

行動履歴に基づく地域の環境要因を考慮した 観光行動モデルの構築とその応用

笠原 秀一^{1,a)} 田村 和範² 飯山 将晃¹ 椋木 雅之³ 美濃 導彦¹

受付日 2015年8月18日, 採録日 2016年2月8日

概要: 旅行ルートを推薦する従来の旅行計画支援システムでは、直前に訪問した観光スポット（以下スポット）や旅行者の嗜好、旅行目的といった観光行動に影響を与える特定の要因を組み込んだ旅行者の観光行動モデルを構築し、モデルが出力する次に訪問するスポットと、旅行者が入力する制約条件に従って作成した旅行ルートを旅行者に提示している。こうした従来の観光行動モデルでは、すべてのスポット間遷移に対して要因が与える影響の度合いは等しいと仮定しているが、実際には遷移ごとに要因の影響度合いは異なる。そこで本研究では、スポット間遷移ごとに影響度合いの大きい要因を用いることで、精度の高い観光行動モデルを構築する。観光行動に影響を与える要因を、旅行者の嗜好という人的要因と、季節や時間帯、天気、訪問スポットといった地域の環境要因に分類する。環境要因モデルでは、地域の環境要因によって変動するスポット間遷移の傾向を、過去の行動履歴から学習した条件付き確率が最大となる要因を用いて記述する。人的要因モデルでは、旅行者の嗜好によって影響されるスポットの訪問傾向を、協調フィルタリングを用いて記述する。また、提案モデルの応用先としてルート推薦システムを構築した。写真共有サイトから得た行動履歴を用いた実験により、提案する観光行動モデルとルート推薦システムが、従来手法に比べて過去の旅行者の実際の旅行ルートに近いルートを生成できることを示した。

キーワード: 旅行ルート生成, マルコフモデル, 協調フィルタリング, 行動履歴

Tourist Behavior Model Construction Based on Tracks of Tourists Using Regional Environmental Factors and Application

HIDEKAZU KASAHARA^{1,a)} KAZUNORI TAMURA² MASAOKI IYAMA¹ MASAYUKI MUKUNOKI³
MICHIIHIKO MINOH¹

Received: August 18, 2015, Accepted: February 8, 2016

Abstract: In this research, we aim at constructing a tourist behavior model using human factor and regional environmental factor, and implementing travel route recommendation system using the output of the tourist behavior model. Previous route recommendation systems recommend the travel route based on tourist behavior model using a certain factor that affects tourist's transition to spots like visited spots. Although the previous models assume that the affecting degree of factors is the same for all transitions, the affecting degrees of factors are actually different for each transitions. Therefore, we can improve the accuracy of model by using combination of multiple factors for each transition. In this research, we classify the factor that affects the tourist behavior into human factor and regional environmental factor, and construct 2 tourist behavior models considering the 2 factors. As human factor model, we describe tourist's taste based on collaborative filtering. As environmental model, we describe tourist's spot transitions based on the environmental factor group like seasonality, time, amount of rainfall, visited spots. In addition, as the application of the proposed model, we implement the route recommendation system that recommends travel route using the output of the tourist behavior model. We conducted experiments for evaluating our model and system. The experimental results showed that our method can produce a trip route that is more similar to the route past tourist actually pursued than the route produced by methods without using the tracks.

Keywords: route planning, Markov model, collaborative filtering, activity history

1. はじめに

観光地を旅行するとき、一般に旅行者は旅行計画を立てる。旅行計画では、自分自身の嗜好に合った観光スポット（以下スポット）を見つけ、旅行可能な時間内にどのスポットをどのような順番で訪れるかを示す旅行ルートを組み立てる。しかし、旅行者にとって観光地は訪問経験が少ない土地なので、自分の嗜好を適切に反映しているか、時間内にすべてのスポットを訪問できるか、訪問する順番は適切なものか、季節や時間帯といった点で適切なスポットなのかなど、旅行計画の妥当性に不安を持つことも多い。旅行者の嗜好、季節や時間帯に応じた旅行ルートを推薦するシステムがあれば、こうした不安を解消できる。

旅行者がスポットを訪れることを観光行動と定義し、主体となる旅行者、次に訪問する対象となるスポット、季節や時間帯、天候、これまでに訪問したスポットといった観光地が属する地域の環境要因、旅行者の嗜好という人的要因の4つの要素によって構成されるものとする。観光行動をモデル化して記述することで、旅行者が次に訪れるスポットを推定できる。観光行動のモデル化としては、環境要因や人的要因を考慮して旅行者が次に訪問するスポットへの遷移を記述するモデルが提案されている [1], [2], [3], [4], [5]。本研究ではこうした遷移に影響する要因を遷移影響因と呼ぶ。従来研究では、すべての遷移において遷移影響因が与える影響の度合いは等しいと仮定している [1], [2]。たとえば、単純マルコフモデルを用いた手法では直前の訪問スポットのみを遷移影響因としている場合が多い。しかし、実際には遷移影響因がスポット間遷移に与える影響の度合いはスポットごとに異なる。それゆえ、複数の遷移影響因の組合せとしてスポットごとにモデルを記述すれば、より精度の高い観光行動モデルを構築できると考えられる。

そこで本研究では、スポットごとに組み込む遷移影響因を考慮した観光行動モデルを構築し、さらにこの観光行動モデルの出力をもとに旅行ルートを推薦するシステムを提案する。これらの観光行動モデルと旅行ルート推薦システムは互いに独立であり、別のモデルやシステムに置換できる。遷移影響因は、旅行者の嗜好という人的要因と、季節や時間帯、天候、これまでに訪問したスポットといった環境要因に分けられる。提案する観光行動モデルは、どのような環境の場合にどのようなスポットを訪れやすいかを記

述する環境要因モデルと、どのような旅行者がどのようなスポットを好んで訪れやすいかを記述する人的要因モデルの2つのモデルからなる。旅行ルート推薦システムは、観光行動モデルが出力するスポット間遷移のスコアを用いて旅行ルートを作成し、旅行者が入力した条件を満たす適切な旅行ルートを提示する。

