

携帯端末位置履歴を用いた 階層ディリクレ混合回帰モデルに基づく活動人口予測

下坂 正倫^{1,a)} 早川 裕太^{1,b)} 坪内 孝太^{2,c)}

概要: 近年の携帯端末の普及により、端末の位置履歴情報を用いた都市における人々の活動を解析することが盛んである。特に都市の活動人口を予測することは、都市計画や人員配置の観点から重要となる。この問題に対して、曜日や天気といった要因を説明変数とした回帰手法が近年盛んに研究されている。これまで提案された双線形回帰に基づく手法は複数の特徴の組み合わせを考慮した予測が可能である。一方で解析対象となる地域に対して独立に予測を行っており、例えば地域間の関係性に着目した分析を具現化しようとする、地域毎に独立したモデル化は安定しないことが知られている。そこで、本研究ではこの問題を解決するため、予測モデルの事前分布に階層ディリクレ過程 (HDP) を適用した、HDP 混合回帰モデルを提案する。HDP は複数分布に共通するパラメータを生成が可能な性質を持つため、パラメータとデータをそれぞれ共有し安定性を高めつつ、解析対象の地域ごとに特化した予測モデルの構築が可能となる。携帯電話により取得された 3200 万件を超える位置履歴データを用いて、大規模かつ高精細なメッシュ上での活動人口予測の比較実験を行い、提案手法が既存手法と比較し安定した学習が可能であり、特にデータの少ない地域に対しても高精度なモデリングを可能にすることを示す。

キーワード: 都市動態予測, 位置履歴情報, 階層ベイズモデル, HDP

MASAMICHI SHIMOSAKA^{1,a)} YUTA HAYAKAWA^{1,b)} KOTA TSUBOUCHI^{2,c)}

1. 序論

都市における人々の活動を解析することは非常に有用であり、都市計画、店舗やサービス等の配置の効率化につなげることができる。近年の携帯端末の普及により、端末の位置履歴情報が非常に多く蓄積されており、都市における人々の流れを反映する。そのため端末の位置履歴情報から都市の人々の活動（都市動態）を解析する研究が盛んである。

例えば、Fan ら [4] は 1 日ごとの都市の活動人口のパターンが地域の性質を表現することに着目し、位置履歴情報からその代表的なパターンを抽出した。こういった活動人口パターンの抽出は、都市動態の解析における主要なアプ

ローチであり多くの研究が行われている [7], [11]。パターン抽出結果を利用することで都市ごとの人口動態の特性を把握することができ、効果的な店舗の出店や人員配置、広告配置につなげることが可能となる。

一方で、それぞれの都市における未来の活動人口を予測することも効果的である。Shimosaka ら [10] は、曜日や天気といった要因に対して変化する活動人口の予測に取り組んでいる。こうした活動人口予測は、天候に依存する売り上げ予測や交通量予測 [8] など、より具体的な実応用を可能にする。既存研究では解析対象となる地域に対してそれぞれ独立に予測が行われてきた。しかしながら、一つ一つの地域のデータ数には限りがあり、不安定な学習につながる恐れがある。特に、より詳細な都市動態解析を目的に解析対象の地域を高精細に分割した場合、一つ一つの地域でのデータ不足は顕著になる。分割の細かさと同トレードオフに解析対象地域の数は増加し、パラメータ増加につながる。

こうした問題を解決するために、複数地域を同時に学習し、地域間でデータを共有することが考えられる。扱う地域数を増やした際にもパラメータ数の増加を抑えることが

¹ 東京工業大学 情報理工学院 情報工学系
Department of Computer Science, School of Engineering,
Tokyo Institute of Technology

² ヤフー株式会社 Yahoo! JAPAN 研究所
Yahoo Japan Corporation, Yahoo! JAPAN Research

a) simosaka@miubiq.cs.titech.ac.jp

b) hayakawa@miubiq.cs.titech.ac.jp

c) ktsubouc@yahoo-corp.jp

できれば学習は安定する。

都市動態解析において複数地域を同時に学習する枠組みとして、都市動態パターン抽出のためにテンソル分解を用いた研究 [4], [12], [16], [17], [19] や階層ベイズの枠組みを用いた研究 [7], [11] がある。これらの手法は複数地域でデータを共有しながら学習を行い、各地域に共通する潜在的なパターンを効果的に抽出している。一方でこれらの研究は都市動態予測に焦点を当てたものではない。

そこで本研究では、地域間でデータを共有しながら、未来の都市動態を予測することを目的とし、予測モデルの事前分布に階層ディリクレ過程 (HDP) を適用した、HDP 混合回帰モデルを提案する。HDP は複数分布に共通するパラメータを生成可能な性質を持つため、パラメータとデータをそれぞれ共有し安定性を高めつつ、解析対象の地域ごとに特化した予測モデルの構築が可能となる。さらに、各地域のモデルは潜在的なモデルの混合として表現されるため、潜在的なモデルのパラメータとその混合比のみの学習が行われ、パラメータ削減が実現される。

