

[Original article]

(2014年10月4日 Accepted)

旅行中の観光目的の変化のモデル化に関する一考察

前田 康成, 後藤 文太朗, 升井 洋志, 榊井 文人, 亀丸 俊一, 鈴木 正清

北見工業大学・情報システム工学科

要約: 本研究では, 旅行中の観光客の目的の変化を隠れマルコフモデルを用いてモデル化する. さらに, 観光客の目的を統計的決定理論に基づいて推定する方法を提案する. 観光目的の推定に関する提案方法では, 推定誤り率をベイズ基準のもとで最小化する. また, 観光客が次に訪問する観光施設を統計的決定理論に基づいて推薦する方法を提案する. 推薦に関する提案方法では, 推薦誤り率をベイズ基準のもとで最小化する. 提案モデルおよび提案方法の有効性を数値計算例によって確認する.

キーワード: 観光履歴, 観光施設, 推薦方法, 隠れマルコフモデル, 統計的決定理論

A Note on Modeling for Changes of Tourist Purposes

Yasunari MAEDA, Fumitaro GOTO, Hiroshi MASUI, Fumito MASUI, Shun-ichi Kamemaru,
Masakiyo SUZUKI*Dept. of Computer Science, Kitami Institute of Technology*

Abstract: *In this research we use hidden Markov model in order to represent changes of tourist purposes. We propose an estimation method for tourist purposes based on statistical decision theory. The proposed estimation method minimizes the estimation error rate with reference to the Bayes criterion. We propose a recommendation method for visiting tourist facilities based on statistical decision theory. The proposed recommendation method minimizes the recommendation error rate with reference to the Bayes criterion. We show the effectiveness of the proposed model and methods by some numerical calculation examples.*

Keywords: *tourist record, tourist facility, recommendation method, hidden Markov model, statistical decision theory*

Yasunari MAEDA

165 Koen-cho, Kitami-shi, Hokkaido, 090-8507, Japan

Phone: +81-157-26-9328, Fax: +81-157-26-9344, E-mail: maeda@cs.kitami-it.ac.jp

1. はじめに

近年、情報工学の分野でも観光に関するさまざまな要素技術が検討されている。従来研究[1][2]では旅行者が指定した観光目的に応じた観光施設の推薦方法が検討されている。他方、従来研究[3]では、旅行者による目的の指定が不要で、旅行者の観光履歴から自動的に目的に応じた観光施設を推薦する方法を提案している。従来研究[3]では、第三者の観光履歴情報（過去の第三者の旅行において、どのような観光施設がどのような順番で訪問されたかを記録したデータベース）と旅行中の旅行者の観光履歴（それまでに訪問済の観光施設の系列（訪問順を示す情報））を比較して似ている観光履歴情報中のデータ（観光施設の訪問順を示す、観光施設の系列）を提示することによって、旅行者の旅行中の観光目的の変化に対応した観光施設の推薦を行っている。しかし、従来研究[3]では旅行者の観光目的が変化する仕組みを明確にモデル化しておらず、推薦結果に対して理論的な最適性の保証はない。

他方、多くの工学分野において確率モデルを導入してモデル化することによって、理論的な最適性の保証を実現している。例えば、自然言語処理の文書分類では確率モデルを用いないベクトル空間法では理論的な最適性を保証できないが、確率モデルを導入することによって分類を間違える確率である分類誤り率を理論的に最小化している[4]。

そこで、本研究では隠れマルコフモデル（HMM）を用いて観光目的の変化をモデル化し、次に訪問する観光施設の推薦問題を統計的決定理論[5][6][7]に基づいて定式化し、推薦を誤る確率を最小にする推薦方法を提案する。HMMは自然言語処理の形態素解析[8][9]などで利用されている確率モデルである。形態素解析では単語を観測して品詞を推定することを目的とし、品詞の推定は本研究のモデルの場合には観光目的の推定に相当する。しかし、本研究では次の訪問施設の推薦を目的とし、これは形態素解析のモデルでは次に生成される単語を予測することに相当する。統計的決定理論は統計学の視点から意思決定問題を体系化した理論で、文書分類や形態素解析の従来研究[4][8]では統計的決定理論に基づいて定式化することによって、理論的に最適性を保証している。本研究でも同様に統計的決定理論に基づいた定式化を行い、理論的に最適性を保証する。また、HMMを用いた提案モデルと提案方法の有効性を数値計算例で確認する。

