

6. 多数の潜在クラスを考慮した行動選択モデルの推定に関する研究 — 中心市街地における活動選択を例に —

A study on incorporating many latent classes into behavior choice model - Case of activity choices in central shopping district -

東京大学工学部都市工学科 60158 中村 敏和

Today marketing information about the pedestrians' behaviors is very important because many invigoration projects have been performed to deal with the decline of urban districts. In such situations, the purpose of this research is to develop behavior choice model with incorporating latent classes into it, from a one-class model to a five-class model, in order to consider individual taste heterogeneity. Moreover, this study manages to make the model with latent classes that have been given before by using the other method and proves the advantage of it.

1. 研究の背景と目的

近年、中心市街地の衰退が多く見られる。郊外大型店舗やネット売買などの発展に対抗していくには、有効なマーケティング戦略をとっていく必要がある。そのため、最近になって、ポイントカードの導入など、中心市街地全体での戦略を実施することが多くなってきた。そうした戦略をとる上で、来街者の回遊行動に関する情報をうまく分析・活用し、より効率よく市街地活性化の施策を実施していくことが重要である。また、年々来街者の回遊行動における嗜好は多様化してきており、来街者の嗜好の特徴をより多く加味していくことが求められている。

このような背景のもと、田澤¹⁾は、松山市の中心市街地を対象に、潜在クラスを導入することで嗜好の多様性を考慮した中心市街地回遊行動モデルの構築を試み、2つのクラスの導入に成功するとともに、その有用性を示している。

本研究は、田澤¹⁾の研究を土台にさらなる検討を加えるものである。具体的には、第一に、潜在クラス数を3, 4, 5クラスに増やしたモデルを構築することで、より特徴ある嗜好を選り分けることができることを示す。ただ、潜在クラスを多くすると、モデルの構築が困難になる。そこで第二に、事前に推定された潜在クラスを基にモデルを

構築することで、実用上のデメリットを解消でき、さらに、より特徴あるモデルを構築できる可能性を持つことも示す。

2. 分析データと回遊行動モデルの概要

2.1 使用するデータ

本研究で用いるデータは、2004年に松山市の中心市街地で行われた回遊行動実態調査データである。調査概要を表1に示す。

表1：調査概要¹⁾

対象	松山中心市街地来街者	
調査手法	アンケート調査 配布地点を設定し、無作為に配布	
調査日時	2004年6月18日(土)、19(日) 2004年11月6日(土)、7(日)	
回収率	23.5%	
質問項目	社会経済属性	性別、年齢、同伴人数、来街手段、など
	当日の回遊行動	入口・出口、利用駐車場・利用駅、 回遊開始時刻、移動手段、移動時間、 立寄り施設、滞在時間、購買金額、など
分析対象人数	294人	

2.2 回遊行動モデルの構成

回遊行動モデル全体の構成は田澤¹⁾に倣う。すなわち、中心市街地では状況に応じた意思決定がなされることが多いと考え、意思決定は逐次的に行われるという仮定を置く。初回は対象エリアに達した時点、2回目以降は前活動が終了した時点で意思決定を行うものとし、

1. 帰宅 or 活動継続の選択と活動を継続する

場合の活動内容の選択を記述するモデル

2. 活動内容が決定した後の活動施設の選択を記述するモデル

の2種類のモデルで構成されるものとする。

モデル1では、帰宅or活動継続の選択（上位レベル）と、活動継続の場合における活動内容の選択（下位レベル）をNLモデルの構造（図1）を用いて記述する。活動内容の選択肢は表2に示した8分類を用いる。また、表3に、構築するモデルに組み込む説明変数をまとめる^{*1}。

なお、本研究では、田澤¹⁾と同様に、帰宅or活動継続の選択と活動内容の選択にクラス毎の嗜好の違いが表れると考え、モデル1（以下、「活動選択モデル」とよぶ）に多くの潜在クラスを考慮することを試みるものとする。以下、モデル2については本稿では触れない。

