

類似観光スポットの比較による 魅力発見システムの構築

佐藤 菜摘¹, 難波 英嗣¹, 石野 亜耶², 竹澤 寿幸¹

¹広島市立大学情報科学部

²広島経済大学経済学部

¹{sato, nanba, takezawa}@ls.info.hiroshima-cu.ac.jp

²ay-ishino@hue.ac.jp

概要：本研究では、旅行ブログなどの大量の観光関連文書を、自然言語処理技術を用いて解析し、体験型スポットを自動検出する。体験型スポットとは、観光客が能動的に行動する必要がある観光スポットのことで、例えば、しまなみ海道サイクリングやオタフクソース株式会社のお好み焼き体験教室などが該当する。こうした体験型観光スポットの情報を訪日外国人観光客に提供することで従来とは異なる広島の魅力のアピールすることができる。また、体験型スポットは比較的長時間を要するため、広島の滞在日数の増加、ひいては経済効果にも一定の影響が期待できる。さらに、ある体験に対する類似体験型スポットについても分析することで、さらなる魅力発見を試みる。

Keywords：体験型スポット、旅行ブログ、訪日外国人

1. はじめに

広島は世界的に見ても知名度の高い観光都市である。広島の代表的な観光名所・施設は、「平和記念資料館」「宮島」「広島城」などであり、実際に一日で見てまわることが可能である。そのため、広島市に訪れる外国人観光客は日帰りであることが多く、それほど高い経済効果があるわけではない。広島により長期間滞在してもらい、広島魅力を伝えるには、上記以外の観光名所・施設以外をアピールする必要がある。

そこで、本研究では体験型スポットに着目する。体験型スポットとは、観光客が能動的に行動する必要がある観光スポットのことで、例えば、しまなみ海道のサイクリングなどが該当する。体験型スポットは比較的長時間を要するため、広島の滞在日数の増加、ひいては経済効果にも一定の影響が期待できる。さらに、ある体験型スポットに対する他の地域での類似体験型スポットを見つけ比較することで、ある体験型スポットの売りや改善点を発見し、さらなる訪日外国人観光客の増加につなげる。これらを旅行ブログエントリから検出・分析する。

本論文の構成は以下のとおりである。2節では、

類似する観光地の推薦や旅行ブログの自動分類による行動分析に関する関連研究について述べる。3節では、体験型スポットとそれに対する類似体験型スポットの分析について述べる。4節では、類似観光地を自動検出する手法を提案する。5節で本稿をまとめ、6節で今後の課題について述べる。

2. 関連研究

本研究では、ある体験型スポットに対する類似体験型スポットを抽出し、比較・分析を行うが、これまでに、ある観光地に対して類似性のある観光地を抽出する研究はすでに行われている。

奥村ら[1]は、観光地開発の支援技術として、開発目的地と類似する観光地における長所情報をブログから抽出する手法を提案している。本研究とは、目的が異なっているものの、類似する観光地を利用している点や、ブログエントリを情報源としている点は関連性がある。

上原ら[2]は、ユーザが入力した観光地に対し、類似性の高い観光地を推薦するシステムを構築している。類似する観光地の提案という点で本研究と似ているが、類似性のある観光地の見つけ方で異なる。上原らは情報源として、日本版 Wikipedia、

Yahoo!知恵袋, ブログ, 地図画像を用いている. さらに複数の要素から特徴ベクトルを生成し, 特徴ベクトル間のコサイン値を算出することで類似度を計算し, 推薦システムに反映している. 本研究では, 旅行ブログのみを情報源とし, ある体験型スポットについてのブログエントリからキーワードを抽出することで類似体験型スポットを抽出する.

土田ら[3]は, Word2Vec を用いることで単語が持つ意味の加算・減算を可能にし, これを観光分野に応用する研究を行っている. これにより, ある地域のある観光スポットに対する他の地域の観光スポットとの関係性, つまり類似性を抽出できる. 本研究では, ある体験型スポットについてのブログエントリから, 類似体験型スポットを自動抽出するためにキーワードを抽出するが, この意味演算を用いることで, 類似度の高い類似体験型スポットを抽出することを目指す.

本研究では, 旅行ブログの中でも体験型スポットについてのブログエントリを扱う. Fujii ら[4]は, 分析用データのひとつである英語データの TravelBlog¹から収集した旅行ブログエントリ約 50 万件を利用し, ブログ著者の属性情報(性別, 使用言語), 旅行者の行動の種類(「見る」「食べる」「買う」「泊まる」「体験する」)を自動付与するシステムを構築している. 本研究では, このシステムを用い, 旅行者の行動の種類として「体験する」カテゴリに分類されたブログエントリのみを抽出し, 利用する.

