

旅行プラン自動生成に向けた旅行記からの情報抽出

Extracting information from travelogues for automatic trip plan generation

坂上 知司

Tomokazu Sakagami

広島市立大学 言語音声メディア工学研究グループ

Language and Speech Research Group, Hiroshima City University

概要 近年、新型コロナウイルスの5類移行や全国旅行支援の実施などにより、観光地に徐々に賑わいが戻り、観光客数が増加している。その中で、観光客にとって旅行を快適に楽しむための旅行プラン作成は非常に重要なものとなっている。しかし、旅行プランを作成するには観光地を回る順番や移動手段などを決定する必要があり、馴染みの無い場所についてSNSやWebなどから多くの情報を収集し吟味する作業は観光客にとって負担が大きい。そこで本研究では、旅行初心者や初めての場所へ行く旅行者を対象とした各観光地における典型的な旅行プランについて、旅行記データからクラスタリングを用いて抽出する手法を提案する。そしてさらに旅行者に合った詳細な旅行プラン作成に向けた今後の展開について報告する。

1 はじめに

近年、世界中で新型コロナウイルスの脅威が収束し、観光業界はその影響から回復しつつある。日本でも新型コロナウイルスの5類移行や全国旅行支援の実施などにより、観光地に徐々に賑わいが戻り、観光客数が増加している。

その中で、観光客にとって、旅行を快適に楽しむための旅行プラン作成は非常に重要なものとなっている。しかし、旅行プラン作成にはSNSやWebなどから観光地を回る順番、移動手段など、普段慣れていない場所について多くの情報を収集する必要があり、観光客にとって負担が大きいものとなっている。特に初めて行く観光地や旅行初心者の場合、それはさらに困難となるだろう。

そこで本研究では、旅行初心者や初めての場所へ行く旅行者を対象とした各観光地における典型的な旅行プランについて、旅行記データからクラスタリングを用いて抽出する手法を提案する。そしてさらに旅行者に合った詳細な旅行プラン作成に向けた今後の展開について報告する。

2 関連研究

倉田らは、旅行者の嗜好や制約に基づいた旅行プランの提案を行うCT-Planner[1]というシステムをインターネット上で公開している。

実際に広島市で旅行プランを生成した例を図1に示す。まずトップ画面で観光地と移送手段を選択し、表示されるいくつかの観光地の中から出発地、到着地および旅行スタイルを選択する。そして各項目について選択を完了すると、図1のように旅行プランが生成される。画面右側には推薦旅程、真ん中にはマップとルートが提示されている。左側にあるプランの条件設定欄には、歩く速度や曜日、歩きたくな

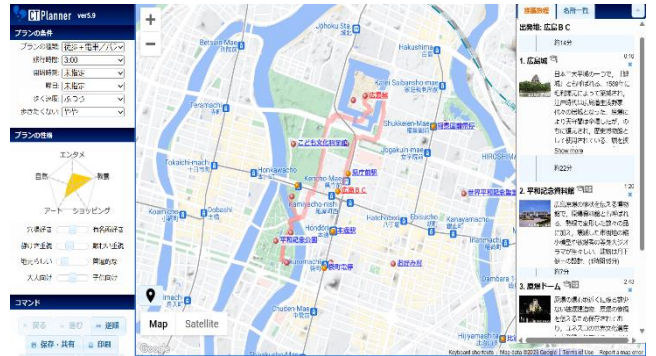


図1: CT-Plannerを用いた旅行プラン生成例

い度合など詳細な設定項目があり、設定を変更するとリアルタイムで推薦旅程に反映される。

このようにCT-Plannerは詳細な旅行プラン作成が可能となっている。しかし、初めて訪れる観光客や旅行初心者にとっては、現地の状況や移動における手間について想像することが難しいため、詳細な設定項目があっても使いこなすことが難しい。また、提案されたプランが一般的に妥当な旅行プランであるかを判断することも容易ではない。

