

## 研究ノート

# テキストマイニングによる観光クチコミ情報の分析

## —山形市を対象として

西 平 直 史

### 1. はじめに

近年、日本において観光業が急速に成長しつつあり、2008年には観光庁が設置された。また、観光立国推進基本計画が策定され、旅行消費額、訪日外国人旅行者数、観光地域の旅行満足度などの目標が設定されている。それに呼応して、各地で観光振興のための種々の取り組みがなされている。

観光業の成長に伴い、観光に関する研究も盛んにおこなわれている。観光に関する基礎的な研究としては、アンケート調査による方法があげられるが、ほとんどが選択肢から選ぶ方式である。アンケートによっては、自由記述回答欄を設けているものもあるが、自由記述データは恣意的に扱われることが多く、データを定量的に分析することは困難であった。最近では「テキストマイニング」と呼ばれる自由記述データを定量的に扱うことができる分析手法が用いられている。<sup>1), 2)</sup>

山形市に注目すると「山形市観光基本計画」が策定されており、平成39年度までに観光消費額を330億円／年、観光客入込数を3,523千人／年にするといった目標が掲げられており、観光振興のための取り組みが推進されている。しかしながら、山形市を対象とした観光の基礎的な研究はほとんどないのが実情である。本稿では、観光クチコミサイトに掲載されている山形市のクチコミ情報を対象としてテキストマイニングによる分析を行い、基礎的な資料を作成することを目的とする。

### 2. データの紹介と分析手法

ここでは、旅行クチコミサイト 4 travel.jp を対象としてクチコミデータを収集した。クチコミデータは、個人が投稿するものであり、観光客の旅行動向や感想が記載されている。実際に集めたデータは、同サイトの「国内旅行ガイド」の「旅行記」と「クチコミ」で山形市として分類されているものであり、2016年3月までに同サイトに登録されたすべてである。「旅行記」とは、同サイトのガイドラインによると、「トラベラー会員が旅先で撮った写真をアップロードして、自分だけのオリジナル旅行記を作ることができるサービスです」となっている。また、クチコミはガイドラインによると、「実際に行った、見た、体験したことがわかるように、また、他のユー

ザーの参考になるように、クチコミの内容は具体的に記述してください」と記されている。このように、「旅行記」を調べることで山形市を含んだ旅行行程のすべてを把握できるし、「クチコミ」は山形市内の特定の観光地や施設の評判を調べることができる<sup>1</sup>。

集めたデータは1つのテキストファイルに保存し、テキストマイニングと呼ばれる手法で分析した。具体的には、kncoder<sup>3)</sup>を用いて抽出語リストの作成ならびに共起ネットワークを作成した。

抽出語リストとは、すべてのテキストデータの中にそれぞれの単語が出現する頻度回数を表したリストである。活用のある語は基本形に直して抽出されるため別の単語としては抽出されない。一方で、複合語に関しては「茶筌」と呼ばれるソフトウェアで複合語の検出を行っているが、同じ単語であっても複合語として1つの単語として抽出される場合と別々の単語として抽出される場合があるため注意が必要になる<sup>2</sup>。

共起ネットワークとは、同じ文章によく一緒に出現する（共起する）抽出語同士を線で結んだネットワークのことである。ネットワークを図示することで特徴語の共起関係を視覚的にとらえることができる。

### 3. 分析結果と考察

4 travel.jp の山形市の「旅行記」および「クチコミ」のデータに対して、kncoder を用いて抽出語リストを作成したものの一部を図1に示した<sup>3</sup>。図のように品詞ごとに抽出語の出現回数がカウントされている。一番左の「タグ」という列に注目してみよう。これは、kncoder が自動で分類する品詞ではなく、ソフトウェアの利用者が注目したい単語や複合語を設定してカウントするものである。例えば、地名は文脈によっては人名としてカウントされることがあったためタグに設定している。また、霞城セントラルは「茶筌」では複合語として分類されなかったため、タグに設定して一つの複合語として頻度をカウントしている。

さて、図1を見ると、山形市の旅行記ならびにクチコミであるため、「山形」が最も多いのは当然である。地名に注目すると「山寺」が1907回で圧倒的に多い。また、「立石寺」(682回)、「五大堂」(356回)、「根本中堂」(264回)と山寺に関連した抽出語も多い。山形市において山寺が観光地として人気があることがよく表れている。他の観光地としては、「霞城公園」(357回)、「蔵王」(351回)、「文翔館」(216回)、「霞城セントラル」(136回)が続いている。

