

# 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<sup>1</sup> 坪内 孝太<sup>2</sup> 下坂 正倫<sup>1</sup>

**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。従来、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。他方、低コストかつ空間・時間的に高精細な人口動態モデリング手法として、GPS 位置履歴を用いた活動人口の解析・予測手法が近年盛んに研究されている。この予測問題を、各時刻、各滞在時間帯を予測する問題として定義することで、既存の活動人口予測手法の拡張により滞在時間のモデリングを行うことが可能である。しかし、これらの予測手法の滞在時間モデリングへの安直な拡張は、観光施設特有の訪問者の滞在時間の有界性を考慮できないため、予測が不安定になりやすい。そこで本研究では、滞在時間の有界性という性質を考慮するため、ディリクレカーネルを用いた手法により、高性能な滞在時間モデリングが可能であることを、有界性を考慮しない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 IWAMOTO<sup>1</sup> KOTA TSUBOUCHI<sup>2</sup> MASAMICHI SHIMOSAKA<sup>1</sup>

## 1. 序論

観光施設における訪問者の滞在時間傾向は、観光資源量や魅力度・その施設の混雑具合に大きく左右され、また、観光地における企画のプランニング等の計画に有用な情報である。例えば、滞在時間が短い時間帯において、イベントやタイムセールといったの施策を実施することで滞在時間を増加させ、それに伴い、物販等の利益を増加させることが可能となる。従って、観光マネジメントの観点から、観光施設における訪問者の滞在時間をモデリング・予測することが重要な問題である。

従来、観光施設の滞在時間傾向の把握に関する研究は、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 [3], [4]。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、同時に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。これに対し、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期間のモデ

リングに関する研究 [5] が存在する。しかしこの手法は、島や町単位での広範囲な地域に対する滞在日数傾向を予測するにとどまり、ある特定の観光施設ごとに滞在時間傾向を細かく把握することはできない。このように、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を用いた分析・予測手法では、特定の休日を対象にしたランドマークごとのモデリングといった、地理的および時間的に詳細な分析を行うことが難しい。

一方で、昨今の GPS 機能を搭載した携帯電話やスマートフォンなどのモバイルデバイスの急速な普及により、都市や観光地における人々の活動の様子を記録した GPS 位置履歴情報がますます蓄積されている。これに伴い、近年、GPS 位置履歴がもつ時空間的特性を用いた活動人口の解析・予測手法が盛んに研究されており [6], [7], [8]、観光地においてもこれらの地理的・時間的に詳細な滞在時間モデリング手法の応用が期待される。これらの手法では、GPU ログを用いて地理的な関係性や時間的な関係性を考慮した回帰モデルを用いて、都市部における高精度かつ高精細な活動人口モデリングを実現した。

この活動人口モデリング手法について、各時間ごとに予

<sup>1</sup> 東京工業大学 Tokyo Institute of Technology

<sup>2</sup> Yahoo! JAPAN 研究所 Yahoo! Japan Corporation

測していた問題を、各時間における各滞在時間を予測する問題と定義することによって、滞在時間のモデリングを行うことが可能である。しかし、滞在時間のモデリングにおいて、既存手法の安直な拡張および適用は、予測の精度劣化に繋がる。何故ならば、GPS 位置履歴には、営業時間外への訪問や、営業時間を越えた滞在時間を示すログが含まれることが多く、これらがモデル構築の際にノイズになりうるからである。つまり、既存手法の単純な適用では、商業施設における始業時間や就業時間を考慮できないため、予測は不安定になりやすい。

そこで本研究では、観光地特有の始業時間や終業時間といった時間に関する特性を考慮したモデリング手法を提案する。これは、営業時間から考えられる滞在時間の上限を制約として導入することで、観光施設における時間制約を考慮しながら滞在時間モデリングを実現する。

本研究の貢献は以下のようにまとめられる。

- 既存の GPS を用いた滞在時間分析において、観光施設における滞在時間に対する制約である営業時間を考慮しないことによる問題を指摘した。
- 観光施設の営業時間の制約を導入したモデリング手法を提案し、観光施設の特性を反映した高精度な滞在時間分析を可能とした。
- 全国の代表的な観光施設 50 箇所における滞在時間推定実験により、観光施設の営業時間を考慮することによって、既存の手法と比べ、高精度に滞在時間推定を行うことが可能であることを示した。

