

GPS 位置履歴を用いた時間制約を考慮した 観光施設の滞在時間モデリング

岩本 睦大¹ 坪内 孝太² 下坂 正倫¹

概要：観光地や商業施設において、利用者の滞在時間の傾向は効果的な観光施策、商業施策を打つなどの、観光業マネジメントにおいて重要な分析対象となっている。また利用者に対しても、観光施設での滞在時間目安がわかることは、旅行や移動プランを立てるといった意思決定に対する大きな要素である。滞在時間に関する分析・研究は、アンケート調査による集計データを基にした研究が盛んに行われてきた。しかしながら、アンケート調査の高コストさ、アンケート実施ごとの結果の不定性などの問題が指摘されている。本研究では、スマートフォンなどの携帯端末から日々得られる GPS 位置履歴データをもちいて、観光施設利用者の滞在時間傾向を曜日、祝日といった外的要因ごとにモデリングする。GPS 位置履歴データを用いることで、これまでのアンケート調査に基づく分析よりも低コストな滞在時間分析を可能とし、より空間・時間的に高精細な分析を可能とする。観光施設における、人々の訪問・滞在時間は、施設の営業時間に対応した有界領域内のデータとなるため、その性質を考慮したモデリングが必要となる。そのような性質を考慮するため、ディリクレカーネルといった非対称カーネルを用いた人口推定手法により、高精度な滞在時間傾向分析が可能であることを、基本的なガウシアンカーネルのような対称カーネルを用いた手法との比較により示す。

キーワード：GPS, 位置情報解析, 滞在時間, ディリクレカーネル

MUTSUHIRO IWAMOTO¹ KOTA TSUBOUCHI² MASAMICHI SHIMOSAKA¹

1. 序論

観光施設の滞在時間傾向は、特に観光マネジメントにおいて多くの分析研究があり、旅行者の意思決定にも大きな影響を与える要因として重要な研究対象となっている [4], [5]。本研究では、観光施設の滞在時間傾向を、GPS ログから取得した、ある訪問時間における滞在時間ごとの人口に基づいてモデリングする。従来、観光施設の滞在時間傾向に関する研究は、アンケートに基づいて収集される滞在時間データを利用して行われてきた [4], [5]。しかし、継続した滞在時間傾向の把握において、アンケートは高コストであり、また、アンケート実施ごとの内容や結果の不安定さが度々問題となる [6]。アンケート以外のデータを用いた研究として、空港のチェックインデータを用いた滞在校期间に関するモデリング [6] があるが、空港のデータである以上、島や町単位での広範囲な地域に対する滞在時間傾

向の把握にとどまり、ある特定の観光施設ごとに滞在時間傾向を把握するには不十分である。アンケートから得られるデータや、空港のチェックインデータでは、ランドマークごと、といった地理的に詳細な分析が困難であり、滞在時間についても、「滞在校期间」(e.g., 宿泊日数) を把握するに止まっている。GPS データを用いた活動人口モデリングに関する先行研究 [8], [9] にあるように、GPS データの活用によって地理的に詳細かつ、時間的にも詳細な滞在時間傾向の把握が可能となる。

滞在時間傾向を、GPS データを用いてモデリングする際には、観光地の営業時間に対応したある有界領域に対するモデリングを考える必要がある。GPS ログには、営業時間外への訪問や、営業時間を超えた滞在時間を示すログが含まれてしまうからである。有界なデータ点のモデリングを考えた時、ガウシアンカーネルなどを用いた基本的な密度推定手法では、無限領域のデータ点に対するモデリングとなり、対象領域外のデータへのフィッティングが起きてしまう。有界なデータ点に対するモデリング手法として、

¹ 東京工業大学 Tokyo Institute of Technology

² Yahoo! JAPAN 研究所 Yahoo! Japan Corporation

Chen ら [2], [3] は、ベータカーネルといった非対称カーネルを用いることにより、より高精度な有界なデータ点のモデリングを実現している。また、そのようなカーネルを用いることにより、観光地の営業時間という実世界の制約を取り込んだモデリングが可能となる。

本研究では、GPS から取得した、観光施設に訪問時間・滞在時間ごとの人口から滞在時間傾向をモデリングする。GPS データを用いたモデリングでは、観光施設の営業時間に起因する訪問・滞在時間の制約により、モデリング対象となるデータが有界領域のデータとなる。滞在時間傾向を、訪問・滞在時間傾向によりモデリングし、さらにディリクレカーネル [1], [7] を用いることで、観光施設の時間制約を考慮したモデリング手法を提案する。

