

# 潜在クラスモデルを用いたアプリオリセグメンテーション技術の開発 — トライアル・リピート観光需要の層別化 —

建設工学専攻  
土木計画研究

こばやし かつゆき  
小林 克行  
指導教員 岩倉 成志

## 1. はじめに

幹線交通の整備によって来訪者の少なかった観光地での大きな需要増加が見られる。幹線交通の整備は観光地の需要増加にきっかけを与えるが、その後、有効な施策を講じないと需要は次第に減少してしまう。これを防ぐには、今まで観光地を訪れたことがなかった観光客と何度も観光地を訪れる観光客の嗜好性を把握して、それに応じた施策を講じることが求められる。そのためには、嗜好性が異なる需要を分離して、分析を行う必要がある。嗜好性が異なる需要を分離する方法として、潜在クラスモデルは多くの研究実績がある。しかし、潜在クラスモデルは尤度関数の最大化に従ってセグメントごとの効用関数を推定して、その結果から、セグメントの特徴を読み取る手法である。そのため、観光客をアプリオリに今まで観光地を訪れたことがなかった観光客と何度も観光地を訪れる観光客に分離することができなかつた。

そこで、本研究では潜在クラスモデルを改良して、アプリオリなセグメント分けを行って、効用関数を推定する方法を提案する。これによって、観光客を「今まで訪れたことがない観光地を好んで訪れる観光客（トライアル層）」と「今まで訪れたことがある観光地を好んで訪れる観光客（リピート層）」の2つのセグメントに分離して、それらの嗜好性を把握する。また、この結果をもとにメンバーシップ関数を構築することで、トライアル層とリピート層の個人属性や旅行中に行っているアクティビティなどを明らかにする。これらから、今まで観光地を訪れたことがない観光客と既に観光地を訪れたことがある観光客に再び訪れてもらうことについて、どのような観光施策が有効であるのかを検討する。

## 2. 需要モデル

### 2-1. 初期パラメータ従属型潜在クラスモデル

通常の潜在クラスモデルを表したのが、式(1)~(4)である。式(1)はセグメント  $s$  の対数尤度関数を表しており、ロジットモデルの対数尤度関数とは、個人  $n$  のセグメント  $s$  への所属確率  $p_n(s)$  を選択結果に乗じている点が異なる。セグメントごとに対数尤度関数を最大化させる  $\beta_s$  を求めることで、セグメントごとの効用関数を推定する。式(2)はセグメント  $s$  に所属する個人  $n$  が選択機会  $t$  に選択肢  $i$  を選択する確率を表した式である。選択機会とは、個人が複数回の選択行動を行っている場合、何回目の選択行動であるのかを示したものである。式(3)は個人  $n$  がセグメント  $s$  に所属する確率を表しており、式(4)はセグメント  $s$  の構成比を表している。

本研究では、アプリオリなセグメント分けを行う方法として、初期パラメータ従属型潜在クラスモデルを提案する。この方法は、セグメントの特徴を表したパラメータ  $\beta_s$  と期待値を代入した構成比  $\pi_s$  を初期値として与える。次に、初期値として与えた  $\beta_s$  と  $\pi_s$  から所属確率  $p_n(s)$  を求める。この結果をもとに、 $\beta_s$  と  $\pi_s$  の推定を行うことで、 $\beta_s$  と  $\pi_s$  が真値に近づく。この作業を尤度の改善が見られなくなるまで繰り返すことで、アプリオリなセグメント分けが行える潜在クラスモデルとなる。

この方法は、初期値として与える  $\beta_s$  が、ある程度意

$$L_s = \sum_{n=1}^N \sum_{t=1}^{T_n} \sum_{i=1}^I \{p_n(s) \cdot y_{ni}(t) \cdot \ln P_{ni}(i|\beta_s)\} \quad (1)$$

$$P_{ni}(i|\beta_s) = \exp(\beta_s X_{nii}) / \sum_{j=1}^J \exp(\beta_s X_{nij}) \quad (2)$$

$$p_n(s) = \prod_{t=1}^{T_n} \prod_{i=1}^I P_{ni}(i|\beta_s)^{y_{ni}(t)} \pi_s / \sum_{u=1}^S \prod_{t=1}^{T_n} \prod_{i=1}^I P_{ni}(i|\beta_u)^{y_{ni}(t)} \pi_u \quad (3)$$

$$\pi_s = \sum_{n=1}^N p_n(s) / N \quad (4)$$

但し、 $y_{ni}(t)$  : 個人  $n$  が選択機会  $t$  に選択肢  $i$  を選択した結果

$\beta_s$  : セグメント  $s$  の未知パラメータ

$X_{nii}$  : 説明変数

図したとおりにセグメントの特徴を表している必要がある。これを満たす初期値を生成するには、まず各セグメントのパラメータごとに感度条件の仮定を設ける。次にロジットモデルを用いて分析を行う。そして、感度が高いと仮定したパラメータは、ロジットモデルの推定値に1より高い定数を乗じたものを初期値とする。感度が低いと仮定したパラメータは、ロジットモデルの推定値に1未満の正の定数を乗じたものを初期値とする。感度条件の仮定を設けていないパラメータは、ロジットモデルの推定値をそのまま初期値とする。これを全ての  $\beta_s$  について行うことで初期値を生成する。

