

「旅のつぶやき」を用いた観光地の満足度評価手法の構築

鈴木 大智¹・藤生 慎²・高山 純一³・中山 晶一朗⁴

¹学生会員 金沢大学 理工学域環境デザイン学類 (〒920-1192 石川県金沢市角間町)
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: fujiu@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、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。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、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、「なずきのおと」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った。

Key Words : twitter, tweet, bigdata, SNS, tourism, evaluation, satisfy

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数で見てみると、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¹⁾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、研究当初からの対象地である石川県金沢市に存在する代表的な観光地を中心に、研究の一般化を図るために全国各地から数箇所ずつ抽出し、満足度評価を行った。

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」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で既存研究の整理と本研究の位置づけの整理、3でTweetデータの基礎分析について述べ、4では満足度評価手法の構築について述べる。また、5では満足度評価結果について述べ、6で結論を述べる。

2. 既存研究の整理と本研究の位置づけの整理

北陸新幹線が開業し、より多くの観光客が金沢市を訪れることが著者らの既往調査により明らかになっている。既往調査の結果から首都圏在住者にとって金沢の魅力は高く、訪れたい観光地の評価で高評価を得ている。しかし、北陸新幹線開業や2015年4月期の朝の連続テレビ小説の効果により金沢市を中心とした観光魅力度は一時的に向上するものの、2016年3月に北海道新幹線の開業を控えており、北陸新幹線の開業効果を持続することが可能であるか否かについては疑問が残る。そこで、本研究では、「旅のつぶやき」データを用いて金沢を訪れた観光客の満足／不満を抽出し、金沢市の観光魅力度を計測し、金沢市の観光魅力度の維持・向上に資する提案を行うことを目的とする。なお、「旅のつぶやき」データは、twitterやFacebookといったSNS (Social Networking Service) から取得することが可能である。これらのデータは、アンケート調査等では得ることのできない時間・場所 (GPS データの活用) を含んだ観光行動中の生の声を取得することが可能である。以上の通り、本研究では、観光客の生の声の定量的な分析を通じて、金沢市における課題を発見し、地域と連携し解決策の提案を行うことで持続的な観光振興を可能とする点が既存研究と比較して新規性のある点である。

さらに、本研究の活動を通じて得られる主たる成果は、金沢市を訪れた観光客の生の声から、魅力度の評価が可能となる点である。従来、観光地の魅力度評価はアンケート調査等を通じて取得したデータに基づく結果であったが、これらでは観光行動中の生の声を把握することはできず、魅力度の評価には限界があった。さらに、観光周遊行動の様子も取得することが可能で、より詳細な魅力度評価が可能になる。

SNS データ (twitter) は国内のみならず海外の観光地のデータも容易に取得することが可能である。したがって、本申請の活動を通じて金沢市の観光地を対象とした分析手法を確立できれば、国内外の観光地の魅力度の計測に本申請で構築した魅力度評価手法を援用することが可能である。さらに、英語版の SNS データも加えて分析することで、金沢を訪問した外国人観光客の魅力度評価も可能となる。

以下に示した5本の既往研究は twitter と観光を対象として行われたものであるが、魅力度の評

価手法の構築など行われておらず、twitter データの利用可能性について論じたものとなっている。

①中嶋ら：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 (2013)

②笛田ら：twitter を用いた観光情報の探索, 岡山大学紀要 (2012)

③環光庁：スマホから発信されるデータの観光地の集客への活用 (2013)

④環光庁：観光ビッグデータを活用した観光振興について (2013)

⑤ゼリングデータコム：「混雑統計」データを用いた観光行動分析 (2013)

3. Tweetデータの基礎分析

3.1 なずきのおと

「なずきのおと」とは、NTT data が提供する有料サービスで、Web 上にあるニュース/ブログ/Twitter の収集・簡易分析を行うことのできるマーケティングリサーチ・ソリューションである¹⁾。なずきのおとを利用することで、ある「物」や「事象」に関して Web 上の評判や認知度を把握することができる。尚、ツイートデータに関しては、日本語及び日本で記載された全量ツイートデータをダウンロードすることができる。以下になずきのおとによる分析結果の一例を記載し、簡単に説明する。

図1はツイートに「21世紀美術館」を含むツイート (=リサーチキーワードが「21世紀美術館」) の内、2014年5月3日～2014年5月6日にツイートされたツイートを対象に分析を行ったものである。¹⁾ リサーチ当たりのリサーチキーワードは最大で16語で、除外ユーザーは最大50まで登録できる。本研究では、同アカウントによる大量の同一ツイートによって満足度の偏りが発生することを防ぐため「BOT/Bot/bot」を除外ユーザーに登録した。また、ツイートデータは最大で13ヶ月前まで遡ってデータの取得ができたので、1年分のツイートデータを用いて分析を行った。

