

位置情報履歴の欠損と周期性を考慮した 訪問系列パターン抽出手法

Transition Pattern Extraction Based on Location History Considering Missing Value and Periodicity

林 亜紀

Aki Hayashi

日本電信電話株式会社 NTT サービスエボリューション研究所 *1

NTT Service Evolution Laboratories, NTT Corporation

hayashi.aki@lab.ntt.co.jp

亀岡 弘和

Hirokazu Kameoka

日本電信電話株式会社 NTT コミュニケーション科学基礎研究所

NTT Communication Science Laboratories, NTT Corporation

kameoka.hirokazu@lab.ntt.co.jp

松林 達史

Tatsushi Matsubayashi

日本電信電話株式会社 NTT サービスエボリューション研究所

NTT Service Evolution Laboratories, NTT Corporation

matsubayashi.tatsushi@lab.ntt.co.jp

澤田 宏

Hiroshi Sawada

日本電信電話株式会社 NTT コミュニケーション科学基礎研究所

NTT Communication Science Laboratories, NTT Corporation

sawada.hiroshi@lab.ntt.co.jp

keywords: location history, pattern extraction, nonnegative matrix factorization (NMF) with Markov-chained bases, missing value interpolation, periodicity

Summary

Understanding of human behavior patterns is important for many tasks such as location-based recommendation, urban design and crowd control. Our purpose is to extract easy-to-grasp multiple transition patterns which have hierarchical periodicities (e.g., weekly, daily) for enhancing above mentioned tasks. Conventional Hidden Markov Model extracts typical transition patterns, though extracted patterns are not always based on explicit periodicities and extracted patterns are independent. Also, even though conventional Nonnegative Matrix Factorization (NMF) approximates human behaviors as the weighted sum of multiple typical visit frequency distributions, it is not able to extract transition patterns. The proposed method solves these problems by extending a kind of extension of NMF called NMF with Markov-chained bases in two perspectives by approximating human behaviors as the weighted sum of multiple typical transition patterns in hierarchical periods. First, the proposed method extracts patterns interpolating missing values that are typical for location history. Second, we add restriction for estimating transition patterns that promotes extraction of transition patterns in hierarchical periods. For example, the proposed method can approximate a user's behavior as the weighted sum of daily transition pattern on working day (e.g. leave home and go to the office in the morning, go to the restaurant and return to the office in lunch break and return home in the evening) and weekly transition pattern (e.g., visit office in week day and go to the other recreation places in weekends).

1. はじめに

近年、モバイルデバイスやソーシャルネットワーキングサービスが普及し、各ユーザの位置情報履歴の観察が可能になった。位置情報履歴を活用して頻出訪問パターンを抽出し、行動を予測する取り組みは、情報配信、都市計画、人流制御による混雑緩和など多くの場面で有用である。位置情報履歴に含まれる行動には、24 時間周期、1 週間周期、1 年周期など、複数の周期性が存在すると考えられる [Hayashi 14]。また、行動は直前の行動に強く影響されることが示されており [Ashbrook 03]、行動の連続性を考慮することが重要である。

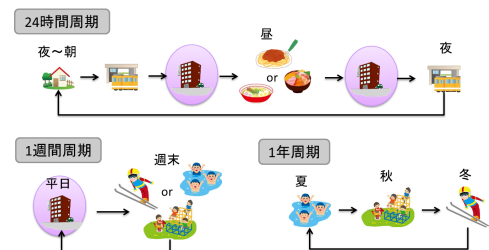


図 1 複数周期に基づく訪問系列パターンのイメージ

本論文の目的は、図 1 に示されるように、複数の周期を持った連続的な訪問系列の重ね合わせで頻出訪問パターンを表現することである。図 1 では、24 時間周期の訪問系列として、朝に家を出て、電車に乗って会社に行き、昼に食事に出かけて会社に戻り、夜には再び電車に乗っ

*1 現所属: 株式会社 NTT ドコモ, aki.hayashi.pk@nttdocomo.com

て家に帰るという連続的行動が表現されている。また、1 週間周期の訪問系列として、平日は会社に行き、週末はスキーやプールや公園に家族を連れて行く行動パターンが描かれている。さらに、同じ週末でも、1 年周期で夏はプール、秋は公園、冬はスキーといった行動変遷があることが表現されている。例えば会社に行くという行動は、1 週間周期では平日に行きやすく、さらに 24 時間周期で朝～昼、昼過ぎ～夜に行きやすいといったパターンの重ね合わせで表現されている。同様に、スキーには 1 年周期で見ると冬に行きやすく、さらに 1 週間周期で見ると週末に行きやすいという異周期パターンの重ね合わせも存在する。

このように、24 時間や 1 週間といった明示的な周期に則した訪問系列をパターン化できれば、パターンの解釈が容易になり、情報配信サービスなどへの活用がしやすくなると考えられる。加えて、複数の周期の重ね合わせで行動を表現することにより、柔軟な行動予測ができると期待される。例えば、図 1 のようなパターンを用いて会社への訪問を予測する場合、平日なら 1 週間周期で考えると少なくとも 1 度は会社に行く可能性が高く、さらにその時間は朝から昼または昼から夜の間に成る可能性が高いといった 2 段階での予測が可能になる。

時系列データから頻出訪問系列パターンを抽出するために、直前の行動の影響が強いというマルコフ性を考慮したマルコフモデル (Markov Model, MM) [Ashbrook 03] や隠れマルコフモデル (Hidden Markov Model, HMM) [Eddy 96] が有効な手法の代表例として挙げられる。HMM では、同時に訪問されやすい複数の場所群への訪問頻度分布を「状態」として捉え、その状態遷移を推定することにより、繰り返し同じ順で訪問されやすい訪問系列パターンを抽出できる。しかしながら、HMM では明示的な周期の考慮はできないため、1 週間・24 時間周期など解釈と活用が容易な系列を抽出できるとは限らない。また、複数の訪問系列パターンの重ね合わせの形で行動を表現することはできない。