### 2-2. セグメント分けの検証

2-1で提案した手法でのセグメント分けは、パラメータに着目して行われており、トライアル層は異なる観光地を多く訪れて、リピート層は同じ観光地を多く訪れる保証はない。そのため、 $p_n(s)$  からトライアル層とリピート層に分けて、実行データにおけるリピート行動に着目した生存時間解析を行う。この結果から、リピート層は短期間に多くリピート行動しているかを検証する。

今回、生存時間解析として用いる Kaplan-Meier 法では、リピート行動発生時を  $0 < t_1 < t_2 < \dots < t_k < \dots < t_L$  としたとき、 $t_k$  における生存率関数は式(5)で表される。式(5)を用いて、全てのリピート行動発生時における生存率を求めて、それを縦軸、観測時間を横軸として描いたのが、生存率曲線である。

$$S(t_k) = \prod_{k=1}^{t_k} \{(r_k - d_k) / r_k\} \quad (5)$$

$$Z = \frac{\left( \sum_{k=1}^L D_{seg1-k} \right) - \left( \sum_{k=1}^L \frac{D_{all-k} \cdot R_{seg1-k}}{N_{all-k}} \right)}{\sum_{k=1}^L \sqrt{\left( \frac{D_{all-k} \cdot R_{seg1-k}}{R_{all-k}} \right) \left( 1 - \frac{R_{seg1-k}}{R_{all-k}} \right) (R_{all-k} - D_{all-k}) / (R_{all-k} - 1)}} \quad (6)$$

但し、 $r_k$  : 観測時間  $t_k$  における観測対象数

$d_k$  : 観測時間  $t_k$  におけるリピート行動発生数

$D_{seg1-k}$  : 観測時間  $t_k$  におけるセグメント1でのリピート行動発生数

$D_{all-k}$  : 観測時間  $t_k$  における全体でのリピート行動発生数

$R_{seg1-k}$  : 観測時間  $t_k$  におけるセグメント1での観測対象数

$R_{all-k}$  : 観測時間  $t_k$  における全体での観測対象数

トライアル層の生存率曲線とリピート層の生存率曲線の差が統計的に有意であるのかを検証するため、ログランク検定を行う。ログランク検定では、式(6)で得られる  $Z$  が漸近的に標準正規分布に従うため、これをもとに2つの生存率曲線の差が有意であるか検定を行う。

### 2-3. メンバーシップ関数

メンバーシップ関数を表したのが、式(7)~(8)である。式(7)は個人  $n$  がセグメント  $s$  に所属する確率を表しており、式(8)は対数尤度関数を表している。

$$M_n(s) = \exp(\theta \cdot W_{ns}) / \sum_{n=1}^S \exp(\theta \cdot W_{nu}) \quad (7)$$

$$L' = \sum_{n=1}^N \sum_{s=1}^S \{\delta_n(s) \cdot \ln M_n(s)\} \quad (8)$$

但し、 $\theta$ : 未知パラメータ  $W_{ns}$ : 説明変数  
 $\delta_n(s)$ : 個人  $n$  のセグメント  $s$  への所属結果 (式(3)で求めた  $p_n(s)$  を1-0化したもの)

## 3. 分析結果

### 3-1. データ概要

分析には、表1に示したアンケートを用いる。分析対象者は、今回の得られたデータのうち、調査対象期間中の旅行回数が2~10回の258人とする。

### 3-2. セグメントの仮定

表2にトライアル層とリピート層のパラメータの感度条件と初期値生成の定数を表した。空欄は感度条件の仮定を設けていないことを示す。観光地魅力度でリピート層の感度が高いのは、魅力的な観光地を何度も訪れるためである。移動費用でトライアル層の感度が低いのは、トライアル層は様々な観光地を訪れ、移動費用が高い観光地も訪れるためである。移動時間変化でトライアル層の感度が高いのは、リピート層は移動時間短縮と関係がないためである。リピート間隔・回数、12ヶ月後リピートダミーでリピート層の感度が高いのは、リピート層は短期間に多くリピートするためである。

### 3-3. 初期パラメータ従属型潜在クラスモデル

潜在クラスモデルの分析結果を表したのが、表3である。この結果から、観光地魅力度と移動費用では仮定を満たしていないが、それ以外では仮定を満たしている。パラメータの  $t$  値や尤度比も良好な結果が得られた。