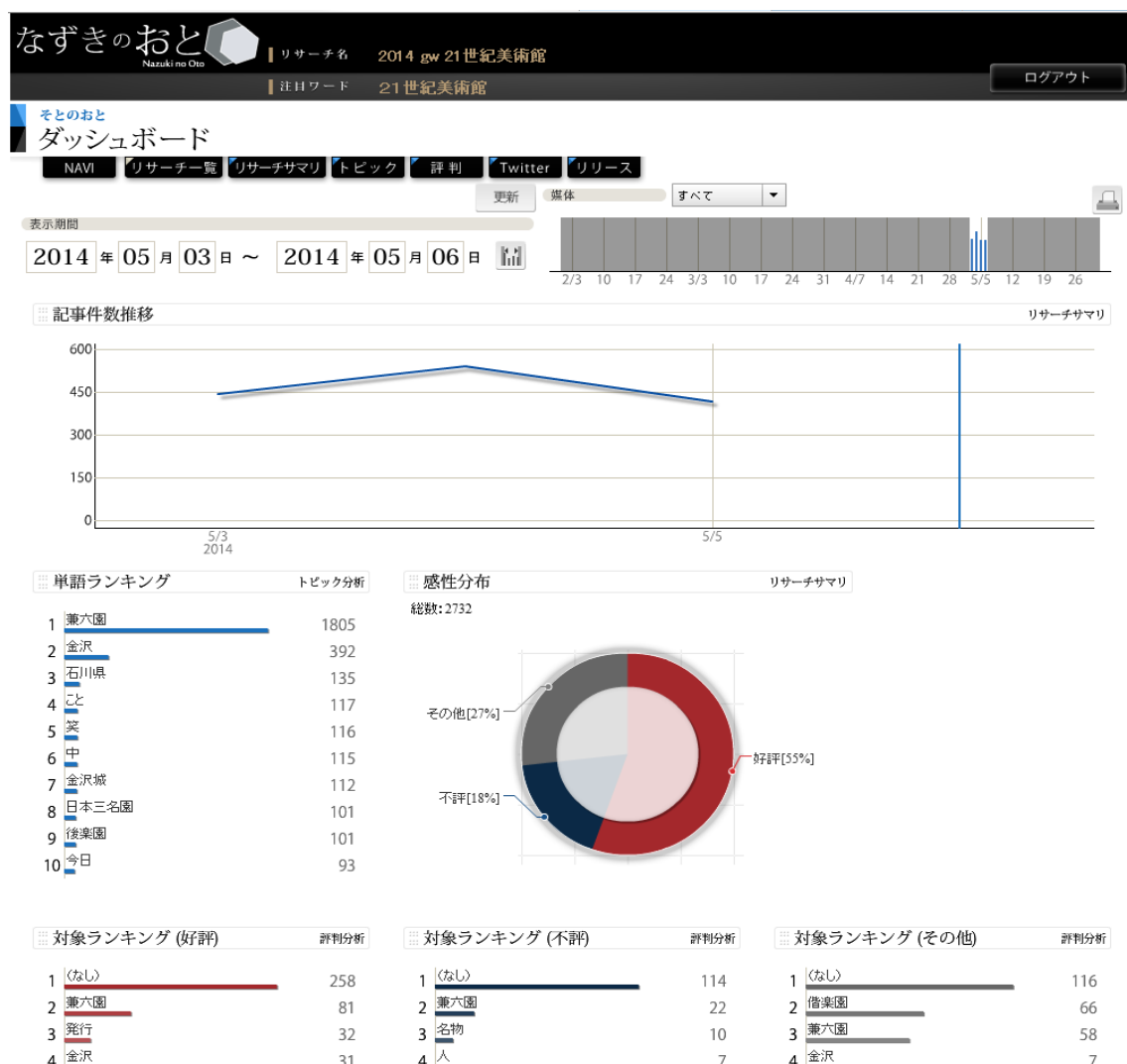


図1 なずきのおとによる分析結果の一例²⁾

3.2 SPSS

SPSSとは統計解析スタンダードソフトのことである。今回はその中でもSPSS Text Analyticsというソフトを使用することでツイートデータを感情解析し、ツイートごとにその感情属性をPositive/Negative/無属性の3種類に分類した。感情解析の基礎となっているのは、Sentiments tapという製品に同梱されている既成のテキスト分析パッケージである。以下にSPSSによる感情分析の一例を記載し、簡単な説明を記す。

ツイート, Negativeツイートの定義を示す。

Positive ツイート：観光地・観光地の評価につながる周辺環境に対し、称賛の意を含んだツイート。

Negative ツイート：観光地・観光地の評価につながる周辺環境に対し、非難の意を含んだツイート。

4. 満足度評価手法の構築

4.1 Positiveツイート/Negativeツイート

本研究では得られたツイートの内、観光地に関するツイートを「観光ツイート」と呼ぶことにし、また観光ツイートの内、満足度に関与するものをそれぞれPositiveツイート, Negativeツイートと呼ぶことにする。以下にPositive

20	20	携帯電話でメールやインターネットができるようになってからは、知りたい情報が外にいてもすぐに入手できるようになった。	その他 - 問い合わせ
21	21	いつでも連絡が取れる	
22	22	家族や友人と気軽に、しかも頻繁に、連絡が取れるようになった。特にメールで。	
23	23	いつでもどこでも電話・メールが出来る	
24	24	出先から、会社や自宅で連絡を取りやすい。	ポジティブ
25	25	個人が持つ事でプライバシーが保てる事。携帯可能により急な用件や待ち合わせに便利になった。	ポジティブ
26	26	コミュニケーションが密になった。	
27	27	公衆電話を探さなくて良い。ひまつぶしができる。	ポジティブ
28	28	人との待ち合わせの時など、以前固定電話しかなかった頃に比べると、いつでも連絡が取れるという利点はかなり大きいと思います。	ポジティブ

図1 SPSSによるテキスト感情分析の例①

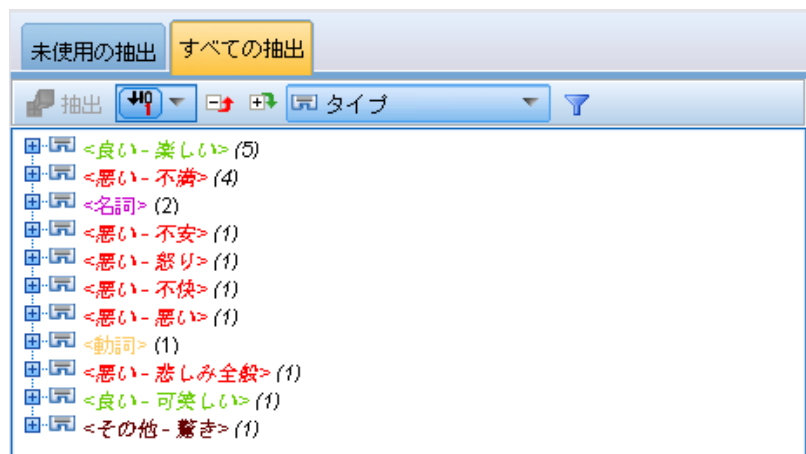


図2 SPSSによるテキスト感情分析の例②

4.2 外部条件

本研究ではツイートの属性を季節ごとに分けるだけでなく、図1のように気温・湿度・降水量の3軸を用いて合計27個のブロックに分け、ツイートに外部条件として付与した。尚、全ての気象条件は気象庁のデータを利用している。

気温（平均気温）/湿度（平均湿度）：各季節ごとの最大値と最小値を3等分しグループ分けを行った。

降水量（合計降水量）：地域によってばらつきはあるものの、降水量=0の日が1年の内約3~5割占めており、降水量=0, 0<降水量≤中央値, 中央値<降水量の3グループに分けた。

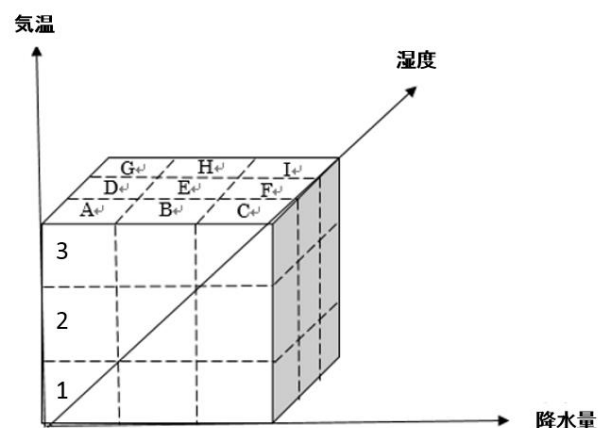


図1 外部条件のブロック分け

4.3 対象観光地・対象期間

満足度評価対象観光地と対象期間は表1の通りである。「なずきのおと」によるツイートデータ取得には時間がかかり、また過去のツイートを遡れるのは13ヶ月までなので、データ取得時期に若干の差がある。各地方人口が最も多い都道府県の県庁所在地の気象データを使用し、対象観光地はその都道府県内でるるぶ.comにおいて上位のものを幾つか抽出した。尚、関東地方について最も人口が多い都道府県は東京都であるが観光ツイートの数