本論文の貢献は以下ようになる。

- 活動人口予測のために地域間でデータを共有しながら学習する枠組みを提案する。地域間のパラメータを共有して学習することでデータ数不足を解消する。従来手法に比べパラメータ数を削減し、過学習のリスクを小さくしている。
- 実際の携帯電話から得られた 3200 万件程度の位置履歴を用いて予測実験を行い、既存の回帰手法や生成的なモデリング手法と比較し安定したモデリング精度が得られることを示す。

以降、2 節では関連研究について述べ、3 章では都市動態モデリングの問題設定、4 章では提案手法について述べる。5 章では携帯電話の位置履歴情報を利用した性能評価実験を行い、6 章で本研究の貢献と将来課題をまとめた。

2. 関連研究

ここでは都市活動人口モデリングの既存研究について述べる。都市活動人口のモデリング手法としては、人口推移の代表的なパターンを抽出する研究が盛んに行われている。パターン抽出を実現する手法として、テンソル分解を用いた手法 [4], [12], [16], [17], [19] や混合モデルを用いた手法 [7], [11] が盛んに研究されてきた。

テンソル分解を用いた手法では、多くの研究が、都市活動量データを地域、時間、日付を軸としたテンソルの形で表現し、分解によって潜在的なパターンを抽出している。Zheng ら [19] は都市で生じる騒音データをテンソルによってモデリングし、テンソル分解によってランク圧縮を行い、そのパターンを算出している。また Takeuchi ら [12] は人々のチェックイン行動をテンソルによってモデリングしテンソル分解によってそのパターンを抽出し、地域と

チェックイン先のカテゴリの関係性について論じている。Fan [4] らは携帯電話の位置履歴ログをテンソルデータとみなし、市街地の基本的なパターンを抽出した。また、東日本大震災前後で被災地でのパターンがどのように変化しているかを分析している。

都市における活動量のパターンモデリングのため混合モデリングを用いた研究 [7], [11] もある。Nishi ら [7] は一日の活動人口を正規分布でモデリングし、ディリクレ過程混合を用いて地域間に共通する活動人口のパターンを抽出している。また Shimosaka ら [11] は混合モデリングにより地域間に共通するパターンを抽出しながら、同時にクラスタリングを行う枠組みを提供している。このクラスタリングにより都市間の類似性を明示的に示すことが可能となっている。これらの混合モデリングを用いた手法もテンソル分解同様、地域間でデータを共有しながら代表的な活動量パターンの抽出を可能にしている。しかしながら、テンソル分解や混合モデルに基づく研究は、過去のデータに潜在的に見られるパターンを抽出することを目的としており、曜日や天気などの条件によって変化する未来の活動量を与えるものではない。未来の値を予測するために、その活動量に影響を与える要因を説明変数としてモデルに組み込む必要がある。

説明変数を用いた都市活動量の予測の研究 [2], [10], [15] も多く行われている。Wang ら [15] は、人口や気候などの説明変数から、負の二項回帰によって交通量予測を行っている。また、Bogomolov ら [2] は人口統計情報からランダムフォレストを用いた犯罪数予測モデルを構築している。都市の活動人口予測の先行研究としては、Shimosaka らの研究 [10] があげられる。Shimosaka ら [10] は活動人口に大きな影響を与える時間的要因とそれ以外の要因の組み合わせを考慮した双線形ポアソン回帰によって予測を行う。組み合わせを考慮することで、平日と休日といった大きく動態が異なるパターンも効果的にモデリングすることができる。一方で、Shimosaka ら [10] は一つの地域に対して一つのモデルを作成しており、予測対象となる地域のデータのみを使用して学習する。そのため、モデルに与えられるデータが非常に少なくなる可能性がある。また、予測対象となる地域数が増えることを想定すると、学習するパラメータ数も線形に増加し学習が不安定になってしまう。これらの問題の解決のため、都市間でデータの共有を行いながら、説明変数を用いて都市活動量予測をするモデルが求められる。

Zhang ら [18] は人々の移動をテンソルによってモデリングし曜日などの説明変数を用いて予測を行っている。隣接するメッシュ間の活動量データを畳み込みニューラルネットワークを用いて畳み込むことでデータ共有を図っている。しかしながら畳み込みの効果により都市の大まかな特徴を捉えるモデルとなっており、それぞれの地域に関して

の詳細な予測を行うものではない。また、隣接しない地域とのデータ共有を行うことができない。

Okawa ら [8] は、交通量予測の問題において、双線形ポアソン回帰のパラメータを潜在的なクラスタのパラメータに分割し路線間に共通するパターンを抽出している。そして、そのパラメータの混合比を路線固有のものとする事で路線固有の予測モデルを構築している。一方でパラメータ分割の結果、複数の特徴量の組み合わせを考慮できなくなっており、その組み合わせが重要な都市動態予測において適用が難しい。より安定した予測を実現するために、説明変数を用いた予測と、隣接するかによらない地域間でのデータ共有、学習の安定性向上の両方を行うモデルが求められる。

本研究ではこれを実現するため混合モデルに回帰を組み合わせたモデルを提案する。地域ごとに混合比を学習しつつ最適な混合数を自動で決定するために HDP を事前分布として用いる。提案手法については4章で述べる。

