

テキストマイニング手法を用いた観光評価と観光行動の関係性分析 ～北陸新幹線沿線観光地を対象として～

長野工業高等専門学校 学生会員 ○金森 真梨
長野工業高等専門学校 正会員 柳沢 吉保

長野工業高等専門学校 正会員 轟 直希

1. はじめに

近年、地域振興として全国各地で観光に対する期待が高まり、地域経済に及ぼす影響が見込まれている。一方で日本人の国内旅行客数に大きな変化は見られておらず、その原因として地域の観光資源の特徴と観光客のニーズの理解の不足が挙げられる。そこで、現代の生活の一部となっている ICT を活用し、観光需要を的確にとらえることで地域の活性化や観光誘客の推進につながると考えられる。

以上を踏まえ、内閣府が運用している RESAS（産業構造や人口動態、人の流れなどを可視化したビッグデータ）に注目する。RESAS などの ICT と観光行動に関する研究として小林¹⁾らは、情報認知により観光意欲が移動負担を上回る可能性を示したが、どのような情報が観光に影響を与えるかは明確になっていない。本研究では、web 上に投稿された観光地の口コミデータの活用を検討する。口コミは観光客の生の声であり、それらを集約することでより具体的な観光地への意見を得ることができる。そして得られた意見を分析し、定量化することで新たな観光地評価を取得することを目的とする。

2. 調査方法

(1) 調査対象エリア

2015 年 3 月 14 日の北陸新幹線延伸(長野-金沢間)効果が予想される長野県、新潟県、富山県、石川県を対象エリアとする。そのうち新幹線停車駅のある 12 沿線市町村の中から RESAS の目的地分析から得た各県上位の観光地の口コミの集計を行う。

(2) 口コミサイトの有効性検証

使用する口コミサイトとして、じゃらん net、TripAdvisor、4traberl.jp の 3 サイトを挙げ口コミ数と入込客数との相関を求めると有効性の検証を行った。対象期間を 2014 年から 2019 年の長野県の北陸新幹線沿線市町村にある観光地から RESAS の目

表-1 口コミ数と入り込み客数の相関

サイト名	じゃらん net	TripAdvisor	4traberl.jp
相関係数	0.7205	0.7079	0.6445

的地分析ランキングを用い、順位換算法で上位 3 位を決定した。善光寺、戸隠神社、上田城跡の順で結果が得られたが、善光寺は 2015 年の御開帳による変動、戸隠神社は立地場所が二次交通の利用が必須であり、新幹線延伸効果が十分に分析できない可能性があることから除外した。従って、長野県上田市に所在する上田城跡の口コミサイトの有効性の検証を行う。

前述の 3 サイトにおいて月別口コミ件数を取得し、月別入り込み客数との相関を求めた結果を表-1 に示す。表-1 より、じゃらん.net で最も大きい相関を得ることができたため、本研究では口コミサイトはじゃらん.net を用いて分析を行う。

(3) 口コミ観光評価の抽出手法

口コミサイトからのデータ抽出のために無料 web スクレイピングツールの Octoparse を使用した。スクレイピングとは、情報収集の方法の一つであり、スクレイピングツールとはそれらを行うサービスのことである。Octoparse では、取得先の URL を入力することでデータが自動的に検出される。取得したデータは、Excel や HTML、CSV のような構造化フォーマットで指定のデータベースに保存することができる。コードを書くことなくインターネット上にある無限の情報の中から必要なデータを必要な形で入手できるという利点を持っている。

3. 口コミ観光評価の整理

収集したデータの分析を行うために、分析ソフトウェア KH Coder を使用した。これはテキスト型データを統計的に分析するためのソフトウェアであり、アンケートの自由記述他、様々な社会調査データを分析するためのものである。今回は抽出後リストを使用し、口コミ内の多頻出語句の抽出を行う。