一方、時系列データを複数の典型的訪問頻度分布の重み付き和として表現する方法として、非負値行列分解 (Non-negative Matrix Factorization, NMF) [Lee 01, 澤田 12] の有効性が示されている。NMF は非負値ベクトルの時系列を並べたデータ行列を基底行列と係数行列の積に分解する手法である。位置情報履歴におけるパターン抽出においても、NMF の行列をテンソルに拡張した非負値テンソル分解 (Nonnegative Tensor Factorization, NTF) [Shashua 05] の有効性が示されている [Fan 14]。上記手法では、位置情報履歴をユーザー × 場所 × 時間の 3 次元テンソルで表現した上で、そのテンソルを 3 つの行列の積の形で近似することにより、各時間の行動を基底数と同じ数の訪問傾向 (訪問頻度分布) に分類する。複数の典型的な訪問頻度分布が検出可能だが、一連の行動遷移に対応する、順序立った訪問系列を抽出することはできない。

NMF や NTF において、Factorial Hidden Markov Model (FHMM) [Ghahramani 97] の概念を導入した可変基底 NMF [Nakano 10, Ozerov 09] が音響信号の解析を題材として提案されている。従来の NMF では、各基底に時間変化する係数を乗じてデータに混在する各成分を表現する。音楽音響信号解析において、各楽音のスペクトルが音量以外に時間変化しないと仮定できれば、各分解成分は楽音のスペクトル系列に対応する。しかしながら、実際にはアタック、サステイン、ディケイなどの「状態」の変化に伴って楽音のスペクトルの形状は時間変化する。そこで、可変基底 NMF では、各基底の基底行列が状態遷移に応じて変化することを許容したモデルを採用している。これは、スペクトル系列を、複数の HMM からの出力系列の和でモデル化していることに相当する。

位置情報ログにおいても可変基底 NMF が適用できれば、各基底で順序立った訪問頻度分布の遷移を表現した上で、行動全体をそれらの重み付き和として近似できるため、図 1 で示したような階層的な周期を持つ系列パターンを抽出可能になると期待される。本論文では位置情報ログの特性を考慮して可変基底 NMF を拡張し、階層的な周期を持つ訪問系列パターンの抽出を実現する^{*2}。位置情報ログへの適用で考慮すべき特徴として、欠損の多さと、曜日・時刻など絶対時間による周期性が挙げられる。

可変基底 NMF の有効性が示されている音楽音響信号に比べ、位置情報ログは欠損が多い。例えば、位置情報履歴の中でもチェックインサービスのログなど、ユーザによる能動的な作業により訪問履歴が残される形式のログでは、実際にどこかを訪問していても訪問が記録されない場合や、1 箇所に数時間滞在していても到着時にしかログが残らない場合があることに起因する。データの欠損が多いと、全ての場所への訪問確率が 0 となる有用性の低い頻度分布が多く時刻で抽出される可能性や、一度だけ訪問した場所が抽出パターンに強く作用する可能性がある。提案手法では、パターン抽出時の繰り返し計算中に、更新前のパターンを用いた補間ステップを導入することにより、データの傾向を考慮した欠損値補間を行い、有用性の高いパターンの検出を促進する。

周期性について、従来の可変基底 NMF による各基底における状態推定では、通常の HMM と同様、データの絶対時間軸上の周期性の明示的考慮はない。提案手法では、位置情報データで想定される 1 週間周期、24 時間周期などの外部情報を与え、可変基底 NMF における状態推定に制約を与えることにより、絶対時間上の周期性を考慮できるように可変基底 NMF を拡張する。

本論文では、従来の可変基底 NMF と、位置情報ログの特性を考慮して可変基底 NMF を拡張した提案手法を説明し、実際のチェックインログを用いて提案手法が階層的な周期性を持つ行動パターンを抽出できることを示す。