3. 活動人口予測の問題設定

私たちは、ある領域の1日の活動人口の推移をモデル化する問題を扱う。具体的には、1日を S の時間帯に分割し、領域 l におけるある時間帯 τ の GPS のログ数を活動人口とみなして、時間帯 $\tau = \{1, \dots, S\}$ についての活動人口 $y_{l,\tau}$ の推移をモデリングをする。

既存のモデリング手法としては、活動人口予測とパターン抽出に大別される。活動人口予測では、その活動人口を時間帯 $\tau = \{1, \dots, S\}$ や領域 l 、他の条件（例えば天気、曜日、祝日など） c を説明変数として用いて、日にち n 、時間帯 τ の活動人口 $y_{c,\tau}^{(n,l)}$ を予測する。パターン抽出は1日の活動時間の推移 $\mathbf{y}_c^{(n,l)} = (y_{c,1}^{(n,l)}, \dots, y_{c,S}^{(n,l)})^\top$ をパターンと定義し、データから地域間に共通する潜在的なパターンを抽出する問題である。

3.1 双線形ポアソン回帰による活動人口予測

活動人口予測問題の先行研究として双線形ポアソン回帰による予測 [10] がある。このモデルでは活動人口 $y_{c,\tau}^{(n,l)}$ が以下のようにポアソン分布に従うと仮定している。

$$p(y_{c,\tau}^{(n,l)}) = \mathcal{P}(y_{c,\tau}^{(n,l)} | \lambda_{c,\tau}^{(l)}) = \frac{1}{y_{c,\tau}^{(n,l)}!} \lambda_{c,\tau}^{(l) y_{c,\tau}^{(n,l)}} \exp(-\lambda_{c,\tau}^{(l)}) \quad (1)$$

時間を表す特徴量とその他の特徴量を大別し、その組み合わせを考慮した予測を行っており、平日と休日といったような条件の差によるピークの差をうまくとらえることを可能としている。

時間を表す特徴量 $\phi(\tau) \in \mathbb{R}^S$ は次のように定式化される。

$$\{\phi(\tau)\}_t = \mathcal{N}(t | \tau, \sigma^2), \quad (2)$$

ここで、 $\mathcal{N}(\cdot)$ は正規分布を表し、 τ はその平均、 σ^2 が

その分散となる。時間を表さない特徴量 $\varphi(\mathbf{d}) \in \mathbb{R}^M$ は、曜日、祝日、天気などを表現しており、1-of-K 表現などによって表される。双線形ポアソン回帰はポアソン分布のパラメータ $\lambda_{c,\tau}^{(l)} > 0$ を重み行列 $\mathbf{W}_l \in \mathbb{R}^{M \times S}$ を用いて、次のように表現する。

$$\ln \lambda_{c,\tau}^{(l)} = \varphi(\mathbf{d})^\top \mathbf{W}_l \phi_\tau(\tau) \quad (3)$$

また Shimosaka ら [10] はこの重み行列を $\mathbf{W}_l = \mathbf{U}_l \mathbf{V}_l^\top$ なる、 $\mathbf{U}_l \in \mathbb{R}^{M \times K}$ と $\mathbf{V}_l \in \mathbb{R}^{S \times K}$ に分解することで低ランク化を図っている。ここで $K \ll M, K \ll S$ である。

低ランク化によって過学習を抑制する効果があるが、対象となる1地域のデータしか使用できず、データ数が不足する場合がある。また、全地域に対するパラメータ数は地域数を L とすると、 $K(M+S)L$ である。地域数に対して線形で増加するためにパラメータは非常に多くなり、学習は不安定になってしまう。

3.2 混合モデリングによるパターン抽出

パターン抽出の枠組みとしては、以下のような混合モデルを用いた研究 [7], [11] が行われている。

$$\mathbf{y}_c^{(n,l)} \sim \sum_k \pi_k p(\mathbf{y}_c^{(n,l)} | \theta_k) \quad (4)$$

全地域に対して一つのモデルを考えることで、地域間でデータを共有することが可能になっている。また、基本的にパラメータの数は混合数 K に比例する。ここで地域数 L に対して $K \ll L$ であるため、パラメータも削減できる。

Nishi ら [7] はこの混合モデルの混合係数の事前分布としてディリクレ過程 (DP) を考えることで、混合数を自動的に決定している。Shimosaka ら [11] は、事前分布として結合階層ディリクレ過程 (cHDP) を考え、パターン抽出と地域のクラスタリングを同時実現している。

本研究では地域間でのデータ共有を行いながら、地域固有の予測モデルを構築することが目的となる。DP では複数の地域について単体の回帰の割り当てを行うことになり、制約が強い。cHDP はこの割り当てについて混合比を考慮する方式であるが、DP と同様、地域間のわずかな差を表現するには制限が強い。パラメータ数の増加を抑制しながらも地域間のわずかな差を適切に考慮する方策が求められる。地域間でデータ共有をしつつ、地域間に固有のモデルを提供する事前分布としては HDP が知られている。本研究において、次章で述べる提案手法では HDP [13] を事前分布として用いた予測モデルを構築する。

