

携帯電話から得られる大規模な位置履歴情報を用いた 都市動態モデリング

Modeling urban dynamics with large-scale spatio-temporal data from mobile phones

下坂 正倫^{*1}
Masamichi Shimosaka

築地 毅^{*1}
Takeshi Tsukiji

坪内 孝太^{*2}
Kota Tsubouchi

西 賢太郎^{*1}
Kentaro Nishi

前田 啓輔^{*1}
Keisuke Maeda

^{*1}東京大学
The University of Tokyo

^{*2}Yahoo! JAPAN 研究所
Yahoo! JAPAN Research

This paper describes a hierarchical Bayesian method to model daily people activity patterns in a city (urban dynamics) with large-scale spatio-temporal data obtained from mobile phones. Thanks to the superiority of hierarchical Bayesian modeling, we can extract appropriate number of activity patterns among cities. In addition to existing models using a Dirichlet process (DP) as a prior distribution, we construct latest models using a hierarchical Dirichlet process (HDP). HDP models can consider the mixture rates of patterns in each area unlike DP models. The results of our experiment with large-scale spatio-temporal data, about 40 million logs per day show that our HDP models have better performance than existing DP models.

1. 序論

都市の振る舞いは様々な人々の活動によって織りなされている。どの時間にどれくらいの人が活動しているか、という情報はその都市の特性を表しており、その特性を把握することで、都市に応じて適切な都市設計、商業活動、災害からの復旧などが可能になる。近年の携帯端末の普及に伴い、GPSによる位置履歴データが大量に蓄積されてきており、それを利用した都市における人々の活動パターンを解析、モデル化する研究が盛んに行われている [Fan 14][Witayangkurn 13]。本研究では特に都市毎の時間変化する人口の推移に注目し、これを「都市動態」と定義しそのモデリング手法について論じる。

本研究のように時間変化する活動量を解析する類似の取り組みとして、時間帯とセンサ反応回数関係のモデリングでは、ガウス・コックス過程 [Moller 98] といったポアソン過程によるモデル化が提案されている。この手法では単一のパターンで活動量推移を表現するため、場所・曜日などの違いを考慮できない、モデルのパラメータが複雑になり大規模なデータへの適用が難しいという問題がある。そこで場所や曜日といった外的な要因により、活動量の推移に複数のパターンが存在するとしたモデリング手法が提案されている。Witayangkurnらは場所、曜日毎に人口の推移をモデル化し、都市での異常検知を行った [Witayangkurn 13]。また Fanらは非負値テンソル分解 (NTF) [Shashua 05] を用いて、都市に共通する動態パターンを抽出した [Fan 14]。しかしこれらの手法ではモデル学習時に、予めパターン数が決められてしまう問題がある。曜日毎では7つで固定であり、NTFではデータ毎に適切なパターン数を評価関数をもとに手動で決めなければならない。Nishiらは階層ベイズモデルの一つである、ディリクレ過程 (Dirichlet process, DP) [Antoniak 74] を事前分布とした混合モデルにより動態パターンの抽出を行った [Nishi 14]。このモデルは都市間に共通する動態パターンを、その数も含めて自動的に推定することが可能である。

一方、Nishiらのように DP を事前分布としたモデルを用いて、全ての都市でパターンを共有して学習をすると、各パターンの重みは全ての都市で共通なものになる。都市毎にパターンの重みに関して差異を考慮できれば、モデルを用いた都市毎の動態予測の精度の向上が期待できる。しかし都市毎に差異を持たせるために、都市毎にモデルを構築すると、モデル毎に扱えるデータが減り精度が下がるリスクがある、というジレンマがある。

本研究では階層ディリクレ過程 (hierarchical Dirichlet process, HDP) [Teh 06] に注目し、都市動態パターンをモデル化する枠組みを提案する。提案する枠組みでは、都市間で共通するパターンを共有しつつ、パターンの重みを都市毎に持つことができるため、予測時の精度の向上が期待できる。また携帯端末から得られる大規模な位置履歴データを用いて、提案する枠組みの性能を評価し、既存の DP によるモデルより高性能な都市動態のモデル化手法であることを示す。

2. 階層ベイズモデルによる都市動態モデリング

2.1 問題設定

本研究では都市動態パターンとして、新宿、渋谷といった、駅などを中心とした地域における1日の時間毎の人口の推移をモデル化の対象とする。そこである地点を中心とした矩形領域 (例 500m×500m) を解析対象の領域とし、土地の空間的な連続性は考慮せず、その領域内での推移を議論する。ここでこの領域のことを POI (Point Of Interests) と定義する。

