

観光の形態に基づいた 旅行プランの自動生成

柴田 有基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{shibata, 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に対して自然言語処理技術を用いて解析し、観光の形態に基づいて自動分類をする。近年、旅行者は従来の娯楽を追求するのみの観光だけではなく、様々な観光の形態が誕生し、現在もその多様化は進んでいる。例えば、健康回復や維持、増進につながる観光はヘルスツーリズム、災害被災跡地や戦争跡地など、人類の死や悲しみを対象にした観光はダークツーリズムと呼ばれる。このような観光の形態を旅行ブログから自動判定することで、世界各地の観光地でどのような形態の観光が可能か調べることができる。また、特定の形態に基づいた観光地の推薦や旅行のプランニングも可能になると考えられる。そこで本研究では、9種類の観光の形態を定義し、機械学習を用いて旅行ブログエントリをこれらの観光の形態に自動分類する手法を提案した。1168件の英語旅行ブログデータを用いて実験を行った結果、精度 0.287、再現率 0.790 が得られた。

Keywords：旅行ブログ、観光の形態、ツーリズム

1. はじめに

広島は世界的に有名な観光都市であり、連日多くの観光客が訪れている。広島の代表的な観光名所・施設には、「平和記念資料館」や「原爆ドーム」などがあるが、これらは負の遺産と呼ばれており、これらを対象にした観光は一般的にダークツーリズムと呼ばれている。ダークツーリズムのように、近年の観光では、ただ「楽しむ」といった娯楽の要素だけではなく、「学ぶ」や「考える」といった娯楽以外の要素も含まれることが多い。このように、旅行者の観光の形態は多様化しており、こうした情報を知ることは、観光事業を計画する上で非常に重要である。

ある観光地における旅行者のニーズや観光の形態を知るための方法の一つとして、旅行者に対してアンケートを実施する方法がある。しかし、この方法には多くの時間やコストがかかるという問題が存在する。そこで、近年では Web 上で公開されている旅行記、すなわち旅行ブログエントリを収集、分析するという方法が広まりつつある。本研究では、旅行ブログエントリを対象に、旅行者の観光の形態を判定し、その情報を用いて旅行プ

ランを自動生成するシステムを開発する。

本論文の構成は以下のとおりである。2節では、旅行ブログの自動分類による行動分析や旅行プランの推薦に関する関連研究について述べる。3節では、観光の形態とそれに基づいた旅行ブログエントリの分類について、4節では、観光の形態に基づいた旅行ブログエントリの分類における実験と考察、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、旅行ブログを観光の形態に基づいて自動で文書分類を行い、その結果を反映した旅行プランを推薦するが、これまでも旅行ブログのような文書をカテゴリ別に分類する研究や旅行プランを推薦する研究はすでに行われている。

カテゴリ別の分類について、野沢ら[1]は訪問サイトを「観光」「ビジネス」「食事」「購買」の4つのクラスに自動分類する手法を提案している。また藤井ら[2]は旅行ブログを「買う」「食べる」「体験」「泊まる」「見る」に分類する手法を提案している。これらの分類の観点は、旅行者の行動、つ

まり旅行者が何をやっているのかに基づいたものである。本研究で実施しようとしている「観光の形態」に着目した分類と関連性はあるものの、基本的には別の観点であると考えられる。本研究と藤井らの関係を下記の表 2.1 に示す。また、藤井らの分類手法に加えて、本研究で提案する観光の形態に基づく分類が可能になれば、「カルチュラルツーリズム」で「食べる」に関する情報を調べる、といったように、よりきめ細かい検索が可能になると考えられる。

表 2.1：本研究と藤井ら[2]の対応関係

本実験	藤井ら
インフラ、ハードツーリズム	「買う」 「食べる」 「体験」 「泊まる」 「見る」
ヘルスツーリズム	「食べる」 「体験」 「泊まる」 「見る」
スポーツツーリズム	「体験」 「見る」
グリーンツーリズム	「食べる」 「体験」 「見る」
ヘリテージツーリズム	「見る」

旅行プランの推薦では、新井ら[3]は Twitter を用いて、観光スポットにおける旅行者の観光体験に基づくツイートを収集し、それらのツイートに含まれる観光情報を利用して観光ルートを推薦する手法を提案している。観光ルートの推薦に用いられた観光情報は、旅行者の観光ルートから得られた観光スポットの共起頻度と観光スポットを訪れた時間帯、「食事」、「景観」、「土産」、「行動」の4つのカテゴリに分類した結果を観光ルート推薦の指標にしている。本研究では観光の形態に基づいた旅行プランを推薦する点で異なる。