なお、実際の旅行では、1人旅の場合、家族旅行の場合、友人同士のグループ旅行の場合などによって観光目的の変化の仕方も異なることが想定される。しかし、本研究では観光目的の変化の仕組みをHMMでモデル化する基礎検討に注力し、上記のような詳細な区別は行わない。また、実際の旅行では単に観光目的のみではなく、観光施設までの距離、移動手段、コスト（移動コスト、入場料など）なども考慮して次の訪問施設を決めることが想定される。しかし、本研究では上記のとおりHMMによる観光目的の変化のモデル化に注力し、その他の制約の考慮あるいはモデルの拡張については今後の課題とする。

2. 準備

本研究で用いる各種記号などの定義を行う。

θ_i^* , $\theta_j^* \in O$ は i 番目の観光目的を示し、 O , $O = \{\theta_1^*, \theta_2^*, \dots, \theta_{|O|}^*\}$ は観光目的の集合である。観光目的がHMMの状態に相当する。 $p(\theta_j^* | \theta_i^*)$ は旅行者の観光目的が θ_i^* から θ_j^* に変化する確率で、HMMの状態遷移確率に相当する。本研究では確率分布 $p(\theta_j^* | \theta_i^*)$ は既知と仮定する。 s_i , $s_j \in S$ は i 番目の観光施設を示し、 S , $S = \{s_1, s_2, \dots, s_{|S|}\}$ は観光施設の集合である。 $p(s_j | \theta_i^*)$ は観光目的 θ_i^* のもとで観光施設 s_j を訪問する確率で、 $\sum_{s_j} p(s_j | \theta_i^*) = 1$ である。観光目的を示す記号 θ_i^* は確率分布を支配する真のパラメータも兼ねており、真のパラメータは既知と仮定する。

$p(\theta_i^*)$ は観光目的の事前確率で既知である。 y^T は旅行者の観光履歴 $y_1 y_2 \dots y_T$, $y_i \in S$ を示す。 $y_i \neq y_j$, $i \neq j$ である。 x^T は観光履歴に対応する観光目的の系列 $x_1 x_2 \dots x_T$, $x_i \in O$ で未知（観測不可）である。本研

究では、観光履歴 y^T を与えられたもとで、次の訪問先の観光施設を推薦する (y_{T+1} を推定する) 方法を統計的決定理論に基づいて提案する。

3. 提案方法

3.1 観光目的の推定

本研究の主目的ではないが、参考のため観光目的の推定について本節で検討する。統計的決定理論に基づいて、観光目的の推定を定式化する。損失関数を以下のように定義する。

$$L_1(d_1(y^T), x^T, y^T, x_T) = \begin{cases} 1, & d_1(y^T) \neq x_T; \\ 0, & d_1(y^T) = x_T, \end{cases} \quad (1)$$

ただし、 $d_1(\cdot)$ は観光履歴 y^T を受け取って、 T 期目の観光目的の推定結果を返す決定関数である。損失 L_1 は T 期目の真の観光目的が x_T のもとで、観光履歴 y^T を受け取って決定関数 $d_1(\cdot)$ を用いて観光目的の推定を行った場合の 0-1 損失である。

次に、初期の真の観光目的が x_1 という条件のもとで決定関数 $d_1(\cdot)$ を用いた場合の損失 L_1 の期待値であるリスク $R_1(d_1(\cdot), x_1)$ を次式で定義する。

$$R_1(d_1(\cdot), x_1) = \frac{1}{m_1} \sum_{y^T \in S(T)} \sum_{x_2^T} p(y_1|x_1) \times \prod_{i=2}^T p(x_i|x_{i-1}) p(y_i|x_i) L_1(d_1(\cdot), x^T, y^T, x_T), \quad (2)$$

$$m_1 = \sum_{y^T \in S(T)} \sum_{x_2^T} p(y_1|x_1) \prod_{i=2}^T p(x_i|x_{i-1}) p(y_i|x_i), \quad (3)$$

$$x_2^T = x_2 x_3 \cdots x_T, \quad (4)$$

ただし、 $S(T)$ は観光施設の集合 S の異なる T 個の要素で構成される長さ T の系列の集合である。リスク R_1 は観光目的の推定を間違える確率である推定誤り率に相当する。式(2)右辺の $\sum_{x_2^T} p(y_1|x_1)$ 以降の部分では長さ

