

観光地名無しツイートにおける観光地に関する感想の 自動抽出手法の提案

渡邊 小百合¹ 吉野 孝¹

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きている。そのため、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。先行研究において、位置情報付きツイートや観光地名を含むツイートの前後ツイート等を用いて観光地名を含まないツイートから観光地に関する感想を収集する手法を提案したが、収集した情報が観光地に関する感想であるかの判断は全て人手で行っており、収集に時間と手間がかかるという問題があった。そこで本研究では、観光地名無しツイートにおける観光地に関する感想の自動抽出手法を提案する。観光地名を含むツイートを用いて、観光地に関する感想を含むツイートを判定する分類器を構築し、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自動抽出を行う。構築した分類器の精度検証を行った結果、以下の2点を明らかにした。(1) 観光地名入りツイートに含まれる観光地に関する感想を高精度(正例再現率 > 0.8)で分類できる。(2) 観光地名入りツイートを用了分類器によって、観光地名無しツイートも分類できる可能性がある。

Proposal of Automatic Extraction Methods of Impressions about Tourist Spots in Tweets not Containing Tourist Spot Names

SAYURI WATANABE¹ TAKASHI YOSHINO¹

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、外国人旅行者が 2012 年から年々増加しており、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、観光客の観光行動は変化してきている。しかし、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ても、観光地側がそれに気づいてい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 [4] では、位置情報付きツイートや観光地名を含むツイートの前後ツイート等を用いて観光地名を含まないツイートから観光地に関する感想を収集する手法を提案した。しかし、収集した情報が観光地に関する感想であるかの判断は全て人手で行っており、収集に時間と手間がかかるという問題があった。

そこで本研究では、観光地名無しツイートにおける観光地に関する感想の自動抽出手法を提案する。観光地名を含むツイートをを用いて、観光地に関する感想を含むツイートを判定する分類器を構築し、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自動抽出を行う。