が多すぎデータ取得に時間がかかりかかってしまうため、次点で人口の多い神奈川県を対象観光地とした。同様な理由でツイートの取得を中断した観光地が数箇所ある。また、研究着手当初は石川県金沢市の観光地に対して満足評価を行う予定だったので、中部地方の観光地は石川県としており、ツイートの取得期間も他の観光地に比べ早くなっている。

表1 対象観光地と対象期間^⑧

地方名	都道府県名	観光地名	地方名	都道府県名	観光地名
北海道	札幌	支笏湖	近畿	大阪	空中庭園(展望台)
		白い恋人パーク			梅田ジョイポリス
		美瑛の丘			
東北	宮城	仙台城址	中国	広島	ガラスの里
		日和山公園			大和ミュージアム
		鳴子峡			
			四国	愛媛	タオル美術館
					四国カルスト
関東	神奈川	横浜大世界			
		万葉公園	九州	福岡	マリンワールド海の中道
					太宰府天満宮
中部	石川	兼六園			
		21世紀美術館			分析対象期間
		近江町市場		緑字: 2013/09/01～2014/08/31	
		金沢城		赤字: 2013/10/01～2014/9/30	
		長町武家屋敷		青地: 2013/11/01～2014/10/31	

4.4 評価手法

本研究では満足評価の手法として以下に示す 2 通りの方法を使用した。

① 満足度=Positive ツイート数-Negative ツイート数

② 満足度= $\frac{\text{Positive ツイート数}-\text{Negative ツイート数}}{\text{Positive ツイート数}+\text{Negative ツイート数}} \times 100$

上式①において 基本的に Positive ツイートの数は Negative ツイートの数より多いため、このような式順番になった。また①式により、あるブロックにおける観光客の観光地に対しての注目度とその感情ベクトル（図 1 左図）を可視化し、②式により、観光客の満足度評価の偏りを可視化する（図 1 右図）。

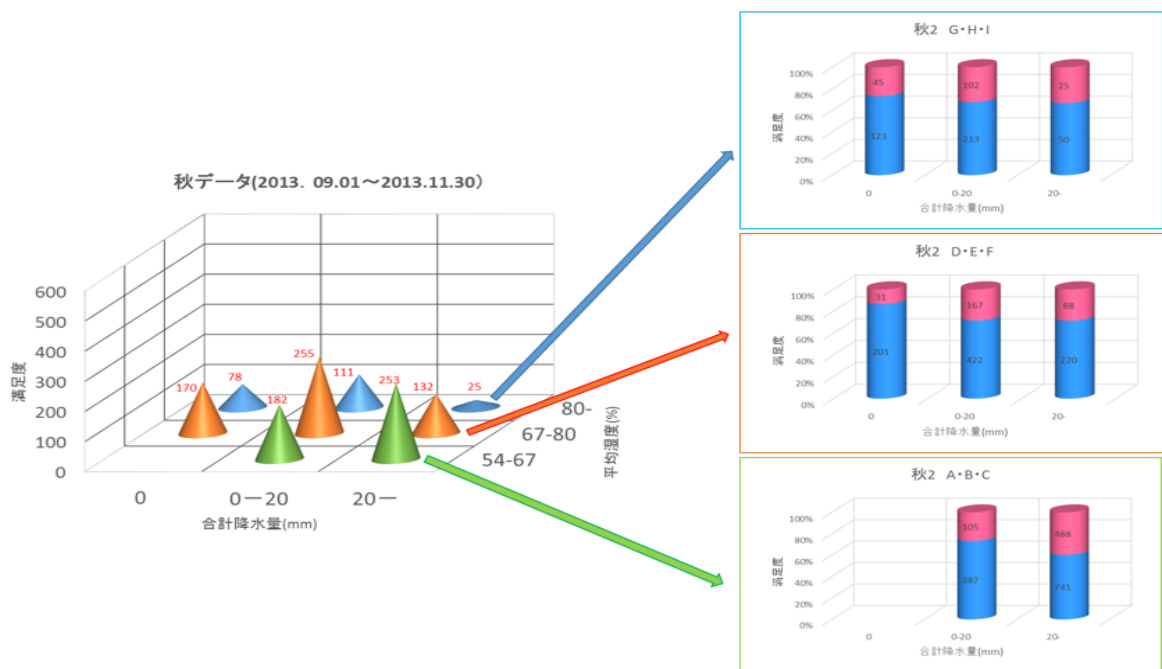


図1 満足度評価サンプル図

5. 満足度評価分析結果

この章では満足度分析結果を記載し、ひとつの季節毎に分析結果に関する評価を述べる。以下に、中部地方の満足度対象地域である兼六園の満足度評価の分析結果を記載し、その他の地域の満足度評価はページの都合上割

愛する..グラフは秋→冬→春→夏の順に記載し、各シーズン気温が高いグループ（気温グループ3）から記載することとし、グラフごとのデータの個数をデータラベルで表示した。尚、秋から記載するのはデータ取得の時期都合上、最も古いツイートデータが秋（9月/10月/11月）から始まっているためである。