## 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 [3] 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ても変化に富み、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 [4] が存在し、アンケートから得られた滞在時間を生存分析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 [5] 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 [8] 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 [8] 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のモデリングを目的として、観光施設における 1 日の訪問時間、滞在時間ごとの活動人口予測モデルを考える。GPS 位置情報のログから得られる個人の位置履歴を用いて移動履歴を集計し、観光施設を覆うような矩形領域内での移動履歴の最初のタイムスタンプを訪問時刻とし、最後のタイムスタンプまでの時間を「滞在時間」と定義する。

1 日を  $S$  個の訪問時間帯  $s$ 、 $T$  個の滞在時間帯  $\tau$  に分割し、この時間帯に分割し、訪問時間帯と滞在時間帯を表現する。このとき、ある観光施設  $l$  における日付  $d$  に対して、訪問時間帯と滞在時間帯を表すデータ点  $t = (s, \tau)$  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 位置履歴から活動人口のモデリングを行う先行研究として、Shimosaka ら [8] の低ランク双線型ポアソン回帰モデルがある。この手法では、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、ある訪問時間  $t$  における活動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。

$$y_t^{(l,c)} \sim \text{Pois}(y_t^{(l,c)} | \lambda_s) = \frac{\lambda_t^{(l,c)} y_t^{(l,c)} \exp(-\lambda_t^{(l,c)})}{\Gamma(y_t^{(l,c)} + 1)}. \quad (1)$$

この手法では、天気や曜日、平日か休日かといったコンテキスト  $c$  を表現する説明変数  $\varphi(c) \in \mathbb{R}^M$  と  $T$  個のデータ点  $t$  を表現する時間的特徴量  $\phi(t) \in \mathbb{R}^T$  を用いた回帰によって  $y_t^{(l,c)}$  を推定する。

ある日付  $d$  のコンテキストを表現する説明変数は、曜日や平日か休日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 1-of-k 符号化法により表現する。また、時間的特徴量  $\phi(t)$  は以下のように定式化される。

$$\phi(t) = \{t_s | t_s = \mathcal{N}(s | \tau, \sigma^2), s = 1, \dots, T\}, \quad (2)$$

ここで、 $\mathcal{N}(\cdot)$  は正規分布を表し、 $\tau$  はその平均、 $\sigma^2$  がその分散である。

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。

$$\ln \lambda_t^{(l,c)} = \varphi(\mathbf{c})^\top \mathbf{W}_l \phi(t) \quad (3)$$

Shimosaka ら [8] 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、学習対象のパラメータ数を抑制することで学習の安定性を向上させている。

しかしこの手法は、ある訪問時間帯のみにおける活動人口の予測に焦点を当てた研究であり、本研究の目標である滞在時間帯も考慮した活動人口の予測は行えない。

### 3. 提案手法: 営業時間の有界性を考慮した訪問・滞在時間帯別活動人口の予測モデル

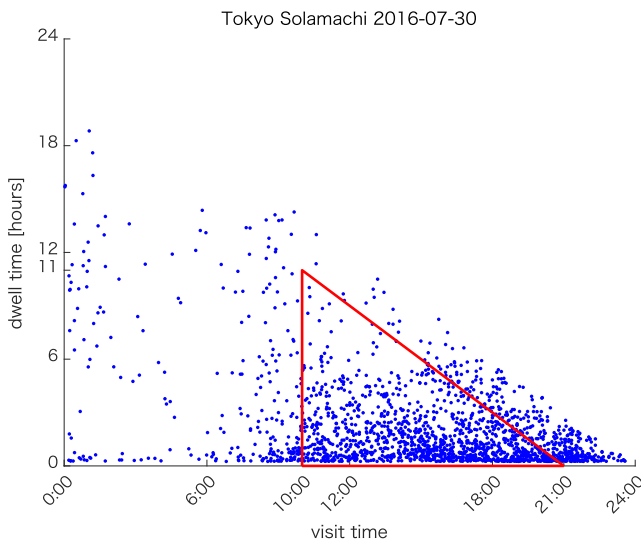


図 1 施設利用者ログの訪問時刻・滞在時間ごとのプロット

観光地の営業時間の有界性を考慮した滞在時間モデリングを実現するために、ディリクレカーネル [1], [2] を用いた滞在時間帯別の活動人口予測手法を提案する。

#### 3.1 滞在時間帯を説明変数への追加による双線型ポアソン回帰の拡張

滞在時間を考慮した活動人口の予測手法の構築のため、2.1 章で述べた双線型ポアソン回帰の拡張を行う。

訪問時間帯  $s$ 、滞在時間帯  $\tau$  に対応するデータ点  $\mathbf{t} = (s, \tau)$  が  $j$  番目のデータ点であるとき、それに対応する時間特徴