2. 関連研究

2.1 ツイート文を用了情報抽出に関する研究

眞野らは、ユーザ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 [5]。同じクラス

¹ 和歌山大学
Wakayama University Sakaedani 930, Wakayama, 640-8510 Japan

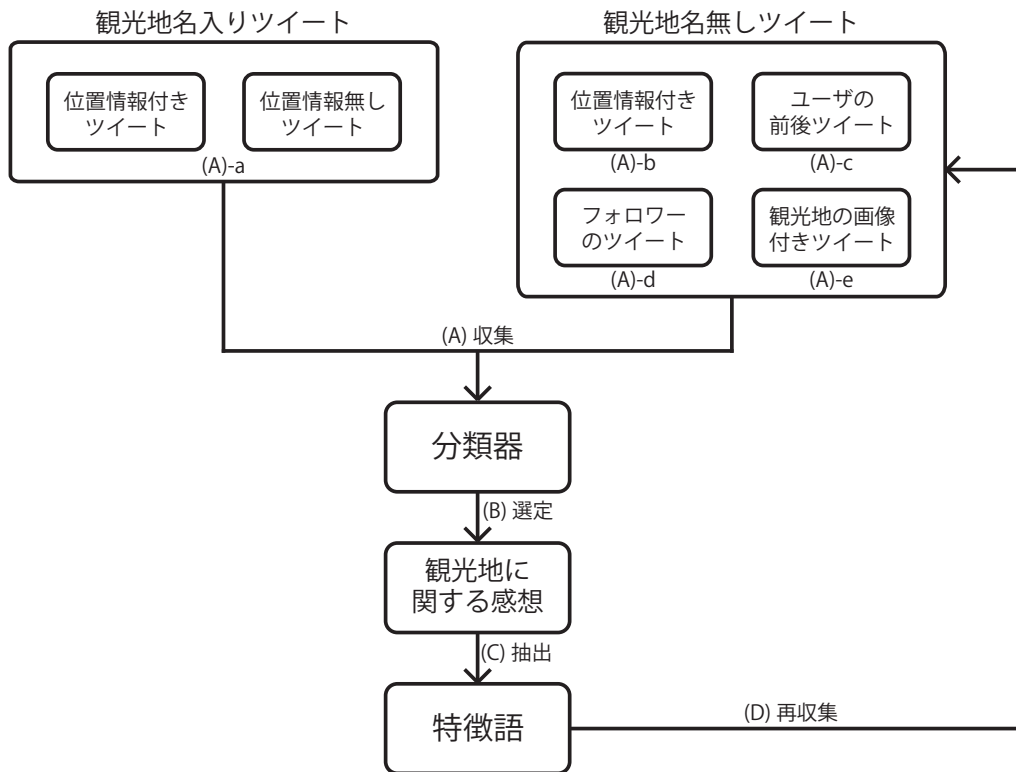


図 1 観光地に関する感想の自動抽出手順

タに属するユーザ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 [6]。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 [7]。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、観光地に関する情報を地図上に表示する。Alan 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 [8]。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係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情報の抽出を行う。

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 [9]。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、クラスタリングを行う。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 [10]。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の推薦を行っている。Ryong 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 [11]。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時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極めて少なく、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、そのユーザの人間関係等を用いて、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。また、ツイートの付加情報からだけではなく、観光地名入りツイートを用いた分類器によって、観光地名を含まないツイートから観光地に関する感想を含むツイートの抽出を行う。

表 1 観光地名入りツイートに含まれていた感情を表す語の例とその数

| 単語 | 含まれていた数 |
|---------|---------|
| 楽しい | 386 |
| 良い | 371 |
| 美味しい | 170 |
| 綺麗だ | 170 |
| 好き | 170 |
| 凄い | 139 |
| 可愛い | 134 |
| 最高 | 74 |
| 嬉しい | 60 |
| 怖い | 56 |
| 面白い | 54 |
| やばい | 38 |
| 素敵だ | 37 |
| 驚く・びっくり | 36 |
| 気持ちいい | 29 |
| 幸せだ | 26 |
| かっこいい | 25 |
| 辛い | 20 |
| 美しい | 20 |
| 無理 | 19 |

3. 観光地に関する感想の自動抽出手法

本研究の目的は、観光地名無しツイートからの観光地に関する感想の収集を自動化することである。

本研究は、観光地名無しツイートからの観光地に関する感想の抽出手法において、機械学習を用いて収集した情報が観光地に関する感想かどうかの選定を行う。これより、観光地に関する感想の収集を自動化することを目指す。

図 1 に観光地に関する感想の自動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の収集

位置情報付きツイートと位置情報無しツイートから、観光地名を含むツイートを収集する(図 1(A)-a)。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 [10] を用いて、観光地に関することが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)-b)。

(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の前後 3 時間以内のツイートを収集する(図 1(A)-c)。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、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ため、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、収集する時間の幅を観光地や観光地の種類ごとに設定する必要があるが、今は検討中である。

(3) フォロワーのツイートの収集

観光地名を含むツイートをしたユーザのフォロワーのツイートを収集する(図 1(A)-d)。観光地名入りツイートをしたユーザのフォロワーの中に、その時一緒に行動していた人がいる可能性があるため、ユーザのフォロワーのツイートから、観光地名入りツイート付近のツイートを収集する。この時、(2) と同じく、観光地名入りツイートの前後 3 時間以内に呟かれたフォロワー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)-e)。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) 観光地に関する感想の選定

(1)~(4) において収集したツイートの中から観光地に関する感想を選定する(図 1(B))。観光地に関する感想の選定には、観光地名入りツイートをを用いて構築した分類器を用いる。この分類器については、詳しくは後述する。

(6) 特徴語の抽出

(5) で選定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))。形態素解析システム JUMAN^{*2} を用いてツイート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。

(7) 特徴語を含むツイートの収集

位置情報無しツイートから、(6) で抽出した特徴語を含むツイートを収集する(図 1(D))。この収集では、特徴語を含む位置情報付きツイートや画像付きツイート、特徴語入りツイートの前後ツイート、特徴語入りツイートをしたユーザのフォロワーのツイートを収集対象とし、すでに収集したツイートは除外する。

