

行動予測における Web 閲覧ログの活用可能性に関する実験的検討

愛媛大学 賛助会員 ○隅田和樹 愛媛大学 正会員 倉内慎也
愛媛大学 正会員 白柳洋俊

1. 研究の背景と目的

近年、経路検索サービスのデータなどを用いて交通需要予測を行っている例は多々ある。また、Web 関連技術の発展により実際の行動以外にも個人の嗜好に関する膨大な個人データが Web 上に蓄積されるようになっている。それらのデータは、オンラインショッピングの商品購入ログのような、ユーザーの意思決定により何かしらの選択が行われた履歴データと、ある特定ページをある個人がどの程度閲覧しているか、何回閲覧したかというような閲覧ログの 2 種類に大別され、前者のデータは情報の信頼性が高く、それらを基に Web 上で行動予測やマーケティングを行っている例は多い。一方後者のデータは、データ量は膨大なものの情報の信頼性は不明瞭であり、行動予測にもあまり活用されていないが、少なからず個人の嗜好に関する情報を含むものと考えられる。そこで本研究では、Web 上で SP 実験を行い、閲覧ログが個人の嗜好の推定に影響があるのかどうかを検証する。これにより、個人の行動予測における Web 閲覧ログの活用可能性を探ることを目的とする。

2. Web 上での SP 実験

本来であれば、民間企業などが保有している膨大な量の個人の閲覧ログを用いて分析するのが理想であるが、個人情報保護の問題や、それらのデータは企業の重要な資産であることから、提供を受けることは難しい。そのため、まずは個人ごとに閲覧ログを取得できるような SP 実験システムを独自に開発した。

2.1. 調査対象と調査方法

今回は被験者として、現在 1 人暮らしをしている愛媛大学の学部生と大学院生に、研究室の 1 台のパソコンを用いて一人ずつ SP 実験を行った。実験内容は、比較的學生が回答しやすい 2 者択一形式の物件選択アンケートで、回答人数は 28 人、そのうちエラーデータを除いた 26 人が有効サンプルである。

2.2. 実験概要

SP 実験を実施するにあたり、仮想の物件を 2 つ提示し、被験者が望ましいと思う物件を選択してもらうような web アンケートシステムを構築した。各物件に関する属性は不動産会社の HP を参考に、表 1 のように 11 個選定した。ここで、11 個すべての属性について、閲覧時間等のログを取得することは、被験者にとって操作性に劣るため現実的でない。そこで、表 1 のように似たような属性をグルーピングし、最終的に「家賃・敷金・礼金」、「物件の場所」、「間取り・部屋向き・階数」、「部屋設備」、「築年数」の 5 つの属性カテゴリーに分類し、その属性カテゴリーごとに閲覧ログを収集することとした。

表 1 SP 実験における物件の属性とカテゴリー設定

属性 カテゴリー	家賃 敷金・礼金	物件の場所	間取り 部屋向き 階数	部屋設備	築年数
属性	賃料（管理費込） 敷金・礼金・保証金	物件～愛媛大学までの徒歩時間 物件～大街道までの徒歩時間	間取り 部屋向き 階数	浴室 室内洗濯機置き場 独立洗面台	築年数

なお、SP 設問数は、部分要因配置法を用いて 1 個人当たり 12 問とし、さらに精度検証用に別途 4 問を生成し、1 個人当たり合計 16 問への解答を要請した。また、今回各設問での物件選択後に、選択の決め手となった属性カテゴリーを 5 つの中から 1 つ選択してもらうこととした。これは、Web サイトにおいてユーザーに何を重視するかを選択してもらうケースを想定しており、閲覧ログを用いた場合と比較するための設問である。

2.3 取得する閲覧ログ

SP 実験で取得する閲覧ログは、「各属性カテゴリーの閲覧時間」、「各属性カテゴリーの閲覧回数」、「最初に閲覧した属性カテゴリー」、「最後に閲覧した属性カテゴリー」の4種類とした。

3. 二項ロジットモデルを用いたモデル推定

本研究では、3つの二項ロジットモデルを用いて各モデルの現況再現性、予測精度を比較した。ある個人 n の t 番目の設問での選択肢 i に対する効用の確定項は以下の式 (1) で表される。

$$V_{int} = \beta_{1n}x_{1nt} + \beta_{2n}x_{2nt} + \cdots + \beta_{kn}x_{knt} \quad (1)$$

