

# 旅行サイトの口コミデータに基づき、 観光者の出身国ごとに対応を調節するための支援システム

木田 唯斗, 高山 毅

尾道市立大学経済情報学部

## 1. はじめに

近年、訪日外国人旅行者が増加している。彼らの口コミデータを分析すると、出身国により傾向に差異が見られることが指摘されている[1]。しかし、そのことをシステムにどう生かすべきかの議論は、充分とは言えない。

本稿では、旅行サイトの口コミデータに基づき、観光者の出身国ごとに対応を調節するための支援システムを提案する。本システムは、以下の3機能を具備する：i) 口コミを出身国別でテキストマイニングした結果を相互比較できる機能、ii) 出身国別で名詞のみ、形容詞のみではなく、様々な品詞で「観光スポットの良し悪しを表示」したり、「観光スポットを検索」できる機能、iii) 対話形式で観光者に出身国を考慮してお薦めの観光スポットを提示する機能である。

## 2. 先行研究

文献[1]では、観光に対する趣味趣向に国民性や民族性による違いがあるのかを、明らかにする方法が提案されている。そこでは、口コミデータを投稿者の居住国別で分析し、*tf-idf* 値が高い名詞や、ある国にのみ出現する頻度の高い名詞を抽出している。

文献[2]では、訪日外国人旅行者と、国内旅行者が持つ印象の、共通点や相違点の把握を試みている。そこでは、旅行者のネット上の口コミをコレスポネンス分析し、視覚化している。

文献[3]では、口コミに含まれる感情語を利用して観光地分類を行うことを検討している。そこでは、ユーザからの口コミデータについて、「楽しい」や「きれい」といった感情や、関連性の高いものとの結びつきを示す特徴語に着目している。

## 3. 旅行サイトの口コミデータに基づき、観光者の出身国ごとに対応を調節するための支援システム

### 3.1 実現方式

今回は、環境に依存せず利用できるように、Webアプリケーションで開発する。

A support system to adjust correspondence depending upon each national origin of tourist based on online reviews of travel site  
Yuito Kida, Tsuyoshi Takayama  
Faculty of Economics, Management, and Information Science,  
Onomichi City University

### 3.1.1 口コミデータの収集

本稿では、TripAdvisor[4]の口コミデータを収集し、分析する。TripAdvisorは、口コミ投稿者の居住地情報があり、7億件以上の口コミ情報が寄せられている世界最大級の旅行サイトである。本稿で収集する言語を、日本語、英語、中国語、イタリア語、スペイン語の5言語とする。収集する観光スポットの口コミ期間は2018年1月1日から8月31日までとし、観光スポットごとに50件の口コミを収集する。

### 3.1.2 分析手順

本稿では、前処理や語の抽出及び集計といったテキストマイニングに、KH Coder[5]を利用する。手始めに日本人にわかるように開発するので、出力されたデータを日本語へ翻訳する。ここでは、シソーラスの一種である日本語WordNet[6]や、Chinese Open Wordnet[7]を利用する。また、多言語オンライン辞書としてGlosbe[8]等を利用する。