武信らも、倉田らと同様に観光スポットと移動経路に対する嗜好を考慮した観光ルート推薦システム[2]を構築している。このシステムでは対象観光地、出発地点、終着地点、観光時間を固定したうえで、利用者に対する個別的な推薦を行っている。本研究では全国各地を扱えるものを目指す。

大澤ら[3]は、リピーターに向けた観光周遊プラン作成支援システムを提案している。この研究はリピーター等の特定の観光地に精通している観光者を主な対象としている。

このように旅行プラン生成についての研究は行われているが、旅行初心者や初めて行く観光地を対象としている研究は見当たらない。そこで、本研究ではまず典型的な旅行プランを提示してから、個別の条件の設定を行うシステムを目指す。その際、ユーザーが設定した観光地に対してプランを推薦できるよう、インターネットから収集した旅行記から対象観光地の典型的な旅行プランを自動生成することを目指す。

3 提案手法

本研究では、株式会社地球の歩き方が国立情報研究を通して提供している「地球の歩き方旅行記データセット」[4]の「旅行記」および旅程を記した「旅スケジュール」のテキストデータを分析対象として、観光地の典型的なプランの作成を行う。旅スケジュールとは、旅行記に付随して任意で入力されるテキストであり、主に移動経路に関する情報が

1 日目 2021 年 10 月 15 日(金)
...
08:19 - 08:29 郡山駅/東北新幹線
08:29 - 09:40 移動 (電車)
09:40 - 09:46 会津若松駅/磐越西線
09:46 - 10:20 若松駅前バスターミナル
10:20 - 11:07 移動 (バス)
...
2 日目 2021 年 10 月 16 日(土)
00:00 - 09:25 庄助の宿 瀧の湯
09:25 - 09:32 東山温泉入口 (瀧の湯前) バス停
09:32 - 09:36 移動 (バス)
...

図 2 : 旅スケジュールの例

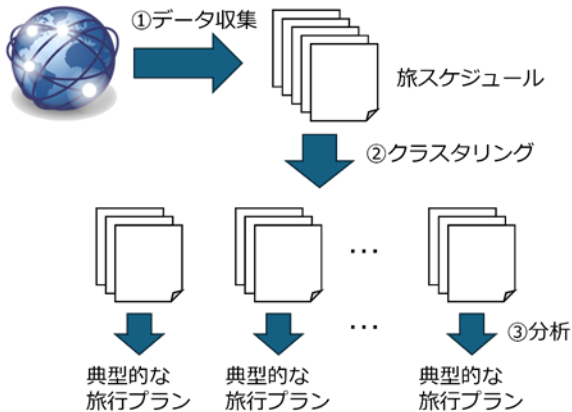


図 3 : 提案手法の流れ

記載されている。旅スケジュールの例を図 2 に示す。

旅スケジュールから典型的な旅行プランを作成する手順を図 3 に示す。

1. 対象観光地の旅スケジュールのテキストを収集する
2. 各旅行者による旅スケジュールを単語ベクトルに変換し、クラスタリングによってプランを分類する
3. クラスタリングの結果から、集まったプランの特徴や典型的なプランがあるかどうかを分析する

さらに本研究では、「旅行記」と「旅スケジュール」の両方に記載されている観光スポットを「旅行者がより印象に残った観光スポット」として抽出する。また、「旅行記」内で観光スポット名の前後に存在する文についてネガポジ分析を行うことで観光スポットに対する旅行者の印象を分析する。

4 分析実験

4.1 典型的なプラン生成のためのクラスタ分析

実験に用いたデータ

本研究では広島県と京都府を観光している「旅スケジュール」と「旅行記」を分析対象とする。実験に用いたデータは広島県が 40 件、京都府が 70 件である。本稿では、広島県の実験結果について紹介する。

表 1 : 広島県の典型的な旅行プランの特徴

プランの特徴	件数
① 広島市内のみを観光	5 件
② 宮島のみを観光	2 件
③ 広島市内&宮島	15 件
④ (広島市内&宮島)+呉, 尾道, 福山	11 件
⑤ 他と観光地のかぶりが少ない	7 件