4. 提案手法:HDP 混合回帰モデル

先述の通り、安定した活動人口予測を実現するためには、地域間でデータを共有しながら学習しつつもパラメータ数を抑えた、説明変数を用いた予測を行うモデルが求められ

ている。本研究ではこれを実現するために混合モデルに回帰を組み合わせたモデルを考える。各地域のモデルが、潜在的な回帰モデルの混合によって表現されていると仮定し、地域ごとに異なる混合比をもつような潜在変数を与える。こうした潜在的な配分を扱う代表的なモデルとしてはLDA[1]やHDP[13]が知られている。本研究においてはこのHDPを使用し、HDP混合ガウス回帰による活動人口予測を提案する。

HDPを用いた学習では、ベイズの枠組みを用いて学習する。予測モデルの先行研究で用いられているポアソン回帰はベイズ推定を解析的に行うことが難しい。ベイズ推定を行うため近似を行った手法[3]も提案されているが、0に近い値での近似精度が十分ではなく、精度劣化を引き起こす可能性がある。本研究では解析的にベイズ推定が可能なガウス回帰を予測モデルとして用いる。ガウス回帰がポアソン回帰と同等の精度で予測ができることは実験で別途確認する。

4.1 HDP 混合ガウス回帰

ここではHDPを用いた混合モデルによる定式化を示す。HDPを実現する棒折り過程は[14]同様Sethuramanによる表現[9]に倣い、前章で定義した活動人口 $y_c^{(n,l)}$ が属するクラスタ m を表現する潜在変数を $z_{l,n,m}$ とする。ここで $z_{l,n,m} \in \{1, 0\}$, $\sum_m z_{l,n,m} = 1$ であり、クラスタ m は地域に固有の仮クラスタとなる。仮クラスタは後述する変分推論を解析的に行うために定義される。 $z_{l,n,m} = 1$ のとき、地域 l 、の n 日目のサンプルに割り当てられたクラスタが m であることを示す。次に地域間で共有のクラスタ k と地域 l のクラスタ m の関係を表現する潜在変数を $r_{l,m,k}$ とする。同様に $r_{l,m,k} \in \{1, 0\}$, $\sum_k r_{l,m,k} = 1$ である。 $z_{l,n,m}$ 同様に、 $r_{l,m,k} = 1$ のとき、地域 l のクラスタ m が全体のクラスタ k に割り当てられていることを示す。

混合モデリングを用いた学習の際に、パターンの形に注目して学習を進めるために、スケーリング項を用いて活動人口を次のように表す。

$$\tilde{y}_{c,\tau}^{(n,l)} = \frac{1}{\eta_l} y_{c,\tau}^{(n,l)} \quad (5)$$

η_l は訓練データを用いて $\eta_l = \frac{1}{N} \sum_{n=1}^N \sum_{\tau=1}^T y_{c,\tau}^{(n,l)}$ と表される。

$\tilde{y}_{c,\tau}^{(n,l)}$, $z_{l,n,m}$, $r_{l,m,k}$ の同時分布は次のように表される。

$$p(\tilde{y}_{c,\tau}^{(n,l)}, z_{l,n,m}, r_{l,m,k}) = \prod_m \pi_{l,m}^{z_{l,n,m}} \prod_k \rho_k^{r_{l,m,k}} \mathcal{N}(\tilde{y}_{c,\tau}^{(n,l)} | \lambda_{c,\tau,k}, \sigma_k^2)^{z_{l,n,m} r_{l,m,k}} \quad (6)$$

ここで $\pi_{l,m}$, ρ_k はそれぞれ、地域ごとの混合比、全地域で共通する混合比である。 $\lambda_{c,\tau,k}$ は以下のように時間特徴量

$\phi_\tau(\tau)$ と時間以外を表現する特徴量 $\varphi(c)^\top$ を用いて表現される。

$$\lambda_{c,\tau,k} = \varphi(c)^\top \mathbf{W}_k \phi_\tau(\tau) \quad (7)$$

また $\pi_{l,m}$, ρ_k は棒折り過程によって以下のように生成される。

$$p(\pi_{l,m}') = \mathcal{B}(1, \beta_0) \quad (8)$$

$$p(\rho_k') = \mathcal{B}(1, \gamma_0) \quad (9)$$

$$\pi_{l,m} = \pi_{l,m}' \prod_{s=1}^{t-1} (1 - \pi_{l,s}') \quad (10)$$

$$\rho_k = \rho_k' \prod_{j=1}^{k-1} (1 - \rho_j') \quad (11)$$

また潜在変数 $z_{l,n,m}$ や $r_{l,m,k}$ 、パラメータ \mathbf{W}_k の事前分布はそれぞれ次のように表される。

$$p(z_l | \pi_l) = \prod_{n,m} \pi_{l,m}^{z_{l,n,m}} \quad (12)$$

$$p(\mathbf{r}) = \prod_{l,m,k} \rho_k^{r_{l,m,k}} \quad (13)$$

$$p(\mathbf{W}_k) = \mathcal{N}(\text{vec}(\mathbf{W}_k) | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (14)$$

$$p(\sigma^2) = \text{Gamma}(\sigma^2 | a_0, b_0) \quad (15)$$

一般にHDPを用いたモデルにおいて、パラメータや潜在変数の事後分布は解析的に求めることができない。その場合、事後分布を近似的に求めるため、ギブスサンプリング[5],[6]等のサンプリング手法や変分推論[14]が用いられる。本研究では上記のパラメータ及び潜在変数の事後分布の学習において、変分推論[14]を適用し、事後分布のパラメータを近似的に求めた。