$T-1$ の系列 x_2^T の全候補に対して、

$$p(y_1|x_1) \prod_{i=2}^T p(x_i|x_{i-1}) p(y_i|x_i) L_1(d_1(\cdot), x^T, y^T, x_T)$$

の足し合わせを行っている。

初期の真の観光目的 x_1 は未知なので、観光目的の事前確率 $p(x_1)$ に対するリスクの期待値であるベイズリスク $BR_1(d_1(\cdot), p(x_1))$ を次式で定義する。

$$BR_1(d_1(\cdot), p(x_1)) = \sum_{x_1 \in O} p(x_1) R_1(d_1(\cdot), x_1). \quad (5)$$

ベイズリスク BR_1 を最小にする決定関数がベイズ基準のもとで損失 L_1 を最小にするという意味で最適な観光目的の推定方法である。観光履歴 y^T を受け取ったもとの最適な観光目的の推定は次式のようになる。

$$d_1^*(y^T) = \arg \max_{\hat{x}_T \in O} p(\hat{x}_T | y^T), \quad (6)$$

ただし、

$$p(x_i | y^i) = \frac{p(x_i | y^{i-1}) p(y_i | x_i)}{\sum_{x'} p(x' | y^{i-1}) p(y_i | x')}, \quad (7)$$

$$\begin{aligned} p(x_{i+1} | y^i) &= \sum_{x_i} p(x_i | y^i) p(x_{i+1} | x_i) \\ &= \sum_{x_i} \frac{p(x_i | y^{i-1}) p(y_i | x_i)}{\sum_{x'} p(x' | y^{i-1}) p(y_i | x')} p(x_{i+1} | x_i), \end{aligned} \quad (8)$$

$$p(x_1 | y^0) = p(x_1). \quad (9)$$

式(6)右辺の \hat{x}_T は、真の観光目的 x_T に対する推定用の変数である。式(7)右辺の分母の x' は、分子と同様の数値 $p(x' | y^{i-1}) p(y_i | x')$ を全観光目的について足し合わせるために導入した変数である。式(8)右辺の第1行では、観光目的 x_i の全候補について $p(x_i | y^i) p(x_{i+1} | x_i)$ の足し合わせを行っている。真の観光目的 x_T に対する推定用の変数である。上記の提案方法は、観光目的の推定を間違える確率である推定誤り率がベイズ基準のもとで最小になることが保証された推定方法である。

3.2 次の訪問施設の推薦

統計的決定理論に基づいて、次の訪問施設の推薦を定式化する。損失関数を以下のように定義する。

$$L_2(d_2(y^T), x^{T+1}, y^{T+1}, y_{T+1}) = \begin{cases} 1, & d_2(y^T) \neq y_{T+1}; \\ 0, & d_2(y^T) = y_{T+1}, \end{cases} \quad (10)$$

ただし, $d_2(\cdot)$ は観光履歴 y^T を受け取って, 次の訪問施設の推薦結果を返す決定関数である. 損失 L_2 は真の観光目的が系列 x^{T+1} のように遷移したもとの, 観光履歴 y^T を受け取って決定関数 $d_2(\cdot)$ を用いて次の訪問施設を推薦し, 旅行者が実際には次に施設 y_{T+1} を訪問した場合の 0-1 損失である.

次に, 初期の真の観光目的が x_1 という条件のもとで決定関数 $d_2(\cdot)$ を用いた場合の損失 L_2 の期待値であるリスク $R_2(d_2(\cdot), x_1)$ を次式で定義する.

$$R_2(d_2(\cdot), x_1) = \frac{1}{m_2} \sum_{y^T \in S(T)} \sum_{x_2^T} p(y_i | x_1) \times \prod_{i=2}^T p(x_i | x_{i-1}) p(y_i | x_i) \sum_{x_{T+1}} p(x_{T+1} | x_T) \times \sum_{y_{T+1} \in (S-y^T)} p(y_{T+1} | x_{T+1}) L_2(d_2(y^T), x^{T+1}, y_{T+1}), \quad (11)$$

$$m_2 = \sum_{y^T \in S(T)} \sum_{x_2^T} p(y_i | x_1) \prod_{i=2}^T p(x_i | x_{i-1}) p(y_i | x_i) \sum_{x_{T+1}} p(x_{T+1} | x_T) \sum_{y_{T+1} \in (S-y^T)} p(y_{T+1} | x_{T+1}), \quad (12)$$

ただし, $S - y^T$ は観光施設の集合 S 中の未訪問の施設集合である. リスク R_2 は次の訪問施設の推薦を間違える確率である推薦誤り率に相当する. 式(11)右辺の 2 行目の $\sum_{x_{T+1}} p(x_{T+1} | x_T)$ 以降の部分では観光目的 x_{T+1} の全候補に対して,

$$p(x_{T+1} | x_T) \sum_{y_{T+1} \in (S-y^T)} p(y_{T+1} | x_{T+1}) L_2(d_2(y^T), x^{T+1}, y_{T+1})$$

の足し合わせを行っている.

