

「旅のつぶやき」を用いた観光地の評価 ～石川県金沢市を事例に～

鈴木 大智¹・藤生 慎²・高山 純一³・中山 晶一朗⁴

¹非会員 金沢大学 理工学域環境デザイン学類 (〒920-1161 石川県金沢市鈴見台)
E-mail: dsds0912@gmail.com

²正会員 金沢大学助教 理工学域環境デザイン学類 (〒921-1192 石川県金沢市角間町)
E-mail: fujju@se.kanazawa-u.ac.jp

³正会員 金沢大学教授 理工学域環境デザイン学類 (〒921-1192 石川県金沢市角間町)
E-mail: takayama@staff.kanazawa-u.ac.jp

⁴正会員 金沢大学教授 理工学域環境デザイン学類 (〒921-1192 石川県金沢市角間町)
E-mail: nakayama@staff.kanazawa-u.ac.jp

概要：近年, TwitterやfacebookなどのSNSやブログ等のサービスを用いて情報を発信する個人や団体が増えており, その中には旅行者にとって非常に有益な情報が含まれている. Twitter等のシェアの広いSNSサービスからデータを取得することで, 大規模なデータを使って, リアルタイム, 季節ごとの様々な条件下で, より細かな満足・不満足を明らかにすることができ, 新たな魅力の発見や, 季節ごとの各観光地における課題の発見などが可能となる. また, ガイドブックのように一様な質による観光地の紹介ではなく, 季節ごとの魅力, 穴場の発見・紹介, モデル観光コースなどの提供が可能となり, 金沢に訪れる観光客へきめ細かい情報提供が可能となる. 本研究では, Twitterにおいて, 旅行者のtweetによく見られる特徴を把握し, 「なすきのおと」のサービスを用いて観光ツイートを収集し, それを基に観光地評価を行う術を記した.

Evaluation of sightseeing area using tweet data ～focused on Kanazawa city～

Key Words : twitter, 観光地, 評価, 金沢

1. はじめに

近年, 特に若者を中心にTwitterやブログなどのサービスを利用し, 自分の周囲で起きた出来事や経験したことなどを周囲に発信する人が増加している. Twitterについては, 日本における利用者は2014年2月の時点で約2070万人存在し, これは国民10人の内1～2人にあたる数値となっている. また, 全世界では2009年には約1億1000万人だった利用者が, 2013年の時点で約2億4100万人まで増加した. tweet数で見ると, 黎明期の2007年はTwitterの1日のツイート数は約5000回ほどだったが, 2009年には1日のツイート数が250万回を超えるようになった. さらに2010年に入ると前年比1,400%の伸びとなる1日3500万回を記録するなど利用者・利用頻度共に増加していることがわかる. 小さく軽量化が進んだタブレットやノートパソコン, スマートフォン等の普及またフリーwifiエリアの拡大などにより更なる利用者の増加が予想される. Twitterでは, 思ったことを誰宛にでもなくその場で独り言

を「つぶやく」ような手軽さで自由に発信できるため, 例えば旅行先での景観に対する感情や現地での体験したことなどについてリアルタイムで「つぶやく」利用者も多い. 実際の体験に基づいたこれらの情報には, リアルタイム性や直感性といった点において非常に有用なものが多く含まれている. しかし, Twitterの検索機能を用いても旅行者のツイートを発見することは難しく, 膨大な「つぶやき」の量が情報収集の妨げとなっている.

本稿では, 石川県金沢市に存在する代表的な観光地「兼六園」「21世紀美術館」「近江町市場」「東茶屋街・西茶野街」「金沢城」の5つの観光地を含む観光ツイートをそれぞれ集め, 人の目によりtweet内にポジティブワードやネガティブワードが存在するか確認してその特徴の把握までを記載し, それを用いた金沢の観光地の評価を目的とする.