POIでの位置履歴は点データとして与えられる。本研究では1日を H 個の時間帯に分割し、時間帯 $h = \{1, \dots, H\}$ の位置履歴データの総数 x_h を集計し、これを活動人口と定義する。そして x_h を並べたヒストグラム $\mathbf{x} = \{x_1, \dots, x_H\} \in \mathbb{R}_+^H$ として、1日の活動人口の推移を表現する (図1)。これにより時間帯ごとの独立したパラメータをモデル化することになり、後述するディリクレ混合モデルを適用する際に、各種計算における計算の複雑化を防ぐことができる。

以降の説明では、POIのインデックスを $i = \{1, \dots, m\}$ 、日にちのインデックスを $d = \{1, \dots, n\}$ とする。すると、 i 番目の POI の d 日目の1日のデータは $\mathbf{x}_{i,d} = \{x_{i,d,1}, \dots, x_{i,d,H}\}$

連絡先: 下坂正倫, simosaka@ics.t.u-tokyo.ac.jp

連絡先: 築地毅, tsukiji@ics.t.u-tokyo.ac.jp

連絡先: 坪内孝太, ktsubouc@yahoo-corp.jp

連絡先: 西賢太郎, nishi@ics.t.u-tokyo.ac.jp

連絡先: 前田啓輔, maeda@ics.t.u-tokyo.ac.jp

と表現できる．

$x_{i,d,h}$ を生成する確率分布としてはポアソン分布，ガンマ分布，対数正規分布等が考えられる．本研究では，人口データの非負であるという制約を満たしていないが，パラメータに共役事前分布を設定でき，推論と学習が容易になること，そして時間帯毎の人口のばらつきを考慮できることを重視して，その確率分布をガウス分布とし，1日の活動人口推移を以下のように定式化する．

$$p(x_{i,d}|\mu, \lambda) = \prod_{h=1}^H \mathcal{N}(x_{i,d,h}|\mu_h, \lambda_h^{-1}) \quad (1)$$

ただし $\mu_h \in \mathbb{R}, \lambda_h > 0$ は時間帯 h におけるガウス分布の平均パラメータ，精度パラメータを表す．したがって，本研究は POI 毎の活動人口の推移を求めるため，POI 毎に $x_{i,d}$ を生成しやすいパラメータ μ, λ を求めることが問題となる．

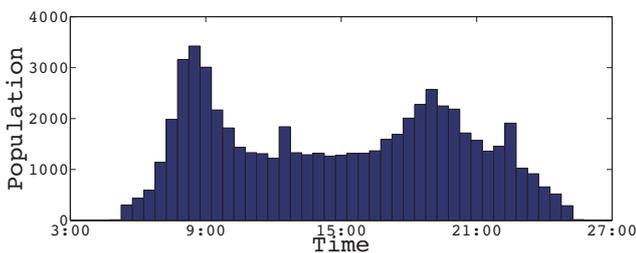


図 1: 1 日の人口推移

2.2 階層ベイズ法によるモデル設計の指針

本研究では序論で述べたように，階層ベイズ法の枠組みでのモデリングを行う．本節では階層ベイズ法における事前分布として，既存のモデルで用いられたディリクレ過程 (DP)，および提案モデルで用いる階層ディリクレ過程 (HDP) について述べる．

2.2.1 ディリクレ過程 (DP)

DP は階層ベイズ法の一つで，混合モデルの事前分布に組み込むことで，そのモデルの混合数も推定することが可能となる．DP によるモデルを用いて動態パターンの抽出を行うと，全 POI・日付のヒストグラムデータをクラスタリングしパターンを抽出する．パターンの重みは全データに対するものである．つまりその情報のみでは，ある POI ではどのパターンが出やすいかという情報を得ることはできない．図 2 の左側の図は，DP を用いたモデルのパターン抽出の様子を示している．この図では 1 つ目のパターンの重みは 0.15，2 つ目のパターンの重みは 0.4 となっているが，これは全 POI 間で共通である．このモデルを用いて予測を行う場合，各 POI において，実際の各パターンの出やすさがこの共通の重みに近ければ，精度の高い結果を期待することができる．しかし，例えば 2 つ目のパターンがほとんど出ない POI の場合でも，予測時にはそのパターンの重みは 0.4 として計算されてしまう．このように全体の傾向から外れた動態特性を持つ POI では，予測性能が下がるおそれがある．

