

最尤推定による回遊行動モデリング Modeling of Rambling Activities Based on Maximum Likelihood Estimation

鈴木優伽[†]
Yuka Suzuki

斉藤和巳[†]
Kazumi Saito

風間一洋[‡]
Kazuhiro Kazama

1. はじめに

本研究では、観光地等における回遊者の将来行動予測を妥当な精度で可能にする回遊行動モデルの構築と、その有効性の評価を基本課題とする。具体的には、従来研究 [1] と同様、回遊者の行動プロセスが LevyFlight [2] に従うと仮定し、観光地間の距離や人気度などの属性に対するパラメータを導入することで、回遊行動モデルを構築する。本研究の特色として、モデルの解析的特性からではなく、統計的機械学習でパラメータを推定する。これにより、ある程度複雑で多様なパラメータを組み込むことが可能となり、従来研究より柔軟なモデルの構築が可能である。さらに、回遊者の行動予測性能とパラメータ推定精度の安定性を用いて評価し、提案モデルの妥当性を示す。

2. 回遊行動モデル

観測される観光スポット集合を $\mathcal{S} = \{s, t, v, \dots\}$ 、スポット s からスポット t までの距離を $d(s, t)$ とする。このとき、回遊者の行動プロセスが LevyFlight に従うとすると、スポット s から t への推移確率 $p_1(t|s; \theta_1)$ は次式となる。

$$p_1(t|s; \theta_1) = \frac{d(s, t)^{-\theta_1}}{\sum_{v \in \mathcal{S}} d(s, v)^{-\theta_1}} \quad (1)$$

ここで、 θ_1 は距離に対するパラメータであり、 $p_1(t|s; \theta_1)$ はベキ則に従う。また、スポット t の人気度を $f(t)$ 、人気度に対するパラメータを θ_2 とすると、スポット t の選択確率 $p_2(t; \theta_2)$ は次式で表せる。ただし、スポット $v \in \mathcal{S}$ とする。

$$p_2(t; \theta_2) = \frac{f(t)^{\theta_2}}{\sum_{v \in \mathcal{S}} f(v)^{\theta_2}} \quad (2)$$

いま我々は、回遊者は式 (1), (2) に従って回遊行動を行うと仮定し、回遊行動モデル $p(t|s; \theta_1, \theta_2)$ を以下で定義する。

$$p(t|s; \theta_1, \theta_2) = \frac{p_1(t|s; \theta_1)p_2(t; \theta_2)}{\sum_{v \in \mathcal{S}} p_1(v|s; \theta_1)p_2(v; \theta_2)} \quad (3)$$

また、距離と人気度以外の属性 n に対するパラメータを θ_n とし、各属性に対するパラメータベクトルを $\Theta = (\theta_1, \theta_2, \dots, \theta_N)^T$ とする。ただし、ここで $(\theta_1, \theta_2, \dots, \theta_N)^T$ は $(\theta_1, \theta_2, \dots, \theta_N)$ の転置を表す。このとき、回遊行動モデルは次式のように一般化できる。

$$p(t|s; \Theta) = \frac{p_1(t|s; \theta_1) \prod_{n=1}^N p_n(t; \theta_n)}{\sum_{v \in \mathcal{S}} p_1(v|s; \theta_1) \prod_{n=1}^N p_n(v; \theta_n)} \quad (4)$$

本研究では、モデル構築の第一歩として、距離と人気度のみを考慮し、パラメータベクトル $\Theta = (\theta_1, \theta_2)^T$ とする。

3. パラメータ推定

機械学習アプローチでパラメータを推定する。詳細は以下の通りである。回遊者の集合を $\mathcal{M}$ とし、回遊者 $m \in \mathcal{M}$ がステップ h で訪れたスポットを $r(m, h)$ 、訪れた総スポット数を H_m とすると、回遊者 m が訪れたスポット系列 $r(m) = \{r(m, 1) \rightarrow \dots \rightarrow r(m, h) \rightarrow \dots \rightarrow r(m, H_m)\}$ と書き表せる。また、全回遊者が訪れたスポット系列の集合を $\mathcal{R} =$

$\{r(1), r(2), \dots, r(M)\}$ とする。この時、観測データ $\mathcal{R}$ に対する対数尤度 $L(\mathcal{R}; \Theta)$ が最大となるような Θ を推定する。ただし、 $h = 1$ の時、 $p(r(m, 1)|r(m, 0))$ は $p_2(r(m, 1); \theta_2)$ に従うとする。

$$L(\mathcal{R}; \Theta) = \log \prod_{m=1}^M \prod_{h=1}^{H_m} p(r(m, h)|r(m, h-1); \Theta) \quad (5)$$