観光ツイートの収集は初めフリーtweet検索ソフトtopsyを用いることを試みたが, 検索できる母集団が少ないこ

とや作業に時間がかかりすぎてしまうため、最終的にはNTT dataの「なずきのおと」^[1] というシステムを利用し研究を行った。「なずきのおと」ではweb上にあるニュース・ブログ・Twitterの収集が可能で、日本語及び日本で記載された全量tweetデータを対象として集計及び原文をダウンロードすることが可能である。

本稿の構成は以下のとおりである。2で関連研究、3では観光ツイートの収集手法について述べ、4ではそれを用いたポジティブツイート・ネガティブツイートそれぞれの特徴把握について述べる。5では今後の観光地評価の手法について述べる。

2. 関連研究

本節では、twitterから観光情報の抽出や、それを用いた観光情報の推薦に関する研究について紹介する。

2.1 観光情報の探索に関する研究

文献 [1] では、Streaming APIによりTwitter社が一般に向けて提供している全量tweetの1/10のtweetを利用し、あるキーワードを含むtweetの量を時系列順に見ていくことで、事後的に話題になっている事柄やその情報をTwitterのデータから読み取るということをしてきた。キーワードが少しマイナーなものだったためにあまり対象となるデータが得られず、また関係ないtweetを完全にシャットアウトできず研究が困難化したことが記されていた。

2.2 観光ルートの推薦に関する研究

文献 [2] では、FousquareとInstagramという2つのアプリケーションを介してつぶやかれたtweetをTwitterAPIを用いて収集し、観光ツイートを「食事」「景観」「行動」の3つのカテゴリーに分けて分析している。対象となる観光ツイートの付加情報をtweetしたユーザーのタイムラインから取得し、キーワードを時系列順に並べることでユーザーが訪れた場所を予測して各ルートの選定を行う。こちらの研究でもツイートの量をその基準としていた。

3. 観光ツイートの収集手法

ここではtopsyを用いた観光ツイート収集と「なずきのおと」を用いた観光ツイートの収集手法について述べる。観光ツイートとはある観光地についての思いやそこで体験した感情などを記したtweetのことである。本研究では、金沢の観光地の名称を含むtweetを観光ツイートとし、その全てを人の目によってチェックし、対象となるtweetかどうかの判断を行い研究を進めた。

3.1 利用したサービスとその特徴

本研究では先述のように、初めはtopsy、後半からは「なずきのおと」を用いて観光ツイートを収集した。

topsyは図1のようなフリーtweet検索サイトで、指定したキーワードを含むツイートを収集することができる。数年前のtweetまで遡ってtweetデータを取得することが可能で、ユーザーがつぶやきに添付した写真もその場で見ることができる。しかし、一度に表示できるtweet数が10件と少なくtweetの原文を取得するためにはサイトに表示されているtweetをコピーするという作業が必要であり、~botなどの個人やある団体が定期的にニュースや情報などを投稿する自動プログラムのtweetを排除することができない。そのため大量のtweetデータを取得することには向いておらず、研究の途中で利用を中止した。

次に、途中から利用した「なずきのおと」について紹介する。「なずきのおと」はNTT dataが提供する有料サービスで、今回は主に観光ツイートを収集するために使用した。「なずきのおと」はweb上にある過去13ヶ月のニュース・ブログ・Twitterのデータを分析することができ、Twitterのデータについては必要であればその原文をダウンロードすることが可能である。また、指定したキーワードを含む対象tweetに対し、簡易的な感情分析は可能である。しかし、ユーザーがつぶやきに添付した写真を見るためにはダウンロードしたtweetデータのURLを使い、一度Twitterにログインしてからページを開く必要があり、手間がかかるため、一見しただけではユーザーが何に興味・関心を持ったのかわからない。「なずきのおと」の分析イメージを図2に示す。



図-1 topsyのトップページ



図-2 「なずきのおと」の分析イメージ

表-1 ネガティブツイートのキーワードと代表tweet

ネガティブツイート	
キーワード	全文
<兼六園>	
寒い	寒いのと、人増えてきたのと、韓国人旅行者があまりにも酷いので、兼六園はそこそこにして退散。
混む	しかし寒い…雨降りそうだし、風が冷たい。 半袖は無謀だったんや…
羨える	兼六園も 金沢も混んでる(。) 兼六園付近でデモやってて羨えてる
停める+られる+ない	兼六園で和を感じようとしたけど混んでて車停められん。 諦めー!!
壊す	兼六園は隣の公園のマーチングバンドが雰囲気ぶち壊しだったし、金沢城は見るとこ少なかったし
疲れる	@hachiraku1 久しぶりの兼六園ですが人が多く疲れました(。)
暗い	兼六園暗(。口) GEO寄ろっかな??
わかる+ない	@yanasee 兼六園付近わからぬです！とりあえず片町ってとこ行けばイロイロお店ありますのでそのあたりが良いかもですね
渋滞	兼六園はメッチャ混んでて超渋滞してんのそこら近い石引商店街が商店街として機能してないから人全くないの本当に勿体無い
勿体無い	