初期の真の観光目的 x_1 は未知なので, 観光目的の事前確率 $p(x_1)$ に対するリスクの期待値であるベイズリスク $BR_2(d_2(\cdot), p(x_1))$ を次式で定義する.

$$BR_2(d_2(\cdot), p(x_1)) = \sum_{x_1 \in O} p(x_1) R_2(d_2(\cdot), x_1). \quad (13)$$

ベイズリスク BR_2 を最小にする決定関数がベイズ基準のもとで損失 L_2 を最小にするという意味で最適な次の訪問施設の推薦方法である. 観光履歴 y^T を受け取ったもとの最適な推薦方法は次式のようになる.

$$d_2^*(y^T) = \arg \max_{\hat{y}_{T+1} \in (S-y^T)} \sum_{x_{T+1} \in O} p(x_{T+1} | y^T) p(\hat{y}_{T+1} | x_{T+1}). \quad (14)$$

上記の提案方法は, 次の訪問施設の推薦を間違える確率である推薦誤り率がベイズ基準のもとで最小になることが保証された推薦方法である.

4. 数値計算例

HMM によるモデル化と提案方法の有効性を確認するために行った数値計算例の 2 例を報告する.

4.1 オホーツク圏の例

オホーツク圏の例を説明する. 観光目的はいろいろなものが想定されるが, この数値計算例では $O = \{\theta_1^* = \text{科学}, \theta_2^* = \text{歴史}, \theta_3^* = \text{景色}\}$ とする. 本研究では, 観光目的 θ_i^* のもとで観光施設 s_j を訪問する訪問

確率 $p(s_j | \theta_i^*)$ を既知と仮定しており, 数値計算を行う

には訪問確率を何らかの方法で事前に推定しておく必要がある. この数値計算例では, 各観光施設が各観光目的の訪問対象として含まれるかどうかを著者らが主観的に判定した結果を訪問確率の推定に利用する. 想定した観光施設と各観光目的に含まれるかどうかを著者らが判定した結果を表1に示す. 表1中の観光目的の下に書かれた 0 および 1 は, 当該観光施設が当該観光目的に含まれると著者らが判定した場合に 1, 含まれないと判定した場合に 0 とした. この判定結果を利用して, 観光目的に含まれる観光施設の訪問確率が含まれない観光施設の訪問確率の 10 倍の大きさになるように訪問確率を設定した. その訪問確率を表2に示す.

表2の各観光目的のもとの訪問確率の全施設での総和が 1 にならないのは, 表に掲載するにあたり訪問確率を小数点以下第四位で四捨五入したからである. 各観光目的のもとで目的に含まれる施設の訪問確率が含まれない施設の 10 倍の確率になるように設定したが, 各観光目的に含まれる施設の数的大小によって, 訪問確率が異なるという特徴がある. つまり, 当該観光目的に含まれる施設が少ない目的のもとの訪問確率は大きく, 含まれる施設が多い目的のもとの訪問確率は小さい. 著者らの主観より, 観光目的の事前確率 $p(\theta_i^*)$ は等確率とした. また, 同様に著者らの主観より HMM での観光目的の遷移確率 $p(\theta_i^* | \theta_j^*)$ は, $i \neq j$ の場合に 0.2, $i = j$ の場合に 0.6 とした.