ここで、 $x_{m,h,v,1} = -\log d(r(m, h-1), v)$ 、 $x_{m,h,v,2} = \log f(v)$ とし、 $\mathbf{x}_{m,h,v} = (x_{m,h,v,1}, x_{m,h,v,2})^T$ と表すと、 $p(r(m, h)|r(m, h-1); \Theta)$ は以下のように書き表せる。

$$\begin{aligned} p(r(m, h)|r(m, h-1); \Theta) &= \frac{p_1(r(m, h)|r(m, h-1); \theta_1)p_2(r(m, h); \theta_2)}{\sum_{v \in \mathcal{S}} p_1(v|r(m, h-1); \theta_1)p_2(v; \theta_2)} \\ &= \frac{d(r(m, h-1), r(m, h))^{-\theta_1} f(r(m, h))^{\theta_2}}{\sum_{v \in \mathcal{S}} d(r(m, h-1), v)^{-\theta_1} f(v)^{\theta_2}} \\ &= \frac{\exp(\Theta^T \mathbf{x}_{m,h,r(m,h)})}{\sum_{v \in \mathcal{S}} \exp(\Theta^T \mathbf{x}_{m,h,v})} \end{aligned}$$

すなわち、

$$\begin{aligned} L(\mathcal{R}; \Theta) &= \Theta^T \sum_{m=1}^M \sum_{h=1}^{H_m} \mathbf{x}_{m,h,r(m,h)} \\ &\quad - \sum_{m=1}^M \sum_{h=1}^{H_m} \log \left(\sum_{v \in \mathcal{S}} \exp(\Theta^T \mathbf{x}_{m,h,v}) \right) \end{aligned}$$

となり、以下の更新式で Θ を算出する。

$$\Delta \Theta = - \left(\frac{\partial^2 L(\mathcal{R}; \Theta)}{\partial \Theta \partial \Theta^T} \right)^{-1} \left(\frac{\partial L(\mathcal{R}; \Theta)}{\partial \Theta} \right) \quad (6)$$

4. データセット

本研究では、オンライン写真共有サイト Flickr に投稿された位置情報付き写真群を学習データとして用いる。ただし、投稿された位置情報付き写真群を Mean Shift クラスタリングした結果を使って、スポットと人気度を定義する。具体的には、多くの写真が撮られた地点ほど重要なスポットと考えられるため、各クラスターの極大点を本研究におけるスポット t とし、各クラスターの総撮影数を人気度 $f(t)$ とする。

5. 評価実験

5.1. パラメータ推定精度

本研究における回遊行動モデルを、スポット系列をリサンプリングした際のパラメータ推定精度の安定性から評価する。M 人のスポット系列をサンプリング割合 α でリサンプリングした集合を $\mathcal{R}_\alpha$ とし、 $\mathcal{R}_\alpha$ におけるパラメータを式 (5) から学習する。ただし、パラメータの推定精度は α に依存するため、 α に応じた推定パラメータ $\hat{\Theta}$ の値を図 1 に示す。詳細には以下のアルゴリズムで $\hat{\Theta}$ を算出する

St1: $\mathcal{R}_0 \leftarrow \emptyset$, $\alpha = \beta = 0.1$ とする。

St2: $\mathcal{R} - \mathcal{R}_{\alpha-0.1}$ からランダムに βM 人を選択する。リサンプリングしたスポット系列の集合を $\mathcal{Q}$ とし、 $\mathcal{R}_\alpha \leftarrow \mathcal{Q} \cup \mathcal{R}_{\alpha-0.1}$ を構成する