POI 毎にパターンの重みを考慮する方法として，POI 毎に DP によるモデルで学習を行う方法が考えられる．この方法では POI 間でパターンを共有せず，パターン及び重みは POI 毎に独立のものとなるが，学習で使えるデータが少なくなり，過学習により精度が悪くなるおそれがある．

一方，パターンを POI 間で共有しつつ POI 毎に重みを持たせるには，例えばモデル学習後，POI 毎にパターンの出現率

を重みとして持たせるなどが考えられる．この方法では重みとパターンは別の枠組みで求められることになるが，同一の枠組みで両者を取得できる方が望ましい．そこで本研究では，それを達成するために階層ベイズ法の枠組みの一つである階層ディリクレ過程 (HDP) に注目する．次節では HDP 及びそれを用いたモデルについてその特徴を述べる．

2.2.2 階層ディリクレ過程 (HDP)

HDP は DP を階層的に組み込んだ確率過程で，これを事前分布として利用すると POI 間で共通するパターンの抽出とともに，POI 毎にパターンの重みの取得ができる．POI 毎のパターンの重みの推定は学習時にパターンの推定と同時にされる．図 2 の右側の図は，HDP を用いたモデルのパターン抽出の様子を示している．この図では POI1 においては，1 つ目のパターンの重みが 0.6，2 つ目のパターンの重みが 0.4 であることを表しており，POI 毎に個別の重みを持つことがわかる．このモデルを用いた予測では，各 POI 毎にこの個別のパターンの重みを考慮して計算が行われるので，POI 毎の性質を上手く考慮したモデルであるといえる．このように POI 間でパターンを共有しつつ，POI 毎に重みを持つ特性から，DP を用いたモデルに比べ精度の高い予測ができることが期待できる．

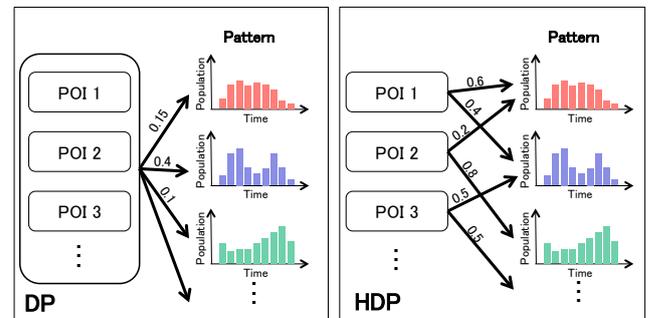


図 2: DP(左) と HDP(右) によるパターン抽出

2.3 階層ディリクレ過程による都市動態パターンモデルの定式化

2.2 節で述べた，階層ベイズ法を用いた都市動態パターンモデルの定式化を行う．ここでは特に新規に提案，構築した HDP によるモデルについて述べる．

潜在変数として， i 番目の POI の d 日目のデータが属する，その POI 内のパターン t を示す $Z = \{z_{i,d,t} | z_{i,d,t} \in \{0, 1\}, \sum_t z_{i,d,t} = 1\}$ ，そのパターン t が属する POI 間で共通のパターン k を示す $C = \{c_{i,t,k} | c_{i,t,k} \in \{0, 1\}, \sum_k c_{i,t,k} = 1\}$ を定義する．POI に共通するパターンの混合比，及び POI 毎のパターンの混合比を生成するパラメータを ρ', π' とする．

パターン k の時間帯 h のデータを生成するガウス分布の平均パラメータ $\mu_{k,h}$ ，精度パラメータ $\lambda_{k,h}$ に対する共役事前分布はガウス・ガンマ分布であり，事前分布のパラメータを m_0, δ_0, a_0, b_0 とすると，

$$p(\mu_{k,h}, \lambda_{k,h}) = \mathcal{N}(\mu_{k,h} | m_0, (\delta_0 \lambda_{k,h})^{-1}) \mathcal{G}(\lambda_{k,h} | a_0, b_0) \quad (2)$$