2. 従来の観光行動モデルと旅行ルート推薦

2.1 環境要因に基づく観光行動モデル

一般的な旅行者が次のスポットをどれほど訪れやすいかを記述するモデルとして、単純マルコフモデルが広く利用されている [1], [2]。単純マルコフモデルでは、次に訪れるスポットは現在のスポットのみに依存すると考える。しかし、実際の遷移はそれ以前に訪れたスポットや、季節や時間帯といった環境要因の影響を受けている。単純マルコフモデルでは、現在のスポット以外の環境要因を考慮できない。これに対し、Kurashima ら [3] は、単純マルコフモデルとトピックモデルを融合したモデルを提案している。トピックモデルにより、旅行者が現在訪れているスポットだけではなく、旅行者がそれまでに訪れたスポットの情報も利用している。また、山崎ら [4] や Canneyt ら [5] は、日時を考慮して遷移先を推定している。これらの研究は、観光行動に影響を与える特定の環境要因に注目し、観光行動モデルに組み込んだものと解釈できる。本研究では、より多くの環境要因を考慮するため、遷移先への条件付き確率が最も高くなる環境要因の組合せをモデルに組み込む。その際、遷移元のスポットごとに条件付き確率が高くなる組合せが異なるため、過去に蓄積された旅行者の行動履歴を利用して最も有効な組合せをモデルに用いる。

2.2 人的要因に基づく観光行動モデル

旅行者の嗜好という人的要因に着目したモデルとしては、協調フィルタリング [6] に基づいて次に訪問するスポットを記述するモデルがある。協調フィルタリングは情報推薦システムで広く用いられている。このモデルでは、他の旅行者との類似性からある旅行者が訪問したいスポットを推定する。また、旅行者のスポットに対する嗜好を直接あるいは間接的にスコア付けし、スコアに基づいてルートを推薦する手法もある。Maruyama ら [7] が提案するシステムでは、全スポットの重要度評価を旅行者に要求する。Vansteenwegen ら [8] が提案するシステムでは、旅行者が自身の嗜好を入力し、入力内容と事前に入力されたスポット情報から間接的にスポットをスコア付けする。Kurata ら [9] は、旅行者は自身が求める旅行ルートの条件をすべて認識しているわけではないと考え、対話的なやりとりを通して旅行者の嗜好を把握し、スポットをスコア付けする。こうした直接あるいは間接的にスポットに対する嗜好をスコア付けする手法は、多かれ少なかれ旅行者の入力を必要

¹ 京都大学学術情報メディアセンター
Academic Center for Computing and Media Studies, Kyoto University, Kyoto 606-8501, Japan

² 京都大学大学院情報学研究所
Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan

³ 宮崎大学工学教育研究部情報システム工学科
Faculty of Engineering, University of Miyazaki, Miyazaki 889-2192, Japan

a) hidekazu.kasahara@mm.media.kyoto-u.ac.jp

としており、旅行者の負担となる。そのため、本研究では人的要因モデルに協調フィルタリングを利用する。

2.3 旅行ルート生成におけるスポットの訪問順序

先行研究 [7], [8], [9] では、スポットに対するスコアのみを考慮して旅行ルートを生成している。しかし、スポットの訪問順序も旅行ルートの良し悪しに関わる重要な要素である。スポットの訪問順序を考慮して旅行ルートを生成する手法として、過去の旅行者がたどった旅行ルートをスポットの列と見なし、列そのものを直接分析することで、旅行者に提示する旅行ルートを獲得する手法がある。Okuyama ら [10] は、同じスポットを含む過去の旅行者の旅行ルートを2つ選び、それらの前半と後半をつなげることで、提示する旅行ルートを獲得している。また、Arase ら [11] は、旅行ルートを時間情報が付いた列 (Temporary Annotated Sequence, TAS) ととらえ、TAS のマイニング手法を応用することで、頻出する旅行ルートを抽出している。これらの手法は、獲得した旅行ルートの中から、旅行者の入力に合致するものを推薦しているので、条件に合致する旅行ルートがない場合は推薦できない。これに対し Lu ら [12] は、スポット間遷移のスコアを、それらのスポットを順番に遷移した旅行者数として定義し、スポット間遷移をスコア付けすることで、スポットの訪問順序を考慮する手法を提案している。旅行者数は過去の旅行者の行動履歴から抽出している。提案する旅行ルート推薦システムでは、Lu ら [12] と同様にスポット間遷移をスコア付けすることで、スポットの訪問順序を考慮したルートを推薦する。

3. 観光行動モデルと行動履歴を用いた旅行ルート推薦システム

3.1 提案手法の概要

本研究では対象となる観光地での観光行動モデルを、特定の環境要因における旅行者の遷移傾向を記述する環境要因モデルと、旅行者の訪問スポットに関する嗜好を記述する人的要因モデルとして記述する。入力は、現在訪問しているスポット、必ず訪問したい要求スポット、訪問したくない除外スポットである。出力は、環境要因の組合せに対する現在訪問しているスポットから条件付き確率に基づいて算出した、次に訪問するスポットへの遷移のスコアと、スポット間遷移時間、各スポットでの所要時間である。観光行動モデルは過去にある観光地を旅行した多くの旅行者の行動履歴から抽出する。行動履歴は、旅行者が、いつ、どこを訪れたかを表す情報である。行動履歴には行動の日時が含まれているので、旅行者がどのような環境要因で行動したかを把握できる。以下、過去にある観光地を旅行した多くの旅行者の行動履歴を過去履歴と呼ぶ。

スポットの観光に必要な所要時間とスポット間遷移時間も、過去履歴を用いて算出する。過去履歴を利用すること

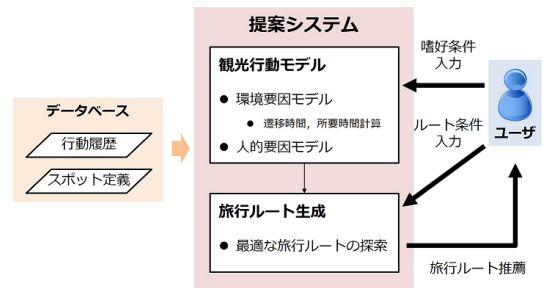


図 1 提案する旅行計画支援システムのフレームワーク
Fig. 1 Framework of travel planning support system.