^{*2} 本論文は、著者らによる研究会予稿 [林 16a] の内容を発展させたものである。

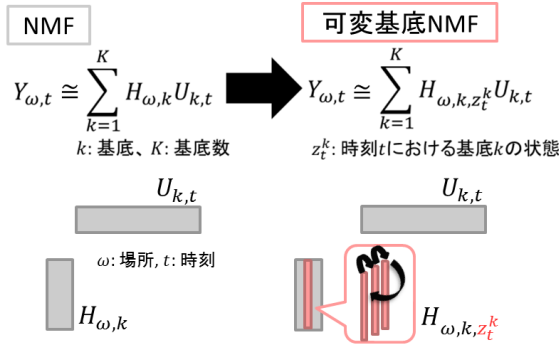


図2 NMFと可変基底NMFの概念図

2. 関連研究

まず、可変基底 NMF [Nakano 10] について説明する。図 2 は、従来の NMF と可変基底 NMF の概念図である。可変基底 NMF では、状態遷移に応じて NMF における基底行列 $\mathbf{H}$ （本論文では場所別の訪問頻度分布に該当）が時間変化することを許容する。したがって、NMF では観測行列 $\mathbf{Y}$ を $Y_{\omega,t} \simeq X_{\omega,t} = \sum_{k=1}^K H_{\omega,k} U_{k,t}$ で近似するのにに対し、可変基底 NMF では $Y_{\omega,t} \simeq X_{\omega,t} = \sum_{k=1}^K H_{\omega,k,z_t^k} U_{k,t}$ と近似する。 $\mathbf{H}, \mathbf{U}$ は各要素を正の実数とする行列である。 z_t^k は時刻 t における基底 k の状態を、 K は基底数を表す。係数行列 $U_{k,t}$ は、各基底のアクティベーションに該当し、時刻 t における基底 k の基底行列の重みを示す。状態 z_t^k は状態 z_{t-1}^k の影響を受けて推定される。すなわち、 $\pi_{i,j}^{(k)}$ は基底 k における状態 i から j への遷移確率とすると、 $z_t^k | z_{t-1}^k \sim \pi_{z_{t-1}^k, z_t^k}^{(k)}$ である。ここで、 $\pi_i^{(k)}$ を基底 k において初期状態 ($t=1$ の時の状態) が状態 i である確率とする。

可変基底 NMF では、観測行列 $\mathbf{Y}$ を最もよく近似する $\mathbf{X}$ を得るために、 $\mathbf{Y}$ と $\mathbf{X}$ の誤差を目的関数として、これを最小化する。I-divergence [Csiszár 75]

$$I(y|x) = y \log \frac{y}{x} - y + x$$

を距離尺度として用いると、可変基底 NMF の目的関数は以下のように書ける。

$$\mathcal{I}(\mathbf{H}, \mathbf{U}, \mathbf{Z}) = \sum_{\omega,t} I(Y_{\omega,t} | X_{\omega,t}) - \log P(\mathbf{Z})$$

$$\log P(\mathbf{Z}) = \sum_k \left(\log \pi_{z_1^k}^{(k)} + \sum_{t=2}^T \log \pi_{z_{t-1}^k, z_t^k}^{(k)} \right)$$

目的関数 $\mathcal{I}$ を最小化する $\mathbf{H}, \mathbf{U}, \mathbf{Z}$ を求めることにより、観測行列 $\mathbf{Y}$ を最もよく再現する $\mathbf{X}$ が得られる。 $\mathcal{I}$ からは、解析的に $\mathbf{H}, \mathbf{U}, \mathbf{Z}$ の値を求めることはできないが、Jensen の不等式を用いて設定した上界関数

$$\mathcal{I}^+(\mathbf{H}, \mathbf{U}, \mathbf{Z}, \boldsymbol{\lambda}) = \sum_{\omega,t} \left(Y_{\omega,t} \log Y_{\omega,t} - Y_{\omega,t} \sum_k \lambda_{\omega,k,t} \log \frac{H_{\omega,k} U_{k,t}}{\lambda_{\omega,k,t}} - Y_{\omega,t} + X_{\omega,t} \right) - \log P(\mathbf{Z})$$

を最小化することにより、間接的に $\mathbf{H}, \mathbf{U}, \mathbf{Z}$ の値を最適化する補助関数法が有効である。紙面の都合上証明は省略するが、 $\mathcal{I}^+$ を $\mathbf{H}, \mathbf{U}, \mathbf{Z}$ に関してそれぞれ最小化するステップと、 $\mathcal{I}(\mathbf{H}, \mathbf{U}, \mathbf{Z}) = \mathcal{I}^+(\mathbf{H}, \mathbf{U}, \mathbf{Z}, \boldsymbol{\lambda})$ となるように $\boldsymbol{\lambda}$ を更新するステップを繰り返すことで、 $\mathcal{I}$ を単調に降下させることができる。 $\mathcal{I} = \mathcal{I}^+$ となる $\boldsymbol{\lambda}$ は

$$\lambda_{\omega,k,t} = \frac{H_{\omega,k,z_t^k} U_{k,t}}{\sum_{k'} H_{\omega,k',z_t^{k'}} U_{k',t}}$$

である。

観測行列 $Y_{\omega,t}$ に対して、以下のような手順で可変基底 NMF を適用することができる。

- (i) NMF の結果の $H_{\omega,k}, U_{k,t}$ を使い、 $H_{\omega,k,z_t^k}, U_{k,t}$ の初期値を求める。
- (ii) 補助変数 $\boldsymbol{\lambda}$ の初期値を計算する。
- (iii) Viterbi アルゴリズムにより $\mathbf{Z}$ を更新する。
- (iv) H_{ω,k,z_t^k} と $U_{k,t}$ を更新する。
- (v) 収束するまで (iii) と (iv) を繰り返す。
- (vi) $\mathcal{I} = \mathcal{I}^+$ となるよう補助変数 $\boldsymbol{\lambda}$ の値を更新する。
- (vii) (iii) から (vi) を収束するまで繰り返す。
- (viii) 目的関数 $\mathcal{I}$ の値が十分に小さくなれば、各基底の状態数を順に増加していく。

(i) の初期値を作成する際に、各基底内で類似した状態が設定されるようにするために、 $H_{\omega,k,1}$ には $H_{\omega,k}$ の値を入れ、 $H_{\omega,k,i}$ (i は $1 \leq i \leq N_k$ を満たす整数) には $H_{\omega,k}$ をランダムに微小変化させた値を入れるなどの方法が考えられる。ここで N_k は基底 k の状態数を表し、 N_k の初期値は 2 とする。