表1 観光施設と観光目的の関係

施設	施設名	θ_1^*	θ_2^*	θ_3^*
s_1	天都山	0	0	1
s_2	能取岬	0	0	1
s_3	能取湖	0	0	1
s_4	モヨロ貝塚	0	1	0
s_5	常呂遺跡	0	1	1
s_6	鹿鳴の滝	0	0	1
s_7	オシンコシンの滝	0	0	1
s_8	知床峠	0	0	1
s_9	美幌峠	0	0	1
s_{10}	知床五湖	0	0	1
s_{11}	北海道立北方民族博物館	0	1	0
s_{12}	オホーツク流水館	1	0	1
s_{13}	白滝遺跡	1	1	1

表2 訪問確率

施設	施設名	θ_1^*	θ_2^*	θ_3^*
s_1	天都山	0.032	0.020	0.089
s_2	能取岬	0.032	0.020	0.089
s_3	能取湖	0.032	0.020	0.089
s_4	モヨロ貝塚	0.032	0.204	0.009
s_5	常呂遺跡	0.032	0.204	0.089
s_6	鹿鳴の滝	0.032	0.020	0.089
s_7	オシンコシンの滝	0.032	0.020	0.089
s_8	知床峠	0.032	0.020	0.089
s_9	美幌峠	0.032	0.020	0.089
s_{10}	知床五湖	0.032	0.020	0.089
s_{11}	北海道立北方民族博物館	0.032	0.204	0.009
s_{12}	オホーツク流水館	0.323	0.020	0.089
s_{13}	白滝遺跡	0.323	0.204	0.089

観光履歴 $x^3 = s_{13}s_5s_3$ の場合の目的の推定結果を表3,

次の訪問施設の推薦結果を表4に示す. この観光履歴は白滝遺跡, 常呂遺跡, 能取湖という順番で観光した履歴である. 表1より, 白滝遺跡は科学, 歴史, 景色の観光目的に含まれ, 常呂遺跡は歴史と景色, 能取湖は景色のみである. よって, 表1から白滝遺跡や常呂遺跡を訪問している時の観光目的は歴史や景色の可能性が高く, 能取湖を訪問しているときは景色の可能性が高いことがわかる.

表3 $x^3 = s_{13}s_5s_3$ の場合の目的の推定

目的	目的名	固定	HMM
θ_1^*	科学	3位	3位
θ_2^*	歴史	1位	2位
θ_3^*	景色	2位	1位

表4 $x^3 = s_{13}s_5s_3$ の場合の施設の推薦

施設	施設名	固定	HMM
s_1	天都山	4位	4位
s_2	能取岬	4位	4位
s_4	モヨロ貝塚	1位	2位
s_6	鹿鳴の滝	4位	4位
s_7	オシンコシンの滝	4位	4位
s_8	知床峠	4位	4位
s_9	美幌峠	4位	4位
s_{10}	知床五湖	4位	4位
s_{11}	北海道立北方民族博物館	1位	2位
s_{12}	オホーツク流水館	3位	1位

各提案方法では候補の第1位を算出するが, 各表には参考のため, 2位以下も掲載した. 表中のHMMが前章の提案方法の結果を示す. 表中の固定は提案方法との比較のために観光目的が変化しない場合について前章同様に統計的決定理論に基づいて導出した方法の結果である. 以下で各結果について説明する.

最初に表3の固定について見る. 固定の方法では, 観光履歴に対して観光目的が固定として観光目的を推定する. よって, 固定の方法を単純に観光履歴中の多数決と解釈すると, 今回の場合, 観光履歴中の白滝遺跡, 常呂遺跡, 能取湖の3施設に共通する景色の可能性が高い. しかし, 数値計算結果では歴史が1位である. これは, 観光目的の景色に含まれる観光施設は11と多く, 歴史に含まれる観光施設は4と少ないことが影響して, 白滝遺跡と常呂遺跡の歴史のもとでの訪問確率が景色のもとでの訪問確率よりも大きいためである.

他方, 表3のHMMについて見ると, 景色が1位である. 数値計算に利用している訪問確率は固定の場合と同じである. しかし, HMMでは観光目的の変化を考慮している. 観光履歴中の最後の訪問施設の能取湖が観光目的の景色のみに含まれていることが影響して, 観光目的の変化を考慮したHMMでは景色が1位という数値計算結果になっている.

旅行中の観光目的の変化のモデル化に関する一考察

表1および表3より, 歴史の観光目的で白滝遺跡と常呂遺跡を訪問し, その後, 観光目的が歴史から景色に変化したもとで能取湖を訪問したと解釈するのが妥当だと考えられる. このように, 観光目的の推定に関する提案方法では, 観光目的の変化をHMMを用いてモデル化することによって, 観光目的の変化を考慮した観光目的の推定を実現できている.