量  $\phi(\mathbf{t})$  を、カーネルを用いて計算されるグラム行列  $\mathbf{K}$  の  $j$  列目の列ベクトルを  $\mathbf{k}_j$  と定義することで、カーネルによる非線形写像空間上での一般線型回帰モデルとして以下で定式化できる。

$$\ln \lambda_t^{(l,c)} = \varphi(\mathbf{c})^\top \mathbf{W}_l \mathbf{k}_j \quad (4)$$

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帯  $s$  と滞在時間帯  $\tau$  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。

#### 3.2 ディリクレカーネルを用いた営業時間の有界性を考慮した活動人口の予測モデル

観光地の営業時間の有界性を考慮するために、データ点の有界性を明示的に扱えるようなカーネルを用いて、説明変数を構築することを考える。有界データに対するカーネルを用いた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 [1], [2]。そこで本研究では、非対称カーネルの一種である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 simplex なベクトル  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$  に対して (5) で定義できる。

$$k_D(\mathbf{x}, \mathbf{y}; b) = \frac{\Gamma(\|\boldsymbol{\alpha}\|_1)}{\prod_{i=1}^{d+1} \Gamma(\alpha_i)} \prod_{i=1}^{d+1} \beta_i^{\alpha_i - 1}, \quad (5)$$

where,

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

$$\boldsymbol{\beta} = (y_1, \dots, y_d, 1 - \|\mathbf{y}\|_1)^\top.$$

$\mathbf{x}$  が simplex なベクトルであるとは  $\sum_i x_i = 1, x_i \geq 0$  であることを指し、 $b$  はバンド幅を表し、ハイパーパラメータとして与える。

ここで、観光施設  $l$  の始業時間が  $T_o^{(l)}$  であり、終業時間が  $T_c^{(l)}$  であるとした時、訪問時間  $s$  および滞在時間  $\tau$  が指し示すデータ点をディリクレカーネルの導入のため、 $\mathbf{t} = \left( \frac{s - T_o^{(l)}}{T_c^{(l)} - T_o^{(l)}}, \frac{\tau}{T_c^{(l)} - s} \right)$  として、営業時間に対応する指定領域を simplex な空間にスケールして表す。

最終的に、ポアソン分布のパラメータ  $\ln \lambda_t^{(l,c)}$  は以下で表される。

$$\ln \lambda_t^{(l,c)} = \varphi(\mathbf{c})^\top \mathbf{W}_l \phi(\mathbf{t}) \quad (6)$$

先行研究 [8] 同様、重みパラメータ  $\mathbf{W}_l$  に対して、 $\mathbf{W}_l = \mathbf{U}_l \mathbf{V}_l^\top$  なる低ランク近似によって  $\mathbf{U}_l \in \mathbb{R}^{M \times K}$