5.1 兼六園の評価

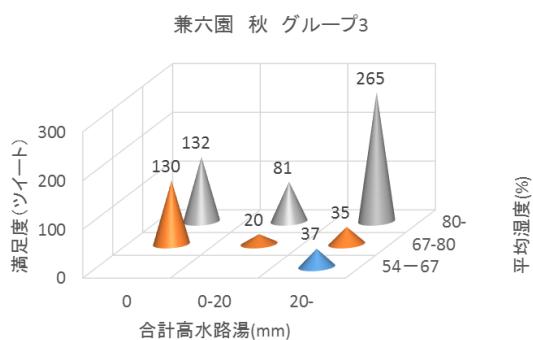


図 1-1 評価手法① 気温グループ3

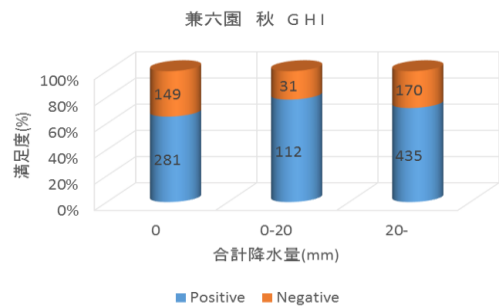


図 1-2 評価手法② 気温グループ3 GHI

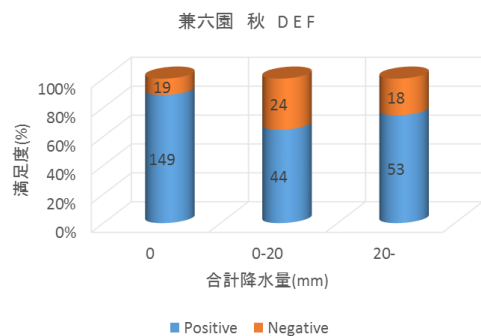


図 1-3 評価手法② 気温グループ3 DEF

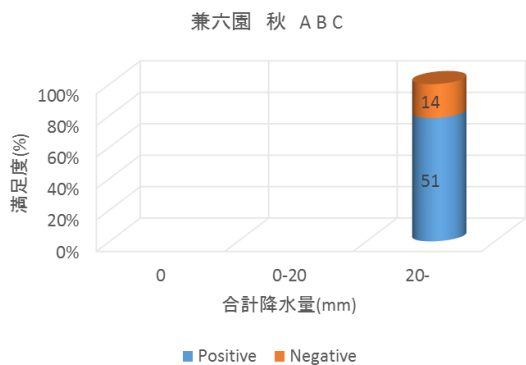


図 1-4 評価手法② 気温グループ 3 ABC

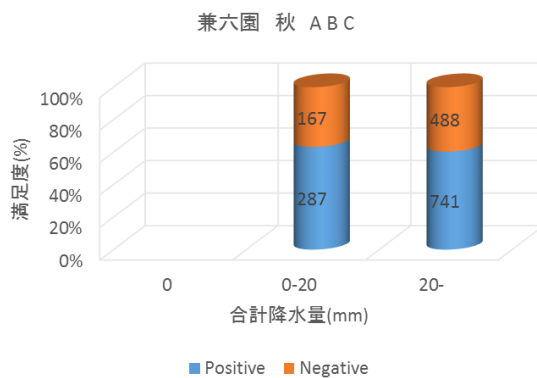


図 2-4 評価手法② 気温グループ 2 ABC

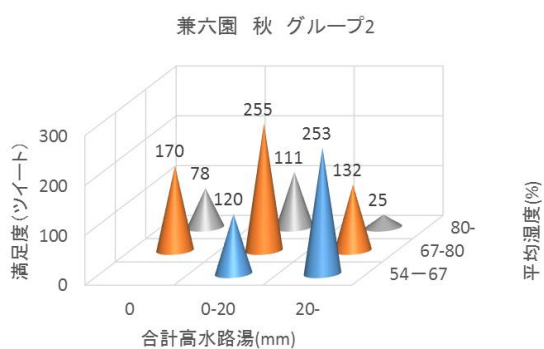


図 2-1 評価手法① 気温グループ 2

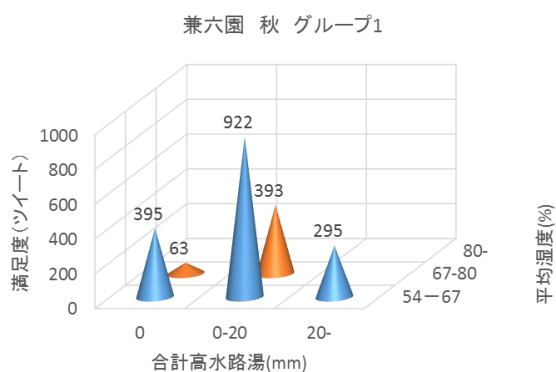


図 3-1 評価手法① 気温グループ 1

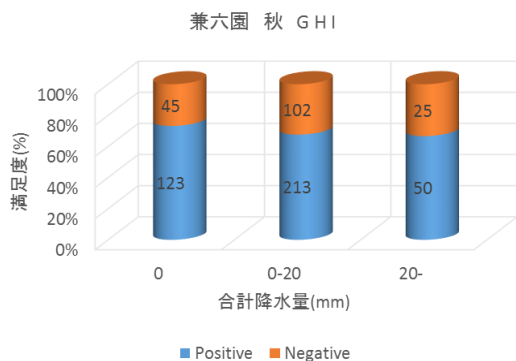


図 2-2 評価手法② 気温グループ 2 GHI

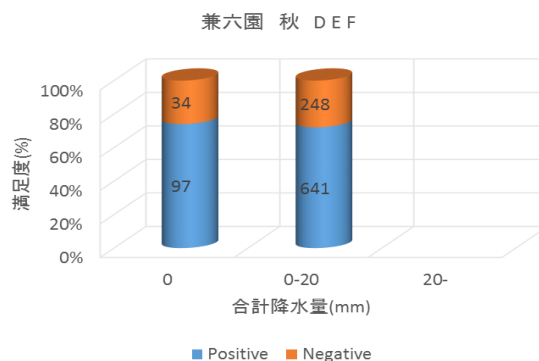


図 3-2 評価手法② 気温グループ 1 DEF

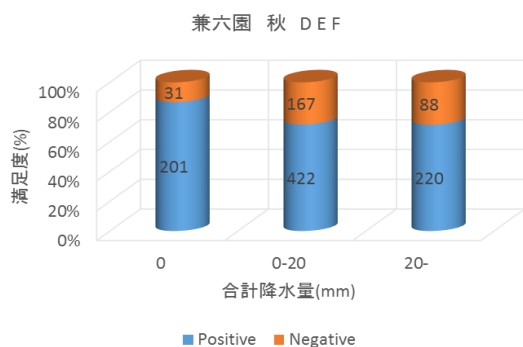


図 2-3 評価手法② 気温グループ 2 DEF

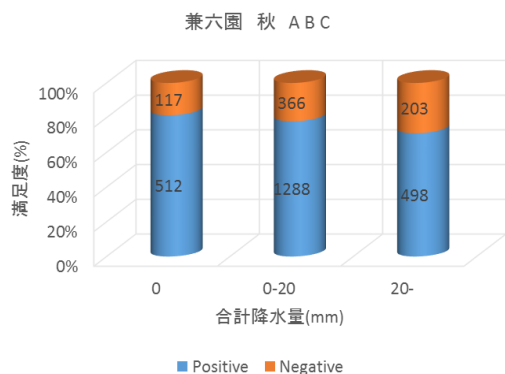


図 3-4 評価手法② 気温グループ 1 ABC

兼六園は春の桜の開花時期、秋の紅葉の時期に無料開放を行っており、ツイートには秋の無料開放の際に訪れたと推測されるものが多く、そのほとんどが兼六園に対して **Positive** な評価をしたものであった。また、評価手法②において若干 **Negative** の割合が高いためツイートを見ると、この時期に選挙活動がされており演説の音が兼六園の閑静で情緒ある景観を壊しているといった内容のものが多かった。他には無料開放による観光客増加に伴う駐車場不足に関するものや渋滞など、交通環境に対しての不満も多く見受けられた。