表 2 : 典型例の割合

クラスタ数	5	10	15	20
典型例が見つかったクラスタの数	4	8	13	15
典型例の割合	80.0%	80.0%	86.6%	75.0%

実験条件

旅スケジュールテキストを単語ベクトルに変換する手法には BERT[5]を用いる。本実験では 768 次元のベクトルに変換する。また、クラスタリング手法には階層的クラスタリングを用いる。クラスタ数は 5, 10, 15, 20 に設定し、実験を行った。

実験結果

クラスタリング結果に対して作業員 1 名が分析を行った結果、広島県の旅行プランは大きく 5 つの典型的旅行プランに分けられると判断した。表 1 にクラスタリングによって得られた広島県の典型的旅行プランの特徴を示す。表中の件数は広島県観光の旅スケジュール 40 件が各旅行プランに分類された件数を表す。「①広島市内のみを観光」は、「原爆ドーム」をメインに、その近くにある「ひろしま美術館」や「平和記念公園」などを観光しているプランである。「②宮島のみを観光」は、「厳島神社」をメインに、ロープウェイに乗って「弥山」に登り、獅子岩展望台から景色を楽しむというようなプランとなっている。「③広島市内&宮島」は一番件数が多く、「原爆ドーム」と「厳島神社」という広島にある 2 つの世界遺産を巡るプランとなっている。「④(広島市内&宮島)+呉, 尾道, 福山」は「原爆ドーム」、「厳島神社」に訪れ、さらに広島市外の人気観光地である呉の「大和ミュージアム」や尾道の「千光寺」、福山の「鞆の浦」などに訪れているプランである。「⑤他と観光地のかぶりが少ない」は、「竹原」や「世羅町」など、満足度は高いものの地域外の人にはあまり知られていない観光地を訪れるプランとなっている。

この結果から、本論文の分類手法によって広島県の基本的な人気観光地が入っている典型的なプランが生成されることが確認できた。これらの典型的なプランを参考に、ユーザーはそれぞれのプランが自分に合っているか判断し、旅行プラン作成に役立てることができる。

続いて典型例が見つかったクラスタの数の割合を計算する。計算式を式 I に示す。

典型例が見つかったクラスタ数の割合 =

$$\frac{\text{典型例が見つかったクラスタ数}}{\text{総クラスタ数}} \times 100 \quad \dots \text{式 I}$$

典型例は表 1 の①～④までのプランとして、⑤のプランのみが含まれるクラスタは数えない。表 2 に実験結果を示す。

クラスタ数を変化させた実験では、データ数が少ないため4つとも高い割合となっているが、クラスタ数 15 で実行した際が一番高い割合を示した。一方、クラスタ数をさらに5増やしたクラスタ数 20 での結果は最も低い値となった。クラスタ数 5 と 10 で実行した際の割合は同じだが、クラスタ数が低いほど、1つのクラスタに含まれるプランの数が多くなるため、典型例を見つけやすいのはクラスタ数 10 の方である。この結果から、クラスタリングによって典型的プランを抽出するには、適切なクラスタ数の選択が重要であることが示唆された。

4.2 観光スポットに対する印象情報の抽出

実験条件

観光スポットに対する印象情報を抽出するため、「旅スケジュール」に記載のある観光地について「旅行記」と文字列の完全一致で対応付けを行う。本研究では、両方に記載のある観光地について TF-IDF を用いて重要度の高い観光地を調べる。

実験結果

表 3 に TF-IDF 値の高い上位 10 件の観光スポットを示す。対応付けについては、重要度の高い観光スポットとして「厳島神社」と「原爆ドーム」が出てきており、旅行者の印象に残っていることが分かる。また、エリアで見ると、「廿日市市」エリアに含まれる観光スポットが 4 つ、「広島市」エリアに含まれる観光スポットが 3 つと、この 2 つのエリアが半分以上を占めている。このエリア以外の観光スポットを見ると、「錦帯橋」や「出雲大社」など同じ中国地方の人気観光スポットが上位に出てきている。この結果から広島県だけでなく他県も巡りたいユーザーに、より詳細なプランを生成する場合に役立てることができる。