で、環境要因による影響を含めた所要時間や遷移時間が得られる。

提案する旅行ルート推薦システムは、観光行動モデルが出力したスポット間遷移のスコア、スポット間遷移時間、各スポットでの所要時間を入力とし、旅行者入力を制約条件とするスポット間遷移スコアの和の最大化問題を解くことにより、最も適切な旅行ルートを旅行者に提示するシステムである。旅行者の入力は、旅行可能時間、要求スポット、除外スポット、出発・到着スポット、スポット数である。

提案システムのフレームワークを図 1 に示す。データベースとしてあらかじめ用意されるのは、過去履歴とスポット定義である。スポット定義は、ルート推薦の対象となるスポットの位置と名称からなる情報である。

3.2 過去履歴収集とスポット定義

本節では、過去履歴をどのように収集するかと、スポットをどのように定義するかを述べる。

過去履歴を、写真共有サイトにアップロードされた位置情報付き写真から抽出する。旅行者の旅行者 ID ごとに撮影時刻順に位置情報付き写真を並べたものを、その旅行者の写真撮影履歴と呼ぶ [3]。写真撮影履歴は、旅行者の行動履歴と見なすことができる。写真撮影履歴を蓄積したものを過去履歴として用いる。

観光行動のモデル化にあたり、写真撮影履歴における撮影位置を訪問場所に対応付けることで、旅行者の観光行動を訪問場所間の遷移として記述する。撮影位置と訪問場所を対応付けるには、訪問場所をある程度の広さを持った地域として扱う必要があるため、スポットの概念を導入する。スポットの設定方法として、人手で与える方法とクラスタリングで自動抽出する方法 [3], [13] がある。本研究では手動でスポットを定義しており、スポットに対応する領域を地図平面上の多角形領域として指定している。

4. 観光行動モデル

4.1 環境要因モデルと人的要因モデル

本節では観光行動モデルの構築について述べる。観光行動において、スポットに対する嗜好は旅行者ごとに異なる

が、環境要因ごとに訪れやすいスポットの傾向は旅行者に依存しないと考えられる。そこで、両者を独立と仮定し、それぞれを人的要因モデルと環境要因モデルに分けてモデル化する。環境要因モデルでは、訪問したスポットの地理的属性や天候、季節、時間帯といった地域の環境要因によって変動するスポット間遷移の傾向を記述する。また、スポット間遷移時間とスポットの所要時間も算出する。人的要因モデルでは、旅行者の嗜好によって影響されるスポットの訪問傾向を記述する。

4.2 環境要因モデル

環境要因モデルとして、観光行動に影響を与える特定の遷移影響因に注目し、環境要因モデルに組み込んだモデル [3], [4], [5] が提案されている。本研究ではスポットごとに遷移に影響する環境要因は異なると考え、スポット間遷移に影響を与える環境要因を列挙し、それらの環境要因の集合ごとに遷移先の条件付き確率を計算する。その際、遷移元のスポットごとに最も条件付き確率が高くなる集合が異なるため、過去履歴から条件付き確率を計算し、遷移元スポットごとに最も条件付き確率が高くなる環境要因を探索し、得られた環境要因の集合をモデルに組み込む。

4.2.1 環境要因モデル

モデルに組み込む候補となる環境要因の集合を \mathcal{F} とし、1つの環境要因を $f \in \mathcal{F}$ で表す。提案する環境要因モデルには、 \mathcal{F} から環境要因をいくつか選んだ環境要因部分集合 $F \subset \mathcal{F}$ を組み込む。考慮する環境要因部分集合 F は遷移元のスポットごとに異なる。環境要因はそのときどきに応じて、日時など何らかの具体値を持つ。扱いやすくするため、具体値を環境要因 f の有限個の状態のうちの一つにマッピングし、これをマッピング状態 d_f とする。具体値に対応付けるマッピング状態が何通りの値をとりうるかを環境要因 f の細分数 b_f と呼ぶ。なお、時刻や角度のような連続値は均等に細分して対応付ける。さらに、環境要因部分集合 F の各要素 f のマッピング状態 d_f をまとめたものを、マッピング状態集合 D_F と呼ぶ。

考慮する環境要因部分集合が F 、それに対するマッピング状態集合が D_F であるとき、スポット s_i にいる旅行者がスポット s_{i+1} へ遷移する確率を $P(s_{i+1} | s_i, D_F)$ と表す。ベイズの定理より、

$$\begin{aligned} P(s_{i+1} | s_i, D_F) &= \frac{P(s_i, s_{i+1}, D_F)}{P(s_i, D_F)} \\ &= \frac{P(s_i)P(s_{i+1} | s_i)P(D_F | s_i, s_{i+1})}{P(s_i, D_F)} \end{aligned} \quad (1)$$

である。 $P(s_i)$ と $P(s_i, D_F)$ は s_{i+1} と独立なので、定数 $K(s_i, D_F)$ を用いると、

$$\begin{aligned} P(s_{i+1} | s_i, D_F) &= K(s_i, D_F)P(s_{i+1} | s_i)P(D_F | s_i, s_{i+1}) \end{aligned} \quad (2)$$

となる。 $P(D_F | s_i, s_{i+1})$ は、各環境要因のマッピング状態 $d_f \in D_F$ の独立性を仮定すると、

$$P(D_F | s_i, s_{i+1}) = \prod_{d_f \in D_F} P(d_f | s_i, s_{i+1}) \quad (3)$$

となる。確率 $P(s_{i+1} | s_i)$, $P(d_f | s_i, s_{i+1})$ は以下の最尤推定式で求められる。

$$P(s_{i+1} | s_i) = \frac{N(s_i, s_{i+1})}{\sum_{s \in S} N(s_i, s)} \quad (4)$$

$$P(d_f | s_i, s_{i+1}) = \frac{N(s_i, s_{i+1}, d_f)}{N(s_i, s_{i+1})} \quad (5)$$

