

SNS 情報を用いた京都市の観光地に関する人気要因分析

01306410 同志社大学 *津田 博史 TSUDA Hiroshi
山田 奏人 YAMADA Kanato

1. はじめに

京都市は国内有数の観光都市であり、英国の有力な旅行雑誌「Wanderlust」にて2年連続でベストシティ部門1位を獲得している。このような背景から京都市への年間観光客数は国内・訪日観光客数を含め、2013年より5年連続で5,000万人を維持している。京都市内の有名観光地である伏見稲荷大社は世界最大の旅行サイト「TripAdvisor」の日本の人気観光スポットランキングで2017年、2018年と2年連続で首位を獲得している。同旅行サイトが2013年末に12,225人を対象に旅行の計画・決定における口コミが与える影響について調査^[1]を行った。回答者の半数以上(53%)の人は、他の旅行者の口コミを読むまでは、ホテルをはじめとする旅行に関わる予定を確定させないこと、8割以上(83%)の人がクチコミを参照することで、旅行の決定に自信を持ち、旅行をより良いものにすることが出来ると回答している。また、近年、SNS (Social Networking Service)が急速に広がっていることから、観光地の人気に対してSNS上の評判が大きく影響を与えると推測される。

そこで、本研究は、京都市産業観光局が毎年行っている京都市内の観光統計^{[2][3]}、TwitterのTweet情報、Google Trends情報を使用し、SNS上の観光地に対する人気の要因に関して考察することを目的とする。

2. 使用データ

本研究では、SNS情報として、京都市産業観光局の観光統計において観光地訪問割合が公開されている観光地名を本文中に含むTwitter上のTweetを使用する。分析対象データとして、2018年11月より2週間以前にTweetを投稿したユーザーの過去のTweetを収集し、その中から観光地名を含むTweetのみ抽出した。

観光地に関するTweetがポジティブな意見を持つものか、ネガティブな意見を持つものか、どちらでもないニュートラルな意見を持つものかの判別を行うために、ポジティブ・ネガティブ・ニュートラル・どれも無い等のラベルが付与されたTwitter日本語評判分析データセット^[4]を教師データとして用いた。データセットのうち、収集出来た230,029件のTweetを使用した。

Twitter日本語評判分析データセットにはTweetの対象に対する評判ラベルが付与されている。それぞれのラベルに対して極性有無と感情スコアを表1のように定める。

表1 評判ラベルと極性有無と感情スコア

評判ラベル	極性有無	感情スコア
対象に対してポジティブ・ネガティブな意見を持つ	有	ニュートラル
対象に対してポジティブな意見を持つ	有	ポジティブ
対象に対してネガティブな意見を持つ	有	ネガティブ
対象に対してニュートラルな意見を持つ	有	ニュートラル
対象に対して無関係なTweetである	無	

観光地のウェブ上での人気度を表す指標として、Google社のGoogle Trendsを使用した。Google Trendsは同社の検索エンジンで使用されたキーワード検索頻度を表示するサービスである。観光地毎に2013年12月22日から2018年12月22日の261週分のGoogle Trendsデータを用いた。

3. 分析方法

3.1 分析のための前処理

収集したTweetを用いて、分析を行うために前処理を行った。Tweet本文にはウェブページのURLやリツイート情報、会話相手の情報など本文に無関係な文字が多く含まれている。Tweet本文から、無関係な情報を除去し、英数字の半角統一、カタカナの全角統一、絵文字の除去、記号の除去を行った。Tweet本文の単語頻度表の作成を行うために、形態素解析を行い単語の原形を取得した。形態素解析とは自然言語処理技術の一つで、自然言語(文章)を形態素(言葉が意味を持つ最小単位)に分割を行うものである。単語の原形を使用することで、Twitter上の文字表記のゆらぎによる単語数増加を防ぐ。本研究では、形態素解析を行うツールとして、MeCab^[5]を使用している。MeCabの辞書には固有名詞を多く登録しているmecab-ipadic-NEologd^[6]を採用した。形態素解析時に、全形態素で構成される形態素リストと、名詞・動詞・形容詞のみ抽出した形態素リストの2つのリストを作成した。