### 3-4. セグメント分けの検証

トライアル層とリピート層の生存率曲線を図1に表し、表1) アンケート概要

調査方法	インターネット調査
調査時期	2005年6月 回答数 343人
サンプル条件	○居住地: 東京都, 神奈川県, 埼玉県, 千葉県, 茨城県 ○年齢: 30歳以上
アンケート内容	○観光行動履歴に関する質問 ・調査対象期間: 2000年4月~2005年5月 ・訪問地 ・アクティビティ ・交通手段など ○観光地魅力度の定量化に関する質問 ○個人属性

表2) セグメントの仮定と初期値生成の定数

	トライアル層		リピート層	
	感度	定数	感度	定数
観光地魅力度	小	0.9	大	1.1
移動費用(円)	小	0.9	大	1.1
移動時間(分)		1.0		1.0
移動時間変化(分)	大	1.1	小	0.9
リピート間隔(箇月)	小	0.9	大	1.1
リピート回数(回)	小	0.9	大	1.1
12ヶ月後リピートダミー	小	0.9	大	1.1
海外旅行ダミー		1.0		1.0

表3) 潜在クラスモデル分析結果

	トライアル層	リピート層
EM回数	7回	
観光地魅力度	3.84E-01 ***	3.37E-01 ***
移動費用	-8.29E-05 ***	-8.93E-06
移動時間	-3.68E-03 ***	-3.89E-03 ***
移動時間:変化	-2.26E-02 ***	-5.63E-03
リピート間隔	3.34E-03	4.90E-02 ***
リピート回数	8.38E-01 ***	9.96E-01 ***
12ヶ月後リピートダミー	1.04E+00 ***	2.71E+00 ***
海外旅行ダミー	-1.32E-01	2.73E+00 ***
構成比	0.53	0.47
尤度比	0.22	0.33

但し、\*10%有意 \*\*5%有意 \*\*\*1%有意

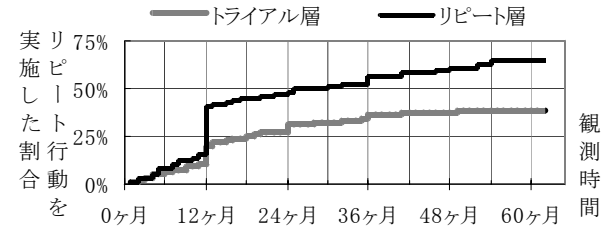


図1) 生存率曲線

表4) メンバーシップ関数分析結果

層	変数	係数
トライアル層	個人属性:男性ダミー	7.97E-01 *
	個人属性:40歳代ダミー	7.60E-01 *
	アクティビティ「温泉浴」実施割合	1.43E+00 **
	アクティビティ「名所・旧跡」実施割合	1.24E+00 *
定数項		3.68E+00 ***
リピート層	旅行目的「帰省・訪問・家事」割合	1.47E+00 *
	旅行実施時期:6月割合	3.33E+00 *
	平均旅行日数	9.56E-01 ***
	交通手段「航空」利用割合	5.59E+00 ***
	アクティビティ「海水浴」実施割合	2.00E+00 *
	アクティビティ「キャンプ」実施割合	3.94E+00 *
尤度比		0.48

但し、\*10%有意 \*\*5%有意 \*\*\*1%有意

た。この結果から、リピート層の方が多くリピート行動を行っていることがわかる。2つの生存率曲線についてログランク検定を行ったところ、 $Z$  は-5.827になった。このことから、統計的に2つの生存率曲線はほぼ100%異なるため、正しくセグメント分けできたとと言える。

### 3-5. メンバーシップ関数

メンバーシップ関数の分析結果を表したのが、表4である。この結果から、以下のことが読み取れる。

男性ダミーがトライアル層の正の特性変数であるので、男性は異なる観光地を訪れ、女性は同じ観光地を訪れる傾向にある。また、平均旅行日数がリピート層の正の特性変数なので、リピート層の確保には女性向けの長期滞在型宿泊施設を整備することが有効であると言える。温泉浴実施割合がトライアル層の正の特性変数なので、「癒し」がリピート行動に直結しないことがわかる。

航空利用割合がリピート層の正の特性変数であり、リピート層は移動費用の感度が低く、移動時間の感度が高いことが、3-3の結果からわかる。このことから、魅力ある観光地に高速交通体系を整備すれば、都心から離れた観光地でもリピート層を確保できることがわかる。

## 4. まとめ

今回、アプリアリにセグメント分けする潜在クラスモデルの提案し、その有効性が確認された。また、この分析結果からメンバーシップ関数の構築を行うことで、トライアル層とリピート層の個人属性や旅行中に行っているアクティビティなどを把握することができた。