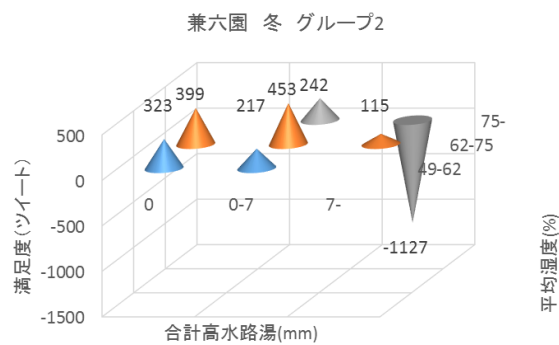


図 5-1 評価手法① 気温グループ 2

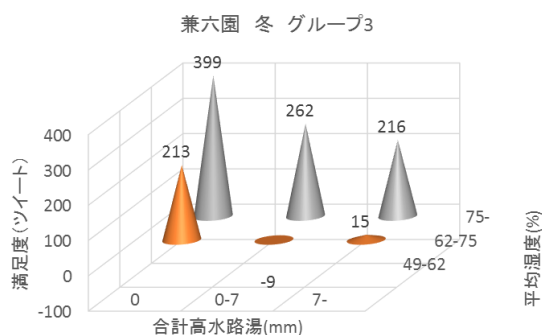


図 4-1 評価手法① 気温グループ 3

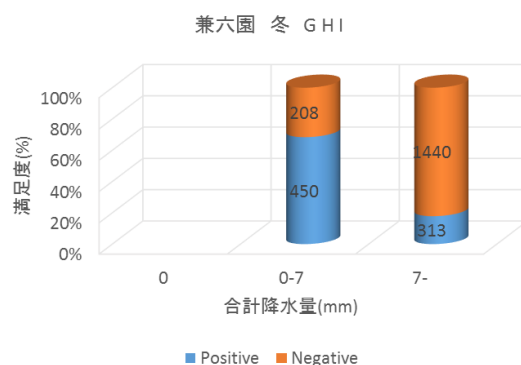


図 5-2 評価手法② 気温グループ 2 GHI

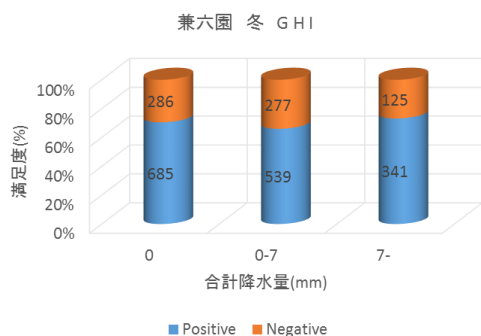


図 4-2 評価手法② 気温グループ 3 GHI

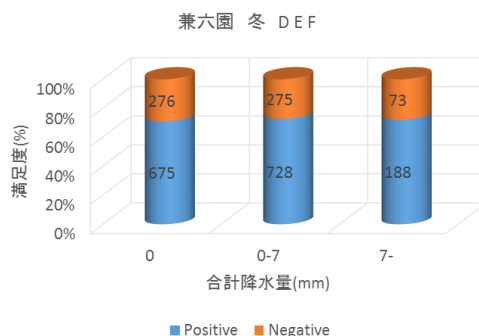


図 5-3 評価手法② 気温グループ 2 DEF

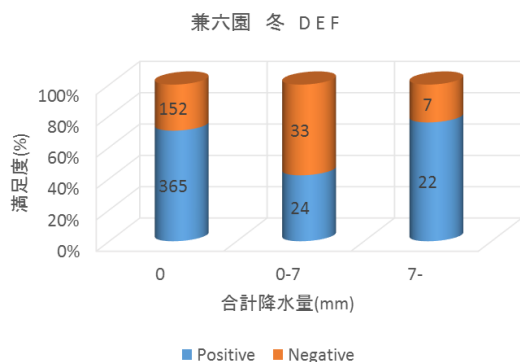


図 4-3 評価手法② 気温グループ 3 DEF

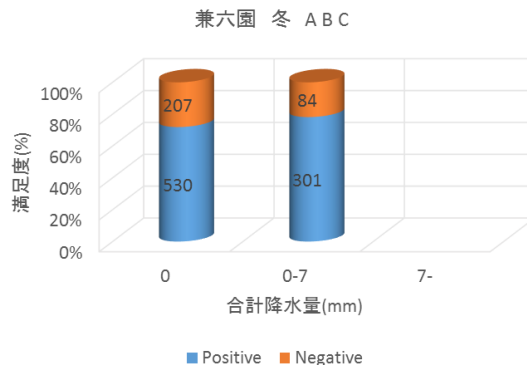


図 5-4 評価手法② 気温グループ 2 ABC

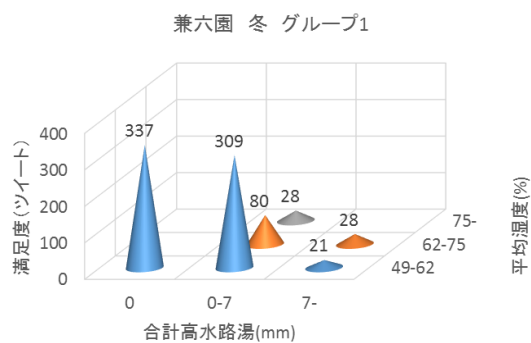


図 6-1 評価手法① 気温グループ 1

冬は評価手法①においても評価手法②においても気温グループ 2 のブロック I が異状に **Negative** な評価となっている。ツイートを見たところ、最も大きな要因は兼六園付近で熊が出没し、その情報がツイートによって広く拡散しておりそれに関して「怖い」など恐怖や驚きを表すツイートが大多数を占めていた。またこれに関して、熊が出没したという情報ばかりが拡散されており、それに対する措置や事態の収集状況などの情報がほとんどツイートされていなかった。特にこのような普段の生活から逸脱した事象に対し、ツイートによる情報の不確か性と欠如性、また一通性を感じた。

