

## Twitter データを用いた外国人観光客の観光周遊行動のモデル分析

広島大学大学院	非会員	○大田	高滉
復建調査設計株式会社	正会員	福永	晃彦
広島大学大学院	正会員	力石	真
広島大学大学院	正会員	藤原	章正

## 1. 研究背景・目的

訪日外国人観光客数は年々増加傾向にある中で、中国地方ではインバウンド消費が全国最下位となっており、外国人観光客の増加が中国地方の活性化に結び付いていないことが問題となっている。原因の一つとして、関西圏等から日帰り観光できるため、宿泊客数が相対的に少ない点が指摘されている。本研究では、以上の観光行動は、観光地の魅力度と観光地へのアクセス性（交通コスト）のトレードオフを踏まえた意思決定の結果として現れたものと考えられる。

これまでも観光地の魅力とアクセス性を考慮した観光周遊行動モデルが構築されてきたが、既存モデルの多くは、(1)観光周遊パターン数を少数に限定した上でモデル化している点、(2)観光資源数など、季節や時間帯（昼間と夜間）により変動する観光地の魅力を反映したモデルとなっていない点、が課題として残る。そこで本研究では、(1)Twitter データにより季節別・時間帯により異なる観光魅力度を表現するとともに、(2)選択肢を列挙する必要のない Recursive logit<sup>1)</sup>モデルを援用することにより、膨大な数の観光周遊パターン（実証分析では  $47^{13}=5.46 \times 10^{21}$  個の選択肢）を再現するモデルを構築する。

## 2. 使用データ

本研究では観光庁「訪日外国人消費動向調査」、航空局「国際航空旅客動態調査」、国籍別出国者数に関する法務省「出入国管理統計月報」を組み合わせて作成された訪日外国人流動データ（FF-Data）、国土交通省が提供する総合交通分析システム「NITAS」をもとに OD 別代表交通機関ごとの所要時間が最小となる経路・所要時間・費用・距離を整備した OD 別交通サービス水準および株式会社ナイトレイが提供する位置情報データを解析・付与した Twitter のデータの 3 種類を用いる。

## 3. モデルの概要

本研究では観光周遊行動モデルを構築する際に Recursive logit モデルを援用する。まず、訪日外国人観光客は到着地/到着日、出発地/出発日という時空間制約が与えられた条件下で周遊行動を決定すると考える。さらに、膨大なパターンが存在し得る観光周遊行動を表現するために、観光地での滞在リンクと観光地間を移動する移動リンクから構成される時空間ネットワーク（Fig. 1）<sup>2)</sup>を考える。今回使用している Twitter データには時刻情報が格納されており、時間は連続量であるが、5時から17時59分を「昼」、18時から4時59分を「夜」として2区分に時間を離散化する。またノード間の移動は昼、夜いずれかの時間帯で発生するものとする。移動リンクには移動に伴う負の効用、滞在リンクには観光によって得られる正の効用が発生すると考え、外国人観光客は周遊行動全体の総効用を最大にするように周遊パターンを決定すると仮定する。この行動原則の下で、観光周遊行動を図1の時空間ネットワーク上の経路選択問題とみなすことで、動的計画法を活用した解法が利用可能になる。

本研究では、ネットワーク上流側リンク  $k$  から隣接する下流側リンク  $a$  に移動する際に得られる効用  $u(a|k)$  を以下のように定義する。

キーワード 観光周遊行動 訪日外国人観光客 Recursive logit Twitter FF-Data

連絡先 〒739-8529 東広島市鏡山 1-5-1 広島大学大学院国際協力研究科

TEL 082-424-4693

$$u(a|k) = v(a|k) + V^d(a) + \mu\varepsilon(a) \quad (1)$$

ここで、 $v(a|k)$ は上流側リンク $k$ から下流側リンク $a$ の移動の際に得られる瞬時効用、 $V^d(a)$ はリンク $a$ から目的の地ダミーリンク $d$ までの期待最大効用、 $\varepsilon(a)$ は標準ガンベル分布に従う誤差項、 $\mu$ はスケールパラメータである。また、 $A(k)$ はリンク $k$ から流出する下流側リンクの集合である。式(1)をもとに、ランダム効用最大化に基づく行動原理を仮定すると、リンク $k$ からリンク $a$ に移動するリンク選択確率は以下のロジットモデルで定義される。

$$P(a|k) = \frac{e^{\frac{1}{\mu}(v(a|k)+V^d(a))}}{\sum_{a' \in A(k)} \delta(a'|k) e^{\frac{1}{\mu}(v(a'|k)+V^d(a'))}} \quad (2)$$

下流側の期待最大効用は以下のベルマン方程式により定義する。

$$V(k) = E \left[ \max_{a \in A(k)} \left( v(a|k) + V^d(a) + \mu\varepsilon(a) \right) \right] \quad (3)$$