(8) 観光地に関する感想の再収集

(5)~(7) 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から観光地に関する感想を再収集する。

4. 観光地に関する感想の分類器

先行研究 [4] においては、収集した情報が観光地に関する感想であるかの判断は全て人手で行っており、収集に時間と手間がかかるという問題があった。そこで、本研究で

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

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

表 2 収集する対象の観光地とツイート数および正例数

| 地域 | 観光地名 | ツイート数 | 正例数 |
|------|----------------------|-------|-----|
| 北海道 | 藻岩山 (北海道札幌市) | 72 | 46 |
| 東北 | 瑞鳳殿 (宮城県仙台市) | 35 | 24 |
| 関東 | 横浜みなとみらい 21(神奈川県横浜市) | 101 | 50 |
| 中部 | トヨタ産業技術記念館 (愛知県名古屋市) | 22 | 21 |
| 近畿 | 道頓堀 (大阪府大阪市) | 496 | 163 |
| 中国 | 広島平和記念資料館 (広島県広島市) | 34 | 27 |
| 四国 | 松山城 (愛媛県松山市) | 136 | 73 |
| 九州 | 大濠公園 (福岡県福岡市) | 118 | 67 |
| 沖縄 | 首里城 (沖縄県那覇市) | 82 | 32 |
| 日本全体 | 伏見稲荷大社 (京都府京都市) | 568 | 389 |

表 3 観光地名を含むコーパスの一部

| 正例 (1)/負例 (0) | ツイート内容 ^{※1} |
|---------------|---|
| 1 | <p>T1: トヨタの「紡績→自動車」って話は名古屋の トヨタ産業技術記念館 行けば学べるし、あそこ行けば格好良い車が沢山あるし <u>楽しい</u>。</p> <p>T2: 今日は数年ぶりに <u>松山城</u> 登ってききました！八重桜も満開で、空もとても <u>綺麗</u>でした！！<u>松山城</u> も久しぶりに見れて <u>良かった</u>！行けて <u>良かった</u>～ < URL > ^{※2}</p> <p>T3: 伏見稲荷 ちょちゃちょちゃ観光客が溢れてたけど、山頂付近まで参拝している人は少なく、その辺まで行くと写真撮るのが <u>好き</u> そうな人達が譲り合って写真撮ってるの何か <u>面白かった</u>。人写ってるのも人写ってないのも撮るのが <u>楽しい</u> よね。人ががやがやしてる神社も <u>好き</u> だよ。神社は人の想いが集まる場所だし。</p> |
| 0 | <p>F1: @username せやな！高体連で富士山登れるの本当 <u>楽しみ</u> だな！今まで練習で使ってきた <u>藻岩山</u> とは全然違うけど、俺らの今までの成果を高体連でみせよーぜ！！ ^{※3}</p> <p>F2: 昨日の <u>道頓堀</u> でとったさきちゃんっ さきちゃんのダンス歌声 <u>大好き</u> なんです いつも話しかけてくれてありがとうございます また、お話ししましょう < URL > ^{※2}</p> <p>F3: 首里城 で沖縄サミットの出席お歴々を森総理やクリントン大統領が並んでいる中、プーチン大統領は当時から国のトップって <u>驚いた</u> わ。そういえばってなった</p> |

※1 観光地名と感情を表す語を下線で示している

※2 個人情報のため URL を < URL > と表記している

※3 個人情報のためユーザ名を @username と表記している

は、観光地に関する感想の自動分類器の構築を行う。

4.1 コーパス

本研究では、2017年4月13日～4月24日のうちの1週間^{*3}に呟かれた観光地名を含む約43000ツイートから、感情を表す語を含むツイートを1000件取得し、それらのツイートが観光地に関する感想を含んでいるかどうかを手で評価してコーパスを作成した。表1に観光地名入りツイートに含まれていた感情を表す語の例と数を示す。感情を表す語は手によって判断しており、1つのツイートに複数個含まれている場合もある。コーパスには、トリップアドバイザー^{*4}において2017年4月22日時点で、北海道・東北・関東・中部・近畿・中国・四国・九州・沖縄の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^{*5}において1位である伏見稲荷大社の、合計10箇所の観光地についてのツイートを使用する。表2に収集対象とした観光地とツイート数、および正例数、表3にコーパスの一部をそれぞれ示す。表3については、観光地名と感情を表す語を下線で示している。例えば、T1やT2、T3のように観光地名と「楽しい」「良かった」「好き」という感情を表す語が含まれており、観光地に関する感想のツイートは正例、F1、やF2、F3のように観光地名と「楽しみ」「大好き」「驚いた」という感情を表す語が含まれてはいるが、観光地に関する感想ではないツイートは負例となっている。