と書ける．ただし $\mathcal{G}(x|a, b)$ はガンマ分布 $\mathcal{G} \propto x^{a-1} e^{-bx}$ を表す．観測データ $X = \{x_{i,d}\}$ が与えられた時に， X と各パラメータの確率は以下のように定式化される．

$$p(\mathbf{X}|\mathbf{Z}, \mathbf{C}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \prod_{i,d,t,k} \mathcal{N}(x_{i,d,h} | \mu_{k,h}, \lambda_{k,h})^{z_{i,d,t} c_{i,t,k}} \quad (3)$$

$$p(\mathbf{Z}|\boldsymbol{\pi}') = \prod_{i,d,t} \left\{ \pi'_{i,t} \prod_{s=1}^{t-1} (1 - \pi'_{i,s}) \right\}^{z_{i,d,t}} \quad (4)$$

$$p(\mathbf{C}|\boldsymbol{\rho}') = \prod_{i,t,k} \left\{ \rho'_k \prod_{j=1}^{k-1} (1 - \rho'_j) \right\}^{c_{i,t,k}} \quad (5)$$

$$p(\boldsymbol{\pi}') = \prod_{i,t} \mathcal{B}(\pi'_{i,t} | 1, \alpha) \quad (6)$$

$$p(\boldsymbol{\rho}') = \prod_k \mathcal{B}(\rho'_k | 1, \beta) \quad (7)$$

$$p(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \prod_{k,h} \mathcal{N}(\mu_{k,h} | m_0, (\delta_0 \lambda_{k,h})^{-1}) \mathcal{G}(\lambda_{k,h} | a_0, b_0) \quad (8)$$

一般に階層ベイズ法を用いたモデルの事後分布は解析的に求めることができない．そこで本研究では，変分推論 [Blei 05] により事後分布のパラメータの計算を近似的に行った．

3. 性能評価実験

提案した HDP による都市動態パターンモデルの性能を評価するために，実際の位置履歴データを用いた性能評価実験を行った．

3.1 実験概要

実験データ

Yahoo! JAPAN の防災アプリ^{*1}により取得された，匿名化ならびに統計情報として加工された位置履歴データを利用する．位置履歴データは携帯端末が移動している時，すなわち端末を持っている人が動いている時に取得されるので，本データセットは人々の活動をよく表しているものといえる．データは { タイムスタンプ, 緯度, 経度 } の要素からなり，データの規模は 4000 万件/日である．利用するデータ期間は 2013 年 7 月 1 日から 2014 年 6 月 30 日の 1 年分である．本実験では都市として関東にある駅，観光地から $m = 150$ 地点を選び，その地点を中心とした 900m 四方の範囲を POI とする．1 日を活動人口が最も減る 03:00 から翌日 02:59(26:59) までとし，1 日の分割数は $H = 48$ として，30 分毎の人口推移のモデル化を行う．人口規模が異なっても似た形状のパターンの共有が進むように，データの正規化を行う．本実験では POI 毎に 1 日の総活動人口の平均をスケールパラメータと定義し，学習データは都市毎にスケールパラメータで割ったものを利用する．

評価指標

評価指標は新規データに対する対数尤度，ピーク時刻誤差 [分] を用いる．対数尤度の計算時は，与える新規データも都市毎にスケールパラメータで割ったものを利用する．ピーク時刻誤差は，予測される 1 日の活動人口推移とテストデータの活動人口が最も多くなる時刻のずれの絶対値を計算する．本実験では 1 日を 48 の時間帯に分割しているので，ピークとなる時間帯の誤差を計算し，その値を 30 倍することで分単位の出力を行う．

比較手法

本実験では比較手法として，DP を用いた 2 つのモデルを用いる．一つは POI 間でのパターン共有の有無による性能を比べるため，POI 毎に DP による学習を行い，パターンを POI 間で共有しないモデル (DP-i) である．もう一つは POI 毎に重みを持つかどうかによる性能差をみるため，DP により全 POI でパターンを共有して POI 毎に重みを持たないモデル (DP-g) である．

実験設定

本実験では訓練日数の違いによる精度の違いを評価するため，訓練データの期間を 30 日，60 日，...，180 日と変えて実験を行う．テストデータの期間はいずれの訓練期間でも 180 日である．訓練データ，テストデータは 365 日分のデータセットから重複のないようにランダムに抽出した．各期間毎にテストを 5 回を行いその平均を求めた．