4. 口コミ観光評価と観光行動の関係性分析

(1) 多頻出語句の抽出

今回は、長野県の北陸新幹線沿線観光地である善光寺を対象とし、先述した抽出語リストを用いて訪問時期が2014~2019年の口コミにおける語句の抽出を行った。口コミ内容の傾向として名詞や固有名詞は多頻出しやすく、書き手によって用いる動詞は差が生じやすいことが挙げられる。そこで今回は抽出語句を物の性質や状態を表す言葉である形容詞、形容動詞に限定し、合わせた上位30の語句を抽出した。抽出語の出現回数上位20位までの単語を、形容詞は灰色の枠、形容動詞は白色の枠で表-2に示す。

抽出された語句の形容詞、形容動詞の数の偏りはみられず、「良い」「楽しい」といった比較的好印象な語句が多くあることがわかった。

(2) 観光地評価

抽出リストより取得した語句から、観光地の評価を得る。分析にはクラスター分析を用い結果として、抽出した30の語句は5つのクラスターに分別された。次に、再び抽出語リストを用い、今度はクラスターごとの多頻出語句を取得し、最後に観光客自らが設定可能な観光地評価である星の数の平均値をクラスターごとで取得し、口コミからの観光地評価を得た。結果を表-3に示す。

抽出された語句を基にそれぞれのクラスターに5つの印象を設けた。観光地に訪れる前と後で異なる意見が抽出された「訪問前後の印象」、その観光地のできる「活動についての印象」、観光地で受けた店員や施設の整備などから受ける観光地からの「対応についての印象」、観光地の立地場所や充実度などの「施設に対する印象」、実際に訪問して「施設での体験からの印象」である。サンプル数が1である「対応についての印象」分析対象から除外し、結果として「活動についての印象」「施設に対する印象」「施設から受けた印象」では、点数が4を超える高評価を得ることができた。これは、特に「施設での体験からの印象」で抽出されている語句において観光地の様子を想像することが容易であることが考えられる。

一方、「訪問前後の印象」について5つのクラスターの中で最も低い評価となった。これは、抽出語句から「大変」という語句が確認でき、善光寺に訪れた

表-2 出現回数

順位	抽出語	出現回数	順位	抽出語	出現回数
1	多い	217	11	素晴らしい	63
2	良い	201	12	少ない	54
3	楽しい	170	13	綺麗	51
4	真っ暗	136	14	荘厳	43
5	有名	112	15	早い	42
6	大きい	106	16	非常	41
7	立派	95	17	すてき	40
8	美味しい	89	18	寒い	39
9	大変	72	19	きれい	37
9	広い	72	20	怖い	36

表-3 口コミから取得した観光地評価

クラスターの印象	サンプル数	星の数の平均(点)	上位抽出語句
訪問前後比較	20	3.8	大変/有名/楽しい
活動	152	4.3	多い/楽しい/真っ暗
対応	1	5.0	綺麗/親切/無い
施設全般	13	4.3	多い/綺麗/悪い
施設での体験	22	4.5	立派/良い/楽しい

際に観光客が何かしらの困難を感じている可能性があると考えられる。加えて、他のクラスターでは確認できないネガティブな印象を与える語句が多頻出していることで評価が低下していると考えられる。

5. まとめ

本研究では、口コミと観光地評価の関係性を検証した。観光地の雰囲気や様子が想像しやすい語句があれば評価が高くなる一方、ネガティブな印象を与える語句があれば評価が低くなる傾向がみられた。今後、これらのクラスターに性別や年齢などといった個人属性を付加し、入り込み客数との関係性を分析することで観光行動の取得も期待できる。

参考文献

- 1) 小林碩、轟直希：ICTによる情報入手プロセスを考慮した観光意思決定モデルの検証、長野高専専攻科修士論文、2019
- 2) 松本義之：SNSから収集した近接地域のテキストマイニングによる分析、バイオメトリック・フェイ・システム学会誌、Vol8、No2、pp41-48、2016
- 3) 野守耕爾：クチコミビッグデータに人工知能を応用した地域観光の次世代マーケティング、人工知能学会全国大会論文集、2016