

アクセシブルツーリズムにおける着地型観光情報の有効性
～Q&A サイトのテキストデータからの重要な観光情報の因子抽出～

村山朗† 辻靖彦† 山田恒夫†

†放送大学大学院文化科学研究科情報学プログラム

1. 背景と目的

アクセシブルツーリズム(以下 AT)において、旅行者ごとに異なるバリアに対して最適な観光情報を提供することはハード面の整備と同等に重要な施策である[1]. 森山ら(2001)[2]は移動困難などを伴う障害者がどのような観光情報を求めているかをアンケート調査によって明らかにした. また, 観光庁の調査(2023)[3]では, 障害者が旅行に行きたくなる条件として, バリアフリー関連の観光情報の整備などがあると報告された. しかし, 観光困難者にとって重要な観光情報の詳細は明らかにされていない.

一方, 適切な観光情報をネット上のテキストデータから抽出する試みは, 観光情報推薦システムの開発などにおいて数多く行われており, また対応分析やトピックモデルなどテキストマイニングの手法を用いて分析し旅行者が満足する要因を明らかにした研究などもある(例えば北中ら, 2018[4]). しかしこれらの研究で AT を取り上げた事例は少ない. そこで本研究では Q&A サイトのテキストデータを分析し, どのような情報が観光困難者にとって重要であるかを抽出する試みを行った.

2. 分析の方法

分析の目的は, 観光困難者にとって重要な AT 観光情報の因子を Q&A サイトのテキストデータからトピックモデルを用いて抽出し, AT 観光情報の提供に役立てることである. 分析対象とするデータは Yahoo!知恵袋の書き込みデータで, 国立情報学研究所の許諾を受けて入手した. 2018年4月1日～2021年3月31日の間に書き込まれた Yahoo!知恵袋サービスの質問文から10%の無作為抽出をしたもので, 回答文はそれらの質問文に紐づくデータである. 分析のステップとしては以下で行った.

1. 一般旅行者データの準備と分析
2. 観光困難者・質問+回答データの準備と分析
3. 観光困難者・回答データの準備と分析

3. 分析の結果

3-1. 一般旅行者データの準備と分析

入手した Yahoo!知恵袋の全データから「地域, 旅行, お出かけ>国内」のカテゴリーに属するデータを抽出し, 質問(データ数 32,762)に各質問に対する回答(データ数 83,441)をすべてマージして一般旅行者データとした. このデータの傾向を頻出語・共起ネットワーク・対応分析等によって把握したところ, 以下の特徴が見られた.

- 観光旅行全般に関わる単語が頻出している
- ホテル, 食事, テーマパーク, 観光地のおすすめなど, 観光旅行全般についてのアドバイスを求める質問とそれに対する回答が多い

続いてこのデータから, トピックモデルによって AT 関連のトピックを分類・抽出する試みを行った. データを TF-IDF に加工しモデルは LDA を用いたが, perplexity が低い≒分類予測性能の高いモデルは構築できず, 前処理や品詞選択などに検討の余地を残した. またトピック数が90以上で AT 関連のトピックが分類・抽出されたが, これは Yahoo!知恵袋において AT 関連の質問と回答が少ないことに起因すると考えられる.

3-2. 観光困難者・質問+回答データの準備と分析

次に, 一般旅行者データから AT と関連するキーワードが含まれているテキストを抽出し, 観光困難者のデータセットとした. AT 関連キーワードは障害・車椅子・バリアフリー, およびこれらの単語と Word2Vec を用いて類似度が高いと推定された単語群(計 23 単語)として, 311(全体の1%以下)のテキストを抽出した. その傾向を3-1と同じ方法で分析し, 以下の特徴を見出した.

- 一般旅行者と共通する単語に加えて, AT における利便性に関わる単語が頻出している
- 一般旅行者と共通する質問と回答に加えて, ディズニーランドやアトラクションの利用, 障害者手帳等についての質問と回答が見られる
- 公園, レンタカー, 市内, 電車, 新幹線, 乗り物, レストラン, アトラクションなど, 観光旅行の要素に困難さを感じている傾向が見られる

このデータから観光困難者にとって重要な AT 観光情報の因子を抽出する試みを, 長谷川ら(2018)[4]を参考に, NMF(非負値行列因子分解)と BM25^{注1)}による手法を用いて行った. ここで文書集合は質問と回答をマージしたデータであり,

Validation of the effectiveness of landed tourist information in accessible tourism - Factor extraction of important tourism information from text data of Q&A site

† Informatics Program, Graduate School of Cultural and Social Sciences, The Open University of Japan

各文書のラベルは質問の先頭 10 文字とした。また、再構成誤差のプロットで明らかなエルボーは見られなかったため、解釈性の高いトピック数 9 を採用した。その結果、重要な AT 観光情報因子(図 1)および因子と関連のある単語群(図 2)として以下が抽出された注 2)。