ここで、式(3)で定義される期待最大効用は以下のログサム形式に書き換えることができる。

$$V(d) = \begin{cases} \mu \ln \sum_{a \in A(k)} \delta(a|k) e^{\frac{1}{\mu}(v(a|k)+V^d(a))} & \forall k \in A \\ 0 & k = d \end{cases} \quad (4)$$

式(4)の両辺指数化することでベルマン方程式を線形方程式で表現できるため、逆行列を使った連立方程式の求解<sup>3)</sup>、または、後ろ向き帰納法<sup>4)</sup>により式(4)を効率的に計算することができる。本研究では、後ろ向き帰納法を用いて $V(k_0)$ の計算を行い、計算された $V(k_0)$ を用いて最尤推定法を適用しパラメータを推定する。

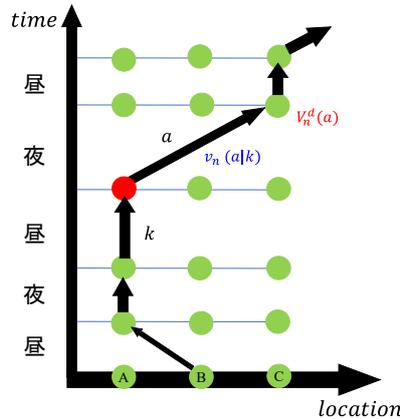


図 1. 観光周遊行動の時空間ネットワーク

#### 4. 実証分析

モデルの詳細と推定結果を以下の表 1, 表 2 に示す。Model 1 は観光地の魅力度変数としてトリップアドバイザーのデータを用いており、観光地数(口コミ数)を各トリップの目的地に応じて設定されたモデルである。Model 2 以降は観光魅力度を訪日外国人観光客の Twitter のツイート数で定量化している。Model 1 と比べて Model 2 の最終対数尤度が大きいことから、観光魅力度を Twitter のコメント数で定義したモデルの方が観光周遊行動の再現性が高いモデルであることが分かる。

Model 3 では、Model 2 とは異なり、説明変数を昼夜に分けて定義している。Model 3 に比べて Model 2 の最終対数尤度が高いことから、昼夜共通の観光魅力度を用いた場合の精度が高いことが分かる。この原因として、(1) 必ずしも観光地訪問中にリアルタイムでツイートがなされるとは限らないこと、(2) 昼夜別により標本数が減少し、特に訪問客数の少ない都道府県の観測誤差が大きくなること、が考えられる。

Model 4 の推定結果より、時間帯別の観光魅力度に着目してみるとパラメータの値が昼の観光魅力度より夜の観光魅力度の方が大きくなっていることが確認できる。このことから、観光周遊行動に対して夜の観光魅力度が大きく影響を与えている可能性があることが示唆される。

Model 5～Model 16 では、データを季節別に分割してモデル推定を行っている。説明変数に関してそれぞれ Model 5～Model 8 は昼夜別／季節別、Model 9～Model 12 は昼夜共通／季節別、Model 13～Model 16 は昼夜共通／全季節としたモデルである。Model 5～Model 8, Model 9～Model 12, Model 13～Model 16 それぞれの最終対数尤度の和をとると、Model 9～Model 12 の最終対数尤度の和の方が大きいことが分かり、モデル推定の精度が良いことが分かる。このことから、説明変数に関して昼夜分けずにモデルを構築したほうが良いことが分かる。しかし、使用データが季節別のモデル (Model 5～Model 8, Model 9～Model 12, Model 13～Model 16) の最終対数尤度の和がいずれも、使用データが全季節である場合の最終対数尤度より小さいため、推定を行う場合は使用データが全季節のモデル構築を行ったほうが適切であることが分かった。その原因として、データ分割に伴う標本数の減少により観測誤差が大きくなったことなどが考えられる。

次に季節ごとのモデルの特徴に着目すると、Model 5～Model 8, Model 9～Model 12, Model 13～Model 16 の推定結果より春のモデルの一般化費用パラメータが最も大きいため観光周遊行動を行う際の交通費用抵抗が小さいことが分かる。一方、冬のモデルの一般化費用パラメータが最も小さいため、交通費用抵抗が大きいことがいえる。このことは、冬は多くの観光地を巡らずに少数の観光地を訪問するにとどまる傾向にあることを示している。一方、春は各所で桜が見られ、移動自体が観光行動として成立する可能性が高く、交通費用の影響が小さくなる傾向にある可能性がある。