ここで、 $N(s_i, s_{i+1})$ は学習データ中において旅行者がスポット s_i の直後にスポット s_{i+1} を訪れた回数、 $N(s_i, s_{i+1}, d_f)$ は環境要因 f のマッピング状態が d_f であるときに、スポット s_i の直後にスポット s_{i+1} を訪れた回数、 S は全スポットの集合である。ただし、データ数が少ないことによる影響を緩和するため、additive smoothing [14] を行う。このスムージングにより、サンプルに現れない事象の出現確率がゼロになってしまうことを防ぐ。additive smoothing を用いると、式 (4), (5) は以下のように変形される。

$$P(s_{i+1} | s_i) = \frac{N(s_i, s_{i+1}) + 1}{\sum_{s \in S} N(s_i, s) + |S|} \quad (6)$$

$$P(d_f | s_i, s_{i+1}) = \frac{N(s_i, s_{i+1}, d_f) + 1}{N(s_i, s_{i+1}) + b_f} \quad (7)$$

ここで、 $|S|$ はスポット集合 S の要素数を表す。式 (2), (3), (6), (7) より、 $P(s_{i+1} | s_i, D_F)$ を求めることができる。これにより、マッピング状態集合が D_F であるときにスポット s_i を訪れている旅行者は、下式で表されるスポット s^* に遷移すると予測することができる。

$$s^* = \arg \max_{s \in S} P(s | s_i, D_F) \quad (8)$$

4.2.2 スポット間遷移に影響を与える環境要因

本項では、複数ある環境要因 f がそれぞれどのように観光行動に影響を与えうるかについて議論し、モデルに組み込む、すなわち 4.2.1 項において定義した環境要因集合 \mathcal{F} に含める候補について述べる。

本研究では、環境要因として、天候、季節、時間帯、訪問したスポットを候補とする。これらの候補は、入手の容易さとスポット間遷移への影響の程度を考慮して決定した。季節や時間帯、天気は旅行者の行動に影響を与える要因として一般に認識されており、先行研究でも利用されている [4], [5]。また、ある旅行者は互いに比較的近いスポットのみを訪れ、他の旅行者は遠く離れたスポットにも訪れるというような、過去に訪れたスポットの地理的な性質に依存した移動の傾向が観察されており、旅行者が訪れたス

ポットの地理的な性質も環境要因の1つといえる。この仮説に基づき、直前の訪問スポットから地域、方角、距離を抽出し、組み込むべき環境要因の候補とする。

- **季節**：季節によるスポットの魅力度の変化を示すため、旅行時の季節をモデルに組み込む。
- **時間帯**：時間帯による観光行動の変化を示すため、訪問の時間帯をモデルに組み込む。
- **天気**：スポットのタイプによる降雨時の行動変化を示すため、降水量を天気としてモデルに組み込む。
- **地域**：スポットが属する大まかな地域を示すため、スポット集合をクラスタリングし、直前に訪れたスポットが属するクラスタをモデルに組み込む。
- **方角**：現在のスポットから直前に訪れたスポットへの方角をモデルに組み込む。旅行者は移動効率の良いルートを選択しやすく、Uターンするようなルートはあまり選択しないという仮説に基づく。
- **距離**：現在のスポットと直前に訪れたスポットとの距離をモデルに組み込む。近いスポットに遷移した旅行者は次も近いスポットに遷移しやすく、遠いスポットに遷移した旅行者は次も遠いスポットに遷移しやすいという仮説に基づく。

4.2.3 有効な環境要因の探索

スポット間遷移ごとに有効な環境要因 f は異なる。また、同じ環境要因でも細分数 b_f が異なればモデルへの組み込み方も異なるものになることから、各環境要因における細分数 b_f を適切に定める必要がある。そこで、最も予測精度が高い環境要因部分集合 F を見つけるための探索をスポットごとに行う。予測精度は、環境要因部分集合 F を組み込んだモデルを用いて、旅行者が次に訪れるスポットを予測したときの条件付き確率で評価する。

有効な環境要因の探索問題を定式化する。モデルに組み込む候補となる環境要因の集合 \mathcal{F} において、同じ環境要因名がついていても、細分数 b_f が異なれば違う環境要因として扱う。 \mathcal{F} から環境要因をいくつか選んだ環境要因部分集合 $F \subset \mathcal{F}$ を探索の対象とする。ただし、 F は同じ環境要因名を持つ環境要因を複数含むことはないとする。この条件下で、モデルに組み込んだ場合の予測精度が最大となる環境要因部分集合 F を見つけ出すことが、探索の目的である。

環境要因の候補の集合 \mathcal{F} の中から、どの環境要因を選ぶかをすべて試す全探索を行うと、環境要因数を m 、各環境要因の細分数のパターンを $e_1 \dots e_m$ ($i = 1, \dots, m$) 個とすると、組合せの数は $\prod_{i=1}^m (e_i + 1)$ 通りとなり、これが探索範囲となる。そこで、現実的な時間内で探索を終わらせるため、探索空間を削減した探索を行う。これは、モデルに組み込むと予測精度が向上する環境要因を見当付けすることで、探索の対象を絞る手法である。環境要因の候補の一部

Algorithm 1 探索空間を削減した探索における再帰関数 $func$

Recursive Function $func$ for Search in Search Space Reduction

```

1: function  $func$ (要因の候補集合  $F$ )
2:   if  $F$  の要素数が 1 以下 then
3:     return  $F$ 
4:   end if
5:   候補集合  $F$  を二分して  $F_L$  と  $F_R$  とする
6:    $F' \leftarrow func(F_L) \cup func(F_R)$ 
7:    $F'$  に含まれる要因のみを対象にして、要因の選び方を
   すべて試す探索を行う
8:    $F_1, \dots, F_{N_{brute}} \leftarrow$  上記の探索で得られた予測精度が大
   きい要因集合上位  $N_{brute}$  件
9:   return  $F_1 \cup \dots \cup F_{N_{brute}}$ 
10: end function

```

のみを対象にした探索を繰り返すことにより、探索空間を小さくする。

探索空間を削減した探索は、Algorithm 1 で示される再帰関数 $func$ によって処理される。関数 $func$ は、環境要因の候補集合 F を引数として受け取る。まず、 F の要素数が 1 以下である場合、 F を返す。そうでない場合、 F の要素を F_L と F_R に二分する。 F_L と F_R をそれぞれ引数として、関数 $func$ を再帰的に実行し、その結果得られた要因の集合 $func(F_L) \cup func(F_R)$ を F' とする。続いて、 F' に含まれる環境要因のみを対象にして、環境要因の選び方をすべて試す探索を行う。ここでは、 F' の部分集合のすべての組合せに対して予測精度を計算する。予測精度の推定は式 (8) に基づいて算出した遷移先と、学習データとして持っている正解の遷移先とを用いて行う。