本研究の貢献は以下のようにまとめられる。

- GPS データを用いたモデリングによって、従来のアンケートを基にした滞在時間分析の問題点であった高コスト、結果の不安定さだけでなく、より地理的・時間的に詳細な滞在時間傾向分析の枠組みを提案。
- GPS データを用いた滞在時間分析において発生する、観光施設の営業時間の制約を考慮したモデルによって、適切で、高精度な滞在時間傾向分析を可能とした。

2 章では、滞在時間傾向モデリングの問題設定と、前提手法を、3 章では、提案手法の定式化を行う。4 章では、実際の GPS データを用いたモデリング精度評価を、営業時間制約を考慮しない場合との比較によって行う。5 章は結論として、本研究のまとめ、貢献、将来課題についてのべる。関連研究

ここでは、観光施設等の滞在日数や傾向に関する調査、および GPS 位置履歴を用いた活動人口動態のモデリングに焦点を当てた研究について述べる。

Decrop ら [4] は、観光における意思決定のなされ方や傾向を明らかにするために、アンケートによる調査を行っている。この研究では、ベルギーの 25 家庭を対象に、夏季休暇中の過ごし方における意思決定をアンケートにより調査し、その意思決定の過程はとも変化に富み、決定者の意思に強く影響されやすいことを示している。また同様のアンケートをベースにした滞在時間に関する調査研究には、Golovali らの研究 [5] が存在し、アンケートから得られた滞在時間を生存分席の手法を用いて解析している。しかし、アンケートを用いた滞在時間に関する研究は、滞在時間の傾向を掴む程度にとどまり、安定した滞在時間予測を提供するには至っていない。

これに対し、Montano らの研究 [6] では、空港のチェックインデータを用いた滞在日数予測手法を提案している。この研究ではスペインのパレアリック島の空港における到着と出発が記録されたチェックインデータから、島の訪問者がどの程度の滞在日数を有していたか回帰分析を用いて予測する手法を提案している。しかし、滞在時間のモデ

リングに焦点を当てた時、この研究を適用することはできない。

これに対し、GPS 位置履歴が持つ時空間的特性を用いて、都市や観光地における活動人口のモデリングに焦点を当てた研究 [9] が存在する。OO らの手法では、テンソル分解を用いて観光地の活動人口の動態に関する解析を行なっているが、これは予測に焦点を当てた研究ではない。これに対して、Shimosaka らの手法 [9] では、地域や天気、曜日といった説明変数から活動人口を予測する双線型ポアソン回帰を提案している。この手法では、説明変数に滞在時間を追加することで、滞在時間別の訪問者の予測に容易に拡張可能であるが、訪問者の滞在時間の有界性により、単純な説明変数の追加では予測が不安定になる。

2. 問題設定と前提手法

本研究では、観光施設の滞在時間傾向モデリングとして、1 日の訪問時間、滞在時間ごとの人口予測モデルを考える。GPS から得られる個人の位置履歴から移動履歴を集計し、観光施設を覆うような長方形領域内での移動履歴の最初の時刻を訪問時間とし、最後の時刻までの時間を滞在時間と定義する。ある観光施設 l のある 1 日 d に対して、訪問時間帯と滞在時間帯を表す点 t に対応する人口 $y_{d,t}^{(l)}$ をモデリングする問題である。

2.1 双線型ポアソン回帰による活動人口モデリング

GPS 位置履歴から人口のモデリングを行う先行研究として、下坂ら [9] の低ランク双線型ポアソン回帰モデルがあり、ある 1 日の活動人口の推移を曜日や祝日か否かといったコンテキスト c を考慮し、あるデータ点 t における人口がポアソン分布に従うと仮定したモデリングを行なっている。

$$p(y_t^{(l,c)}) = \mathcal{P}(y_t^{(l,c)} | \lambda_s) = \frac{\lambda_t^{(l,c) y_t^{(l,c)}} \exp(-\lambda_t^{(l,c)})}{\Gamma(y_t^{(l,c)} + 1)}. \quad (1)$$

この手法では、コンテキスト c を表現する説明変数 $\varphi(c) \in \mathbb{R}^M$ と T 個のデータ点 t を表現する説明変数 $\phi(t) \in \mathbb{R}^T$ を用いた回帰によって $y_t^{(l,c)}$ を推定する。