まずモデル1は、「 $\beta_{knt} = \beta_k$ 」として同質な個人を仮定した上で選択結果のみを用いて推定する標準的なモデルである。次にモデル2は、選択結果に加え、被験者が回答した決め手となった属性カテゴリーをパラメータの説明変数として導入したモデルで、「 $\beta_{knt} = \beta_k' + \beta_k'' \times$ (個人 n が t 番目の設問で回答した決め手の属性ダミー)」としてモデルを推定する。最後にモデル3は、選択結果の他に、4種類の閲覧ログをパラメータの説明変数として用いたモデルで、「 $\beta_{knt} = \beta_k' + \beta_k'' \times$ (個人 n の t 番目の設問での閲覧時間) $+\beta_{kn}''' \times$ (個人 n の t 番目の設問での閲覧回数) $+\beta_{kn}'''' \times$ (個人 n の t 番目の設問で最初に閲覧した属性カテゴリーダミー) $+\beta_{kn}''''' \times$ (個人 n の t 番目の設問で最後に閲覧した属性カテゴリーダミー)」としてモデル推定を行った。

4. モデルの推定結果と現況再現性

表2にモデルの推定結果を示す。まず、モデル2では11個の決め手ダミーパラメータのうち、4個が有意に推定され、モデル3では44個の閲覧ログパラメータのうち9個が有意に推定された。次に、現況再現性として各モデルの適合度を表すAICを比較してみると、モデル3で最も適合度が高く、次いでモデル2、モデル1という結果になった。これより個人の行動予測にWeb閲覧ログは十分活用できる可能性があると考えられる。

表2 各モデルの推定結果

属性	説明変数	モデル1		モデル2		モデル3	
		推定値	t値	推定値	t値	推定値	t値
家賃・共益費(万円)	定数	-1.37	-3.24	-1.47	-4.52	-4.48	-1.91
	閲覧回数(回)					-0.886	-2.98
	最後に閲覧した属性カテゴリー(回)					-2.82	-3.74
愛大からの徒歩時間(時間)	定数			-5.63	-2.40		
大街道からの徒歩時間(時間)	決め手ダミー			13.89	1.93		
間取り(畳)	定数	0.228	2.33	0.328	3.02		
	閲覧時間(分)					3.09	2.99
部屋向き(北:0, 南:1)	定数	1.91	3.58				
	決め手ダミー			3.38	4.69		
	閲覧回数(回)					1.71	4.90
階数(階)	決め手ダミー			0.761	3.95		
	定数	1.59	5.13			2.14	3.90
浴室(ユニットバス:0, バストイレ別:1)	決め手ダミー			4.34	4.26		
	閲覧時間(分)					10.70	2.38
独立洗面台(無:0, 有:1)	定数	0.929	2.96				
	閲覧時間(分)					10.40	3.96
室内洗濯機置場(無:0, 有:1)	定数	0.514	1.96				
	定数	-0.447	-3.45	-0.660	-3.49	0.532	2.17
築年数(年/10)	閲覧回数(回)					-0.958	-3.81
	最初に閲覧した属性カテゴリー(回)					-1.39	-1.79
	最後に閲覧した属性カテゴリー(回)					-1.22	-2.57
サンプル数		26		26		26	
AIC		298.233		229.660		223.091	
自由度調整済み尤度比		0.306		0.466		0.481	

5. 予測精度の検証

モデル推定に用いなかった4問の回答に対する予測精度の検証を行った。3つのモデルを対数尤度で比較すると、モデル1では-41.702、モデル2で-83.218、モデル3で-41.590と、モデル3で最も精度が高い結果となった。ゆえに、予測精度においても、閲覧ログを用いた場合が最も望ましいと言える。一方、モデル2の予測精度は大きく劣っている。現況再現性はモデル3に僅かに劣る程度であったことを踏まえると、決め手となった属性カテゴリーについての被験者回答は状況依存적である上に、多大に主観が含まれ、必ずしも信頼性の高いものではないと言えよう。従って、わざわざ被験者に属性の重要度を回答してもらわなくとも、閲覧ログを用いることで嗜好の推定は自動化できるものと言え、今後そのようなシステムの実装がユーザーの意思決定支援やサイト運営側のマーケティングや行動予測に有益な効果をもたらすものと考えられる。