この探索で得られた予測精度が大きい環境要因部分集合上位 N_{brute} 件を、それぞれ $F_1, \dots, F_{N_{brute}}$ とする。最後に、 $F_1 \cup \dots \cup F_{N_{brute}}$ を関数 $func$ の結果として返す。 $F_1 \cup \dots \cup F_{N_{brute}}$ に含まれる環境要因は、モデルに組み込むと予測精度が向上すると見当付けされた環境要因の集合といえる。また、Algorithm 1 の手順 8 における要因集合の絞り込みによって探索範囲を絞り込むことができる。関数 $func$ は要素数が 1 以下となるまで再帰的に呼び出されるため、要因候補数に比例した回数（候補数を n とすると $2n - 1$ 回）だけ呼び出される。得られた環境要因の集合を、4.2.1 項における環境要因部分集合 F として用いる。

4.3 スポット所要時間とスポット間遷移時間の算出

本節では、スポットにおける観光に必要な所要時間と、スポット間の遷移に必要な時間を求める。写真共有サイトから得られる過去履歴から、スポットへ到着した時刻は、そのスポットで最初に写真を撮影した時刻、スポットを出

発した時刻は、そのスポットで最後に写真を撮影した時刻を抽出し、その差として所要時間を推定する。スポット間遷移時間も過去履歴から抽出する。過去履歴が利用できない場合、地図サービスの経路案内を用いる。ただし、経路案内では環境要因による遷移時間の変動を把握できないので、過去履歴から得られる遷移時間を用いて補正することで、環境要因による変動を考慮したスポット間遷移時間を求める。スポット所要時間とスポット間遷移時間はともに平均値を用いている。個人差が大きい場合についての処理は特に行っていない。

4.4 人的要因モデル

旅行者が入力した要求スポットには、旅行者の嗜好が反映されていると考えられる。そこで、要求スポットの情報をを用いて旅行者の各スポットへの嗜好を推定するため、協調フィルタリング [6] を用いる。協調フィルタリングでは、似たような旅行者は同じアイテムに対して似たような評価を与えるという仮定の下、対象旅行者の各アイテムに対する評価値を推定する。

協調フィルタリングにより、スポットに対する旅行者の嗜好を推定するため、過去履歴を用いる。過去にある観光地を旅行した多くの旅行者全体の集合を U 、旅行者 $u \in U$ が訪れたスポットの集合を S_u とする。また、旅行者 u がスポット s を訪れたかどうかを表すバイナリ値 $a_{u,s}$ を以下で定義する。

$$a_{u,s} = \begin{cases} 1 & \text{if } s \in S_u \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

協調フィルタリングを適用するためには、観光行動モデルの対象となる旅行者を u^* としたとき、対象旅行者 u^* と他の旅行者 u がどれほど似ているかを表す類似度 $w_{u^*,u}$ を定める必要がある。本研究では、類似度 $w_{u^*,u}$ を、旅行者 u が訪れたスポット S_u と対象旅行者 u^* の要求スポット $S_{preferred}$ をそれぞれ集合ととらえ、その集合間がどの程度似ているかによって定める。集合間の類似度を測る尺度として、ジャッカード係数を用いる。ジャッカード係数を用いると、 $w_{u^*,u}$ は以下のように求まる。

$$w_{u^*,u} = \frac{|S_{preferred} \cap S_u|}{|S_{preferred} \cup S_u|} \quad (10)$$

ただし、 $|S_{preferred} \cap S_u|$ は、 $S_{preferred}$ と S_u に共通して含まれるスポットの数であり、 $|S_{preferred} \cup S_u|$ は、 $S_{preferred}$ と S_u の少なくともどちらか一方に含まれるスポットの数である。

協調フィルタリングにより、対象旅行者にとっての各スポット s に対するスコア v_s を以下で求める。

$$v'_s = \begin{cases} 0 & \text{if } s \in S_{preferred} \\ \sum_{u \in U} w_{u^*,u} a_{u,s} & \text{otherwise} \end{cases} \quad (11)$$

$$v_s = \frac{v'_s}{\sum_{s \in S} v'_s} \quad (12)$$

式 (11) のように、 $s \in S_{preferred}$ の場合、スポット s は必ず旅行ルートに含まれるので、スコアを 0 とする。 $s \notin S_{preferred}$ の場合、各旅行者 $u \in U$ に対して、対象旅行者 u^* との類似度 $w_{u^*,u}$ を求め、それを重みとして $a_{u,s}$ を足し合わせた値をスコア v'_s とする。ただし、式 (12) で示したとおり、すべてのスポットに対するスコアの和が 1 になるように正規化したものを v_s とする。これにより求められたスコア v_s が、スポット s に対する対象旅行者 u^* の嗜好を表す人的要因モデルとなる。

5. 旅行ルート生成

5.1 定式化

4 章で述べた観光行動モデルの出力と旅行者の入力を用いて、旅行ルートをスコア付けし、スコアが最大となる旅行ルートを提示する。観光行動モデルからの出力は、スポット間遷移のスコア、スポット間遷移時間、スポット所要時間である。旅行者の入力は、旅行可能時間、要求スポット、除外スポット、出発・到着スポット、スポット数である。

旅行ルートを生成する問題を、制約付き最適化問題として定式化する。旅行ルート $R = (s_1, \dots, s_{|R|})$ に対する目的関数 g を以下で定義する。

$$g(R) = (1 - \alpha) \sum_{i=1}^{|R|-1} P(s_{i+1} | s_i, D_{F_i}) + \alpha \sum_{i=1}^{|R|} v_{s_i} \quad (13)$$