(iii) において、観測行列 $Y_{\omega,t}$ に対して $\mathcal{I}^+$ を最小にする最適状態系列 $\{z_t^k\}_{t=1}^T$ を k ごとに動的計画法により効率的に解くことができる。 T は時刻 t の最大値である。まず、 $\mathcal{J}_{k,t}^+(s) := \sum_{\omega} (-Y_{\omega,t} \lambda_{\omega,k,t} \log \frac{H_{\omega,k,s} U_{k,t}}{\lambda_{\omega,k,t}} + H_{\omega,k,s} U_{k,t})$ と置き、 $\delta_1^k(s) = \mathcal{J}_{k,1}^+(s) - \log \pi_s^{(k)}$ とする。次に $t=2, \dots, T$ について、逐次的に $\delta_t^k(s)$ を $\delta_t^k(s) = \min_{s'} [\delta_{t-1}^k(s') + \mathcal{J}_{k,t}^+(s) - \log \pi_{s',s}^{(k)}]$ により計算していくことができる。各ステップで選択される状態番号を $\psi_t^k(s) = \operatorname{argmin}_{s'} [\delta_{t-1}^k(s') - \log \pi_{s',s}^{(k)}]$ に記憶しておくことで、 $t=T$ まで到達後に $z_{t-1}^k = \psi_t^k(z_t^k)$ ($t=T, \dots, 2$) により選択された状態番号を辿っていくことができ、最適経路 $z_1^k, \dots, z_T^k$ を得ることができる。この際、推定された状態系列における遷移確率を元に $\pi_{i,j}^{(k)}$ を更新する。

上記手順の収束後、基底行列 H_{ω,k,z_t^k} では、基底 k の、各状態 z_t^k における各場所 ω に対する訪問確率分布が表現される。係数行列 $U_{k,t}$ では、各時刻 t 毎にどの基底がアクティベートされるか（有効になっているか）が表現される。例えば、時刻 $t=3$ において、 $k=1$ のとき $U_{1,3}$ が 1 で、 k が他の値の時 $U_{k,3}$ が全て 0 になっていた場合、 $H_{\omega,k,z_t^k} U_{k,t}$ の分布は $H_{\omega,1,z_3^1} U_{1,3} = H_{\omega,1,z_3^1}$ となる。ここで z_3^1 は時刻 $t=3$ における基底 $k=1$ の状態を示す。

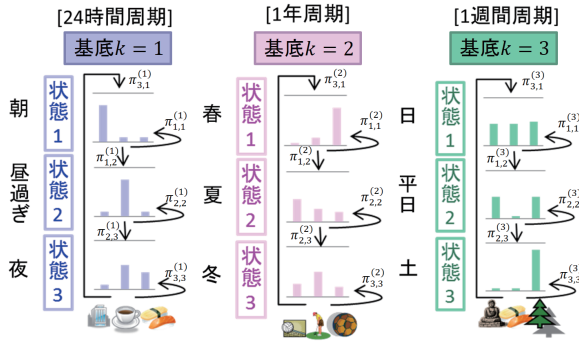


図3 階層的な周期パターン抽出のイメージ

同一時刻において、各基底の状態は1つの状態に定まるものとする。

3. 提案手法

3.1 観測行列の作成

本稿ではユーザ ID, 年月日時刻, 場所の情報を含むチェックインログを想定する。特定のユーザの位置情報履歴に対して可変基底 NMF を適用する場合の前処理として、場所 $\omega \times$ 時刻 t の観測行列 $Y_{\omega,t}$ を作成する。時刻は日付単位, 時間帯単位など任意の単位時間で区切り, 単位時間内に各場所へのチェックイン履歴があれば $Y_{\omega,t} = 1$, ない場合には $Y_{\omega,t} = 0$ とする。

3.2 提案手法によるパターン抽出のイメージ

図3に、提案手法によるパターン抽出イメージを示す。可変基底 NMF の枠組みを導入しているため、従来の NMF における各基底が、それぞれ独立な「状態」と「遷移確率」をもつ left-to-right の HMM に基づいて時間変化する。ここで、「状態」は場所群への頻度分布を表す。例えば基底 $k=1$ の状態1は、会社（ビル）に行きやすい状態、状態2はカフェに行きやすい状態、状態3はカフェと寿司屋の両方に行きやすいが、ややカフェの方が確率が高い状態を示す。時間変化する状態の系列と係数行列との積が、前述の連続行動パターンを形成する。基底 $k=1$ では24時間周期、基底 $k=2$ では1年周期、基底 $k=3$ では1週間周期など様々な階層的周期性を持った訪問パターンの抽出を実現する。

3.3 補間ステップの導入による欠損の考慮

チェックインはユーザの能動的操作により記録されるため、 $Y_{\omega,t} = 0$ となっても、その場所への滞在や他の場所への到着・滞在があった可能性がある。本稿では $Y_{\omega,t} = 0$ となっている箇所を欠損と呼ぶ。 $Y_{\omega,t}$ における非ゼロ要素率は手元のデータで1%を下回り、欠損が多い。提案手法では、可変基底 NMF で計算を繰り返していく際に、抽出されたパターンを考慮して欠損箇所を補間

するステップを導入する。位置情報ログの欠損に対応する方法として、同一時間帯の同ユーザの訪問傾向や、他ユーザの同場所への訪問傾向を使ってデータを補間しながら NTF を適用する手法が提案されている [Wang 14] が、同一時間帯や同一場所の訪問傾向には必ず類似性があるという強い仮定が入っており、異なる時間帯に起こる類似した傾向の考慮や、同一時間帯でも複数の傾向が混在する状況に対応した補間は実現できていない。