上述のHDPを用いた混合モデリングによって従来手法と比べパラメータ数が抑えられていることを確認する。既存の予測モデル[10]では、パラメータの大きさは低ランク化の次元 K 、時間特徴量の次元 M 、その他の特徴量の次元 S に対して、前述の通り $K(M+S)L$ であった。例えば、予測対象の地域数を $L = 100 \times 100 = 10000$ これは各特徴量の次元として、 $M = 28$, $S = 48$, 低ランク化の次元として、 $K = 5$ があたえられる時、3.8Mのパラメータが必要である。提案手法では、パラメータ \mathbf{W}_k が地域間に共通するクラスタ数 B に依存する。混合比を与えるパラメータの数を考えても、その総数は $BMS + LT + B$ であり、それぞれ全体の最大クラスタ数 $B = 50$ 、地域ごとの最大クラスタ数を $T = 50$ とすれば、567K程度に抑えられる。さらに、扱う地域数 L を増加させた際にこの差は大きくなることから、提案手法がスケーラビリティに優れたモデルであることがわかる。

4.2 HDP 混合ガウス回帰を用いた予測

全地域に渡る訓練データ $\mathbf{Y} = \{\tilde{y}_{c,\tau}^{(n,l)}\}_{n,l,\tau}$ を与えたとき

の事後分布を,

$$p(\boldsymbol{\rho}, \boldsymbol{\pi}, \boldsymbol{r}, \boldsymbol{z}, \boldsymbol{W}, \boldsymbol{\sigma} | \boldsymbol{Y}) = q(\boldsymbol{\rho})q(\boldsymbol{\pi})q(\boldsymbol{r})q(\boldsymbol{z})q(\boldsymbol{W}, \boldsymbol{\sigma}) \quad (16)$$

なる変分近似を用いて表現する. このとき, HDP 混合ガウス回帰の予測分布は以下のように表される.

$$p^*(\hat{y}_{c,\tau}^{(l)*} | \boldsymbol{Y}) \simeq \sum_m \mathbb{E}_{q(\boldsymbol{\pi})}[\pi_{l,m}] \sum_k \mathbb{E}_{q(\boldsymbol{r})}[r_{l,m,k}] \mathbb{E}_{q(\boldsymbol{W}, \boldsymbol{\sigma})}[p(\hat{y}_{c,\tau}^{(l)*} | \lambda_{c,\tau,k}, \sigma_k^2)] \quad (17)$$

予測分布は混合分布であり多峰性関数である. そのため本研究では予測分布の最頻値に η_l を用いてスケールを戻した,

$$\hat{y}_{c,\tau}^{(l)*} = \eta_l \arg \max_{\hat{y}_{c,\tau}^{(l)*}} (p^*(\hat{y}_{c,\tau}^{(l)*} | \boldsymbol{Y})) \quad (18)$$

を予測値として用いる.

5. 大規模活動人口予測における比較実験

都市活動人口予測における, 提案手法の性能を評価するため, 既存の都市活動人口モデリング手法との比較実験を行った.

5.1 実験データ

本実験で用いるデータは Yahoo! JAPAN のアプリケーション, 防災速報*1 によってユーザ同意のもとで取得された, 位置履歴情報である. データはユーザ情報を取り除くことで匿名化し, さらに統計情報化したものであるため, 個人を特定できるデータではない. データはタイムスタンプと緯度経度の情報を持ち, 日本全国のユーザから取得されている. データ数は東京圏のみで 1 日に約 1500 万件にも及ぶ. データは基本的に移動している状態の端末や基地局が切り替わった端末から取得されるため, 人々の動きを表したデータと言える. 利用するデータの期間は 2013 年 7 月 1 日から 2014 年 6 月 30 日までの一年間である. 図 1 に示す, 3km 四方の領域内に落とされた位置履歴情報を, 各メッシュ 30 分間隔でカウントし, 訓練・評価データとする.

実験では領域内を分割するメッシュのサイズを, 図 1 に示す 600m 四方と, 図 2 に示す 200m 四方の 2 通り用意し, それぞれで実験を行う.

実験 1 では, 既存の予測手法との比較を行い, 提案手法が既存手法と比較して良い精度で予測が可能であることを確認する. 実験 2 では, データ共有を行う生成的なパターンモデリング手法との比較を行い, 同程度の予測精度が得られることを確認する.