表-2 ポジティブツイートのキーワードと代表tweet

ポジティブツイート	
キーワード	全文
<兼六園>	
よい	@chachami_0919 兼六園よかった！けど美術館がやばみ！岡山は広島行くのにちらみするくらいやな……………ことは……………
やばい	
すごい	兼六園の噴水、自然の噴水ってすごいなって思った。
美しい	世界で最も美しい駅「金沢駅」の写真を撮ってきた http://t.co/9mrLfhQwHI @zapaさんから 金沢と言えば日本三名園の一つ兼六園が有名だけど、金沢駅そのものもすげーんだよね。自動改札ないけど。
情緒+ある	軽の右翼カー！初めて見た。雨で兼六園どうかしら？兼六園は雨の方が情緒ありそうやけど。
見事	@masato_imascg 人生二度目の兼六園ですん。相変わらず見事な日本庭園ですん
綺麗	兼六園を散歩中。さすが文化財指定庭園、つくりが綺麗で心が癒される(。▽)
癒す	
すげえ	兼六園スゲエ http://t.co/d00ekUFdU
風情+ある	兼六園！金沢城の外郭の庭。文化財指定庭園なだけあってとても立派で風情のある庭園でした。とにかく広かった、そして緑豊か。庭園というかもはや自然公園かと思うレベル http://t.co/LrnDCnGTgt

4. ポジティブツイートとネガティブツイートの特徴の把握

Twitterを用いて観光地の評価を行うに際してその基準となるポジティブツイートとネガティブツイートの特徴を把握することは不可欠であり、今回はGW期間中の5/3-5/6の間につぶやかれたtweetを対象に特徴の把握を行った。対象としたキーワード及びそれぞれのキーワードを含むtweetの数は以下のとおりである。

「金沢駅【262件】」「兼六園【1613件】」「21世紀美術館【1033件】」「近江町市場【358件】」「東茶屋街・西茶野街【262件】」「金沢城【540件】」

これらのtweet全てに目を通しキーワード探しと観光ツイートの特徴の把握を行った。その一例として表-1、表-2はそれぞれ対象期間内に「兼六園」を含むtweetを、「なすきのおと」を用いて取得したものの中からポジティブワードやネガティブワードを含む観光ツイートを抽出したものである。今回はスペースの都合上、それぞれ10個のキーワードのみ表に表した。表1から、GW期間中ということもあってか「渋滞」や「混む」などのネガティブワードが見受けられ、道路や駐車場が大変混雑してい

たことが分かり、交通整備が十分でなかったと予想できる。

また、「羨える」というキーワードや、他の観光地に対しての観光ツイートでは「つむ（何らかの障害が生じたためにやりたかったことができず、不満な気持ち。うなだれる様子）」などの感情を表す若者言葉の確認もできた。次に、表-2であるがやはり日本三大庭園のうちの一つである兼六園の素晴らしさを表す「美しい」や「見事」というキーワードや、日本独特の歴史ある美しさを感じさせる「情緒ある」や「風情ある」などのキーワードが見受けられた。また、他の観光地に対するポジティブツイートの中には「やばい（文脈から良いことを表すワードと認知）」や「おされ（おしやれの意）」などの若者言葉も存在した。

5. 今後の観光地評価の手法

観光ツイートを詳細に分析すると、兼六園や茶屋街などの観光地に対し、雨が降っているほうが趣があり良いというtweetや、季節の言葉を含んだものが見つかり、天気や季節、また連休などの外部条件によって、観光