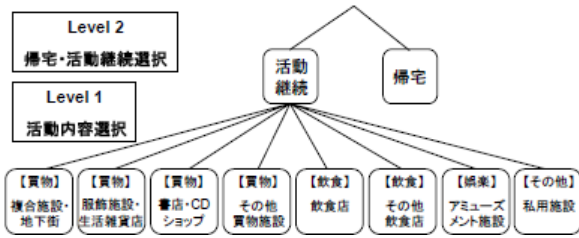


図1：モデル1（活動選択モデル）の構造¹⁾

表2：活動内容8分類¹⁾

活動内容	略記	対応施設数
複合施設・地下街【買物】	FB	8
服飾施設・生活雑貨店【買物】	HB	91
書店・CDショップ【買物】	SB	9
その他買物施設【買物】	OB	35
飲食店【飲食】	IE	60
その他飲食店【飲食】	OE	33
アミューズメント施設【娯楽】	AA	18
私用施設【その他】	OO	36

表3：説明変数リスト

選択レベル	導入した説明変数
Level2	帰宅 活動継続
Level1	今までの市街地滞在時間(分) ログサム変数 今までの活動数(回)、今までの活動有無、 昼時ダミー、1回目活動ダミー

3. 潜在クラスモデルの定式化

潜在クラスモデルでは、クラス q の効用関数パラメータベクトル $\vec{\beta}_q$ と、個人のクラスへの帰属確率 π_{nq} が算出される。各クラスに異なるパラメータを推定することで、クラス毎の嗜好の違いを反映

することができる。

まず、個人 n がクラス q に属し、機会 t において選択肢 i を選択する確率は次のように表される。

$$P_t(i|\vec{x}_n, q) = f_i(i|\vec{x}_n, \vec{\beta}_q) \quad (1)$$

\vec{x}_n ：個人 n の説明変数ベクトル

これより、個人 n がクラス q に属するときの選択確率の尤度は以下のように表される。

$$f(y_n|\vec{x}_n, \vec{\beta}_q) = \prod_{t=1}^{T_n} \prod_{i=1}^{C_n} f_t(i|\vec{x}_n, \vec{\beta}_q)^{d_{ni}(t)} \quad (2)$$

$d_{ni}(t)$ ：個人 n が機会 t に選択肢 i を選択した場合1、選択しなかった場合0

C_n ：個人 n の選択肢数

T_n ：個人 n の機会数

よって、個人 n のクラス q への帰属確率 π_{nq} を用いて、潜在クラスモデルの対数尤度関数は以下のように表される。

$$\begin{aligned} \ln L &= \sum_{i=1}^N \sum_{q=1}^Q \ln [f(y_n|\vec{x}_n, \vec{\beta}_q) \cdot \pi_{nq}]^{d_{nq}} \\ &= \sum_{i=1}^N \sum_{q=1}^Q [d_{nq} \ln f(y_n|\vec{x}_n, \vec{\beta}_q) + d_{nq} \ln \pi_{nq}] \quad (3) \end{aligned}$$

d_{nq} ：個人 n がクラス q に属している場合1、属していない場合0

また、本研究では帰属確率 π_{nq} を個人属性からロジット型の確率式で求める。

$$\pi_{nq} = \frac{\exp(\vec{\delta}_q \cdot \vec{z}_n)}{\sum_{q=1}^Q \exp(\vec{\delta}_q \cdot \vec{z}_n)} \quad (4)$$

\vec{z}_n ：個人 n の個人属性ベクトル

$\vec{\delta}_q$ ：クラス q のメンバーシップパラメータベクトル

いま、 d_{nq} は観測されない架空の変数であるため、欠損値として扱う。推定にはEM法を用いる。まずE-Stepにおいて、 \vec{z}_n 、 $\vec{\delta}_q$ より個人がクラスに所属する確率 W_{nq} （事後確率）を計算する。

$$W_{nq} = \frac{\pi_{nq} \cdot f(y_n|\vec{x}_n, \vec{\beta}_q)}{\sum_{q=1}^Q \pi_{nq} \cdot f(y_n|\vec{x}_n, \vec{\beta}_q)} \quad (5)$$

その後M-Stepにおいて、 d_{nq} を W_{nq} で置き換え、最尤推定を行う。それによって $\vec{\beta}_q$ 、 $\vec{\delta}_q$ を求める。このプロセスを、(3)式の尤度の向上が見られ