Topic 2: 障害者手帳によるチケットの割引情報 → ディズニーリゾート, USJ, 太秦映画村など

Topic 6: ホテルの部屋予約情報 → バリアフリー, 車椅子, 東京駅周辺, お台場など

Topic 7: アトラクションのサービス, 待ち時間や混み具合に関する情報 → ディズニーランド, USJ, 車椅子, 高齢者

Topic 8: ディズニーランドで車椅子を借りれるかなどの情報 → ディズニーランド

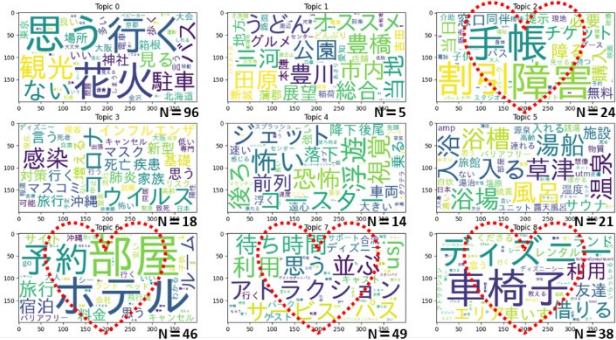


図 1-NMF により抽出された因子(トピック)



図 2-各因子(トピック)と関連のある単語群

3-3. 観光困難者・回答データの準備と分析

続いて質問項目に対して、目視で AT と直接関連する 46 件に絞り込み、その回答である 112 のテキストデータを用いて因子情報の抽出を行った。解釈性の高いトピック数 6 を採用し、図 3 の因子の解釈を行った。並行してどのトピックが重要であるかを判断するため、トピックごとのベストアンサー率を算出したところ、N 数が十分(=19)でベストアンサー率が高いのは Topic1 の登山観光(52.6%)であった。そこで Topic1 のテキストにキーワードを付与してラベルとし、因子とラベルの関連を NMF で見たところ、登山

注 1) ただし分析の結果 BM25 に負の値が含まれていたため、単語の出現頻度の重みづけには TF-IDF を用いている。
注 2) 各図中のハートマークが AT 関連のトピックとその関連単語を示す。

観光は車椅子で行けるか、駐車場はあるかなどが重要な情報であると解釈された(図 4)。



図 3-観光困難者・回答データの因子解釈

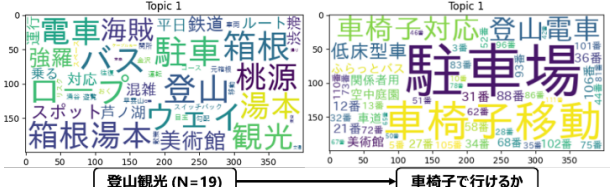


図 4-登山観光因子と関連のある単語群

4. 結果のまとめと今後の課題

Yahoo!知恵袋の「地域, 旅行, お出かけ>国内」カテゴリにおいては、観光旅行全般についてのアドバイスを求める質問と回答が多数を占め、AT に関連した質問は少なかった。AT 関連単語で抽出した質問および回答は、アトラクションの障害者割引, 車椅子の利便性など、AT に関連したものが多数見られた。特に観光困難者は移動・アトラクション・飲食店などで困難さを感じており、重要な観光情報因子としては、テーマパークにおける割引や車椅子利用の情報, ホテルの部屋のバリアフリー情報などが抽出された。また、登山観光は車椅子で行けるか、駐車場はあるか、なども重要な情報であることが明らかになった。

今回、分析対象としたデータは AT 関連の質問の比率が少なく、また、データ抽出の際に用いた関連用語の選択次第で観光困難者にとって重要と思われる共通因子が左右され、結果が安定しなかった。この問題を解消するためには、AT 関連の書き込みが豊富であるテキストを対象とすることに加え、トピックモデルや LLM など関連用語によるデータ抽出以外の方法を検討する必要があると考えられる。

謝辞

Yahoo!知恵袋のデータは国立情報学研究所と LINEヤフー株式会社よりご提供いただきました。

参考文献

- ユニバーサルツーリズムに対応した観光案内の実践方策, 国土交通省 (2017)
- 障害者用設備の情報提供が身体障害者の観光地訪問意向に及ぼす影響, 森山昌幸, 藤原章正, 杉志頼幸, 木佐幸佳, 土木計画学研究, Vol.18 pp.957-962 (2001)
- ユニバーサルツーリズムに関する調査業務報告書, 観光庁 (2023)
- ホテル利用客の満足度に対するテキキストマイニングによる分析, 北中英明, 香坂千佳子, 伊藤宗彦, 経営情報学会全国大会 (2018)
- 食における言語資源に基づいたシズル感に関する因子情報の抽出, 長谷川永奈, 小宮香乃, 齊藤史哲, 石津昌平, 日本感性工学会論文, 17-2, pp.299-308 (2018)