4.2 素性と学習アルゴリズム

4.1節で述べたコーパスを用いて、観光地に関する感想を含むツイートであるかどうかを判定する分類器を構築した。形態素解析にはJUMANを用いており、素性には観光地名と、2文字以上の名詞、形容詞、形容動詞、動詞を用いた。表4に分類器に用いた素性の一部を示す。観光地

^{*3} ツイートの収集に Twitter REST API を用いており、収集開始時から過去1週間以内に呟かれたツイートのみ収集される。観光地によって収集開始日が異なっているため、各観光地について収集されたツイートは4月13日～4月24日のうちの1週間分となっている。

^{*4} <https://www.tripadvisor.jp/>

^{*5} <https://www.tripadvisor.jp/Attractions-g294232-Activities-Japan.html>

表 4 分類器に用いた素性の一部

| ツイート内容 | 素性 |
|---|-----------------------------------|
| 首里城から街を眺めるのがすごい好きだったなあ。ホント居心地が良かった。 | 観光地名 すごい好きだ ホントだ 居心地 良い |
| @username おお！素敵やん！松山城に八重桜キレイやな！ノーブランでブラブラが一番贅沢な旅な気がする！※1 | 素敵だ 観光地名 八重 キレイだ ノーブラン 贅沢 |
| 札幌2日間楽しかったなあ 色んなところに行けたしいつか見たいと思ってた藻岩山から見る札幌の夜景が綺麗すぎてほんとに感動した。また行きたいな！<URL>※2 | 札幌 楽しい ところ 観光地名 札幌 夜景 綺麗だ ほんとだ 感動 |

※1 個人情報のためユーザ名を@usernameと表記している

※2 個人情報のためURLを<URL>と表記している

表 5 観光地名入りツイートの判定精度

| ラベル | 適合率 | 再現率 | F 値 |
|--------|------|------|------|
| 正例 (1) | 0.60 | 0.83 | 0.70 |
| 負例 (0) | 0.74 | 0.46 | 0.57 |
| 平均 | 0.67 | 0.65 | 0.63 |

表 6 観光地名無しツイートの判定精度

| ラベル | 適合率 | 再現率 | F 値 |
|--------|------|------|------|
| 正例 (1) | 0.53 | 0.64 | 0.58 |
| 負例 (0) | 0.69 | 0.59 | 0.64 |
| 平均 | 0.62 | 0.61 | 0.61 |

名については、各観光地のツイート数に差があるため、ツイート内の各観光地名を全て「観光地名」に置き換えている。形容詞、形容動詞、動詞については、活用形による変化を無くすため、全て原形を用いている。

これらの素性を用いて、Python ライブラリ scikit-learn による SVM*6 で分類器を構築した。SVM による学習には RBF カーネルを用い、パラメータは 5 分割交差検定とグリッドサーチによって決定した $C=4$, $\gamma=0.125$ を用いている。

5. 実験

本実験では、構築した分類器の精度を検証するため、以下の2つの実験を行う。

(1) 観光地名入りツイートの分類

観光地名入りツイートに含まれる観光地に関する感想を分類できなければ、観光地名無しツイートの分類も困難であると考えられるため、本実験は、構築した分類器によって、観光地名入りツイートが観光地に関する感想を含むツイートかどうか判定できるかを検証する。分類器に入力するデータとしては、コーパスに使用したツイート以外の観光地名と感情を表す語を含む 416 ツイートを用いる。