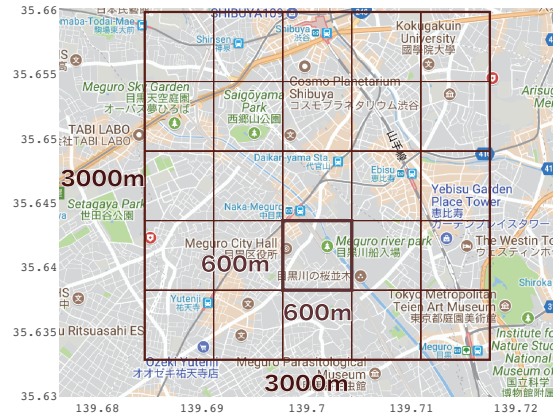


図 1: 解析対象エリア 1 : 600m メッシュ

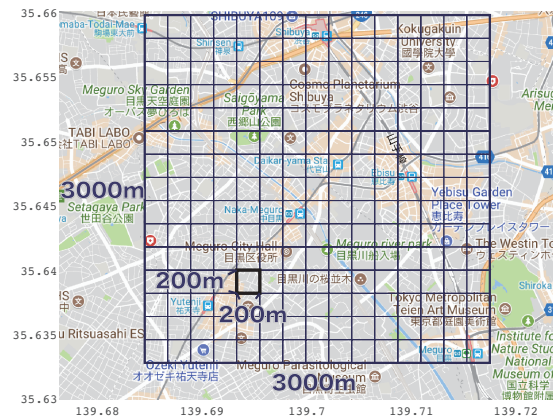


図 2: 解析対象エリア 2 : 200m メッシュ

5.2 評価指標

評価指標としては, 先行研究 [10] でも用いられている, 負の対数尤度 (MNLL: Mean Negative Log Likelihood) および, 絶対予測誤差 (MAE: Mean Absolute Error), 絶対ピーク時間誤差 (MAPE: Mean Absolute Peak Error) を使用する. これらの指標はそれぞれ次のように定義される.

$$\text{MNLL} = \frac{\sum_{n=1}^N \sum_{\tau=1}^S (-\ln p(y_{c,\tau}^{(n,l)} | \lambda_{c,\tau}^{(l)}))}{NS}, \quad (19)$$

$$\text{MAE} = \frac{\sum_{n=1}^N \sum_{\tau=1}^S |y_{c,\tau}^{(n,l)} - \hat{y}_{c,\tau}^{(l)}|}{NS}, \quad (20)$$

$$\text{MAPE} = \frac{\sum_{n=1}^N \left| \arg \max_{\tau \in [1,S]} (y_{c,\tau}^{(n,l)}) - \arg \max_{\tau \in [1,S]} (\hat{y}_{c,\tau}^{(l)}) \right|}{N}, \quad (21)$$

5.3 比較手法

本実験では実験 1 で説明変数を用いた予測を行うモデルの既存手法として双線形ポアソン回帰 (BP) [10], 双線形ガ

*1 <https://emg.yahoo.co.jp/>

ウス回帰 (BG) と MAE, MAPE, MNLL による比較を行い、実験 2 では HDP を用いた生成的なパターンモデリング手法で、回帰を行わないモデル (HDP-gen) との MNLL を用いた比較を行う。双線形ガウス回帰は、提案手法の HDP 回帰同様、正規分布で都市動態をモデリングするもので、HDP-reg と公正に尤度を比較するために導入する。MAE, MAPE の指標において、双線形ガウス回帰が双線形ポアソン回帰と同等の性能を持つことを実験 1 で確認し、HDP-reg と双線形ガウス回帰との比較の妥当性を担保する。

具体的に双線形ガウス回帰は以下の式で与えられる。

$$p(y_{c,\tau}^{(n,l)}) = \mathcal{N}(y_{c,\tau}^{(n,l)} | \mu_{c,\tau}^{(l)}, \sigma^2) \quad (22)$$

$$\mu_{c,\tau}^{(l)} = \boldsymbol{\varphi}(\mathbf{d})^\top \mathbf{W}^{(l)} \boldsymbol{\phi}_\tau(\tau) \quad (23)$$

双線形ポアソン回帰・双線形ガウス回帰は、各エリアに対して、2通りの適用の仕方を考える。

- (1) **BP·BG1forAll**: 全ての解析対象地域に対して一つのモデルを定義する。

学習の際には全メッシュのデータが利用できるため、メッシュを小さくしても安定した学習ができることが予想されるが、全てのメッシュで同じ予測結果が得られるため、地域固有の結果を得ることができない。

- (2) **BP·BG1for1**: 解析対象地域それぞれにモデルを定義する。地域固有の学習をすることで、地域固有のパターンにフィッティングすることができるが、学習に使用されるデータは解析対象域内のデータのみとなり、特に小さなメッシュでは過学習のリスクが大きくなる。