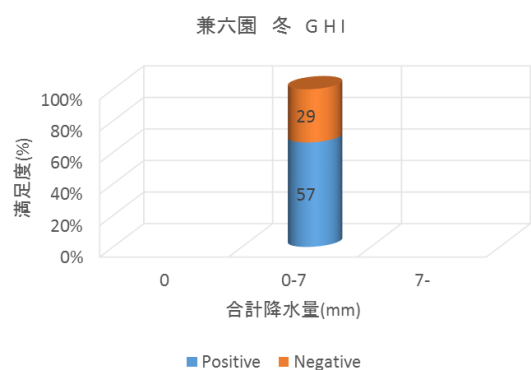


図 6-2 評価手法② 気温グループ 1 GHI

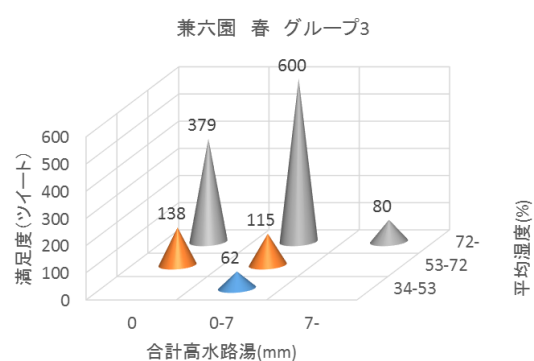


図 7-1 評価手法① 気温グループ 3

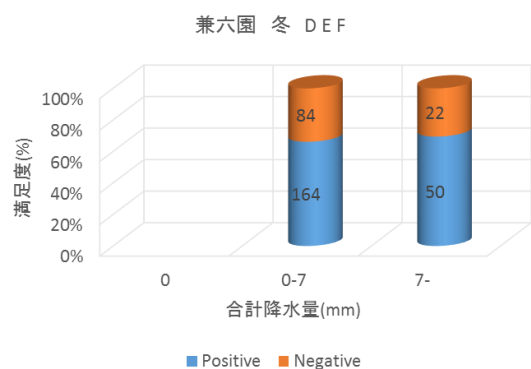


図 6-3 評価手法② 気温グループ 1 DEF

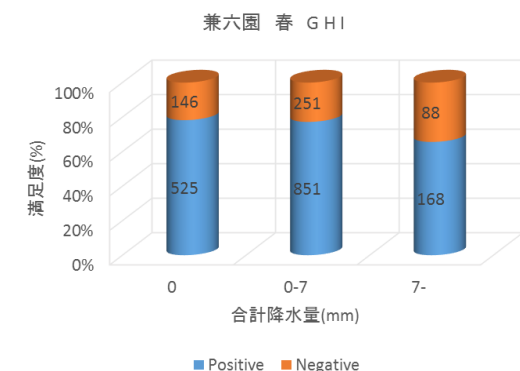


図 7-2 評価手法② 気温グループ 3 GHI

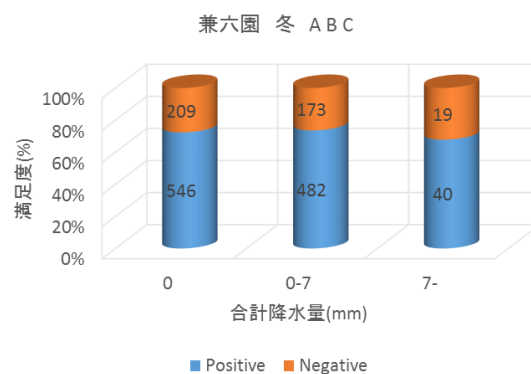


図 6-4 評価手法② 気温グループ 1 ABC

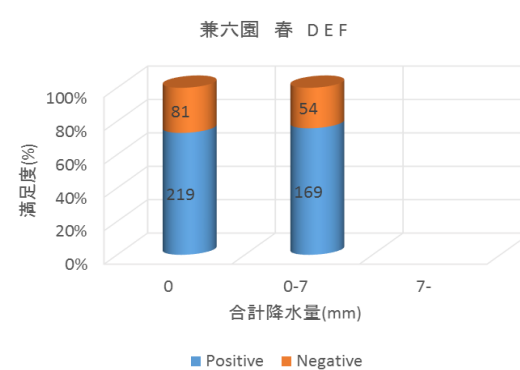


図 7-3 評価手法② 気温グループ 3 DEF

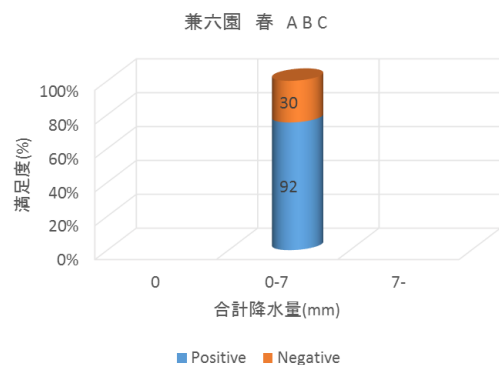


図 7-4 評価手法② 気温グループ 3 ABC

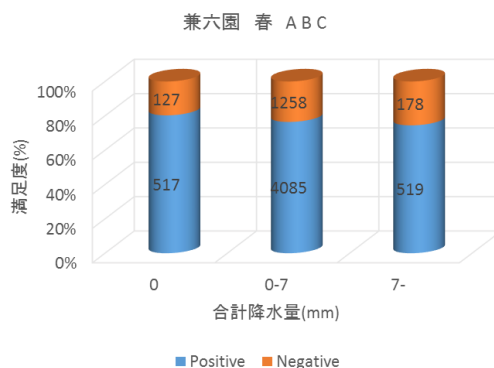


図 8-4 評価手法② 気温グループ 2 ABC

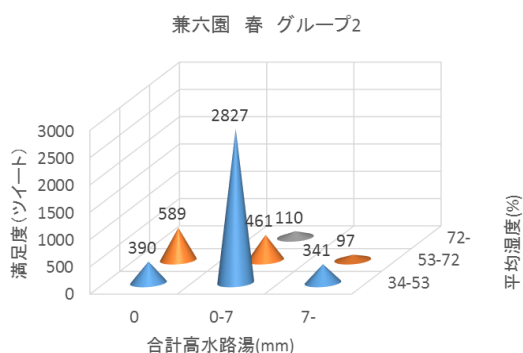


図 8-1 評価手法① 気温グループ 2

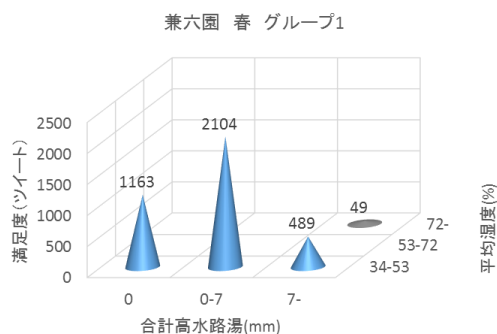


図 9-1 評価手法① 気温グループ 1

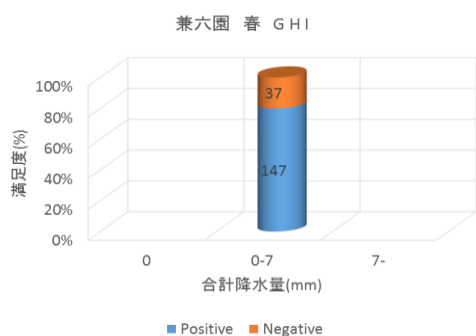


図 8-2 評価手法② 気温グループ 2 GHI

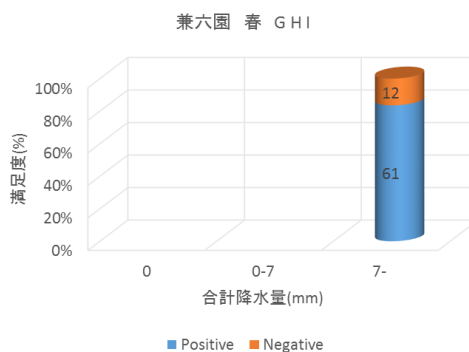


図 9-2 評価手法② 気温グループ 1 GHI

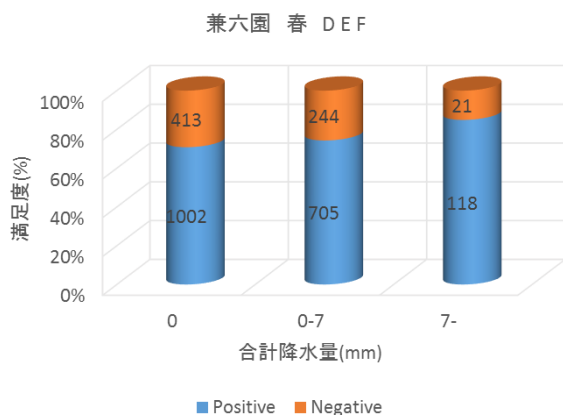


図 8-3 評価手法② 気温グループ 2 DEF

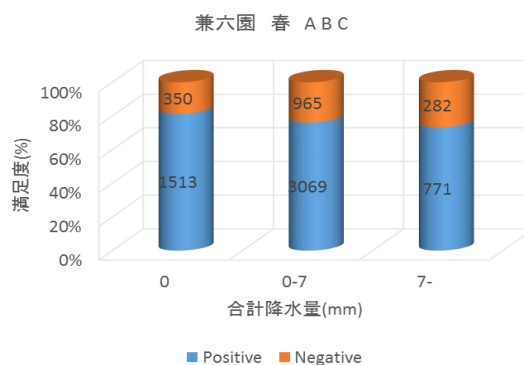


図 9-3 評価手法② 気温グループ 1 ABC

春は桜の時期でツイートも多く、評価手法②のグラフも全体的に **Positive** の割合が高い。また評価手法①において、気温グループ②及び③に関して湿度が低くなるほど **Positive** ツイートの数が増加する傾向にあると分かる。尚、気温グループ③に関しては、このグループは同季節中気温が高いグループのため花見や梅の花を見るには最適な時期となる。そのため湿度による感情へのマイナス影響が相対的に小さくなり、先述した傾向が見られないのだと考えられる。また、少数ではあるが雨上がりの兼六園に対して「趣が増す」や「落ち着く」などのツイートも見られた。