3. 観光の形態に基づいた旅行プランの自動生成

本研究では、旅行ブログエントリーからブログ著者の観光の形態を自動判定し、この形態に基づいた旅行プランを自動生成するシステムを構築することを目的とする。情報源として、旅行ブログエントリーが登録されている TravelBlog¹を利用する。

TravelBlog とは、旅行ブログエントリー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ーが写真とともに掲載されている。本研究では、英語で書かれた旅行ブログエントリーのテキストデータのみを対象に分析する。

3.1. 観光の形態

観光の形態には厳密な定義はないため、使用者によって異なる解釈で用いられる場合がある。本研究では、それぞれの観光の形態に対して、過去の文献[4-8]を参考に独自に定義した。観光の形態とその定義、またその具体例を表 3.1 に示す。本研究では、表 3.1 で示した 9 種類の観光の形態に基づいて旅行者の観光の形態を明らかにするため、個々の英語で書かれた旅行ブログエントリーを自動分類する。分類の方法は、人手で付けた正解ラベルとブログ本文を入力データとし、機械学習によって分類を行う。

3.2. 旅行プランの自動生成

3.1 節で示した観光の形態に基づいて、旅行プランを自動生成する。生成する方法については、分類した旅行ブログを用いて、近隣で同じ形態の観光地を結びつけることで実現する。観光の形態には様々な種類があるが、地域によって過去の歴史や土地などの理由から、偏りがあると推測される。例えば、広島市では「原爆ドーム」や「平和記念資料館」が有名であることから、観光の形態の一つであるダークツーリズムが多くなると考えられる。こうした情報を明らかにすることにより、各訪問地の特徴を活かした観光施策を行うことができると思われる。また、同じ被爆地である長崎も同様に似た傾向になるようであれば、似た観光地同士で観光施策の参考にできることも期待される。下記の表 3.2 に「長崎県」で「グリーンツーリズム」の旅行プランを生成する場合の例を示す。

表 3.2：旅行プランの例

観光の形態	場所	内容
グリーン	おおむら夢 ファームシュシュ	農業体験
グリーン	壱岐市観光連盟	魚釣り体験
グリーン	民宿 安堵家・ 近海荘	海水浴

¹ <https://www.travelblog.org/>

表 3.1 : 観光の形態と具体例

観光の形態	定義	例
インフラ、ハードツーリズム[6]	近代的な建造物や娯楽施設を対象にした観光。	橋、ダム、テーマパーク、ショッピングモール、水族館、博物館、動物園
ヘルスツーリズム[4]	心身を癒すことや散歩などの軽い運動を通して健康維持を目的とした観光。	宗教的巡礼、温泉、ハイキング、トレッキング
ダークツーリズム[5]	災害被災跡地や戦争跡地など、人類の死や悲しみを対象にした観光。	東北大震災、福島原発事故、World Trade Center、原爆ドーム
コンテンツツーリズム[7]	アニメや映画などの聖地巡礼や作品に関連した施設を対象にした観光。	千と千尋の神隠し、坊ちゃん
スポーツツーリズム[4]	スポーツを体験または観戦することを目的とした観光。	MLB、プロ野球、サッカー
グリーンツーリズム[4]	自然と触れ合うことを目的とした観光。	農業（漁業）体験、フルーツ狩り、ピクニック
ヘリテージツーリズム[4]	世界遺産や歴史的な建築物を対象にした観光。	世界遺産、国宝、寺、神社、城
カルチュラルツーリズム[8]	それぞれの地域の生活や文化、民族、伝統などを対象にした観光。	着物体験、神楽、祭り、初詣
スタディツーリズム[4]	留学や修学旅行など学習を目的とした観光。	留学、修学旅行

4. 実験

本実験で使用する旅行ブログエントリーでは、3節で述べた TravelBlog を用いる。TravelBlog には画像データとテキストデータが存在するが、今回はテキストデータのみを使用した。