ある日のコンテキストを表現する説明変数は、曜日や、祝日か否かといったいくつかのコンテキストの組み合わせ M 個に対して、どの組み合わせかを指し示す one-hot エンコーディングによって表現する。また、ある i 番目のデータ点 t_i に対する説明変数は、ガウス分布 $\mathcal{N}(\cdot)$ およびそのパラメータ μ と Σ を用いて $\{\phi(t)\}_i = \mathcal{N}(t_i | \mu, \Sigma)$ で表す。

以上の説明変数と重みパラメータ $\mathbf{W}_l \in \mathbb{R}^{M \times T}$ から、ポアソン分布のパラメータ $\lambda_t^{(l,c)} > 0$ を以下の式から推定する。

$$\ln \lambda_t^{(l,c)} = \varphi(c)^\top \mathbf{W}_l \phi(t) \quad (2)$$

下坂ら [9] はさらに、重みパラメータを $\mathbf{W}_l = \mathbf{U}_l \mathbf{V}_l^\top$ なる低ランク近似によって $\mathbf{U}_l \in \mathbb{R}^{M \times K}$ と $\mathbf{V}_l \in \mathbb{R}^{T \times K}$ (ただし, $K \ll T, K \ll M$) に分解し学習の安定性を向上させている。

3. 提案手法

3.1 有界領域における人口推定

訪問時間・滞在時間ごとの GPS ログをプロットすると、図 1 のようになる。観光施設の始業時間 T_o と、終業時間 T_c を考慮した時、訪問時間 s ・滞在時間 τ に対応するデータ点は、 $T_o \leq s \leq T_c, s + \tau \leq T_c$ で表される有界領域内に存在する。このため、GPS 位置履歴を用いた観光施設の滞在時間傾向のモデリングでは、施設ごとの営業時間を考慮する必要がある。

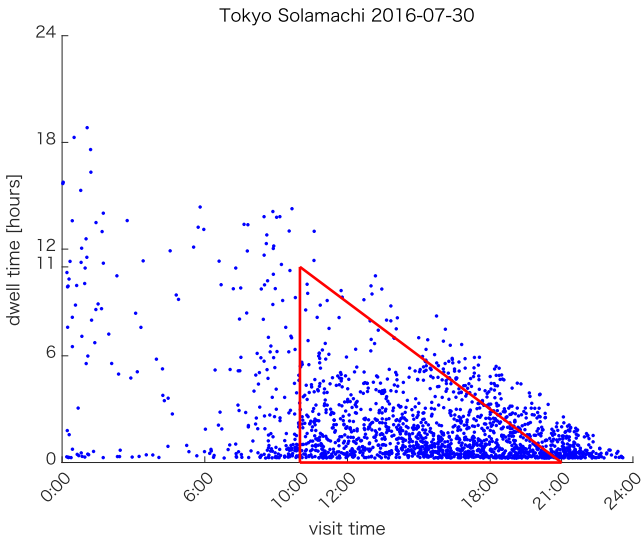


図 1 施設利用者ログの訪問時刻・滞在時間ごとのプロット

有界データに対するモデリング手法として、非対称カーネルを用いる手法が提案されている [2], [3]。本研究では、施設の営業時間に対応した有界領域データのモデリングのため、ディリクレカーネルを組み込んだモデリングを提案する。

ディリクレカーネルはシンプレックスなベクトル $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ に対して以下で定義できる。 $\mathbf{x}$ がシンプレックスなベクトルであるとは $\sum_i x_i = 1, x_i \geq 0$ であることを言い、 b はハイパーパラメータでバンド幅である。

$$k_D(\mathbf{x}, \mathbf{y}; b) = \frac{\Gamma(\|\boldsymbol{\alpha}\|_1)}{\prod_{i=1}^{d+1} \Gamma(\alpha_i)} \prod_{i=1}^{d+1} \beta_i^{\alpha_i-1}, \quad (3)$$

where,

$$\boldsymbol{\alpha} = \left(\frac{x_1}{b} + 1, \dots, \frac{x_d}{b} + 1, \frac{1 - \|\mathbf{x}\|_1}{b} + 1 \right)^\top, \\ \boldsymbol{\beta} = (y_1, \dots, y_d, 1 - \|\mathbf{y}\|_1)^\top.$$

3.2 施設の営業時間制約を考慮した滞在時間モデリング

訪問時間・滞在時間に対応する訪問時間および滞在時間 $s, \tau \geq 0$ に対応するデータ点は、施設の始業・終業時間 T_o, T_c に対し、 $T_o \leq s \leq T_c, s + \tau \leq T_c$ からなる 3 角形の有界データとなる。本研究では、このようなデータの有界性を考慮したモデリングのために、下坂ら [9] の低ランク双線型ポアソン回帰にディリクレカーネル [1], [7] を取り入れたモデルを提案する。