提案手法による補間ステップの詳細を説明する。入力データにおいて0以外の値が入っている箇所を Γ とする。提案手法では、入力データで0以外の値が入っている箇所では $Y_{\omega,t}$ と $X_{\omega,t}$ が近づくように、0の値が入っている欠損箇所では $S_{\omega,t}$ すなわち更新前の $X_{\omega,t}$ の値と $X_{\omega,t}$ が近づくように、最小化させる目的関数を変更する。欠損箇所の補間を行うと、補間箇所に強い影響を受けて、各時刻 t におけるアクティベーション $U_{k,t}$ が多くの基底 k で類似した値になってしまう可能性がある。そこで、基底 k 毎のアクティベーションができるだけスパースになるように誘導する目的で $\sum_{k,t} |U_{k,t}|^p$ という項を加える。この項を含めた最適化問題を $\mathbf{H}, \mathbf{U}, \mathbf{Z}$ の最大事後確率推定問題と捉えれば、 $|U_{k,t}|^p$ は $U_{k,t}$ に関する事前分布の対数の符号を逆転したものと見なせる。ここで p が $0 < p < 2$ のとき、 $\sum_{k,t} |U_{k,t}|^p$ を正則化項として用いることは $\mathbf{U}$ の事前分布として非ガウス分布を仮定していることに相当する。例えば独立成分分析において非ガウス性が独立性を測る指標の一つとして用いられていることから [Hyvärinen 04], この項は基底 k 毎のアクティベーションが互いに独立になるような効果をもたらすと考えられる。

上記の議論から提案法の目的関数を定式化する。まず、欠損していない箇所だけを考えて、提案手法による目的関数は以下のように表現される。

$$\begin{aligned} \mathcal{I}(\mathbf{H}, \mathbf{U}, \mathbf{Z}) \\ = \sum_{(\omega,t) \in \Gamma} \mathcal{I}(Y_{\omega,t} | X_{\omega,t}) + w_p \sum_{k,t} |U_{k,t}|^p - \log P(\mathbf{Z}) \end{aligned}$$

しかしながら、2章のように $\mathbf{Z}$ の更新を Viterbi アルゴリズムにより行うためには、全ての (ω,t) において観測値が必要となる。そこで、目的関数に含まれる ω,t が全範囲になるように、欠損領域における仮想データを表す $\mathbf{S}$ (未知変数) を導入し、以下の補助関数 $\mathcal{I}'$ を考える。

$$\mathcal{I}'(\mathbf{H}, \mathbf{U}, \mathbf{Z}, \mathbf{S}) = \mathcal{I}(\mathbf{H}, \mathbf{U}, \mathbf{Z}) + w_s \sum_{(\omega,t) \notin \Gamma} \mathcal{I}(S_{\omega,t} | X_{\omega,t})$$

$\mathcal{I}(x|y)$ は $x=y$ のときにのみ0となる非負値関数であるため、 $\mathcal{I}(\mathbf{H}, \mathbf{U}, \mathbf{Z}) = \min_{\mathbf{S}} \mathcal{I}'(\mathbf{H}, \mathbf{U}, \mathbf{Z}, \mathbf{S})$ となり、 $\mathcal{I}'$ が $\mathcal{I}$ の補助関数としての要件を満たす。従って、 $\mathcal{I}'$ を最小化するように $\mathbf{H}, \mathbf{U}, \mathbf{Z}$ を更新するステップと、 $\mathcal{I}(\mathbf{H}, \mathbf{U}, \mathbf{Z}) = \mathcal{I}'(\mathbf{H}, \mathbf{U}, \mathbf{Z}, \mathbf{S})$ となるように $\mathbf{S}$ を更新するステップ (すなわち $\mathbf{S} = \mathbf{X}$) を繰り返せば、 $\mathcal{I}(\mathbf{H}, \mathbf{U}, \mathbf{Z})$ を単調減少させることができる。この際、 $\mathbf{S}$ を欠損箇所における観測データとみなして、2章と同様の方法で $\mathcal{I}'$ を最小化

できる．なお、 $\mathcal{I}'$ において、 w_s は補間箇所における I-divergence を考慮する重みを示す定数である．補間された値に $X_{\omega,t}$ が近づき過ぎないように重みづけを行う． w_s の値として、例えば非ゼロ要素の割合を利用できる． w_p は全体の目的関数に対して、どの程度の重みでアクティベーションのスパース化を考慮するかを示す定数である．アクティベーション $U_{k,t}$ のスケールに制約がある場合、 p が $0 < p < 2$ のとき、アクティベーション $U_{k,t}$ の中で 0 となる成分が多いほど $\sum_{k,t} |U_{k,t}|^p$ の値は小さくなる（一つの成分を除く全ての成分が 0 のときに最小となる）．このことから $\sum_{k,t} |U_{k,t}|^p$ は $0 < p < 2$ のときにスパース正則化項としての役割を果たす．

また、 $U_{k,t}$ をスパース化させるための関数 $|U_{k,t}|^p$ の上界関数として、 $|U_{k,t}|^p$ に接する放物線 $\frac{p}{2} |\nu_{k,t}|^{p-2} U_{k,t}^2 + \text{Const.}$ を考えて $\mathcal{I}^+$ を設計する．ここで補助変数 $\nu_{k,t}$ が、更新前の $U_{k,t}$ の値となる時に等号が成立する．