双線形ガウスモデル、双線形ポアソン回帰モデルでは、それぞれ平均値を予測値とする。

HDP-gen は生成的な階層ベイズモデルであり、全地域のパラメータを共有しながら、地域に固有のパターンモデリングを行う。具体的には、 $\tilde{y}_{c,\tau}^{(n,l)}$ と潜在変数の同時分布が以下の式になり、HDP の枠組みで事後分布の近似解を求める。このときパラメータ $\boldsymbol{\mu}$ は条件 \mathbf{d} や時間 τ といった外的要因に影響を受けないパラメータであり、すなわちこのモデルが説明変数を用いた予測を行わないモデルであることを意味する。

$$p(\tilde{y}_{c,\tau}^{(n,l)}, z_{l,n,m}, r_{l,m,k}) = \prod_m \pi_{l,m}^{z_{l,n,m}} \prod_k \rho_k^{r_{l,m,k}} \mathcal{N}(\tilde{y}_{c,\tau}^{(n,l)} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_{l,n,m} r_{l,m,k}} \quad (24)$$

5.4 回帰モデルで使用する特徴量

本実験では回帰モデルには次のような特徴量を与える。時間特徴量 $\boldsymbol{\phi}_\tau(\tau)$ は (2) 式で表される。時間以外の要因として、平日か否かを表す 2 次元の 1-of-K 表現特徴量 $\boldsymbol{\varphi}_1(\mathbf{d}) \in \mathbb{R}^2$ と、天気 (晴れ, 曇り, 雨, その他)

を表す 4 次元の 1-of-K 表現特徴量 $\boldsymbol{\varphi}_2(\mathbf{d}) \in \mathbb{R}^4$ を用い、 $\boldsymbol{\varphi}(\mathbf{d}) = \boldsymbol{\varphi}_1(\mathbf{d}) \otimes \boldsymbol{\varphi}_2(\mathbf{d})$ とする。

5.5 実験 1. 既存の予測手法との性能比較実験

ここでは、既存の予測手法との比較を、図 2, 図 1 に示された、大きさの異なる 2 つのメッシュ上で行う。評価指標は MAE, MAPE, MNLL に対する評価を行う。ここでの比較手法は BP 1 for 1, BP 1 for All, BG 1 for 1, BG 1 for All である。この比較では提案手法の優位性の確認と同時に、state of the art な手法である双線形ポアソン回帰と双線形ガウシアンが MAE, MAPE と同程度の予測性能を持つことを確認する。双線形ポアソンと双線形ガウシアン間では測度が異なるため MNLL による比較は行わない。いずれのモデルも 5 交差検証を行い、学習データとして 30 日分のデータを与え、テストデータとして 180 日分のデータを用いて評価を行った。

実験の結果を表 1, 表 2 に示す。まず、MAE と MAPE の指標について、双線形ガウス回帰と既存の代表的な手法である双線形ポアソン回帰が同程度の性能を持っていることが確認できる。この結果から、HDP-reg と双線形ガウス回帰を比較する正当性が得られる。MAE と MAPE についてはいずれのメッシュでも HDP-reg によって双線形回帰を上回る性能が得られた。メッシュサイズが小さくなると、メッシュあたりのデータ数も小さくなり、MAE の絶対数も小さくなる。しかしながら、その差の割合に着目すると、200 m では HDP-reg と双線形回帰の差が大きくなっていることがわかる。MNLL に関しても双線形ガウス回帰と比較し、HDP-reg が最も良い性能となっている。

5.6 実験 2. 生成モデリングとの尤度における性能比較実験

次に HDP-gen との MNLL における比較実験を行う。解析対象域、及びメッシュサイズは実験 1 と同様のものを用いる。いずれのモデルも 5 交差検証を行い、学習データとして 30 日分のデータを与える。テストデータとして 180 日分のデータを用いて評価を行う。

結果を表 3 に示す。200 m メッシュにおいて HDP-gen と尤度を比較すると尤度が同程度であることがわかる。

結果から提案手法は同程度の性能でモデリングを行えているといえる。HDP-gen は提案手法と異なり、説明変数を用いた予測ができないことに留意されたい。

6. 結論

本研究では、携帯端末の位置履歴情報からの大規模かつ高精細なメッシュ上での都市動態モデリングを行った。高精細に分割したメッシュ上では、各メッシュのデータ量が

*1 測度空間が異なるため、BP1 for 1 と BP 1 for all との間でのみ比較すべき指標である。

表 1: Comparison on 600 m x 600 m in 3 km x 3 km

model	Evaluation Metric		
	MAE	MAPE	MNLL
BG 1 for All	26.0 ± 8.6	114.6 ± 11.8	1.55 ± 0.25
BP 1 for All	26.0 ± 8.6	112.4 ± 12.4	* ¹ 10.0 ± 3.4
BG 1 for 1	24.2 ± 8.7	119.0 ± 7.81	1.44 ± 0.27
BP 1 for 1	23.8 ± 8.9	110.5 ± 9.8	* ¹ 8.73 ± 3.3
Proposed	23.1 ± 7.2	112.5 ± 13.1	1.37 ± 0.25