3. 類似体験型スポット

本研究では, 旅行ブログエントリから体験型スポットに着目し, それに対する類似体験型スポットを検出・分析することで, その体験型スポットの魅力を見つけ, 観光支援につなげることが目的である. 情報源として, 旅行ブログエントリが登録されている TravelBlog を利用する. TravelBlog とは, 旅行ブログエントリをメインとした海外の web サイトであり, 世界各国に関する旅行ブログエントリが写真とともに掲載されている. 本研究では, 旅行ブログエントリのテキストデータのみを対象に分析していく.

3.1. 体験型スポットの分析

TravelBlog 中のブログエントリ 736 件に人手で緯度・経度・タイプを付与した. この中で「体験」カテゴリに属するものは 181 件存在した. これら

を対象に, どの地域でどのような体験ができるのかを人手により分析した. 分析結果の一部を表 3.1 に示す. 日本固有の体験ばかりではなく, 海外でも出来そうな体験も多くみられた.

表 3.1 人手で分析した体験型スポットの例

都道府県	体験
北海道	雪まつり, 札幌ビール祭り, スキー, 富良野トリックアート美術館, 羊蹄山登山, アイスラフティング, 大雪山ハイキング, 旭岳温泉
秋田	乳頭温泉, 孫六温泉, 小安狭温泉
宮城	どんと祭
福島	旅館で日本文化体験
群馬	草津温泉, 流鏑馬
千葉	ディズニーランド・シー, 足湯, 着物体験, リンゴ狩り, いちご狩り
東京	ビール博物館(ビール試飲), 皇居周辺散歩, 東京マラソン, 人力車, メイドカフェ, トリックアート迷宮間
神奈川	箱根湯本温泉, 箱根くらかげゴルフ
福井	スキー, スノーボード
山梨	富士山登山, 富士急ハイランド, ふじやま温泉
長野	スキー, 野沢温泉, トレッキング
岐阜	日本酒試飲, 染色物見学
愛知	着物を着て犬山散歩
三重	伊賀流忍者博物館(忍者体験)
鳥取	プリクラ, スキー, 大山登山, カヤック, キャンプ
広島	酒まつり, マツダ工場見学, しまなみ海道サイクリング, ステンドグラス作り
山口	俵山温泉
徳島	吉野川ラフティング
愛媛	石鎚山登山, 道後温泉, 道後秋祭り, インディゴ染め, お遍路, 大凧合戦
福岡	玉取祭
長崎	ハウステンボスのプール, ランタンフェスティバル
熊本	黒川温泉, 砂浴, 竹瓦温泉
大分	別府温泉, 湯布院, 堀田温泉, サイクリング, 地獄めぐり
沖縄	カヤック, シュノーケリング, ホエールウォッチング, 西表島ジャングルトレッキング

¹ <http://travelblog.org>

3.2. 類似体験型スポットの分析

3.1.節で分析した体験の一部に対し、他の地域で類似体験型スポットがないか、人手により分析した。TravelBlog の検索エンジンに、類似体験型スポットについてのブログエントリが検出できそうなキーワードを入力した。例えば、マツダの工場見学に対しては、「car factory」をキーワードとして検索した。この検索結果の一部を表 3.2 に示す。他の体験に対しても同様に検索した。

この結果から、キーワードとしてそのブログエントリが何を体験したのかを表す言葉を抽出できれば良いことが分かる。また、表 3.2 からキーワードに動詞が含まれているものが多い。したがって、次節で述べる類似観光地の自動抽出には動詞に着目する。

表 3.2 人手で分析した類似体験型スポットの例

体験型スポット	キーワード	類似体験型スポット
マツダ工場見学	car factory	アウディ工場見学
日本酒試飲	wine taste	ワイン試飲
	rum taste	ラム・バハマ祭
登山	climb a mountain	キナブル登山
	hiking	韓国ハルラサン
トレッキング	trekking	ヒマラヤ登山
着物体験	try traditional costume	ベトナム民族衣装体験
しまなみ海道サイクリング	cycling	フランスでのサイクリング

4. 類似観光地の自動検出

4.1. 類似観光地の自動検出システムの構築

本節では、TravelBlog を対象に、ある体験についてのブログエントリの集合と、類似度の高い体験ができる地域を、検索エンジン INDRI²を用いて検出する手法を提案する。類似度の高さは、あらかじめ手がかり語を用意し、ブログエントリの集合が持つ手がかり語の情報と、どれほど一致する情報を持つ地域かで判定する。一致すればするほどより類似度の高い地域といえる。