ここで、 D_{F_i} はスポット s_i での環境要因集合 F_i に対するマッピング状態集合である。また、 α は 0 以上 1 以下の値をとるパラメータである。式 (13) において、第 1 項は環境要因モデルから得られるスコアに対応し、第 2 項は人的要因モデルから得られるスコアに対応する。 α を小さく設定すると環境要因モデルの影響が大きくなり、 α を大きく設定すると人的要因モデルの影響が大きくなる。

本研究では、以下に列挙する制約条件を満たし、目的関数 (式 (13)) を最大化する旅行ルートを探索する。

- スポット所要時間とスポット間遷移時間の和が、入力旅行可能時間の範囲内である。
- 要求スポットに含まれるスポットをすべて含む。
- 除外スポットに含まれるスポットを 1 つも含まない。
- 出発スポットと最終スポットが、ルートの最初のスポットと最後のスポットに等しい。
- 入力されたスポット数がルートに含まれるスポット数と等しい。
- 同じスポットを 2 回以上含まない。ただし、入力において出発スポットと到着スポットが同じであることは許容する。

5.2 最適な旅行ルートの探索

5.1 節で述べた制約条件を満たす旅行ルートの中で、式 (13) で表される目的関数が最大となる旅行ルートを探索する。この問題は NP 困難である。このため本研究では、メタヒューリスティックな探索手法である焼きなまし法を用いて近似的に最適解を探索することで、ルート探索の計算時間を許容可能な範囲に抑える。

6. 実験

提案手法の性能を評価するため、実データを用いた実験を行う。評価の対象は環境要因モデルとルート作成の性能とする。前者は従来手法である単純マルコフモデルと提案手法での遷移先スポットの予測結果を比較する。後者は実際の行動履歴において旅行者がたどったルートとどれだけ類似しているかで評価する。人的要因モデルは協調フィルタリングを用いているため、単独での評価は行わない。

6.1 データ

写真共有ウェブサービス Flickr を用いて、2004 年 1 月から 2014 年 3 月までに京都市で撮影された位置情報付き写真 436,031 枚を収集した。文献 [10] で用いられているフィルタリング手法を用いて、位置情報が不正確な写真を除去した。京都市における主要観光地 30 カ所をスポットとした。各スポットは地図平面上の多角形として定義されており、地理空間アプリケーション ArcGIS の空間検索機能により、位置情報付き写真をスポットに対応付けた。この処理で、撮影位置がスポットに含まれない写真は除去した。最終的に用いた写真枚数は 14,710 枚である。各旅行者の写真撮影履歴における撮影位置をスポットに対応付けた後、スポットを 2 カ所以下しか訪れていない旅行者を除去した結果、旅行者数は 412 となった。

6.2 環境要因モデルの評価

提案する環境要因モデルの性能を評価するために、精度の高いスポット間遷移予測が可能であるかを調査する。予測精度を、旅行者が次に訪れるスポットを正しく予測できた割合として定義する。Leave-one-out 交差検定を行い、平均の予測精度が最も高くなる環境要因集合をスポットごとに探索する。探索の対象となる環境要因について表 1 に示す。

提案手法はスポットごとに環境要因を探索しているので、この性能を評価するため、各スポットで有効な環境要因は同じであると仮定し、全スポットで共通の要因を用いて探索する手法を用意する。ここで、探索空間を削減した探索においてモデルに組み込む環境要因部分集合の数を決めるパラメータ N_{brute} は 5 とした。既存手法との比較のため、環境要因を何も用いない単純マルコフモデルと、現在のスポットに依存せず、単純に多くの旅行者が訪問している人

気スポットに遷移すると予測する多項分布モデルでも探索を行った。表 2 にこれらの予測精度を示す。

表 3 に探索空間を削減した探索によって選ばれた環境要因一覧を示す。表 3 の結果から、スポットによって有効な環境要因は異なることが分かる。なお、仁和寺など 10 スポットでは、有効な環境要因がないという結果になっている。

表 1 探索の対象となる各環境要因についての詳細

Table 1 Details of environmental factors.

分類	遷移影響因名	対応付け方法	細分数 b_f
環境要因	地域	k-means クラスタリング	2,3,4,5,6
	方角	東を 0 度として、角度を均等に細分	2,3,4,5,6
	距離	0~max を均等に細分	2,3,4,5,6
	季節	1~12 月を均等に細分	2,3,4,6
	時間帯	0~23 時を均等に細分	2,3,4,6
	天気	降水量を均等に細分	2,3,4,5

表 2 旅行者が次に訪れるスポットの予測精度

Table 2 Prediction accuracy of spots tourists will visit next.

手法	予測精度
スポットごとに異なる要因を用いて探索	38.3%
全スポットで共通した要因を用いて探索	32.5%
単純マルコフ	27.9%
多項分布モデル	18.2%

表 3 スポットごとの環境要因一覧。括弧内は細分数 b_f

Table 3 Environmental factor sets for each spot.

スポット名称	環境要因部分集合
清水寺	季節 (4), 天気 (4)
京都駅	時間帯 (6)
金閣寺	時間帯 (3), 天気 (3), 距離 (5)
伏見稲荷大社	時間帯 (3), 季節 (6) 天気 (3), 方角 (6), 距離 (2)
八坂神社	季節 (2)
渡月橋	時間帯 (6), 天気 (4)
銀閣寺	時間帯 (6), 季節 (2), 天気 (4)
南禅寺	方角 (3), 距離 (5)
二条城	天気 (3), 距離 (6)
京都御所	時間帯 (3), 季節 (4), 方角 (3)
龍安寺	方角 (5)
平安神宮	時間帯 (3), 天気 (3), 地域 (6)
知恩院	季節 (6), 方角 (6)
北野天満宮	天気 (2), 地域 (5)
高台寺	天気 (3), 方角 (3), 距離 (5)
東福寺	季節 (4), 天気 (4)
京都タワー	時間帯 (4), 季節 (4)
大徳寺	時間帯 (4), 天気 (3)
法然院	季節 (4), 天気 (2)
青蓮院	方角 (6)
仁和寺など 10 カ所	なし