表 2: Comparison on 200 m x 200 m on 3 km x 3 km

model	Evaluation Metric		
	MAE	MAPE	MNLL
BG 1 for All	3.73 ± 0.83	218.8 ± 9.28	2.14 ± 0.19
BP 1 for All	3.73 ± 0.83	217.2 ± 8.21	* ¹ 2.88 ± 0.45
BG 1 for 1	3.45 ± 0.83	218.1 ± 5.38	1.94 ± 0.17
BP 1 for 1	3.42 ± 0.86	215.0 ± 5.24	* ¹ 2.67 ± 0.42
Proposed	3.38 ± 0.68	214.2 ± 7.43	1.92 ± 0.24

表 3: MNLL Comparison with HDP-gen

model	Mesh Size	
	600 m x 600 m	200 m x 200 m
HDP-gen	0.31 ± 0.17	1.33 ± 0.18
Proposed	1.37 ± 0.25	1.92 ± 0.24

小さくなり、地域ごとに予測を行う従来のモデルでは過学習を引き起こす恐れがあった。そこで、地域間でデータを共有しながら学習することで、安定した予測性能を得ることが可能となる。階層ディリクレ過程 (HDP) 混合回帰モデルを提案した。提案手法は、複数分布に共通なパラメータを生成できる HDP をモデルの事前分布に用い、地域間でデータを共有しながら、地域間で異なるモデルを構築することができる。これによりデータ共有が図られ、提案手法と比べパラメータ総数も削減され、データの少ない地域に対しても安定した学習を行うことが可能となる。

東京中心部 3 km 四方の領域上で得られた 3200 万件の携帯端末位置履歴データを用いた実験を行い、既存の予測手法である双線形回帰と比較し絶対予測誤差 (MAE) において 200 m メッシュで 0.59 の改善が見られた。予測を行わない混合モデリング手法である生成的な HDP モデルと比較しても同程度のモデリング精度を持つことを示した。

将来課題として、都市動態の時間的な変換に対応するためにオンラインの変分推論による学習を適用するといったことが考えられる。また、応用上、地域ごとの全人口ではなく、性別、年齢などの属性ごとの予測を行うことは非常に重要である。属性ごとの活動人口予測では特定の条件ごとのデータ不足がより顕著な問題となり、これを解決する手法が求められる。

参考文献

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of machine Learning research*, 2003.

[2] A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi, and A. Pentland. Once upon a crime: towards crime prediction from demographics and mobile data. In *Proc. of ICMI*, 2014.

[3] A. B. Chan and N. Vasconcelos. Bayesian Poisson regression for crowd counting. In *Proc. of ICCV*, 2009.

[4] Z. Fan, X. Song, and R. Shibasaki. CitySpectrum: A non-negative tensor factorization approach. In *Proc. of UbiComp*, 2014.

[5] H. Ishwaran and L. F. James. Approximate Dirichlet process computing in finite normal mixtures: smoothing and prior information. *Journal of Computational and Graphical statistics*, 11(3), 2002.

[6] S. N. MacEachern and P. Müller. Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics*, 7(2), 1998.

[7] K. Nishi, K. Tsubouchi, and M. Shimosaka. Extracting land-use patterns using location data from smartphones. In *Proc. of the First International Conference on IoT in Urban Space*, 2014.

[8] M. Okawa, H. Kim, and H. Toda. Online traffic flow prediction using convolved bilinear Poisson regression. In *Proc. of MDM*, 2017.

[9] J. Sethuraman. A constructive definition of Dirichlet priors. *Statistica sinica*, 1994.

[10] M. Shimosaka, K. Maeda, T. Tsukiji, and K. Tsubouchi. Forecasting urban dynamics with mobility logs by bilinear Poisson regression. In *Proc. of UbiComp*, 2015.

[11] M. Shimosaka, T. Tsukiji, S. Tominaga, and K. Tsubouchi. Coupled hierarchical Dirichlet process mixtures for simultaneous clustering and topic modeling. In *Proc. of ECML-PKDD*, 2016.

[12] K. Takeuchi, R. Tomioka, K. Ishiguro, A. Kimura, and H. Sawada. Non-negative multiple tensor factorization. In *Proc. of ICDM*, 2013.

[13] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 2006.

[14] C. Wang, J. Paisley, and D. Blei. Online variational inference for the hierarchical Dirichlet process. In *Proc. of AISTATS*, 2011.

[15] X. Wang, G. Lindsey, S. Hankey, and K. Hoff. Estimating mixed-mode urban trail traffic using negative binomial regression models. *Journal of Urban Planning and Development*, 140(1), 2013.

[16] J. Yuan, Y. Zheng, and X. Xie. Discovering regions of different functions in a city using human mobility and pois. In *Proc. of KDD*, 2012.

[17] F. Zhang, N. J. Yuan, D. Wilkie, Y. Zheng, and X. Xie. Sensing the pulse of urban refueling behavior: A perspective from taxi mobility. In *ACM Trans. on Intelligent Systems and Technology*, 2015.

[18] J. Zhang, Y. Zheng, and D. Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Proc. of AAAI*, 2017.

[19] Y. Zheng, T. Liu, Y. Wang, Y. Zhu, Y. Liu, and E. Chang. Diagnosing New York city's noises with ubiquitous data. In *Proc. of UbiComp*, 2014.