表-3 作成予定のtweet整理表

tweet原文	観光地名	カテゴリー					感情		外部条件			
		食べ物	文化	景観	交通	再訪願望	positive	negative	天気	連休	季節	
兼六園で和を感じようとしたけど混んでて車止められん。諦めー!! 兼六園の庭園の景色が最高すぎる	兼六園					1			1	1	1	2
ただ一つ残念なことに近所で何かのイベントがあつてラップやらレゲエみたいなズンドコBGMが絶え間無く聴こえて来ることである http://t.co/BAjdFnhFmj	兼六園			1			1	1	1	1	1	2
ナイス @ 近江町市場で海鮮丼！金粉まぶしてあつて2200円！！うっ…まあー！！ → http://t.co/ohTgboSICj #bookmeter	近江町市場		1						1		1	2

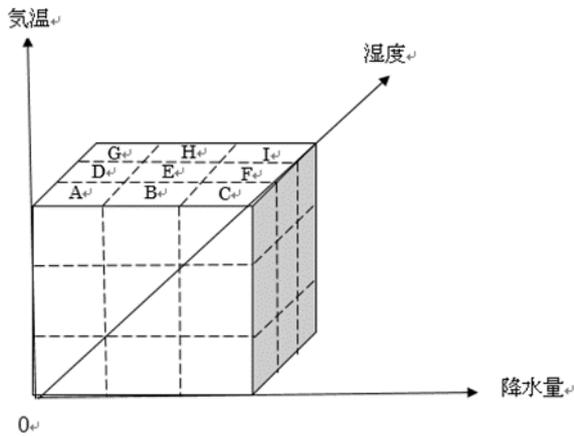


図3 天気に関する外部条件のブロック分け

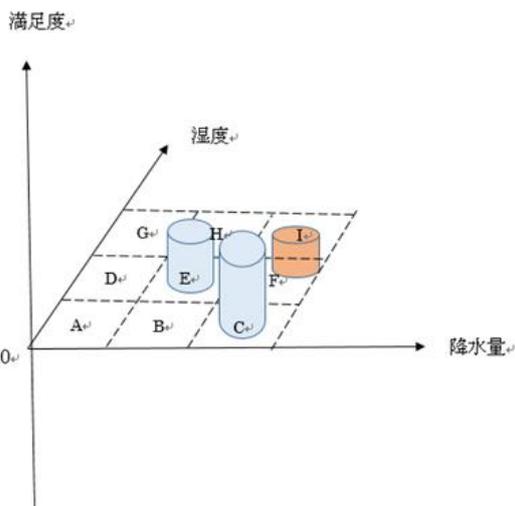


図4 満足度の可視化

ツイートの傾向に変化が見られるのではないかと思います,また,その変化は食べ物や交通,景観などそれぞれのカテゴリーによって違うと思われたため,それらをまとめた表-3の形式でまとめた.1が各観光ツイートが該当するもので,季節などは1~4の数字を用いて表す.カテゴリーや外部条件は例えば,食べ物や文化,天気や季節である.

次に,図3のように季節ごとに(冬:12~2,春:3~5月,夏:6~8月,秋9~11月)気温・降水量・湿度の3軸を用いて9個のブロックに分け各ブロックでの観光地に対する感情評価の解析を行う.一例として,図4のように気温を固定したときの湿度と降水量による満足度の変化を表す.(グラフの分布は実際のデータに基づいておらず,今回はサンプルを示した)

表3を分析することで,金沢の各観光地における天気や連休やイベント等による感情評価・魅力評価を行うことができ,図4より気温や湿度など更に条件を細かく設定した上での観光評価(満足度)の可視化を行う.また,満足度はポジティブワードとネガティブワードの数を用いて表す.また,アンケート調査を実施し,ポジティブワードとネガティブワードの各々の重みを決定した.

参考文献

- 1) 笛田薫,亀田由梨: Twitterを用いた観光情報の探索,日本計算機統計学会論文集(26),67-70(2012)
- 2) 中嶋勇人,新妻弘崇,太田学ぶ: 位置情報付きツイートを利用した観光ルート推薦,情報処理学会研究報告.データベース・システム研究報告 DBS-158(28),1-6(2013)
- 3) NTT data なずきのおと: <https://nazuki-oto.com/>