る。これは、これらのスポットからの遷移に関しては、モデルに環境要因を組み込んでも予測精度が向上しなかったことを意味する。

実験結果から、提案モデルは、マルコフモデルや多項分布モデルに対し、様々な要因をモデルに組み込むことで、予測精度が向上していることが分かる。また、スポットごとに使用する環境要因を変えて探索した場合の方が、共通の要因を用いた探索に比べて、予測精度が高くなっている。この結果はスポットによって有効な要因が異なることを示している。

6.3 スポット所要時間とスポット間遷移時間

4.3 節で述べた手法によって、過去履歴からスポット所要時間とスポット間遷移時間を算出した。旅行時間に影響を与える環境要因としては、「季節」の細分数が4であるもののみを用いた。経路案内で得られるスポット間遷移時間 $\tau(s_i \rightarrow s_{i+1})$ は、Yahoo!路線情報^{*1}により取得した。

6.4 旅行ルート生成の評価

6.1 節において得た 412 名の過去履歴に Leave-one-out 交差検定を行い、提案手法による旅行ルート生成を定量的に評価する。Flickr のデータは提案システムへの入力を前提としていないので、まず、Flickr から抽出した過去の旅行者がシステムを使ったと仮定したとき、システムに入力するであろう情報を過去履歴から抽出する。その入力をもとに旅行ルートを生成し、過去履歴において旅行者がたどったルートとの類似度を計算する。旅行ルート生成の際は、対象となる旅行者以外の過去履歴を用いて観光行動モデルを構築する。提案手法は、環境要因モデルと人的要因モデルを組み合わせた手法であるため、そのままでは既存手法との比較が難しい。そこで本実験では、観光行動モデルを使わず旅行ルートを生成する 2 つの手法（ランダム、最短遷移時間）と比較した。各手法の概要は以下のとおりである。

- ランダム

入力の条件を満たす旅行ルートをランダムに生成する。

- 最短遷移時間

スポット間遷移時間の合計が最小となるルートを生成する。旅行ルートの探索は提案手法と同様に行う。

システムへの入力は、以下のように抽出する。旅行開始日時は、旅行者が最初に写真を撮影した日時とする。旅行可能時間は十分長い時間、スポット数 n_{spot} を旅行者が訪れたスポットの数とする。旅行者は、旅行中に訪れたすべてのスポットで写真を撮影するとは限らない。また、途中でスポットとして定義されていない場所を訪れている場合

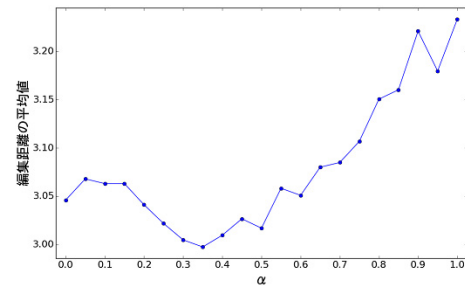


図 2 パラメータ α と編集距離の平均値の関係

Fig. 2 Relationship with α and means of editorial distances.

表 4 各手法における編集距離の平均値の比較

Table 4 Comparison of editorial distances calculated by each method.

手法	編集距離の平均値
提案手法 ($\alpha = 0.35$)	2.998
ランダム	3.466
最短遷移時間	3.344

もある。このため、過去履歴から旅行可能時間を抽出し、システムへの入力として設定すると、過去履歴において旅行者が訪れたスポットの数よりも多くのスポットを訪れる旅行ルートが生成されるため、このような設定とした。要求スポットは、旅行者が最初に訪れたスポットのみとする。除外スポットと出発/終着スポットは入力しないものとする。生成されるルートと、旅行者がたどったルートがどれほど類似しているかを測る評価尺度として、編集距離 [15] を用いる。得られた編集距離の平均値により、旅行ルート生成手法を評価する。本研究で提案する旅行ルート生成の結果は、目的関数におけるパラメータ α に影響を受ける。 α を 0.0 から 1.0 まで 0.05 刻みで動かしたときの、編集距離の平均値を図 2 に示す。ただし、焼きなまし法におけるループ回数は 10,000 とした。実験の結果、 α が 0.35 のとき編集距離の平均値が最小となったので、以降の実験では α を 0.35 とする。この結果は、スポット間遷移モデルとユーザモデルを組み合わせることが、旅行ルート生成の性能向上につながることを意味する。

提案手法と比較手法との比較結果を表 4 に示す。結果より、提案手法は、ランダムや最短遷移時間と比べて、過去履歴から得られた観光行動モデルを活用することで、実際に旅行者がたどるルートに近いルートを生成することができているといえる。

なお、旅行ルートを探索する計算時間を測定したところ、平均の計算時間は 0.34 秒であった。旅行ルートを探索するプログラムは C++ で実装し、Google Compute Engine の仮想マシン (CPU Xeon 2.50 GHz, メモリ 7.5 GB) 上で実行した。

*1 <http://transit.loco.yahoo.co.jp/>

7. 結論

本研究は、スポット間遷移ごとに、スポットごとに最も有効な環境要因の組合せを用いる環境要因モデルと、協調フィルタリングを用いて旅行者の嗜好を記述する人的要因モデルを融合させた観光行動モデルを用いて、旅行ルートを推薦する手法を提案した

実験では、写真共有サイトから得られる過去履歴を用いて提案手法の評価を行った。環境要因モデルの評価では、旅行者が次に訪れるスポットを予測する精度において、マルコフモデルが27.9%であるのに対して提案モデルは38.3%であった。これにより、旅行者がそれまでに訪れたスポットや環境要因を考慮した環境要因モデルが、予測精度の向上に有用であることを確認した。旅行ルート生成の評価では、生成された旅行ルートと、過去の旅行者がたどった旅行ルートがどれほど類似するかを評価した。この評価により、環境要因モデルと人的要因モデルを組み合わせることが、旅行ルート生成の性能向上に有用であることを確認した。

今後の課題としては観光行動モデルの改善があげられる。環境要因モデルに組み込む環境要因を増やしたり、環境要因をマッピング状態へ対応付ける方法を改善したりすることで、より正確にスポット間遷移を記述できると考えられる。また、本研究では人的要因モデルとして要求スポットから旅行者の嗜好を抽出する手法を提案している。旅行者の負担にならない範囲内で、旅行者の嗜好に関する情報を取得する方法を検討する。また、Flickr などから抽出した位置情報付き写真を利用すれば、京都以外の観光地での本モデルの適用は可能である。モデルの有効性を検証するためにも、他の観光地への適用も検討したい。