3.2 分析手順

採用した分析方法は、大きく分けて2つである。

分析方法1

Tweet 情報を定量化するために、2つの形態素リストから、単語頻度ベクトルを作成した。定量化された文章データと Google Trends の関係性を分析するにあたり、文章データを主成分分析することで、単語頻度表の次元を削減し、重回帰分析を行った (Principal Component Regression: PCR 分析)。主成分分析では、主成分の累積寄与率が 90%になるまでの主成分を用いた。

分析方法2

Tweet 情報の感情の有無を調べるモデル(以下、極性判定モデル)、感情がポジティブかネガティブかを推定するモデル(以下、感情スコアモデル)の2モデルを構築し、観光地に関する Tweet の感情とウェブ上の観光地人気度の関係を分析した。モデルの構築には自然言語処理技術の word2vec と RNN (Recurrent Neural Network: 再帰ニューラルネットワーク)の1種の LSTM (Long short-term memory) を使用した。

4. 分析結果

4.1 PCR 分析結果

2013年12月から2018年12月までのSNS上の観光地に関する反応と、ウェブ上の人気度を表す Google Trends の検索ランキングとの関係について PCR 分析を行った。

表2 PCR 分析の決定係数上位5観光地

観光地名	R ²	修正済みR ²
高台寺	0.811	0.597
伏見	0.737	0.588
京都市美術館	0.743	0.558
京都国立博物館	0.784	0.536
大原	0.689	0.513

表2に示すとおり上位5観光地の修正済み決定係数は全て 0.5 を上回っており、分析を行った観光地の修正済み決定係数の平均は 0.146 であった。この結果から、Google Trends の検索ランキングに対し、説明力のあるモデルが構築できたとと言える。

4.2 感情分析結果

PCR 分析と同様に Google Trends の検索ランキングに対して、名詞・動詞・形容詞の形態素リストを使用した極性判定モデルと感情スコアモデルのそれぞれの数値を説明変数にし、重回帰分析を行った。自由度修正済みR²を降順で並べた際の上位10観光地の決定係数をまとめた表3に示す。感情数値との重回帰分析の結果は、PCR 分析の結果と比べ、自由度修正済み決定係数の値が低い。自由度修正済み決定係数の値が低く

なった要因として、観光地に関する各時点の極性判定スコア、観光感情スコアの平均値を説明変数として使用した事が考えられる。

表3 感情分析決定係数上位10観光地

観光地名	R ²	修正済みR ²
西陣織会館	0.261	0.190
八坂神社	0.146	0.134
高台寺	0.131	0.115
京都駅周辺	0.134	0.098
京都タワー	0.104	0.090
東福寺	0.102	0.089
東本願寺	0.093	0.084
西本願寺	0.086	0.066
知恩院	0.074	0.053
大徳寺	0.066	0.047

5. おわりに

本研究において、観光地によって影響を与える単語が異なることを確認できた。季節に関する単語など一時的に検索数が上昇するような単語、観光地に関する単語が観光地の人気度に大きな影響を与えている。また、観光地に対する Tweet の感情がウェブ上の人気に関係していることがわかった。すなわち、極性がある Tweet がウェブ上の観光地人気度に大きな影響を与えており、感情スコアがネガティブであるほど観光地の検索回数の増大に影響を与えることなど新しい知見を得ることができた。

参考文献

- [1] TripAdvisor, “トリップアドバイザー, クチコミ利用の動向に関するユーザー12,000名以上の調査を発表”, プレスリリース, 2014.
- [2] 京都市産業観光局, “京都市観光調査年報”, 2003-2010.
- [3] 京都市産業観光局, “京都観光総合調査”, 2011-2017.
- [4] 芥子郁雄, 鈴木優, 吉野幸一郎, グラムニュービッツ, 大原一人, 向井理朗, 中村哲, “単語意味ベクトル辞書を用いた Twitter からの日本語評判情報抽出”, 電子情報通信学会論文誌, Vol. J100-D, No. 4, pp-530-543, 2017. 4.
- [5] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237 (2004.)
- [6] 佐藤敏紀, 橋本泰一, 奥村学, “単語分ち書き辞書 mecab-ipadic-NEologd の実装と情報検索における効果的な使用方法の検討”, 言語処理学会 第23回年次大会 発表論文集, pp.875-878, 2017.