3.2 実験結果

まず対数尤度に関する実験結果について述べる．図 3 は学習期間と，テストデータに対する対数尤度の平均値の関係を表したグラフである．この図から POI 毎にモデルを作る，つまりパターンを POI 間で共有しない DP-i は，パターンを POI 間で共有する DP-g，HDP に比べて精度が低いことがわかる．パターンの共有の有無で差が広がる原因を考えるため，パターンを POI 間で共有するモデルとしないモデルから抽出される，パターンの特徴に注目する．図 5，図 6 に示しているのは DP-i，及び HDP で抽出されたパターンの一例である．青線が平均，点線が平均周りの標準偏差を表す．両図とも通勤時間帯である時刻に人々の活動が多くなるパターンである．これらの図を比較すると DP-i で得られる POI 毎のパターンは人数の分散，すなわち振幅が小さく，HDP はそれに比べ大きいことがわかる．これらは同じモデルの他のパターンにも見られる傾向であり，DP-i で得られるパターンは他のモデルに比べ分散の小さいパターンが得られる．これは，POI 間でパターンを共有しないため，パターンが POI 毎のデータに過学習してしまっていることを示している．そのため DP-i は DP-g，HDP に比べ精度が低くなったと考えられる．

次にパターンを POI 間で共有する DP-g および HDP によるモデルでの結果を比較する．図 4 は図 3 を拡大し，DP-g と HDP に関する結果のみを示しているグラフである．いずれの訓練期間でも HDP が高い尤度を示した．訓練期間が 30 日の時は有意水準 1% で，訓練期間が 60 日から 120 日までの場合は有意水準 5% とした時に，t 検定により DP-g に対する有意差が認められた．この結果から，POI 毎にパターンの重みを持つことが，高性能なパターンモデルに必要なということになる．また両モデルの差は特に訓練期間の短い時に大きい，これは HDP が POI 毎の重みを考慮して学習をするため，データが少ない場合でも DP に比べ精度よくパターンの抽出ができる，ということにも起因していると考えられる．

次にピーク時刻誤差に関する結果について述べる．図 7 は学習期間とピーク時刻誤差の関係を表したグラフである．学習期間が 30 日の場合は DP-g と HDP がほぼ同一の値だが，残りの学習期間では HDP が他の 2 モデルに比べ誤差が小さかった．対数尤度の時と異なり，POI 間でパターンを共有しない DP-i が，DP-g，HDP と比べてほとんど精度に差がなかった．これはピーク時刻誤差の結果には，時間帯毎の活動人口の平均値の大小関係が影響を与えるため，図 5 のように，過学習を起こし抽出されるパターンの分散が小さくても，それが結果には影響しないからと考えられる．また訓練期間の長さがほとんど

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

ど精度に影響を与えていないが、これは訓練期間が増えモデルの頑強性が高くなったとしても、パターンの活動人口のピーク時刻に関して大きな変化はないためと考えられる。

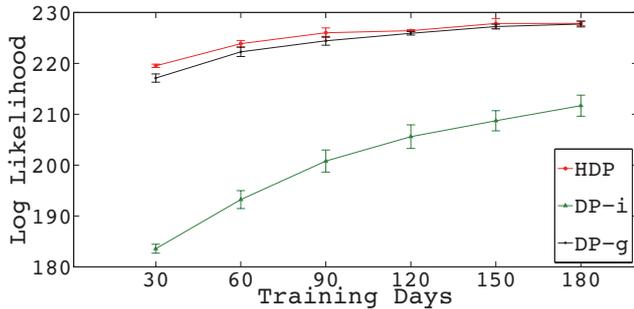


図 3: 学習期間と対数尤度

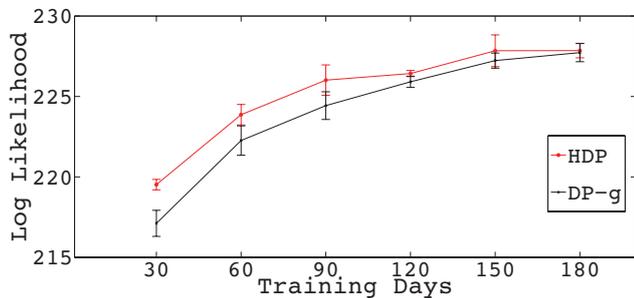


図 4: 学習期間と対数尤度 (図 3 の拡大図)