参考文献

- [1] Ashbrook, D. and Starner, T.: Using GPS to learn significant locations and predict movement across multiple users, *Personal and Ubiquitous Computing*, Vol.7, No.5, pp.275-286 (2003).
- [2] Zheng, Y.-T., Li, Y., Zha, Z.-J. and Chua, T.-S.: Mining Travel Patterns from GPS-Tagged Photos, *Proc. 17th International Conference on Advances in Multimedia Modeling*, Vol.1, pp.262-272 (2011).
- [3] Kurashima, T., Iwata, T., Irie, G. and Fujimura, K.: Travel route recommendation using geotagged photos, *Knowledge and Information Systems*, Vol.37, No.1, pp.37-60 (2013).
- [4] 山崎俊彦, アンドリュウギャラガー, ツーハンチェン, 相澤清晴: 季節・時刻を考慮した大規模位置履歴からの都市内旅行予測・推薦, *電子情報通信学会論文誌*, Vol.J97-D, No.9, pp.1437-1444 (2014).
- [5] Canneyt, S.V., Schockaert, S., Laere, O.V. and Dhoedt, B.: Time-dependent recommendation of tourist attractions using Flickr, *23rd Benelux Conference on Artificial Intelligence (BNAIC)*, pp.255-262 (2011).
- [6] 田中克己, 角谷和俊: 情報推薦システム入門—理論と実

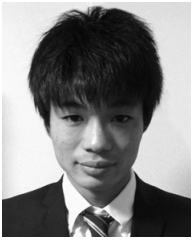
践 (2012).

- [7] Maruyama, A., Shibata, N., Murata, Y., Yasumoto, K. and Ito, M.: A personal tourism navigation system to support traveling multiple destinations with time restrictions, *18th International Conference on Advanced Information Networking and Applications, 2004, AINA 2004*, Vol.2, pp.18-21, IEEE (2004).
- [8] Vansteenwegen, P., Souffriau, W., Berghe, G.V. and Oudheusden, D.V.: The city trip planner: An expert system for tourists, *Expert Systems with Applications*, Vol.38, No.6, pp.6540-6546 (2011).
- [9] Kurata, Y. and Hara, T.: CT-Planner4: Toward a more user-friendly interactive day-tour planner, *Information and Communication Technologies in Tourism 2014*, pp.73-86, Springer (2013).
- [10] Okuyama, K. and Yanai, K.: A travel planning system based on travel trajectories extracted from a large number of geotagged photos on the Web, *The Era of Interactive Media*, pp.657-670, Springer (2013).
- [11] Arase, Y., Xie, X., Hara, T. and Nishio, S.: Mining people's trips from large scale geo-tagged photos, *Proc. International Conference on Multimedia*, pp.133-142, ACM (2010).
- [12] Lu, X., Wang, C., Yang, J.-M., Pang, Y. and Zhang, L.: Photo2trip: Generating travel routes from geo-tagged photos for trip planning, *Proc. International Conference on Multimedia*, pp.143-152, ACM (2010).
- [13] Crandall, D.J., Backstrom, L., Huttenlocher, D. and Kleinberg, J.: Mapping the world's photos, *Proc. 18th International Conference on World Wide Web*, pp.761-770, ACM (2009).
- [14] Chen, S.F. and Goodman, J.: An empirical study of smoothing techniques for language modeling, *Proc. 34th Annual Meeting on Association for Computational Linguistics*, pp.310-318, Association for Computational Linguistics (1996).
- [15] Navarro, G.: A guided tour to approximate string matching, *ACM Computing Surveys (CSUR)*, Vol.33, No.1, pp.31-88 (2001).



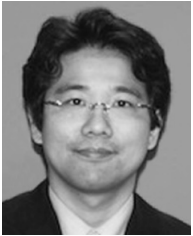
笠原 秀一 (正会員)

平成18年青山学院大学大学院国際マネジメント研究科修了。株式会社ウィルコム等で新規事業開発等に従事。平成25年京都大学大学院情報学研究科博士課程後期指導認定退学。観光の情報化に関する研究に従事。電子情報通信学会、観光情報学会、IFITT各会員。経営管理修士(専門職)。博士(情報学)。



田村 和範

平成 27 年京都大学大学院工学系研究科情報学専攻修了。在学中は旅行者行動のモデル化に関する研究に従事。



飯山 将晃 (正会員)

平成 10 年京都大学工学部情報工学科卒業。平成 15 年同大学大学院情報学研究科博士後期研究指導認定退学。同年同大学学術情報メディアセンター助手。平成 18 年同大学経済学研究科講師。平成 22 年同准教授。コンピュータビジョン・3 次元データ処理の研究に従事。博士 (情報学)。IEEE, 電子情報通信学会各会員。



椋木 雅之 (正会員)

平成 3 年京都大学工学部情報工学科卒業。平成 8 年同大学大学院工学研究科博士後期研究指導認定退学。同年同大学工学部助手。平成 12 年同大学総合情報メディアセンター助手。平成 14 年広島市立大学情報科学部助教授, 平成 21 年京都大学学術情報メディアセンター准教授, 平成 27 年宮崎大学工学教育研究部情報システム工学科教授, 現在に至る。博士 (工学)。画像認識, コンピュータビジョン, 映像メディア処理の研究に従事。電子情報通信学会会員。



美濃 導彦 (正会員)

昭和 53 年京都大学工学部情報工学科卒業。昭和 58 年同大学大学院博士課程修了。同年同大学工学部助手, 昭和 62~63 年マサチューセッツ州立大学客員研究員, 平成元年京都大学工学部附属高度情報開発実験施設助教授, 平成 7 年同教授, 平成 9 年京都大学総合情報メディアセンター教授, 平成 14 年京都大学学術情報メディアセンター教授。画像処理, 人工知能, 知的コミュニケーション関係の研究に従事。工学博士。IEEE, ACM, 電子情報通信学会, 画像電子学会, 日本ロボット学会各会員。電子情報通信学会フェロー, 日本バーチャルリアリティ学会フェロー。