(2) 観光地名無しツイートの分類

本実験は、構築した分類器によって、観光地名無しツイートが観光地に関する感想を含むツイートかどうか判定できるかを検証する。分類器に入力するデータとしては、10箇所観光地について観光地名入りツイートの収集と同じ期間に3章で述べた抽出手法によって収集した 145672 ツイートから、ランダムで各 10 ツイートずつの合計 100 ツイートを用いる。

どちらの実験においても、入力するツイートは形態素解析を行い、分類器の構築に用いたのと同じ 2 文字以上の名

詞、形容詞、形容動詞、動詞の原形となっている。

6. 実験結果と考察

6.1 観光地名入りツイートの分類

表 5 に、分類器による観光地名入りツイートの判定精度の結果を示す。正例は F 値が 0.6 以上となったが、負例の F 値は 0.6 を下回った。その要因としては、負例の再現率が 0.46 と低い結果であることが挙げられる。しかし、負例の適合率は 0.74、正例の再現率は 0.83 と高い結果であることから、観光地に関する感想を含まないツイートを負例として判定し、多くの観光地に関する感想を含むツイートを正例として判定していることがわかる。この結果から、観光地に関する感想を含むツイートを正例として判定する精度が高いため、分類器を用いることで観光地名入りツイートにおいて、観光地に関する感想を含むツイートの自動抽出が可能であることがわかった。

6.2 観光地名無しツイートの分類

表 6 に、分類器による観光地名無しツイートの判定精度の結果を示す。負例は F 値が 0.6 以上となったが、正例の F 値は 0.6 を下回った。この要因としては、正例の適合率が 0.53 と低い結果であることが挙げられる。しかし、負例の適合率が 0.69、正例の再現率が 0.64 と、どちらも 0.6 以上であることから、観光地名入りツイートの分類結果と同様に、観光地に関する感想を含まないツイートを負例として判定し、多くの観光地に関する感想を含むツイートを正例として判定していることがわかる。この結果から、観光地名入りツイートを用了分類器によって、観光地名無しツイートにおいても、観光地に関する感想を含むツイートの自動抽出が可能であると考えられる。

表 7 に、誤判定されたツイートの例を示す。観光地に関する感想を含んでいるが、感想ではないと判定されたツイートは、T1 のようにツイートには「夜桜」といった観光地の特徴となり得る単語が含まれているが、形態素解析

*6 <http://scikit-learn.org/stable/modules/svm.html>

表 7 誤判定された観光地名無しツイートの例

| 本来のラベル | 分類器の結果 | ツイート内容 | 入力した素性 |
|--------|--------|--|---|
| 正 (1) | 負 (0) | T1: @username 夜桜は難しいよね…いいところじゃないと、普通の電灯すぎて桜の色が生えないからね…夜桜とるより、青空をバックに食堂あたりの桜撮るほうがいいかもよ *1 | 難しいいいところ普通だ電灯すぎる生える とる 青空 バック 食堂 あたる 撮る ほう いい |
| | | T2: @username 私は残念ながら肉眼で閲覧したことないので憶測ですが人形という表現は実際の資料や体験談より胸に刻まれるものが多かったとの意見が多いので懸念されている事には同じく不安を覚えますね…*1 | 残念だ 肉眼 閲覧 すること ない のだ 憶測 人形 いう 表現 資料 体験 刻む もの 多い 意見 多い のだ 懸念 同じだ 不安だ 覚える |
| | | T3: 都会の景観でテンション上がったのって神戸と横浜と新宿くらいだな。街の造形美と情報過多にならない程良さが目に楽しい。駅周辺や歓楽街とかは多過ぎる人と看板、電飾、色んな音声で凄く疲れて仕舞うんだよね。物珍しい内は楽しくても自分で吃驚するくらい体調崩すから苦手だわ。 | 都会 景観 テンション 上がる 神戸 横浜 新宿 造形 情報 過多だ 程良い 楽しい 周辺 歓楽 看板 電飾 音声 凄い 疲れる 仕舞う んだ 物珍しい 楽しい 自分 吃驚 する くらい 体調 崩す 苦手だ |
| 負 (0) | 正 (1) | F1: 洋画とか海外ドラマのえっ、お前流れについてこれないってうせやろ？みたいな感じでドコドコ話が進むのほんとすき(で、後から気づいたりする) | 洋画 海外 ドラマ えっ 流れる つく うせる みたいだ 感じる 進む ほんとだ すきだ 気づく |
| | | F2: 吉村さん面白いから良いよ デスマッチの専属司会者でいいよ | 吉村 面白い 良い デスマッチ 専属 司会 いい |
| | | F3: 面白かったー 14 歳なんだなあ… って思いましたw幕間：ダークイルミネイトが見様見真似でチェスをするのを暖かく見守るだけの時間 - 愛とは、全人生をかけてアイドルにチェスを教えること。 < URL > *2 | 面白い んだ 思う ダーク イルミ ネイト 見様見真似 チェス する 暖かい 見守る 時間 人生 かける アイドル チェス 教える こと |