4.3 観光スポットに対するネガポジ分析

実験条件

huggingface/transformers の pipeline クラスを用いてネガポジ分析を行う。モデルとトークナイザには東北大学の BERT[5]モデルを用いた。

実験結果

今回は「旅行記」から原爆ドームについて言及された文章 2 件を対象としてネガポジ分析を行った。結果を表 4 に示す。

ネガポジ分析については、両方ともネガティブと判定さ

表 3 : TF-IDF 値の高い観光スポット

観光スポット	TF-IDF 値	地域
厳島神社	0.139	廿日市市
原爆ドーム	0.121	広島市
広島城	0.050	広島市
広島空港	0.040	三原市
宮島	0.039	廿日市市
錦帯橋	0.035	山口県
平和記念公園	0.034	広島市
弥山	0.033	廿日市市
出雲大社	0.032	島根県
五重塔	0.031	廿日市市

表 4 : 原爆ドーム旅行記のネガポジ分析結果

旅行記テキスト	分析結果
そこだけが時間が止まったような姿で私を迎えてくれました。痛々しくも迫力のある様子は、人間に対する警告や戒めを与えているよう。けれどそこには同時に平和もありました。	ネガティブ score:0.519
建物だけではなく、崩れた瓦礫なんかもそのまま下に残されているんです。テレビや写真で見るドームとはまた違います。なんというか、一気にリアリティが増すというか、その時代に生きていた人が一人一人見えてくるようです。	ネガティブ score:0.603

れた。それぞれの score を見ると約 0.5~0.6 となっておりあまり高くはない。一因として、文章中に悲惨な現実とそこから感じたこと両方についての記述が含まれており、文章全体に対してネガポジ評価を行うのは人間にとっても難しいことが考えられる。

5 今後の展開

本研究では旅行記を用いた旅行プラン自動生成に向けて、3つの実験について取り組んだ。今後の課題として、まず典型的なプラン生成については、データ数の増加やクラスタリング手法の改善を行い、より精度の高い旅行プランの生成を目指す。

また、詳細なプラン生成においては観光スポット名の対応付けが必要となるが、現在は異表記や誤った名称を考慮することはできないため、今後これらを一つの同じ観光地として扱うための改善を行う予定である。

さらに、ネガポジ分析以外にも季節、天候、子供連れなどその時の状況や旅行形態についても情報を抽出し、詳細なプラン生成に役立てることを検討している。

謝辞

本研究は、国立情報学研究所の IDR データセット提供サービスにより株式会社地球の歩き方から提供を受けた「地球の歩き方旅行記データセット」を利用した。

本研究は JSPS 科研費 20K20091 の助成を受けたものである。

6 参考文献

- [1] 倉田洋平:街歩きプラン作成支援ツール CT-Planner5.9 <https://ctplanner.jp/ctp5/> (2024/2/28 アクセス)
- [2] 武信雄平, 奥野拓:観光スポットと移動経路に対する嗜好を考慮した観光ルート推薦システムの構築, 情報処理学会第 83 回全国大会講演論文集 2021(1), 335-336, 2021-03-04
- [3] 大澤吉輝, 高山毅:リピーターのための観光周遊プラン作成支援システム, 情報処理学会第 79 回全国大会講演論文集 2017(1), 991-992, 2017-03-16
- [4] 大内啓樹, 進藤裕之, 若宮翔子, 松田裕貴, 井之上直也, 東山翔平, 中村哲, 渡辺太郎. 地球の歩き方旅行記データセット. 言語処理学会 第 29 回年次大会発表論文集, pp.2920-2924, 2023.
- [5] 大内啓樹, 進藤裕之, 若宮翔子, 松田裕貴, 井之上直也, 東山翔平, 中村哲, 渡辺太郎. 地球の歩き方旅行記データセット. 言語処理学会 第 29 回年次大会発表論文集, pp.2920-2924, 2023.