St3: $\alpha = 1$ ならば処理を終了。

さもなければ $\alpha \leftarrow \alpha + 0.1$ として St2 へ。

[†]静岡県立大学

[‡]和歌山大学

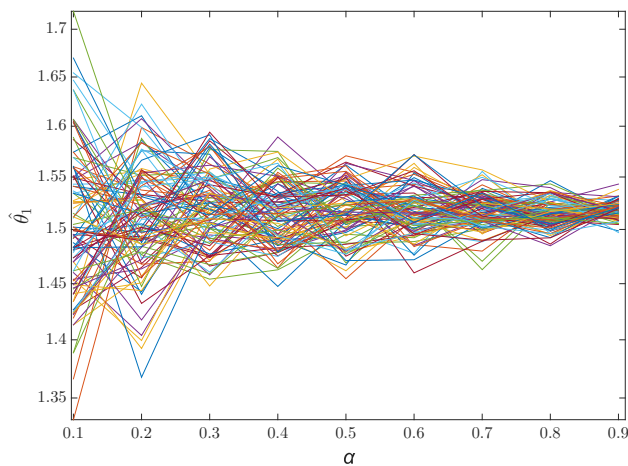


図 1: 京都の距離パラメータ推定結果

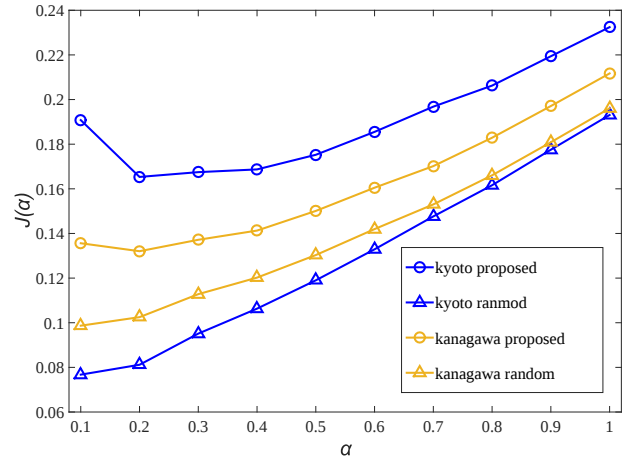


図 2: 各地域における滞在スポット一緻度

5.2. 行動予測精度

本研究における回遊行動のモデル設定が妥当であれば、提案回遊モデルで生成した回遊行動と実際の回遊行動が高い精度で一致すると考えられる。そのため、回遊者の実際の滞在スポットと、モデルに従って生成した滞在スポットとの平均一緻度を用いて、モデルの妥当性を評価する。推定パラメータ $\Theta = (\theta_1, \theta_2)^T$ によって規定される確率分布に従って生成される、回遊者 m が訪れたスポット系列を $\bar{r}(m)$ とする。ただし、 $\bar{r}(m, 1) = r(m, 1)$ とする。また、 $\bar{r}(m, h) (h > 1)$ は、観測されるスポット $s \in \mathcal{S}$ から、 $p(s|r(m, h-1); \Theta)$ に従って生成される。今、 $r(m)$ を回遊者 m が実際に訪れたスポットの集合、 $\bar{r}(m)$ をモデルで生成した回遊者 m の推定滞在スポットの集合とすると、平均一緻度 $J(\alpha)$ は以下で定義される。

$$J(\alpha) = \frac{1}{|\mathcal{R}_\alpha|} \sum_{m \in \mathcal{R}_\alpha} \frac{|r(m) \cap \bar{r}(m)|}{|r(m) \cup \bar{r}(m)|} \quad (7)$$

6. 実験結果

6.1. パラメータ推定精度の結果

図 1 に京都における、 α でリサンプリングした毎のパラメータ推定結果を示す。縦軸に推定パラメータ値、横軸を α とする。ただし、紙面の都合上、本稿では距離パラメータ θ_1 の推定結果のみとする。また、リサンプリングした組合せ全てにおける推定結果をプロットすると膨大となってしまうため、ランダムに選択した 100 通りのリサンプリング結果を異なる色でプロットした。表 1 に各地域における全てのデータを用いた場合のパラメータ推定結果を示す。

表 1: 推定パラメータ Θ の結果

地域	ユーザ数	スポット数	θ_1	θ_2
京都	1601	2060	1.51	0.75
神奈川	2603	4968	1.47	0.82

図 1 から、提案モデルは $\alpha = 0.5$ ほどからパラメータの値が安定することが確認できる。すなわち、全回遊者数に対して半分ほどの回遊者の行動データが存在すれば、比較的高い精度で各パラメータを推定できる。また、この傾向は京都と神奈川の両地域において共通する。以上のことから、提案モデルは、比較的小さいサンプル数でも安定した結果を得ることが出来ると考えられる。

6.2. 行動予測精度の分析

図 2 に京都・神奈川における滞在スポットの一緻度を示す。青線で京都を、黄色点線で神奈川を表し、提案モデルにお

ける一緻度を○マーカ、ランダムに生成した場合の一緻度を△マーカで示す。京都・神奈川の両地域において、ランダムに生成した場合よりも、提案モデルの方が滞在スポットの一緻度が高い傾向にあり、回遊者の行動を妥当な精度でモデル化できていると考えられる。ただし、京都に比べ神奈川の方が提案法での一緻度が低い傾向がある。この理由として、各スポットが京都盆地内に密集している京都に対して、神奈川は、平野部が山岳部に分断され、各スポットが分散して存在していることから、回遊者の行動パターンが多様になったことが考えられる。この問題を解決するためには、地域の特性などを考慮する必要があると考えられる。

7. おわりに

本研究では、観光地等における回遊者の行動モデルを構築し、パラメータ推定精度の安定性や、実際の回遊行動との一緻度からモデルの有効性・妥当性を検証した。結果として、提案モデルが比較的高い安定性をもつこと、実際の回遊行動を妥当な精度でモデル化できていることを確認できた。今後は、距離・人気度以外の属性を追加することで、各回遊者や対象とする地域の特性を考慮したモデルに拡張していく。

謝辞

本研究は、総務省 SCOPE(No.142306004) 及び、科研費 (No.26330345) の補助を受けた。

参考文献

- [1] C.Song, T.Koren, P.Wang, A-L.Barabasi, "Modeling The Scaling Properties Of Human Mobility," Nature Physics 6:818-823, 2010
- [2] D.Comaniciu, P.Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," IEEE Transactions on PAMI 24:603-619, 2002
- [3] D.Brockmann, L.Hufnagel, T.Geisel, "The Scaling Law of Human Travel," Nature Physics 439:462-465, 2006