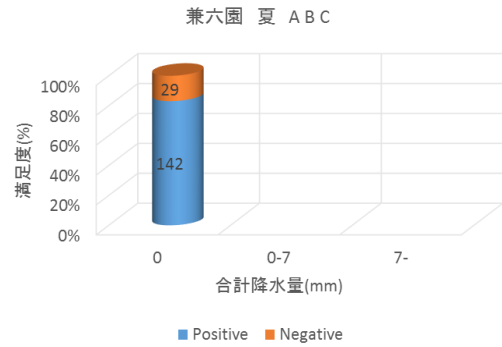


図 23-4 評価手法② 気温グループ 3 ABC

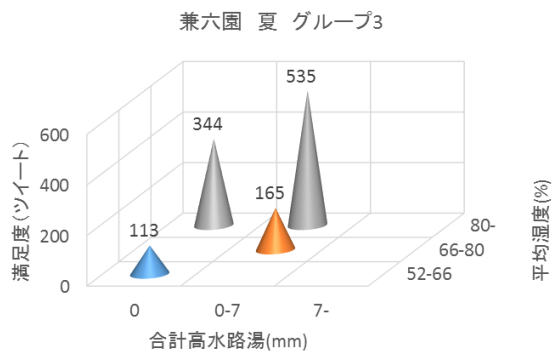


図 10-1 評価手法① 気温グループ 3

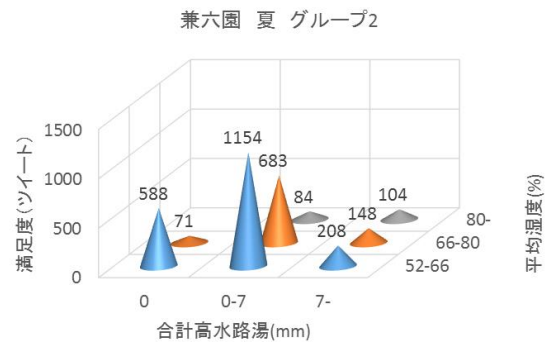


図 24-1 評価手法① 気温グループ 2

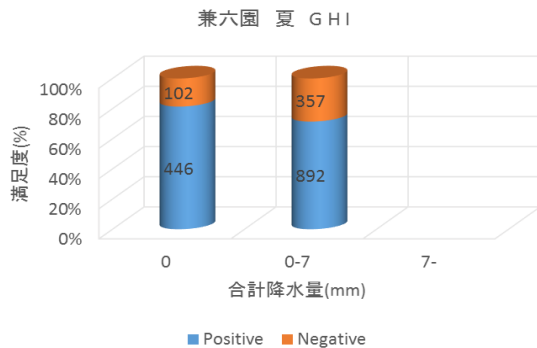


図 23-2 評価手法② 気温グループ 3 GHI

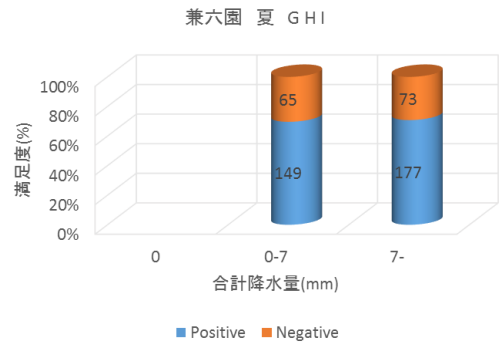


図 24-2 評価手法② 気温グループ 2 GHI

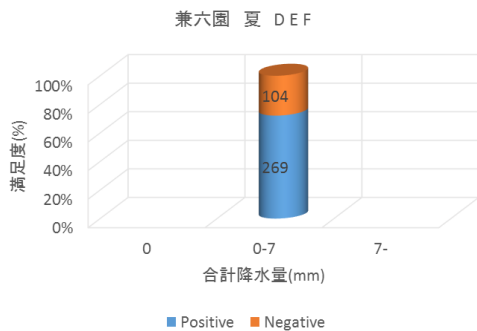


図 23-3 評価手法② 気温グループ 3 DEF

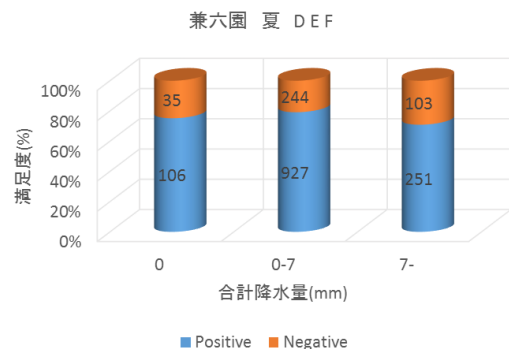


図 24-3 評価手法② 気温グループ 2 DEF

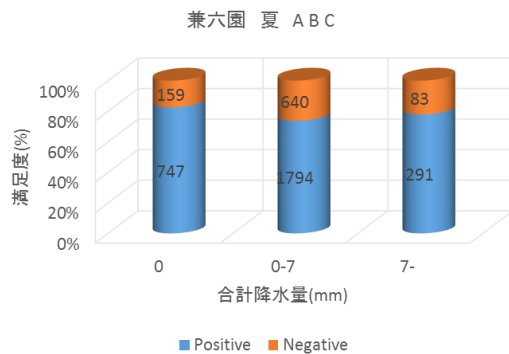


図 24-4 評価手法② 気温グループ 2 ABC

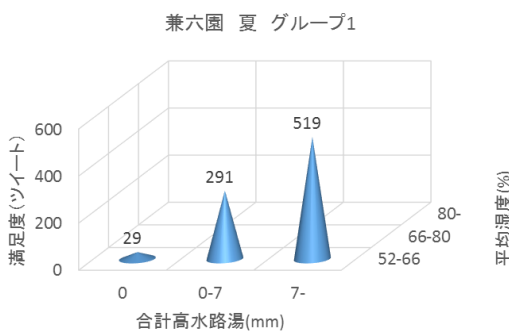


図 25-1 評価手法① 気温グループ 1

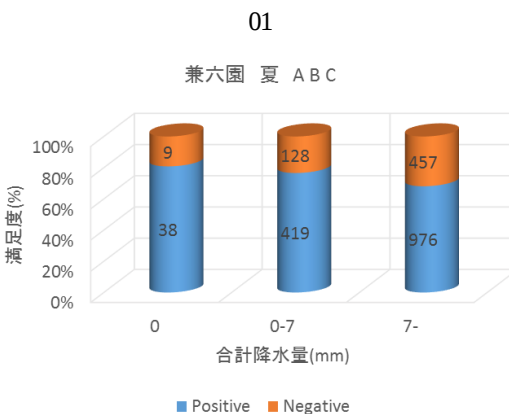


図 25-2 評価手法② 気温グループ 1 ABC

夏も評価手法②による割合評価の満足度は概ね安定して Positive:Negative=4:1 の割合である。また、評価手法①による満足度評価では気温グループ②のブロック B で Positive ツイートがとびぬけて高く、ツイートから百万石祭りや兼六園ライトアップによるものだと分かった。兼六園に関しての Negative なツイートの多くに「駐車場」「渋滞」「混雑」等のキーワードが含まれ、兼六園近郊の交通環境を改善すべきという課題点も見出すことができた。

6. 結論と課題

本研究で Twitter を使った観光地の満足度評価の手法を構築した。また、観光客の満足度と天気（本研究における外部条件）は密接な関係があると断言するには少しデータが乏しいが、少なくとも天気は満足度に影響を与える可能性が高いといえる結果となった。これに関して、より多くの観光地でより多くのツイートをできるだけ精度良く取得できたら更なる結果が期待できると考える。また、今回は天気による満足度の影響を明らかにしようと試みたが、研究途中で各地のイベントに関する感想ツイートはもちろん、情報ツイートや推薦ツイートがかなり多いことに気がついた。金沢市内に熊が出没したときの評価の急激な下落や、兼六園や金沢城等の無料開放時のツイートの急激な増加ツイートなど、Twitter にはその時の「特別な情報」に対し、大きな拡散力と速攻力があることを明らかにできた。また、Twitter は誰でも手軽に利用できるという大きな利点を持っており、3 章で紹介した手段を使えばあらゆる人が容易にデータを集めることができる。本研究で使用した SPSS というテキストマイニングソフトに関しても、テキスト解析能力を多少妥協すればフリーソフトに R というものがある。つまり、資金や特別な能力がなくても本研究のような分析は可能だということであり、Twitter を用いた分析の多様性や容易性を明らかにしたということになる。このことも大きな研究の成果の一つだと考える。

本研究の課題は、データ取得の精度である。本研究で使ったなずきのおとでは先述の通り除外ユーザーを選択することができるが、ビッグデータを用いて行う研究において重要なのは、何度も繰り返しツイートされたりリツイートされたツイートそのものを除外する機能であると考えられる。Twitter 内では、あるユーザー A が発信したツイートが他ユーザー B にリツイートされ、それをユーザー B のフォロワー C が更にリツイートするというサイクルが繰り返されることでツイートは拡散していき、その過程で情報発信者は次々と変わることになる。しかも、検索キーワードや時期、場所が違えば当然対象ユーザーは毎回異なるため、除外ユーザーを選択できるという機能はほとんどその役割を果たしていない。大量に拡散されたりリツイートされたツイートを削除することで、1 ユーザー当たりのツイートの重みが等しくなり、より正確な分析を行うことができる。