表 1. モデル詳細

モデル番号	変数／パラメータの設定	使用データ
Model 1	変数: 昼夜共通, 全季節／パラメータ: 昼夜共通 (観光魅力度: 観光資源数)	全データ
Model 2	変数: 昼夜共通, 全季節／パラメータ: 昼夜共通, 全季節 (観光魅力度: Twitterコメント数)	全データ
Model 3	変数: 昼夜別, 全季節／パラメータ: 昼夜共通, 全季節 (観光魅力度: Twitterコメント数)	全データ
Model 4	変数: 昼夜別, 全季節／パラメータ: 昼夜別, 全季節 (観光魅力度: Twitterコメント数)	全データ
Model 5	変数: 昼夜別, 季節別／パラメータ: 昼夜別, 季節別 (観光魅力度: Twitterコメント数)	春 (4-6月) の旅行者
Model 6	変数: 昼夜別, 季節別／パラメータ: 昼夜別, 季節別 (観光魅力度: Twitterコメント数)	夏 (7-9月) の旅行者
Model 7	変数: 昼夜別, 季節別／パラメータ: 昼夜別, 季節別 (観光魅力度: Twitterコメント数)	秋 (10-12月) の旅行者
Model 8	変数: 昼夜別, 季節別／パラメータ: 昼夜別, 季節別 (観光魅力度: Twitterコメント数)	冬 (1-3月) の旅行者
Model 9	変数: 昼夜共通, 季節別／パラメータ: 昼夜共通, 季節別 (観光魅力度: Twitterコメント数)	春 (4-6月) の旅行者
Model 10	変数: 昼夜共通, 季節別／パラメータ: 昼夜共通, 季節別 (観光魅力度: Twitterコメント数)	夏 (7-9月) の旅行者
Model 11	変数: 昼夜共通, 季節別／パラメータ: 昼夜共通, 季節別 (観光魅力度: Twitterコメント数)	秋 (10-12月) の旅行者
Model 12	変数: 昼夜共通, 季節別／パラメータ: 昼夜共通, 季節別 (観光魅力度: Twitterコメント数)	冬 (1-3月) の旅行者
Model 13	変数: 昼夜共通, 全季節／パラメータ: 昼夜別, 季節別 (観光魅力度: Twitterコメント数)	春 (4-6月) の旅行者
Model 14	変数: 昼夜共通, 全季節／パラメータ: 昼夜別, 季節別 (観光魅力度: Twitterコメント数)	夏 (7-9月) の旅行者
Model 15	変数: 昼夜共通, 全季節／パラメータ: 昼夜別, 季節別 (観光魅力度: Twitterコメント数)	秋 (10-12月) の旅行者
Model 16	変数: 昼夜共通, 全季節／パラメータ: 昼夜別, 季節別 (観光魅力度: Twitterコメント数)	冬 (1-3月) の旅行者

表 2. 推定結果

	Model 1		Model 2		Model 3		Model 4	
	推定値	t値	推定値	t値	推定値	t値	推定値	t値
一般化旅行費用 [100円]	-7.278	-320.90	-7.470	-355.40	-7.387	-349.70	-7.465	-349.85
滞在→移動遷移ダミー	-7.059	-310.99	-6.175	-274.20	-6.253	-278.27	-6.165	-252.00
観光魅力度 (施設数又はツイート数) [1000件]	2.413	118.29	0.197	691.30	0.435	623.66	-	-
観光魅力度 (昼)	-	-	-	-	-	-	0.343	45.05
観光魅力度 (夜)	-	-	-	-	-	-	0.454	45.45
移動→滞在遷移ダミー	1.900	74.15	1.951	78.90	1.921	78.37	1.930	73.58
最終対数尤度	-82924.19		-74629.43		-75260.64		-74671.92	
標本数 (旅行者数)	15,444		15,444		15,444		15,444	
標本数 (通過リンク数)	131,758		131,758		131,758		131,758	
	Model 5		Model 6		Model 7		Model 8	
	推定値	t値	推定値	t値	推定値	t値	推定値	t値
一般化旅行費用 [100円]	-7.297	-115.58	-7.425	-231.25	-7.425	-227.72	-7.628	-147.01
滞在→移動遷移ダミー	-6.351	-111.98	-6.111	-140.65	-6.161	-145.53	-5.601	-103.20
観光魅力度 (施設数又はツイート数) [1000件]	-	-	-	-	-	-	-	-
観光魅力度 (昼)	0.895	23.35	1.469	23.54	1.457	25.40	2.735	25.27
観光魅力度 (夜)	1.305	23.14	2.013	23.45	2.100	26.73	2.493	26.70
移動→滞在遷移ダミー	1.434	23.04	2.070	42.50	2.138	46.39	1.794	30.87
最終対数尤度	-10898.95		-24994.51		-26273.79		-12942.71	
標本数 (旅行者数)	2,410		4,978		5,616		2,440	
標本数 (通過リンク数)	20,113		43,149		47,218		21,278	