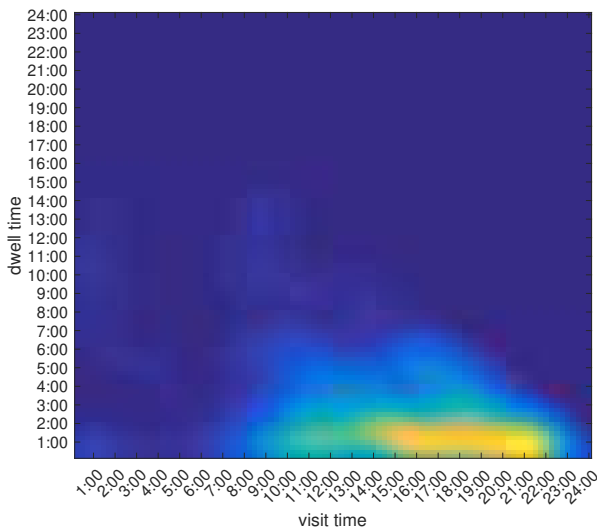


図 2 ガウシアンカーネルを用いたモデリング

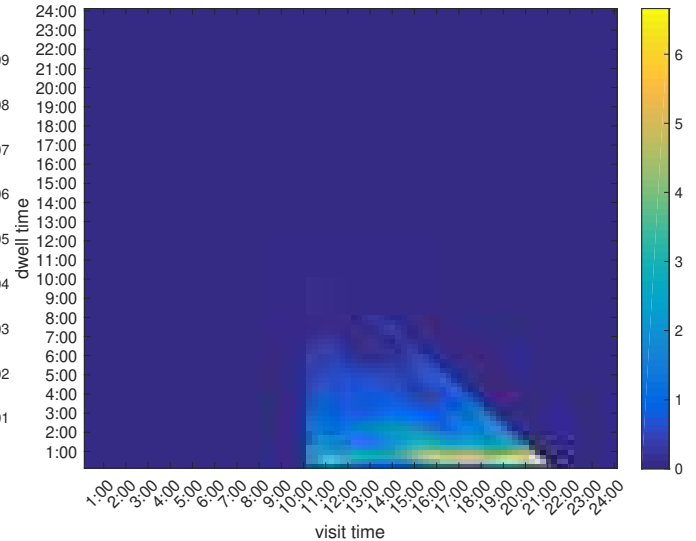


図 3 ディリクレカーネルを用いたモデリング

と  $V_l \in \mathbb{R}^{T \times K}$  ( $K \ll T, K \ll M$ ) に分解し、パラメータ数を削減することによって学習の安定化をはかる。

ディリクレカーネルを導入したモデルによる図 1 から得られる人口の推定結果が、図 3 のようになる。

## 4. 性能評価実験

観光地の滞在時間モデリングにおける提案手法の性能を評価するため、実際の携帯端末の位置履歴データを用いて予測精度の評価実験を行った。本章では、その性能評価実験の目的やデータセット、評価指標、実験結果等に関して述べる。

### 4.1 実験データ

本実験では、携帯電話による GPS 位置履歴データ、具体的には Yahoo!Japan の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 7 月 1 日から 2017 年 9 月 30 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 47 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。

(表)

### 4.2 性能評価指標

本実験では、提案手法の訪問・滞在時間帯別活動人口の予測に対する性能を評価する。そこで性能評価としては、活動人口予測の従来研究にて用いられてきた、平均絶対予測誤差 (MAE: Mean Absolute Error) を用いる。

$$\text{MAE} = \frac{1}{DST} \sum_{d=1}^D \sum_{s=1}^S \sum_{\tau=1}^T \left| y_{s,\tau}^{(l,d)} - \hat{\lambda}_{s,\tau}^{(l,d)} \right|. \quad (7)$$

MAE の値が小さくなるほどモデルの性能が優れていることを表す。

### 4.3 ポアソン回帰モデルの特徴量設計

活動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 1-of-k 符号化法による特徴量設計を行う。すなわち、(6) 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  $\varphi_1(c) \in \mathbb{R}^7$  と祝日か否かを表す説明変数  $\varphi_2(c) \in \mathbb{R}^2$  を用いて、14 次元の one-hot ベクトル  $\varphi(c) = \varphi_1(c) \otimes \varphi_2(c)$  で定義する。

### 4.4 比較手法

比較手法として、3.1 章にて述べた、ガウスク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。このモデルでは、観光地の営業時間に対して滞在時間が有界であることを仮定しない。ガウスクーネルは (8) のように定義でき、時間特徴量は (9) のように定義する。

$$k_G(\mathbf{x}, \mathbf{y}; b) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|_2^2}{b}\right) \quad (8)$$

$$\{\phi(\mathbf{t})\}_{\mathbf{t}'} = k_G(\mathbf{t}, \mathbf{t}') \quad (9)$$

### 4.5 実験結果

観光地の営業時間に対して、滞在時間が有界であることを考慮するディリクレカーネルを用いた提案手法と、4.4 で述べた、滞在時間が有界であることを考慮しない手法で比較実験を行う。