これまでの拡張により、提案する欠損と周期性を考慮した可変基底 NMF を適用する際の処理 (iv) における H 、 U の更新式は以下の通りになる．

$$\begin{aligned} H_{\omega,k,j} &\leftarrow \frac{\sum_t \mathbf{1}[z_t = j] \lambda_{\omega,k,t} A_{\omega,t}}{\sum_t \mathbf{1}[z_t = j] B_{k,t}} \\ U_{k,t} &\leftarrow \frac{\sum_{\omega} \mathbf{1}[\omega, t \in \Gamma] \left\{ -H_{\omega,k,z_t^k} + \sqrt{C_{\omega,k,t}} \right\}}{\sum_{\omega} 2w_p p |\nu_{k,t}|^{p-2}} \\ &\quad + \frac{\sum_{\omega} w_s \mathbf{1}[\omega, t \notin \Gamma] \left\{ -H_{\omega,k,z_t^k} + \sqrt{D_{\omega,k,t}} \right\}}{\sum_{\omega} 2w_p p |\nu_{k,t}|^{p-2}} \\ A_{\omega,t} &= \{ \mathbf{1}[\omega, t \in \Gamma] Y_{\omega,t} + w_s \mathbf{1}[\omega, t \notin \Gamma] S_{\omega,t} \} \\ B_{k,t} &= \{ \mathbf{1}[\omega, t \in \Gamma] U_{k,t} + w_s \mathbf{1}[\omega, t \notin \Gamma] U_{k,t} \} \\ C_{\omega,k,t} &= H_{\omega,k,z_t^k}^2 + 4w_p p |\nu_{k,t}|^{p-2} Y_{\omega,t} \lambda_{\omega,k,t} \\ D_{\omega,k,t} &= H_{\omega,k,z_t^k}^2 + 4w_p p |\nu_{k,t}|^{p-2} S_{\omega,t} \lambda_{\omega,k,t} \end{aligned}$$

ここで、 $\mathbf{1}[\cdot]$ は、 $[\cdot]$ 内の条件を満たす場合には 1、そうでない場合には 0 を取る．

3.4 制約付状態推定による周期性の考慮

位置情報ログでは、行動パターンに時間帯、曜日、月、季節などに応じた周期性が想定される．従来の可変基底 NMF でも、NMF の各基底に left-to-right の HMM を仮定し、周期的パターンの抽出を試みたが、時間帯や曜日など外部情報から得られる絶対時間の考慮は行わず、抽出パターンを絶対時間周期と対応付けて解釈できるとは限らなかった．音響音楽信号の解析時に考慮した微妙なスペクトル形状の時間変化に比べ、暦や時間に応じた多様な周期に基づく訪問頻度分布の時間変化が想定される位置情報データでは、周期性の明示的な考慮が有効であると考えられる．提案手法では、絶対時間軸上の周期性を考慮することにより、従来よりも解釈が容易で、情報配信などへの応用にも効果的なパターンの抽出を試みる．

提案手法では、各基底に異なる絶対時間周期を仮定した上で、HMM の状態推定時に、各周期の最初は特定の

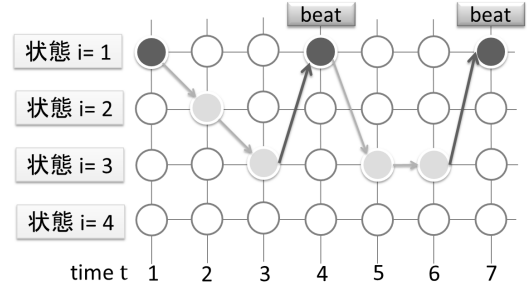


図4 周期性制約のイメージ

状態に固定されるよう制約を与える．周期性による制約のイメージを図4に示す．基底毎に異なる周期の周期性を仮定し、各周期の開始時には状態 i が $i=1$ に固定されるようにする．すなわち、Viterbi アルゴリズムによる状態推定時に、 $t \in \{\text{beat}_k\}$ かつ $i \neq 1$ ならば $\delta_t^k(s) = \infty$ とする． $\{\text{beat}_k\}$ は基底 k に与えた周期の開始時刻 t の集合を示し、図4の例では、 $\{\text{beat}_k\} = \{4, 7\}$ である．

- (1) 各基底に対する周期は、欠損値補間の拡張のみを実施して、アクティベーションのスパース化と周期性の考慮は実施せずに拡張可変基底 NMF を適用した際に、各基底のアクティベーションの時間変化が季節、曜日、時間帯周期を持つ訪問系列を抽出できる可能性が高いかどうかを元に決定する．例えば、毎日複数回アクティベーションが高い場合には、時間帯変化を表現するための 24 時間周期を与える．1 日に 1 回ずつあるいはそれ以下しかアクティベーションが高くなっていない基底では、24 時間の中で状態が変化することはなく、24 時間周期の訪問系列パターンを抽出することは難しいと考えられるが、毎日複数回アクティベーションが高くなっていけば、その中で状態が変化する訪問系列パターンが存在する可能性があるからである．また、1 日 1 回または特定の曜日にアクティベーションが高い場合には、曜日変化を表現するために 1 週間周期を与える．同様にして、特定の季節のみでアクティベーションが高い場合には、季節変化を表現するために、1 年の周期を与える．この時、全ての想定周期に対する集計値が不均一な場合には、周期性を考慮しない．この周期は、周期性を考慮した可変基底 NMF を適用しながら、再度変更していくことも考えられる．決定した周期を用いて拡張可変基底 NMF を適用した際に、特定の状態のみでしかアクティベーションが高くない場合には、周期を小さいものに変更する．一方、アクティベーションが高くなる状態の偏りが少ない場合は、周期の変更を停止する．また、周期が最小となっても状態に偏りが生じた場合は、その基底への周期性付与は行わない．周期の最小値については、データ期間に応じて 1 時間周期、6 時間周期などが考えられる．

現段階の実装では、各基底に最初に与える周期は、4 章で説明するように、欠損値補間の拡張のみを実施した

際の $U_{k,t}$ (例えば図 5) を参照しながら人手で決定している。将来的には、この初期周期の決定を自動化することもあると考えられる。例えば、アクティベーションが閾値以上の値になっている日数と合計回数を数え上げることにより、毎日複数回アクティベーションが高いこと、あるいは 1 日 1 回アクティベーションが高いことが自動的に検知できる。また、同様に、曜日毎または季節毎にアクティベーションが閾値以上になる回数を数え上げることで、特定の曜日または特定の季節のみでアクティベーションが高いかどうかを自動的に判定できる。

