情報通信学会技術研究報告，PRMU，パターン認識・メディア理解，pp.53-58 (2013)。
- [9] 奥健太，橋本拓也，上野弘毅，服部文夫：位置情報付きツイート対応付けに基づく観光スポット推薦システムの開発，ARG第2回Webインテリジェンスとインタラクション研究会，pp.7-12 (2013)。
- [10] Lee, R. and Sumiya, K.: Measuring Geographical Regularities of Crowd Behaviors for Twitter-based Geo-social Event Detection, *Proc. 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, pp.1-10 (2010)。
- [11] 総務省：平成28年情報通信メディアの利用時間と情報行動に関する調査報告書，総務省（オンライン），入手先 (http://www.soumu.go.jp/main_content/000492877.pdf) (参照 2017-07-20)。



渡邊 小百合 (学生会員)

1994年生。2017年和歌山大学システム工学部デザイン情報学科卒業。現在，同大学大学院システム工学研究科システム工学専攻博士前期課程在学中。観光支援に関する研究に従事。



吉野 孝 (正会員)

1969年生。1992年鹿児島大学工学部電子工学科卒業。1994年同大学大学院工学研究科電気工学専攻修士課程修了。博士(情報科学)。現在，和歌山大学システム工学部教授。CSCW, HCIの研究に従事。