なくなるまで繰り返し行う。

4. 潜在クラスモデルと活動選択モデルの同時推定

まず、潜在クラス数を1, 2, 3, 4, 5と変化させて、潜在クラスモデルのパラメータ ($\hat{\delta}_q$) と活動選択モデルのパラメータ ($\hat{\beta}_q$) の同時推定を行った (潜在クラス数=1 のときは潜在クラスを考慮しないモデル)。田澤¹⁾でも3クラス以上のモデルの推定を試みているが、推定されたログサム変数の係数が1を超え、仮定された意思決定構造と矛盾する結果になったとされている。これは1回目も帰宅する可能性があるとしてモデルを推定したためと考えられる。本データでは1回目は活動 (を継続) することが前提であるため、その点を解消することで、上述の矛盾のないモデルを構築することができた。なお、1, 2クラスモデルに関してもモデルを構築し直した。

モデルの評価として、まず、表4にAIC値等を示す。3クラスモデルのAIC値が最も低い値となった。ただ、クラス数の増加に伴い、パラメータの選択が困難になり、EM法における初期値の影響も受けやすくなるため、より最適なモデルが存在する可能性は大いにある。潜在クラスモデルの対数尤度は、クラス数の増加に伴い、指数関数的に増加すると考えられるため、AIC値は、クラスを増やしてもなかなか改善されないといえる。

次に、潜在クラスのクラス毎それぞれの特徴を見てみる。どのモデルも、潜在クラスの数だけ特徴のあるクラス分けが出来た。一例として5クラスモデルの各クラスの特徴を考察する。表5, 6, 7に潜在クラスに関する結果を示す。

活動内容の頻度は、平均と比較して考察した。表7のように、各クラスとも活動数に違いが見取れ、活動内容の特徴も多様である。田澤¹⁾による2クラスのモデルでは多目的なクラスと少目的なクラスの2つに分類されているが、このように、潜在クラスを増やすことで、クラスごとに特徴が

かぶることもなく、特徴あるクラス分けがなされていることから、より来街者の嗜好を考慮できているといえる。

表4：モデルの評価

class	lnL(活動選択モデル)	lnL(クラス選択モデル)	k(活動選択モデルパラメータ数)	k(クラス選択モデルパラメータ数)	AIC
1class	-2414.66	0.00	19	0	4867.32
2class	-2324.07	-79.72	28	9	4881.59
3class	-2276.16	-96.56	36	14	4845.44
4class	-2214.23	-265.56	46	15	5081.57
5class	-2220.70	-280.38	47	17	5130.17

表5：各クラスのクラス割合・平均活動数

クラス	クラス割合(%)	平均活動数(回)
クラス1	36.02	3.78
クラス2	15.70	3.99
クラス3	17.74	3.02
クラス4	11.43	3.47
クラス5	19.11	3.28
計	100.00	3.55

表6：各クラスの活動内容割合(%)

クラス	FB	HB	SB	OB	IE	OE	AA	OO
クラス1	38.11	17.81	6.88	8.66	9.53	9.34	4.60	5.06
クラス2	42.48	18.95	2.69	10.85	8.02	12.42	1.54	3.04
クラス3	39.93	16.97	10.67	11.78	4.59	4.47	1.82	9.78
クラス4	33.38	15.98	7.97	8.57	13.83	8.60	4.32	7.35
クラス5	29.05	17.30	11.10	10.40	10.84	6.98	5.62	8.70
計	37.03	17.59	7.58	9.82	9.23	8.65	3.79	6.32

表7：各クラスの特徴

クラス	特徴
クラス1	活動数がやや多い。どの活動も平均的に行う。
クラス2	活動数が多い。複合施設や喫茶店を多く利用している。
クラス3	活動数が少ない。書店や私事などの用事が多い。
クラス4	活動数は平均的。飲食店を利用する割合が多い。
クラス5	活動数がやや少ない。書店をよく利用し、複合施設を利用する割合が小さい。

5. 事前に推定された潜在クラスを考慮した活動選択モデル

前節では、潜在クラスモデルと活動選択モデル推定を同時に行ったが、これには推定上重要な困難がいくつかある。まず、クラス数が増え、パラメータ数が増えることで、EM法が収束しにくくなり、さらに、初期値の影響を受けやすくなるため、パラメータの選別が非常に困難になる。また、推定に要する時間もクラス数に比例して増えてしまうため、実用上厳しい障害となってしまう。