訪問時間・滞在時間に対応する人口は、以下のように定義できる。

- GPS 位置履歴からは、その有界領域外のデータも得られ、それらを見逃したモデリングが必要となる。
- ガウシアンカーネルを用いたモデリングでは営業時間外の人口を図 2 のように不適切に計上してしまうが、ディリクレカーネルを用いることによって図 3 のように解決。

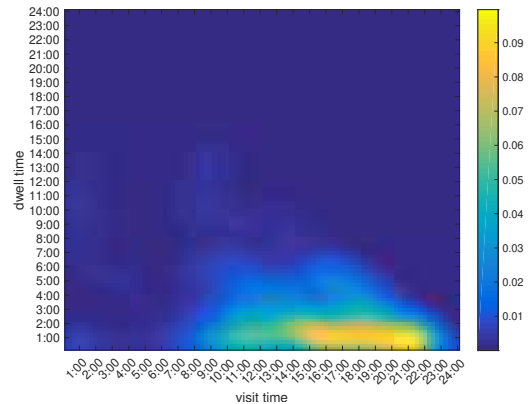


図 2 ガウシアンカーネルを用いたモデリング

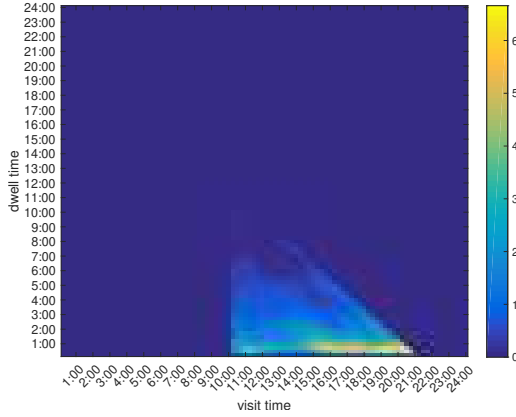


図 3 ディリクレカーネルを用いたモデリング

- 観光施設 l の始業時間が $T_o^{(l)}$ であり、終業時間が $T_c^{(l)}$ とした時、(2) を基に、(4) によってパラメータ $\lambda_{c,s,\tau}^{(l)}$ を推定する。
- ここで、 $\mathbf{k}_{s,\tau}^{(l)} \in \mathbb{R}^{ST}$ は、訪問時間 s および滞在時間 τ が指し示すデータ点が $\mathbf{x} = \left(\frac{s - T_o^{(l)}}{T_c^{(l)} - T_o^{(l)}}, \frac{\tau}{T_c^{(l)} - s} \right)$ で表され、 s, τ に対応するデータ点が $\mathbf{x}$ が j 番目の点であるとき、(3) で定義されるディリクレカーネルを用いて計算されるグラム行列 $K_D^{(l)}$ の j 番目の列ベクトルである。

$$\lambda_{c,s,\tau}^{(l)} = \exp \left(\mathbf{d}_c^\top \mathbf{U} \mathbf{V}^\top \mathbf{k}_{s,\tau}^{(l)} \right) \quad (4)$$

4. 実験

4.1 データセット

Yahoo!防災速報アプリから得られた匿名化された GPS 位置情報履歴を用いた。位置情報履歴には、タイムスタンプと緯度経度情報がある。2016 年 todo 月 todo 日から 2017 年 todo 月 todo 日の間の位置情報履歴から、緯度経度情報をもとに、解析対象 50 箇所の観光施設における滞在履歴を集計したものを利用している。解析対象施設の例を以下に示す。

4.2 性能評価指標

性能評価にあたり、以下のように定式化される、平均絶対予測誤差 (MAE: Mean Absolute Error) および、平均負の対数尤度 (MNLL: Mean Negative Log Likelihood) を用いる。

$$\text{MAE} = \frac{1}{DST} \sum_{d=1}^D \sum_{s=1}^S \sum_{\tau=1}^T \left| y_{s,\tau}^{(l,d)} - \hat{\lambda}_{s,\tau}^{(l,d)} \right|. \quad (5)$$

$$\text{MNLL} = -\frac{1}{DST} \sum_{d=1}^D \sum_{s=1}^S \sum_{\tau=1}^T \ln p(y_{s,\tau}^{(l,d)} | \hat{\lambda}_{s,\tau}^{(l,d)}). \quad (6)$$