### 3.2 試作システム

本稿では、日本の観光スポット20箇所を用いる。操作手順は、以下のとおりである。

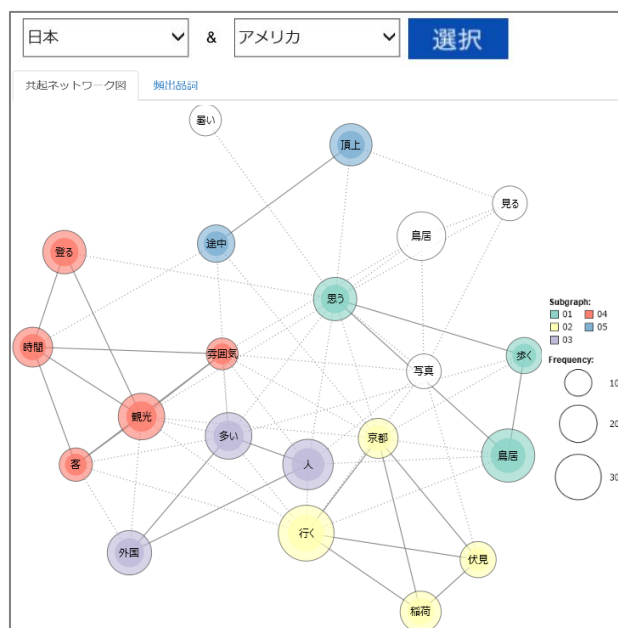


図1 指定した国での頻出語とその共起状況を示す共起ネットワーク。

まず、観光スポットを選択すると、頻出語とその共起状況を、口コミ者の出身国ごとで確認できる(図 1)。ここで頻出語ランキングは、様々な品詞で可能としている(図 2)。対話形式で希望を入力すると、観光者の出身国を踏まえてお薦めスポットが表示される(図 3)。



図 2 指定した品詞での頻出語を、口コミ者の出身国ごとに表示している画面。



図 3 出身国を踏まえてお薦めのスポットを表示した画面。

#### 4. 評価実験

以下の三つのシステムを相対評価する:

- システム i : 提案システム
- システム ii : 5 カ国のデータを区別せずにテキストマイニングしたもの
- システム iii : KH Coder の国別出力データ

被験者は 20 代と 50 代の男女計 22 名とする。

##### 4.1 方法

被験者には、以下の三つの課題に取り組むよう依頼する:

- ◆ 課題 1. 観光スポットの口コミから、国別で観光スポットの感じ方の顕著な品詞を答えさせる
- ◆ 課題 2. 観光スポットの口コミから頻出

している名詞、形容詞を国別で書かせる

- ◆ 課題 3. ある国で、指定した言葉に当てはまる観光スポットを書かせる

また、各課題終了後に、被験者に以下の三項目の五段階定性的評価を依頼する。すなわち、「満足度(使いやすさ)の評価」、「各種データの比較のしやすさ」、「同一スポットに対する各国の違いが論理的に理解できたか」の三つである。

#### 4.2 結果

表 1 課題ごとの平均正解率(%)

システム/課題	1	2	3
i (提案システム)	100.0	100.0	86.3
ii (5 カ国のデータを区別せずにテキストマイニングしたもの)	70.0	10.4	84.6
iii (KH Coder の国別出力データ)	67.9	78.3	45.2

提案システムは課題 1~3 の全てにおいて、平均正解率が最良の結果を得ることができた(表 1)。また、定性的評価でも、紙幅の都合により詳細なデータは省くが、提案システムは 3 尺度でいずれも、最良の評価を得ることができた。

#### 5. 結論と今後の展望

本稿では、旅行サイトの口コミデータに基づき、観光者の出身国ごとに対応を調整するための支援システムを提案した。試作システムによる評価実験の結果、提案システムは有効との結果が得られた。今後の展望として、i) 複数の品詞を組み合わせて観光スポットを検索できる機能、ii) 比較の際、同じ語句の色付けができる機能および、iii) データ量の拡充、等が挙げられる。

#### 参考文献

- [1] 宋 爽, 斎藤 一, 川村 秀憲: 旅行サイトの口コミデータから探る国別の観光ニーズ, 観光情報学会第 15 回研究発表会, pp. 37-40, 2017.
- [2] 大川 慎ほか: テキストマイニングを利用した観光地イメージの比較に関する試み, 観光情報学会第 16 回研究発表会, pp. 74-77, 2017.
- [3] 杉本 祐介, 水野 忠則ほか: 口コミに含まれる感情語を利用した観光地分類の検討, 「マルチメディア、分散、協調とモバイル(DICOMO2014) シンポジウム」, pp. 1345-1350, 2014.
- [4] TripAdvisor: <https://www.tripadvisor.jp/>
- [5] KH Coder: <http://kncoder.net/>
- [6] 日本語 WordNet: <http://compling.hss.ntu.edu.sg/wnja/>
- [7] Chinese Open WordNet: <http://compling.hss.ntu.edu.sg/cow/>
- [8] Glosbe: <https://ja.glosbe.com/it/ja>