最後に挙げるのはユーザーの属性である。Twitter のユーザーは若者中心で高齢者の利用率は低い。大量のデータを使用することで、ある程度この問題は相殺することができる。しかし、今回のようにテーマが観光などの対象が若者ではなく高齢者に重みを置きたいときは、自分が

集めたい高齢者層からの情報は少なく,研究の障害となる若者層からの情報が多くなってしまうという事態に陥ることは免れない,プライバシーの問題もあるので難しいところではあるが,年齢や性別等のカテゴリー情報が付属したツイートの利用が可能になればこの分野の研究の精度および利用用途は飛躍的に向上するものと考ええる。

7. 参考文献

1) ANNUAL REPORT (2015/2/6)

https://materials.proxyvote.com/Approved/90184L/20140328/AR_202076/pubData/source/Twitter,%20Inc%202013%20Annual%20Report.pdf

2) なずきのおと (2015/2/6)

<https://nazuki-oto.com>

3) BINARY(2015/2/6)

<http://www.sophia-it.com/content/Topsy>

4) TOPSY(2015/2/6)

<http://topsy.com/>

5) ツイポーター (2015/2/6)

<http://twport.com/>

6) Twitter (2015/2/6)

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

7) 内田治, 川嶋敦子, 磯崎幸子 : SPSSによるテキストマイニング入門, オーム社

8) るるぶ.com(2015/2/6)

<http://www.rurubu.com/>

Development of satisfaction evaluation method of tourism spot using tweet data

Daichi SUZUKI, Makoto FUJII, Jyunnti TAKAYAMA and Syouichiro NAKAYAMA

There are many people or organization send information by SNS or blogs, and the number of them have been increased. There are useful information some of the information. We can analysis the satisfaction of the tourist spot in greater detail by using the data of SNS or blogs because of the anonymity and free space on a network. For example, discovering a new appeal point, detecting a new problem, proposing a new root of sightseeing. We can do them in real time or seasonal and offer more detailed information to tourist. In this thesis, I perform the satisfaction evaluation of the tourist spot by using the service of “Nazuki No Oto”.