表 1 MAE による精度評価

	MAE
Gaussian kernel	3.35
Dirichlet kernel(Proposed)	3.32

滞在時間が有界であることを考慮した提案手法が、MAE において優っており、有界領域を考慮してモデリングすることの有効性がしめされた。

#### 4.6 定性的評価

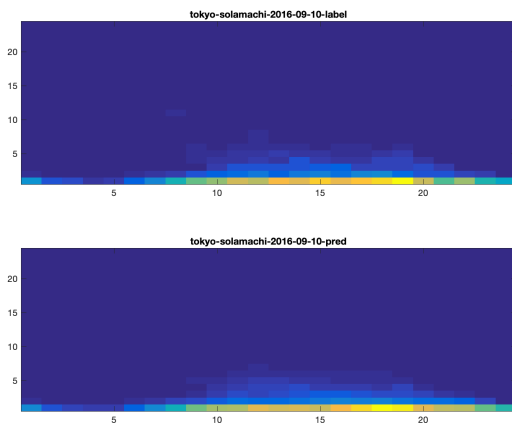


図 4 有界領域を考慮しないモデルによる予測結果. 上図がラベルデータで、下図が予測結果である。横軸が訪問時間、縦軸が滞在時間を表す。

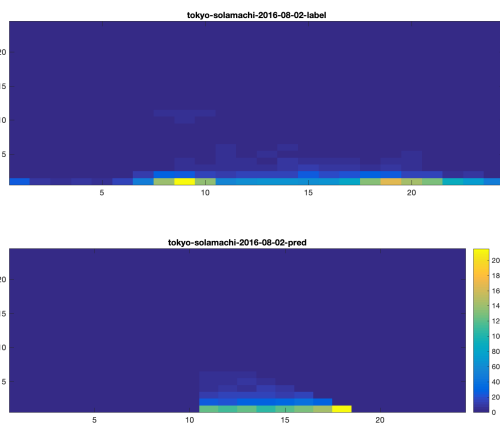


図 5 有界領域を考慮するモデルによる予測結果. 上図がラベルデータで、下図が予測結果である。横軸が訪問時間、縦軸が滞在時間を表す。

図 4 が、有界領域を考慮しない 4.4 で定義したモデルによる、滞在時間傾向の予測結果である。この結果では、有界領域外のデータへのフィッティングが見て取れる。一方で、図 5 は、有界領域を考慮した提案手法による滞在時間

傾向の予測結果である。この結果からは、三角形の有界領域のみにおいてデータのフィッティングができており、滞在時間の有界領域内のみを適切にモデリングできていることがわかる。

#### 5. 結論

本研究では、GPS 位置履歴を用いた、観光地における営業時間の有界性を考慮した滞在時間のモデリングを行った。従来のアンケートや空港のチェックインデータを用いた観光地の滞在期間傾向の把握に関する研究は、データ取得のコストや詳細な滞在時間の分析の難しさという面で課題を抱えていた。そこで、滞在時間帯別の地理的・時間的に詳細な活動人口予測を目的として、GPS 位置履歴を用いた既存の活動人口予測手法を拡張し、さらに観光施設における営業時間の制約により起こる予測性能の劣化が起こることを確認し、その制約により発生する訪問者の滞在時間の有界性を考慮するため、ディリクレカーネルを用いた予測手法を提案した。

全国各地の代表的な観光施設 47 箇所を対象に、実際の携帯電話から取得した GPS 位置履歴を用いた性能評価実験を行い、観光施設における営業時間の制約を考慮しない予測手法と比較し、滞在時間別の活動人口の予測において、平均絶対誤差 (MAE) で 0.03 の精度改善を実現した。また、訪問時間帯・滞在時間帯別の活動人口の予測における定性的な評価により、提案手法が訪問者の滞在時間の有界性を捉えられていることを示した。

将来課題としては、開店前や終業前の行列などにより観光施設が設定している営業時間に必ずしも一致しない現象があり、制約を導入した影響で考慮できない問題が存在するが、これをデータから最適な就業時間を同時推定するなどにより解決することが挙げられる。

#### 参考文献

- [1] Chen, S. X.: Beta kernel estimators for density functions, *Comput. Stat. Data Anal.*, Vol. 31, No. 2, pp. 131–145 (1999).
- [2] Chen, S. X.: BETA KERNEL SMOOTHERS FOR REGRESSION CURVES, *Stat. Sin.*, Vol. 10, No. 1, pp. 73–91 (2000).
- [3] 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).
- [4] 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).
- [5] 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).
- [6] Nishi, K., Tsubouchi, K. and Shimosaka, M.: Extracting Land-Use Patterns Using Location Data from Smartphones, *Proceedings of the First International*

*Conference on IoT in Urban Space*, Brussels, BEL, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), p. 38–43 (online), DOI: 10.4108/icst.urb-iot.2014.257220 (2014).

- [7] 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).
- [8] 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).