次に訪問施設の推薦に関する表4の固定について見る. 固定では, 観光目的の推定結果として歴史が1位だったことが大きく影響して, 推薦結果では観光目的の歴史に含まれている残りの施設であるモヨロ貝塚と北海道立北方民族博物館がともに1位として推薦されている. 3位として観光目的の景色と科学に含まれるオホーツク流氷館が推薦され, 残りの観光施設は全て景色のみに含まれるのでともに4位として推薦されている.

他方, 表4のHMMでは観光目的の推定結果として景色が1位だったことが大きく影響して, 推薦結果では観光目的の景色と科学に含まれるオホーツク流氷館が1位として推薦されている. 2位以下の候補として, 観光目的の景色のみに含まれる7施設が残っている. 他の候補は歴史のみに含まれる2施設である. 表3のHMMによる観光目的の推定結果より, 能取湖訪問後では観光目的の景色の事後確率が最大で, 2番目に大きいのが歴史の事後確率である. 訪問確率については表2より, 観光目的の景色のもとでの訪問確率よりも歴史のもとでの訪問確率の方が大きい. 提案方法HMMでは式(14)により, 施設ごとに目的の事後確率と訪問確率の積を全目的で足し合わせたもの的大小判定を行っている. 今回は歴史の訪問確率の大きさが影響して, 歴史のみに含まれる2施設がともに2位として推薦されている. 固定の場合と同様に景色のみに含まれる残りの観光施設がともに4位として推薦されている.

このように, 観光施設の推薦に関する提案方法では, 観光目的の変化をHMMを用いてモデル化することによって, 観光目的の変化を考慮した観光施設の適応的な推薦が実現できている.

4.2 東京の例

東京の例を説明する. 観光目的を $O = \{\theta_1^* = \text{買物}, \theta_2^* = \text{歴史}, \theta_3^* = \text{景色}\}$ とする. 本節では訪問確率の推定に検索エンジンgooの検索結果を利用

した. 検索エンジンgooによって, 各観光施設と各観光目的のAND検索での件数を調査した. 観光目的 θ_i^* のもとで観光施設 s_j を訪問する訪問確率 $p(s_j | \theta_i^*)$ について, 観光施設 s_j と観光目的 θ_i^* のAND検索での件数と全施設でのAND検索の総和による相対頻度で事前に推定を行った. AND検索の件数を表5に, 訪問確率の推定結果を表6に示す.

表5 AND 検索の件数

施設	施設名	θ_1^*	θ_2^*	θ_3^*
s_1	アメ横	200000	63400	38600
s_2	かつば橋	40600	12800	18000
s_3	浅草仲見世	23200	19900	21100
s_4	東横のれん街	8540	3100	1070
s_5	鳩居堂	10600	4320	1130
s_6	小石川植物園	3910	12100	21100
s_7	上野恩賜公園	30600	60900	82600
s_8	大森貝塚遺跡庭園	457	1560	728
s_9	泉岳寺	12800	19100	16100
s_{10}	吉良邸跡	926	9220	2970
s_{11}	湯島聖堂	2150	10800	8380
s_{12}	六義園	10100	19700	39200

表6 訪問確率

施設	施設名	θ_1^*	θ_2^*	θ_3^*
s_1	アメ横	0.582	0.268	0.154
s_2	かつば橋	0.118	0.054	0.072
s_3	浅草仲見世	0.067	0.084	0.084
s_4	東横のれん街	0.025	0.013	0.004
s_5	鳩居堂	0.031	0.018	0.005
s_6	小石川植物園	0.011	0.051	0.084
s_7	上野恩賜公園	0.089	0.257	0.329
s_8	大森貝塚遺跡庭園	0.001	0.007	0.003
s_9	泉岳寺	0.037	0.081	0.064
s_{10}	吉良邸跡	0.003	0.039	0.012
s_{11}	湯島聖堂	0.006	0.046	0.033
s_{12}	六義園	0.029	0.083	0.156

表6の各観光目的のもとでの訪問確率の全施設での総和が1にならないのは, 表に掲載するにあたり訪問確率を小数点以下第四位で四捨五入したからである. 表5では, 各観光施設の検索件数に大きな差がある. 例え