	Model 9		Model 10		Model 11		Model 12	
	推定値	t値	推定値	t値	推定値	t値	推定値	t値
一般化旅行費用 [100円]	-7.287	-117.18	-7.413	-231.50	-7.425	-229.96	-7.648	-147.23
滞在→移動遷移ダミー	-6.393	-124.28	-6.095	-145.54	-6.196	-153.54	-5.656	-106.71
観光魅力度（施設数又はツイート数） [1000件]	0.534	216.26	0.850	359.44	0.864	389.78	1.317	241.66
観光魅力度(昼)	-	-	-	-	-	-	-	-
観光魅力度(夜)	-	-	-	-	-	-	-	-
移動→滞在遷移ダミー	1.477	25.39	2.064	44.19	2.167	48.97	1.835	32.32
最終対数尤度	-10893.81		-24973.32		-26261.82		-12934.79	
標本数（旅行者数）	2,410		4,978		5,616		2,440	
標本数（通過リンク数）	20,113		43,149		47,218		21,278	

	Model 13		Model 14		Model 15		Model 16	
	推定値	t値	推定値	t値	推定値	t値	推定値	t値
一般化旅行費用 [100円]	-7.373	-116.67	-7.428	-234.48	-7.423	-229.78	-7.481	-146.19
滞在→移動遷移ダミー	-6.343	-119.22	-6.110	-140.50	-6.215	-147.16	-5.683	-103.28
観光魅力度（施設数又はツイート数） [1000件]	-	-	-	-	-	-	-	-
観光魅力度(昼)	0.221	21.29	0.185	22.31	0.185	21.91	0.222	20.57
観光魅力度(夜)	0.226	22.09	0.187	22.40	0.189	22.45	0.229	21.21
移動→滞在遷移ダミー	1.521	24.89	2.060	42.32	2.141	45.86	1.816	30.63
最終対数尤度	-10698.54		-24873.35		-26578.25		-12999.23	
標本数（旅行者数）	2,410		4,978		5,616		2,440	
標本数（通過リンク数）	20,113		43,149		47,218		21,278	

	最終対数尤度の和
Model 5-8	-75109.96
Model 9-12	-75063.74
Model 13-16	-75149.37

## 5. 本研究の整理と今後の展望

本研究では、動的離散選択モデルである Recursive logit モデルを援用することで、(事実上) 無数にある訪日外国人観光客の周遊行動パターンを全て選択肢に内包し、種々の観光政策効果を評価可能な観光周遊行動モデルの構築を試みた。また、観光地魅力度を定義する際に、SNS データ (Twitter データ) を用いることにより、各観光地の季節間及び昼夜間の観光魅力度の違いを反映するモデルに拡張した。実証分析の結果得られた主要な知見として、第一に、昼よりも夜の観光魅力度が観光周遊行動の意思決定への影響が大きいことが示された。言い換えると、夜の観光魅力度の改善が観光客数の増加において重要であることが示唆された。第二に、春は最も移動抵抗が少ない一方、冬は移動抵抗が最も高くなることが示された。ただし、Twitter データの標本数の制約から、観光地魅力度を昼夜別/季節別に分けることによるモデル精度改善はみられなかった。

今後の課題として、第一に、観光魅力度を Twitter のツイート数で定量化しているが、コメント内容を精緻に分析した上で観光魅力度の定量化することが望ましい。第二に、国籍によって観光周遊パターンが異なる可能性があり、訪日外国人観光客の属性に着目した分析を行うことで、より精緻なモデル構築が可能になる可能性が高い。第三に、モデルの精緻化によって、幹線道路の整備や広域公共交通網の整備が観光周遊行動に及ぼす影響や、昼間/夜の観光魅力度の向上が観光周遊行動に及ぼす影響を高い精度でシミュレーションすることができる。また、観光地内の行動パターンを再現する観光周遊行動モデルを構築し、そのログサム変数を通じて各観光地の集計的な魅力を定義することにより、観光地内の周遊行動に関する政策から、観光地間の交通整備に関する政策まで、一貫したフレームワークのもとで政策の評価・比較が可能になることが期待される。

## 6. 参考文献

- 1) Fosgerau, M, Frejinger, E., Karlström A. (2013) A link based network route choice model with unrestricted choice set, Transportation Research Part B: Methodological 56, 70-80.
- 2) 福永晃彦 (2019) 時空間ネットワークを用いた観光周遊行動モデルの構築, 広島大学大学院国際協力研究科, 修士論文.
- 3) Mai, T., Bastin, F. and Frejinger, E. (2016) A decomposition method for estimating recursive logit based route choice models. EURO Journal on Transportation and Logistics, 1-23.
- 4) Oyama, Y., & Hato, E. (2019). Prism-based path set restriction for solving Markovian traffic assignment problem. Transportation Research Part B: Methodological, 122, 528-546.