手がかり語は、Fuji ら[4]が英語旅行ブログのタイプ判定をする際に用いた手がかり語リストから、情報利得を用いて「体験する」タイプのブログに頻出する手がかり語を 100 語収集したデータがあ

るため、その中の動詞のみを用いる。しかし、「get」や「pass」などのように 1 つの動詞に多数の意味を持つ動詞(多義語)も含まれている。手がかり語は、そのブログエントリが何をしたかを明確にできる単語にする必要がある。したがって、得られた動詞の手がかり語に対して、それぞれ WordNet³における語義の数を数え、12 個以上の語義を持つ動詞は削除した。表 4.1 に、手がかり語として用いた動詞一覧を示す。

また、富士山登山と沖縄本島のシュノーケリングのブログエントリの集合に対して、INDRI を用いた時の結果として類似度の高い上位 4 件の地域を表 4.2 に示す。それぞれ上位 4 件の地域のブログエントリを調査した。結果、富士山登山についても沖縄本島シュノーケリングについても似た体験をしているブログエントリであることが確認できた。また、手がかり語で比較をしても、ブログエントリの集合の手がかり語の情報と、類似度が高いとして検出された地域の手がかり語の情報は一見して合致率が高いことが分かった。

表 4.1 手がかり語動詞一覧

climb, hike, dive, jump, swim, trail, soak, trek, snorkel, diving, camp, ski
--

表 4.2 類似度の高い上位 4 件の地域

富士山登山	沖縄本島シュノーケリング
Pitres (Spain)	Hayman-Island (Australia)
Chugchilan (Ecuador)	Cozumel (Mexico)
Göreme (Turkey)	Mount-Gambier (Australia)
huangshan (China)	San-Cristobal-Island (Equador)

4.2. 考察

類似観光地の自動検出について、動詞の手がかり語を用いることで、類似度の高さを判断し、類似度が高い順に、その地域のブログエントリの集合を表示するツールを構築した。4 節で、富士山登山や沖縄本島シュノーケリングに対して類似度が高いとされた地域の体験は似ていることや、手がかり語の情報の合致率が高いことを確認している。さらに、類似度の低い 4 件についても上位 4

² <http://www.lemurproject.org/indri/>

³ <https://wordnet.princeton.edu/>

件の時と同様に、手がかり語の情報という観点から比較をしたが、合致率は低いことを確認した。以上のことからこのツールは有効であるといえる。

5. おわりに

本研究では、体験型スポットに着目した分析を行った。また、それらに対する類似体験型スポットも人手により見つけ比較・分析を行った。これより、類似体験型スポットの検出には、キーワードとして動詞に着目することがよいのではないかという結論を得た。このことから、動詞の手がかり語を用いて、ある体験に対する、類似度の高い地域を検出するツールを構築し、その有効性も確認できた。

6. 今後の課題

本研究では、類似する地域を自動的に検出するツールを構築した。手がかり語の情報として、12個の動詞を用いた。しかし、これら以外の単語については、考慮されていないため、すべての体験に関する旅行ブログエントリに対応できない。よって、手がかり語の検討をする必要がある。さらに、扱う手がかり語を複数単語にも対応させることで、さらにツールを改良できる。

参考文献

- [1] 奥村秀人, 徳久雅人, 村上仁一, 村田真樹, 観光地に対する長所情報の収集と分類の試み, 電子情報通信学会技術研究報告, 思考と言語, Vol. 110, No. 244, pp. 25-30, 2010.
- [2] 上原尚, 嶋田和孝, 遠藤勉, Web上に混在する観光情報を活用した観光地推薦システム, 電子情報通信学会研究報告, 言語理解とコミュニケーション, Vol. 112, No. 367, pp. 13-18, 2012.
- [3] 土田崇仁, 遠藤雅樹, 加藤大受, 江原遥, 廣田雅春, 横山昌平, 石川博, Word2Vecを用いた地域やランドマークの意味演算, 第8回データ工学と情報マネジメントに関するフォーラム, 2016.
- [4] Fujii, K., Nanba, H., Takezawa T., Ishino, A., Okumura, M., and Kurata, Y. Travellers' Behaviour Analysis Based on Automatically Identified Attributes from Travel Blog Entries, In Proceedings of Workshop of Artificial Intelligence for Tourism, PRICAI, 2016.