ば、一般的にアメ横は買物が主な観光目的で、歴史ではない。他方、大森貝塚遺跡庭園は歴史が主な観光目的である。しかし、表5のアメ横と歴史のAND検索件数が6万以上であるのに対して、大森貝塚遺跡庭園と歴史のAND検索件数は1560である。そのため、表6における観光目的の歴史のもとでのアメ横の訪問確率が0.268と大きいのに対して、大森貝塚遺跡庭園は0.007と小さい。このように本節での訪問確率の推定値はインターネット上における当該施設の有名度合い（当該施設に関する記述を含むホームページ数）が大きく影響している。前節と同様に著者らの主観より、観光目的の事前確率 $p(\theta_i^*)$ は等確率、HMMでの観光目的の遷移確率 $p(\theta_j^*|\theta_i^*)$ は $i \neq j$ の場合に0.2、 $i = j$ の場合に0.6とした。

観光履歴 $x^5 = s_8 s_9 s_{10} s_3 s_2$ の場合の目的の推定結果を表7、次の訪問施設の推薦結果を表8に示す。この観光履歴は大森貝塚遺跡庭園、泉岳寺、吉良邸跡、浅草仲見世、かつば橋という順番で観光した履歴である。表5の各施設のAND検索件数の最大値の観光目的のみに着目すると、観光目的の歴史のもとで大森貝塚遺跡庭園、泉岳寺、吉良邸跡を訪問し、観光目的の買物もとで浅草仲見世、かつば橋を訪問した可能性が高い。表6の訪問確率についてもほぼ同様であるが、浅草仲見世の訪問確率は歴史および景色のもとで最大である。よって、表5および表6の両方を考慮すると、浅草仲見世を訪問した際の観光目的の推定は難しい。

表7 $x^5 = s_8 s_9 s_{10} s_3 s_2$ の場合の目的の推定

目的	目的名	固定	HMM
θ_1^*	買物	3位	1位
θ_2^*	歴史	1位	3位
θ_3^*	景色	2位	2位

表8 $x^5 = s_8 s_9 s_{10} s_3 s_2$ の場合の施設の推薦

施設	施設名	固定	HMM
s_1	アメ横	2位	1位
s_4	東横のれん街	7位	7位
s_5	鳩居堂	6位	6位
s_6	小石川植物園	4位	4位
s_7	上野恩賜公園	1位	2位
s_{11}	湯島聖堂	5位	5位
s_{12}	六義園	3位	3位

以下で、表7および表8の結果について説明する。

最初に表7の固定について見る。固定の方法では、観光履歴に対して観光目的が固定として観光目的を推定する。今回の場合、訪問履歴中の5施設のうち3施設の訪問確率が観光目的の歴史のもとで最大になっていることが大きく影響して、歴史が1位である。

他方、表7のHMMについて見ると、買物が1位である。数値計算に利用している訪問確率は固定の場合と同じである。しかし、HMMでは観光目的の変化を考慮している。観光履歴中の最後の訪問施設のかつば橋の訪問確率が観光目的の買物のもとで最大で、かつ前の訪問施設である浅草仲見世の買物のもとでの訪問確率が最大ではないが歴史および景色に近い値であることによって、観光目的の変化を考慮したHMMでは買物が1位という数値計算結果になっている。

次に訪問施設の推薦に関する表8の固定について見る。固定では、観光目的の推定結果として歴史が1位だったことが大きく影響して、推薦結果では観光目的の歴史のもとでの訪問確率の大きな上野恩賜公園が1位、アメ横が2位として推薦されている。歴史のもとでの訪問確率のみを比較すると少しアメ横の方が大きい、観光目的2位の景色のもとでの訪問確率については上野恩賜公園がアメ横の2倍以上であることが影響して上野恩賜公園が1位になっている。

他方、表8のHMMでは観光目的の推定結果として買物が1位だったことと、買物のもとでのアメ横の訪問確率が0.5以上と特に大きかったことが影響して、アメ横が1位として推薦されている。表5から、アメ横の他に東横のれん街や鳩居堂も買物が主な観光目的の施設と考えられる。しかし、インターネット上における両施設に関する記述のあるホームページ数が他の施設に比べて少ないことと、観光目的2位の景色のもとでの両施設の訪問確率が他の施設に比べて小さいことにより、両施設は6位および7位になっている。2位以下の結果については今後の課題とすべき点があるが、1位の推薦結果については、観光目的の変化をHMMを用いてモデル化することによって、観光目的の変化を考慮した観光施設の適応的な推薦を実現できている。以上より、少ない事例であるが、観光目的の変化をHMMでモデル化することによって、観光目的の変化に適応した観光目的の推定や訪問先の推薦が実現できることを確認した。