4.1. 実験条件

【実験に用いるデータ】

1168 件の旅行ブログエントリーを実験に用いた。各ブログエントリーに対し、観光の形態 9 種類のカテゴリに人手で分類し、これらを機械学習の際の訓練用および評価用データとして用いた。人出で分類した結果の内訳を表 4.1 に示す。なお、一つの旅行ブログエントリーには複数の観光の形態が付与されることや逆に一つも付与されないこともありうるという設定にしているため、9つの件数の総和が 1168 より小さくなっている。また、表 4.1 からわかるように観光の形態によっては、ブログ件数に偏りがある。そこで本実験では、比較的ブログ件数が多い表 4.1 中の観光の形態 1, 2, 5, 6, 7 のデータで分類実験を行う。

【機械学習と評価尺度】

機械学習を用いて、観光の形態に基づいた旅行ブ

ログの自動分類を行った。機械学習には fastText² を用いた。fastText とは Facebook が提供する単語のベクトル化とテキスト分類をサポートした機械学習のライブラリである。評価方法については、5 分割交差検定を行い、評価尺度には精度と再現率を用いた。

【実験手法】

fastText による機械学習で分類を行った。分類に用いる素性は、観光の形態それぞれの正解ラベルとそのブログ本文の単語である。また、fastText のパラメータについては、epoch 数は 15、word2vec の次元数は 300、単語は uni-gram とした。

² <https://fasttext.cc/>

表 4.1：人手で分類した結果の内訳

観光の形態	件数
1. インフラ、ハード	28
2. ヘルス	50
3. ダーク	2
4. コンテンツ	4
5. スポーツ	18
6. グリーン	150
7. ヘリテージ	19
8. カルチュラル	2
9. スタディ	1
判定したブログの総数	1168

4.2. 実験結果と考察

fastText による分類により得られた結果は、精度 0.287、再現率 0.790 であった。精度が低い最大の原因は、正解データの数に比べて一つのブログ中に含まれる単語が非常に多いこと(1 ブログあたり平均 699 語)、また、学習に用いられる正例が負例に比べて非常に少ないことが挙げられる。さらに、今回分類する観光の形態は表 4.2 で示す 5 種類であるが、「グリーンツーリズム」が他の観光の形態と比べて 3 倍以上件数が多くなっている。そのため、出力の大部分が「グリーンツーリズム」に分類されてしまっている。

この問題に対し、機械学習を行う際にブログ中のすべての単語を使うのではなく、分類に有効と思われる単語を情報利得や自己相互情報量を使って事前に選定することなどで改善できると思われる。この他、近年では画像解析、特に物体認識の精度が飛躍的に向上しており、この認識結果をテキスト情報とあわせて用いることで、観光の形態をより良い精度で分類することができると考えられる。

5. おわりに

本研究では、旅行ブログエントリを 9 種類の観光の形態に自動分類し、その結果に基づいて、旅行プランを生成する手法を提案した。1168 件の英語旅行ブログを対象に機械学習ライブラリのひとつである fastText を用いて構築した分類器で実験を行ったところ、精度 0.287、再現率 0.790 が得られた。

6. 今後の課題

本研究では、機械学習を用いて旅行ブログを観光の形態に基づいて自動分類を行った。分類には機械学習を用い、素性には人手で分類した正解ラベルとブログ本文を使用した。精度は低い結果と

なったが、学習をする際に有効と思われる単語を選定するなど工夫をすることで改善できるか調べる必要がある。また今後は、分類後の観光プランを生成することにも取り組む。

参考文献

- [1] 野沢悠哉, 遠藤雅樹, 江原遥, 廣田雅春, 横山雅平, 石川博: マイクロブログを用いたユーザの訪問目的と動向の推定, 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)
- [2] 藤井一輝, 難波英嗣, 竹澤寿幸, 石野亜耶, 奥村学, 倉田洋平: 旅行者の行動分析のための旅行ブログエントリの属性推定, 観光と情報, Vol.13, No.1, 83-96 (2017)
- [3] 新井晃平, 新妻弘崇, 太田学: Twitter を利用した観光ルート推薦の一手法, 第 7 回データ工学と情報マネジメントに関するフォーラム (2015)
- [4] 山下晋司: 観光学キーワード, 有斐閣(2011)
- [5] 井出明: ダークツーリズム拡張—近代の再構築, 美術出版社(2018)
- [6] 藤井千賀子: インフラツーリズムガイド 2018, 芸文社(2018)
- [7] 岡本亮輔: 聖地巡礼, 中央公論新社 (2015)
- [8] 中島真美: カルチュラル・ツーリズムによる「住民参加型開発」の有効性, 第 8 回観光に関する学術研究論文入選論文集, 1-17(2002)