*1 個人情報のためユーザ名を@username と表記している

*2 個人情報のため URL を < URL > と表記している

によって「夜」と「桜」に分けられて、素性に使用されなかったことが要因として考えられる。これより、形態素解析の際に、観光地の特徴となり得る単語を崩さないような工夫をする必要がある。観光地に関する感想を含んでいないが、観光地に関する感想であると判定されたツイートは、F1のような映画を見た感想や、F2のような誰かに対する感想といった、観光地以外のものに関する感想であった。これより、素性に観光地特有のものを用いる必要があると考えられる。また、T2やT3、F3のように素性の数が多いものがよく誤判定されており、誤判定したツイートのうち素性が15個以上のツイートが39ツイート中21ツイートであった。これより、素性に用いる要素をより絞る必要があると考えられる。

6.3 考察のまとめ

観光地名入りツイートを用いた分類器において、観光地名入りツイートと観光地名無しツイートの分類精度の検証を行った結果、以下のことがわかった。

- (1) 観光地名入りツイートに含まれる観光地に関する感想を高精度(正例再現率>0.8)で分類できる。

- (2) 観光地名入りツイートを用いた分類器によって、観光地名無しツイートに含まれる観光地に関する感想も分類できる可能性がある。

今後の課題としては、素性の工夫による分類精度の向上や、自動抽出手法を用いた長期間の観光地に関する感想の収集実験を行うことが挙げられる。

7. おわりに

本稿では、観光地名無しツイートにおける観光地に関する感想の自動抽出手法を提案した。観光地名を含むツイートを用いて、観光地に関する感想を含むツイートを判定する分類器を構築し、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自動抽出を行う。観光地名入りツイートと観光地名無しツイートを用いて、構築した分類器の精度検証を行い、以下のことがわかった。

- (1) 観光地名入りツイートに含まれる観光地に関する感想を高精度(正例再現率>0.8)で分類できる。
- (2) 観光地名入りツイートを用いた分類器によって、観光地名無しツイートに含まれる観光地に関する感想も分類できる可能性がある。

今後の課題としては、素性の工夫による分類精度の向上や、自動抽出手法を用いた長期間の観光地に関する感想の収集実験を行うことが挙げられる。

参考文献

- [1] 国土交通省 観光庁:「旅行・観光産業の経済効果に関する調査研究」(2014年版), <http://www.mlit.go.jp/common/001136064.pdf>.
- [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, http://www.mlit.go.jp/kankochou/siryou/toukei/in_out.html.
- [4] 渡邊小百合, 吉野孝: 観光情報可視化システムのための観光地名無しツイートからの観光情報抽出手法の提案, 情報処理学会研究報告, vol.2017-GN-100, No.47, pp.1-6 (2017).
- [5] 眞野裕也, 青山俊弘: ミニブ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).
- [6] 邱 起仁, 樫山 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).
- [7] 免田哲矢, Kryssanov V.V., 林勇吾, 小川均: Twitter を用いたり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).
- [8] Alan Ritter, Mausam, Oren Etzioni, Sam Clark: Open Domain Event Extraction from Twitter, KDD '12 Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp.1104-1112 (2012).
- [9] 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).
- [10] 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).
- [11] Ryong Lee, Kazutoshi Sumiya: Measuring Geographical Regularities of Crowd Behaviors for Twitter-based Geo-social Event Detection, Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, pp.1-10 (2010).