5. まとめと今後の課題

本研究では、旅行者の観光目的の変化をHMMを用いてモデル化し、統計的決定理論に基づいて次に訪問する観光施設を旅行者の観光目的の変化に応じて推薦する方法を提案した。少ない事例ではあるが、提案モデルと提案方法の有効性を数値計算例で確認した。

本研究では各施設の訪問確率、観光目的の事前確率やHMMでの観光目的の遷移確率を既知と仮定して定式化した。現実では未知である。数値計算例では訪問確率を2つの異なる方法で設定した。また、他の確率は著者らの主観で設定した。これらの確率の値によって訪問施設の推薦結果も異なると考えられる。よって、これらの確率の推定方法の検討は今後の重要な課題である。また、趣味などが既知の観光客の実際の観光履歴情報を学習データとして利用できる場合には、上記の各種確率を未知として機械学習の視点から本研究の定式化を拡張することも可能である。この場合には、統計的決定理論に基づいて各種確率が未知のもとでベイズ基準のもとで最適な推薦方法を検討できる。

また、1章で述べた旅行グループの構成（1人旅、家族旅行など）の違いによるモデルの拡張、距離/移動手段/コストなどの制約の考慮などについても重要な今後の課題である。

次に本研究内容のバイオメディカル分野における活用の可能性について述べる。本研究の主目的は観光施設の推薦であるが、本研究を数理モデルの視点から見ると、本研究の主要内容は状態（本研究の観光目的に相当）が未知のHMMにおける次に発生する記号（本研究の訪問施設に相当）の予測である。このようなHMM上の問題設定はバイオメディカル分野でも活用可能だと考える。例えば、患者の真の健康状態をHMMの未知状態、患者の観測可能な症状をHMMの記号とする。この場合、患者の健康状態の遷移確率が本研究の観光目的の遷移確率に相当し、患者の各健康状態のもとでの症状の発生確率が本研究の訪問確率に相当する。患者の症状に関する情報を受け取ったもとで患者の健康状態の推定を行うことによって、患者の診断支援が可能である。健康状態の推定は本研究の観光目的の推定に相当する。また、次の期（例えば次の日）の患者の症状の予測も可能である。次の期の症状の予測が本研究の訪問施設の推薦（次に訪問する施設の予測）に相当する。症状の予測が可能になると、自宅療養中

の患者の家族や、入院施設の看護師に想定される症状（および対応）について事前に情報提供できる。

バイオメディカル分野における本研究内容の活用に関する詳細検討は今後の課題としたい。

参考文献

- [1] 平山篤, 鈴木優, 川越恭二: 観光地推薦におけるR木を用いた目的別推薦の効率化, 情報処理学会第71回全国大会, pp.1-589-1-590, 2009.
- [2] 磯崎紘, 杉本徹: 対話型観光地推薦システムにおける対話機能の改良, 情報処理学会第72回全国大会, pp.2-631-2-632, 2010.
- [3] 黒島理礼, 村本祥吾, 渡邊貴弓, 浦上美佐子, 松野浩嗣: 被災情報共有無線ネットワークの平常時利用: 適応的観光ルート推薦手法の提案と有用性確認実験, 電子情報通信学会技術研究報告, SIS2012-64, pp.103-108, 2013.
- [4] 前田康成, 小原永: 統計的決定理論に基づく電報分類方法に関する一考察, 情報処理学会論文誌, vol.43, no.10, pp.3119-3126, 2002.
- [5] J.O. Berger: Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York, 1980.
- [6] 繁樹算男: ベイズ統計入門, 東京大学出版会, 東京, 1985.
- [7] 松原望: ベイズ統計学概説, 培風館, 東京, 2010.
- [8] Y. Maeda, N. Ikeda, H. Yoshida, Y. Fujiwara and T. Matsushima: A Note on Morphological Analysis Methods based on Statistical Decision Theory, SICE Annual Conference 2007, International Conference on Instrumentation, Control and Information Technology, pp.1563-1568, Takamatsu, 2007.
- [9] 永田昌明: 統計的言語モデルとN-best探索を用いた日本語形態素解析法, 情報処理学会論文誌, vol.40, no.9, pp.3420-3431, 1999.

前田康成 (まえだやすなり)

北見工業大学准教授。

知識情報処理, 自然言語処理の研究に従事。電子情報通信学会, 情報処理学会等各会員。