そこで、本節では、事前に別の方法で行われた潜在クラス推定の結果を元にモデルを構築する

ことで、上記の問題を克服することを目指す。

事前に推定された潜在クラスとして、田澤¹⁾による潜在クラス分析の結果を使う。この分析では、各サンプルの性別、年代、来街人数を外生変数、8分類の活動の実行の有無を顕在変数として、潜在クラス推定を行った。その結果、5クラスに分けるのが最も精度がよかったことから、それを用いる。表8にその特徴を示す（この特徴付けは田澤¹⁾と解釈が異なるが、前節との比較のため、前節と同様の方法で解釈し直した）。

表8：事前に推定された各クラスの特徴

クラス	クラス名
クラス1	平均的な活動数。飲食、喫茶が多い。
クラス2	平均的な活動数。活動内容も平均的。
クラス3	活動数が多め。アミューズメント施設が多い。
クラス4	活動数は少なめ。私用、書店、飲食が多い。
クラス5	平均的な活動数。複合施設利用が多い。

この推定から求められた帰属確率を与えて、活動選択モデルの推定を行う。結果を前節の5クラス同時推定モデルと比較したものが表9である。

表9：モデルの評価

モデルタイプ	lnL	k	AIC
潜在クラス同時推定	-2220.7	47	4535.4
潜在クラス事前推定	-2282.28	45	4654.56

対数尤度、AIC値の両方から、前節のモデルの方が、精度がよいことが読み取れる。ただし、対数尤度に関しては、前節の3クラスモデルと同等であり、AIC値もそれほど差はないといえる。

推定にかかる時間は、同時推定モデルに比べて半分ほどに短縮された。さらに、前者モデルに対し、後者のモデルは、EMアルゴリズムによる推定で不安定になることもほとんどなく、初期値を変えても同じ値に収束するため、パラメータの選別がより実用的であるといえる。パラメータの値に関しても、ログサム変数は0~1の間で安定しており、その他の変数も適切な値を取っている。

また、クラス1では飲食・喫茶に有利に働く昼時ダミー変数のパラメータが大きな値となっているほか、クラス3では活動の有無に関するパラメータが、娯楽に関して高く、1度行ったら他のアミューズメント施設にも行く可能性が高くな

る結果が出ている。他の変数も表8の特徴を表しているものが多い。

これらのことから、潜在クラスを事前に推定しておくことで、パラメータの選別やアルゴリズムの安定性が上がり、より精度のよいパラメータを持つモデルを構築できるといえる。

6. 結論と今後の課題

本研究をまとめると以下ようになる。

- ・潜在クラスを増やすことで、より多様な嗜好を表現できることを示した。
- ・事前に推定された潜在クラスを元に活動選択モデルを推定した場合でも、これらを同時推定するモデルと同等の精度が確保でき、モデルの内容も事前に推定されたクラスの特徴を十分表していることから、モデルの構築をより実用的に行えることを示した。

また、以下に今後の課題をまとめる。

- ・本研究で用いた、事前に行われた潜在クラス推定は、「一連のトリップチェーン全体でどの活動を実行したか」に基づくものであり、選択の逐次性や、複数回の選択など考慮されていない面も多い。この点を克服する潜在クラス推定を事前に行う必要性があると考えられる。
- ・上記の点が改善されれば、事前に潜在クラスを推定しておくことで、クラスごとに特徴にあったモデル構造や説明変数の導入をすることもでき、さらなる精度の改善や実用性の向上が図れる可能性を持っており、その効果等を考察する価値は十分にあると考えられる。

補注

*1 田澤¹⁾では、調査における目的地選択実績に基づく「場所ポテンシャル」を説明変数に加えていたが、本研究では行動結果を説明変数に含めるのは好ましくないと考え、除外した。

参考文献

- 1) 田澤直幹：潜在クラスを考慮した中心市街地回遊行動モデルの構築とシナリオ評価，東京大学大学院工学系研究科都市工学専攻修士論文，2007。