また、現段階の実装では、付与する周期を考慮して、例えば 1 週間周期のパターンを発見したい基底では時間窓を 24 時間ごとに設定するなどのデータの再生成は実施しない。すなわち、全基底で観測行列 $Y_{\omega,t}$ の時間窓数 T は共通の値を取る。行列分解においては、行列の各行における列数は等しい必要があるためである。加えて、例えば 1 週間周期で繰り返される訪問系列パターンの中に、毎週土曜日には朝スポーツジムに行って、昼に公園に行く、というような同一日内の典型的訪問系列が存在する可能性があるからである。このような遷移は、毎日繰り返されない場合必ずしも 24 時間周期を仮定した基底で抽出できるとは限らない。しかしながら、将来的には例えばある基底に 1 週間の周期が存在すると仮定した場合、その基底では同一日内には状態が遷移しないような制約を加えることも考えられる。同一曜日内の典型的訪問系列を抽出することは難しくなるが、曜日毎の 1 日を通しての訪問傾向の分析は容易になる可能性がある。

3.5 拡張可変基底 NMF の処理フロー

提案手法による拡張可変基底 NMF の処理の流れは以下になる。

- (I) 時刻 t ・場所 ω 毎の訪問有無を集計し、 $Y_{\omega,t}$ を作成する。
- (II) $Y_{\omega,t}$ に NMF を適用し、基底 k 毎の場所分布 $H_{\omega,k}$ とアクティベーション $U_{k,t}$ を出力する。
- (III) 周期性の考慮とアクティベーション $U_{k,t}$ のスパース化を行わずに、欠損値補間の拡張のみを実施し、アクティベーション $U_{k,t}$ の出力結果を用いて各基底に与える周期を決定する。
- (IV) (III) で決定した周期を用いて、全ての拡張部分を採用した可変基底 NMF を適用する。
- (V) (IV) で出力された各基底 k の各状態毎の場所分布 H_{ω,k,z_t^k} とアクティベーション $U_{k,t}$ を描画しながら検出されたパターンの観察を行い、行動の分析を実施する。

なお、(IV) では、2 章に記載した可変基底 NMF の適用手順のうち、(iii) の Viterbi アルゴリズムで用いる $\mathcal{J}_{k,t}^+(s)$ を補助関数 $\mathcal{I}^+$ に対応するよう変更した上で、 $\delta_t^k(s)$ の定義も 3.4 節で論じたように変更する。また、(iv) の H_{ω,k,z_t^k} 、 $U_{k,t}$ の更新式を式 (1) と (2) に変更する。なお、与える

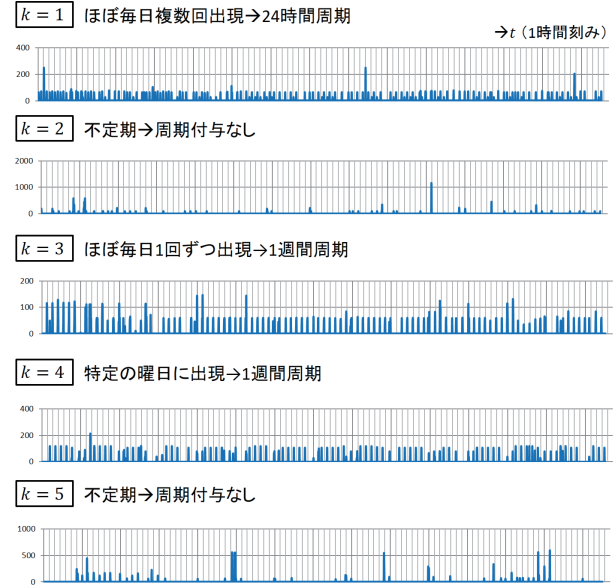


図 5 初期の $U_{k,t}$ と付与する周期の検討例

周期について、(V) でアクティベーションが高くなる状態に偏りが生じている場合には、周期をより短いものに変更して (IV) に戻することも考えられる。

4. 実験結果

実際の位置情報データを用いて、提案手法を検証する。検証実験については、Stanford 大学が以下で公開している Gowalla^{*3}のチェックインデータを利用する。チェックインデータでは、ユーザの能動的操作により、訪問履歴が記録される。基本的に、各地点に到着した際にのみ訪問履歴の記録が実施され、各地点から離れる際の記録は行われないことが多い。チェックインデータでは、各訪問に対してユーザ ID と、日付時刻、同一の緯度経度に割り振られる場所 ID が格納されている。

ここでは、ある特定のユーザのデータ (約 103 日分、全 1496 ログ) のみを抽出して検証実験を行った。チェックイン時刻を 1 時間毎の時間窓に分割した。当該データを集計して $Y_{\omega,t}$ を作成したところ、時間窓数 T は $T = 2438$ 、場所数 Ω は $\Omega = 227$ であった。非ゼロ要素率は 0.269% であった。

図 5 に、初回の可変基底 NMF 適用時 (欠損値補間は行いがアクティベーション $U_{k,t}$ のスパース化と周期性の考慮は未適用) の、アクティベーション $U_{k,t}$ と、決定した初期付与周期を示す。基底数は $K = 5$ とした。各折れ線グラフの横軸は時間窓 t を示し、縦軸は $U_{k,t}$ の大きさを示す。横軸の目盛線は 1 週間の区切りを、補助目盛線は 24 時間毎の区切りを示す。ほぼ毎日複数回のパターンが出現している $k = 1$ では、1 日の中での行動パターンの把握を行うために 24 時間の周期を与えた。また、ほぼ