つぎに図2に作成した共起ネットワークを示した。図で四角で囲まれているのはタグとして登

1 4 travel.jp のガイドラインでは両者が区別されているが、実際の投稿内容を見ると「クチコミ」の方に旅行記に分類した方がふさわしいと思われる内容が記載されている場合もある。そのため、「旅行記」と「クチコミ」は分けずにまとめて分析を行った。

2 例えば、「霞城公園」は「霞城公園」という複合語として抽出される場合と「霞城」と「公園」が別の単語として抽出される場合があった。

3 実際の抽出語リストとは列の順番を入れ替えている。これは、「タグ」列を見やすくするためである。



録した抽出語、丸で囲まれているのはそれ以外の単語である。頻度が大きいほど図形の面積が大きい。また、図形の色は「媒介中心性」（それぞれの語がネットワークの中でどの程度中心的な役割を果たしているかを表すもの）の程度を示している。白い図形の単語は媒介中心性が最も低く、図形の色が濃くなるほど媒介中心性が高い。「山形」、「駅」、「旅行」といった単語が中心的な語となっている様子を見て取れる。図示した共起ネットワークを見ると、他とは独立しているネットワークと他のネットワークと複雑につながっているネットワークがある。

まず、独立しているネットワークに注目してみよう。「面白山」は「面白山高原」と「紅葉」でつながっており、紅葉を楽しむ観光客が多いことがわかる。「霞城セントラル」と「五大堂」はどちらも「眺め」が良い場所で「展望」を楽しむことができる。「米沢」は「牛」、「食べる」と共起関係にあることから、米沢牛を店で食べる旅行者が多いと言える。

つぎに、複数のネットワークが複雑につながっているネットワークに注目してみよう。「東京」からは「新幹線」でネットワークがつながっている。「仙台」からは「仙山線」でつながっている。両地域からの観光客は鉄道利用者が多いことが分かる。一方で、「東北」は「道」と共起関係にあることから、「東北道」を利用して山形に来る観光客もいることがわかる。観光地としては「山寺」、「立石寺」、「根本中堂」が一つのグループになっている。また、蔵王温泉と天童温泉が共起関係にあることから、宿泊地としてはこの二つの温泉地が人気であると言えよう。

#### 4. おわりに

本稿では、観光クチコミサイトに投稿されている山形市のクチコミ情報をテキストマイニング手法を用いて分析し、山形市の観光に関する基礎的な資料を作成した。山形市の観光地としては、山寺の人気の高いと言える。

共起ネットワークを用いることで抽出語の共起関係を図示し分析することができる。しかし、本稿では「抽出語」数を優先したため、共起関係が統計的にやや弱いものも含まれている<sup>4</sup>。今後の課題としては、共起関係の統計的な「強弱」を分析することや広域観光を視野に入れた分析を行うことがあげられる。

#### 参考文献

- 1) 森田・入澤・長塩・野村・塚田・大塚・杉田, 自由記述データを用いたテキストマイニングによる都市のイメージ分析, 土木学会論文集 D 3 (土木計画学), Vol.68, No. 5, pp. I\_315-I\_323, 2012

4 khcorder の共起ネットワークではノード数を優先してネットワークを作成するか、Jaccard 係数を優先してネットワークを作成するかを選択できる。本稿では、ノード数を優先 (120に設定) した。

- 2) 西尾・塚田・森田・湯沢, テキストマイニングによる富岡製糸場の世界遺産登録前における観光まちづくりの把握, ランドスケープ研究, 79 (5), pp. 519-524, 2016
- 3) 樋口耕一, 社会調査のための計量テキスト分析—内容分析の継承と発展を目指して, ナカニシヤ出版, 2014

## **An Analysis of Tourism Information Using Text Mining Methods — A case of Yamagata city**

Naofumi NISHIHIRA

In this paper, an analysis of tourism information using text mining methods is proposed. The collected data for Yamagata city was obtained from a tourist information site. As result of analysis, it was shown that Yamadera is popular among tourists. Furthermore, the data also revealed that Kajo-koen, Zao, Bunshokan and Kajo-central are popular tourist spots.