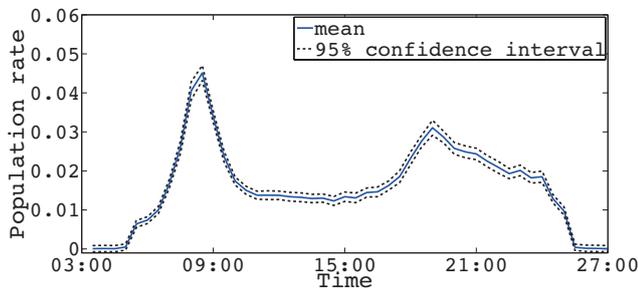


図 5: DP-i で抽出されたパターン例

4. 結論

本研究では、高性能な都市動態のパターンを推定するモデルの構築を目的とし、階層ベイズ法の枠組みでのモデリング手法を提案し、評価実験を行いその性能を検証した。都市間でパターンの共有を可能にする DP によるモデルでは、POI 毎にパターンの重みを持たず、予測時に精度が下がるおそれがあることを指摘し、その解決のため、HDP を用いた POI の動態パターンモデルを構築した。HDP は POI 毎にパターンに対する重みを持ち、パターンは全 POI で共有するモデルである。広範囲で、長期間に渡る大量の位置履歴データを用いて、HDP を用いたモデル、及び DP を用いた複数の枠組みのモデルで性能評価実験を行った。POI 毎にモデルを作る枠組みでは精度が低く、POI 毎にパターンの重みを持たせる HDP の精度が高いという結果となった。このことから、高性能な動態パターンモデルには、パターンを POI 間で共有し、一方でパ

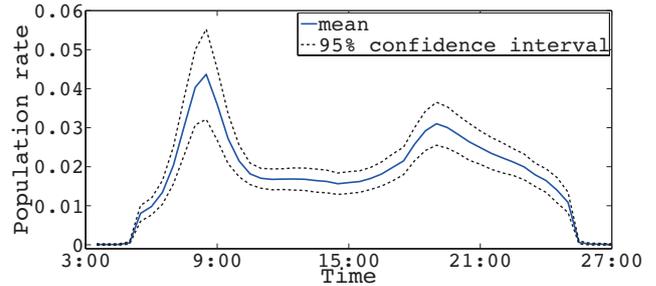


図 6: 提案手法で抽出されたパターン例

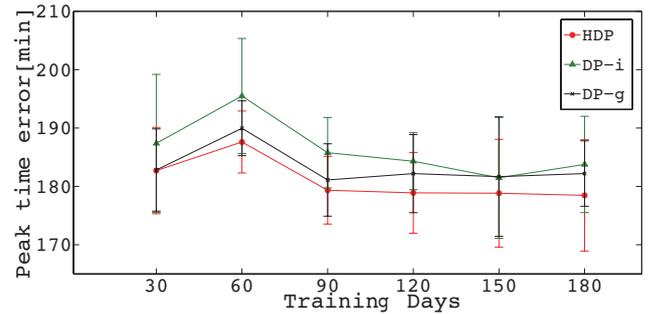


図 7: 学習期間とピーク時刻誤差

ターンの重みは POI 毎に持てるようにする設計が必要であるといえる。

今後の展望として、学習時に都市毎の差異だけでなく、都市の関係性を考慮するモデリングが考えられる。HDP はパターンの重みを POI 毎に個別に学習するため、DP-i と同様に過学習が起きるリスクがある。そこでパターンの重みが似た POI 間で、パラメータを共有して学習する枠組みがあれば、POI 毎の過学習の抑制が期待できる。そのような枠組みを持ったモデルを構築し、今回のような大量の位置履歴データを用いてその性能を検証することが今後の課題である。

参考文献

- [Antoniak 74] Antoniak, C. E.: Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems, *The annals of statistics*, pp. 1152–1174 (1974)
- [Blei 05] Blei, D. and Jordan, M.: Variational inference for Dirichlet process mixtures, *Bayesian Analysis*, Vol. 1, pp. 121–144 (2005)
- [Fan 14] Fan, Z., et al.: CitySpectrum: A Non-negative Tensor Factorization Approach, in *Proc. of UbiComp*, pp. 213–223 (2014)
- [Moller 98] Moller, J., et al.: Log Gaussian Cox processes, *Scandinavian J. of Stats.*, Vol. 25, No. 3, pp. 451–482 (1998)
- [Nishi 14] Nishi, K., et al.: Extracting Land-use Patterns Using Location Data from Smartphones, in *Proc. of URB-IOT*, pp. 38–43 (2014)
- [Shashua 05] Shashua, A. and Hazan, T.: Non-negative tensor factorization with applications to statistics and computer vision, in *Proc. of ICML*, pp. 792–799 (2005)
- [Teh 06] Teh, Y. W., et al.: Hierarchical Dirichlet processes, *Journal of the American Statistical Association*, Vol. 101, No. 476, pp. 1566–1581 (2006)
- [Witayangkurn 13] Witayangkurn, A., et al.: Anomalous Event Detection on Large-scale GPS Data from Mobile Phones Using Hidden Markov Model and Cloud Platform, in *Proc. of UbiComp*, pp. 1219–1228 (2013)