MAE および、MNLL いずれにおいても、値が小さくなるほどモデルの精度が良いことを示す。

4.3 ポアソン回帰モデルの特徴量設定

訪問・滞在時間ごとの人口の予測対象日のコンテキスト情報として、対象日が、何曜日であるか、また祝日であるか否か、を用いた。曜日に関しては 7 通り、祝日か否かについては 2 通りのコンテキストを持ちそれぞれを one-hot エンコーディングによる特徴量設計を行う。すなわち、(2) における、予測対象日のコンテキストに対応する説明変数は、曜日に関する説明変数 $\boldsymbol{\varphi}_1(\mathbf{c}) \in \mathbb{R}^7$ と祝日か否かを表す説明変数 $\boldsymbol{\varphi}_2(\mathbf{c}) \in \mathbb{R}^2$ を用いて、14 次元の one-hot ベクトル $\boldsymbol{\varphi}(\mathbf{c}) = \boldsymbol{\varphi}_1(\mathbf{c}) \otimes \boldsymbol{\varphi}_2(\mathbf{c})$ で定義する。

4.4 比較手法

比較手法として、データが有界であることを想定しない、ガウスカーネルを時間特徴量に用いたポアソン回帰モデルを用いる。ガウスカーネルは (7) のように定義でき、時間特徴量は (8) のように定義する。

$$k_G(\mathbf{x}, \mathbf{y}; b) = \exp \left(-\frac{\|\mathbf{x} - \mathbf{y}\|_2^2}{b} \right) \quad (7)$$

$$\{\phi(t)\}_{t'} = k_G(\mathbf{t}, \mathbf{t}') \quad (8)$$

4.5 性能評価実験

- ディリクレカーネルを用いた提案手法と、時間制約を考慮できないガウスカーネルを用いたモデルとの精度比較実験 (5 交差検定)。
- 観光施設の営業時間内での、訪問・滞在時間ごとの人口予測精度を、MAE, MNLL で評価する。
- 訪問時間別に最大であるカウントデータに対する滞在時間の誤差を MAPE として評価する。

表 1 精度評価

	MAE	MNLL
Gaussian kernel	todo	todo
Dirichlet kernel	todo	todo

5. 結論

- GPS 位置履歴を用いた観光施設における滞在時間傾向のモデリング手法を提案した。
- 滞在時間傾向モデリングを GPS 位置履歴で適切に行うためには観光施設の営業時間制約をモデルに取り込む必要があり、それをディリクレカーネルを用いた方法で実現。

本研究の貢献は以下

- 観光施設における滞在時間傾向を GPS 位置履歴から

得られるデータを用いた予測モデルを構築，滞在時間傾向の分析の低コスト化．

- 観光施設の営業時間を組み込むことによる適切，高精度な滞在時間傾向モデリング手法の提案．

参考文献

- [1] Aitchison, J. and Lauder, I. J.: Kernel Density Estimation for Compositional Data, *J. R. Stat. Soc. Ser. C Appl. Stat.*, Vol. 34, No. 2, pp. 129–137 (1985).
- [2] Chen, S. X.: Beta kernel estimators for density functions, *Comput. Stat. Data Anal.*, Vol. 31, No. 2, pp. 131–145 (1999).
- [3] Chen, S. X.: BETA KERNEL SMOOTHERS FOR REGRESSION CURVES, *Stat. Sin.*, Vol. 10, No. 1, pp. 73–91 (2000).
- [4] Decrop, A. and Snelders, D.: Planning the summer vacation - An adaptable process, *Annals of Tourism Research*, Vol. 31, No. 4, pp. 1008–1030 (online), DOI: 10.1016/j.annals.2004.03.004 (2004).
- [5] Gokovali, U., Bahar, O. and Kozak, M.: Determinants of length of stay: A practical use of survival analysis, *Tourism Manage.*, Vol. 28, No. 3, pp. 736–746 (2007).
- [6] Montaña, J., Rosselló, J. and Sansó, A.: A new method for estimating tourists’ length of stay, *Tourism Management*, Vol. 75, pp. 112 – 120 (オンライン), DOI: <https://doi.org/10.1016/j.tourman.2019.04.009> (2019).
- [7] Ouimet, F.: Density estimation using Dirichlet kernels (2020).
- [8] Shimosaka, M., Hayakawa, Y. and Tsubouchi, K.: Spatiality preservable factored Poisson regression for large-scale fine-grained GPS-based population analysis, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, pp. 1142–1149 (2019).
- [9] Shimosaka, M., Maeda, K., Tsukiji, T. and Tsubouchi, K.: Forecasting urban dynamics with mobility logs by bilinear Poisson regression, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp ’15, New York, NY, USA, Association for Computing Machinery, pp. 535–546 (2015).