*3 <https://snap.stanford.edu/data/loc-gowalla.html>

毎日、1日1回ずつアクティベーションが高くなっている $k=3$ では、曜日によって異なる行動パターンの把握を目的として1週間の周期を与えた。同様に、特定の曜日（主に平日）のみにパターンが出現している $k=4$ についても、各曜日で異なるパターンの抽出を試みるために1週間の周期を与えた。アクティベーションが不定期に高くなっている基底 $k=2$ と $k=5$ については、周期性の付与は行わなかった。

図6に、提案法によるアクティベーション $U_{k,t}$ の出力を示す。訪問系列パターンの1つ目（パターン1）として、 $k=3$ （1週間周期付与）の結果を、2つ目（パターン2）として、 $k=1$ （24時間周期付与）の結果を示す。各ヒートマップの横軸は時間窓 t を示す。

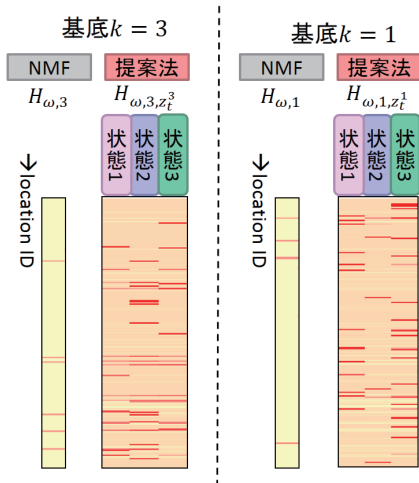
1週間周期を付与したパターン1について、基底行列 H を見ると、状態1には、非習慣的に訪問される店やマンションが、状態2には通勤先と思われる特定地域の娯楽施設やカジュアルな飲食店が、状態3には状態2と同じ地域のバーや居酒屋と、自宅近くと思われる駅や住宅街が分類されていることが分かった。白矢印に示されるように、特定の曜日において状態1と状態2が検出された。また、状態3については、その他の曜日に連続して検出されていることが分かる。また、1週間以内に状態1から状態3、状態2から状態3へと遷移する訪問系列が検出された。状態2には主に昼間に訪問する場所群が、状態3には自宅を含め、夜間によく訪問される場所群が分類されていると捉えることができ、状態2から状態3へと続く行動パターンが特定の曜日に複数回にわたって起こることが分かった。1週間周期で訪問される系列を抽出できたと考えられる。このユーザはほぼ1週間一度状態2にある飲食店を訪問し、その後比較的短い遷移時間で状態3にある場所を訪問することが分かる。こういった傾向が分かると、状態2にある店に行きそうな曜日には、いくつかの通勤先近くの娯楽施設や飲食店の情報を配信し、それらへの訪問が確認されたら、その直後に状態3にある店を紹介するなど、情報配信におけるタイミング向上への活用が考えられる。

24時間周期を付与したパターン2について、基底行列 H を見ると、状態1には通勤時に通る都会の駅とその駅中施設が、状態2には自宅近くの駅近くの商業施設と通勤先近くの飲食店が、状態3には自宅近くの駅と非習慣的に訪問する高級飲食店やエンターテインメント施設が検出された。状態1には、ほぼ毎日通る通勤経路を、状態2と状態3には寄り道をする際に訪問する場所群をそれぞれ抽出することができ、24時間周期の訪問系列をパターン化することができた。例えば、情報配信において、毎日行く場所群と、非習慣的に寄り道をする場所群で情報配信頻度を最適化する活用が考えられる。また、状態3の後、数日以内に状態2の場所を訪問するパターンが2回検出された（緑矢印）。こういった傾向も、情報配信への活用可能性がある。以上のように、パターン1とパ

ターン2の、異なる周期の重ね合わせで習慣的行動と非習慣的行動の訪問系列パターンを表現することができた。

次に、提案法による周期性付与の効果を検証するために、図7に、周期性の付与を行わなかった場合のアクティベーション $U_{k,t}$ の出力を示す。なお、欠損値補間と周期性付与の両方を行わない従来のFHMMを適用すると、全訪問場所への訪問確率が0となった状態からの遷移がほとんど認められず、パターンの抽出が行えなかったため、欠損値補間のみを実施した場合との比較を行う。パターン1として記載した $k=3$ について、周期性の付与を行わないと、提案法の状態1や状態2に抽出された場所のうちの一部がそれぞれ抽出されたが、各状態の出現頻度に明示的な周期性は認められなかった。この結果からも状態1や状態2から状態3への遷移が複数回起きていることは分かるものの、出現頻度が低い状態2が、こういったタイミングに出現しやすいかを予測することは難しく、情報配信への応用は難しい可能性がある。これに対して、提案法では図6に示したように、特定曜日に起きやすい非習慣的行動を状態1と状態2に、またその他の曜日に起きやすい習慣的行動を状態3に分類することができた。パターン2として記載した $k=1$ について、周期を与えないと、状態数4が選択されたが、状態4は一度もアクティブにならず、状態2と状態3には類似した場所群が抽出されていた。周期性の導入により、類似した状態を統合することができたと捉えることができる。

続いて、欠損値補間の効果を確認するために、図8に基底 $k=1, k=3$ における従来手法NMFの基底行列 $H_{\omega,k}$ と、提案手法における各状態の H_{ω,k,z_t^k} の出力結果を示す。前述の通り欠損値補間を行わずに可変基底NMFでパターン抽出を行うことはできなかったため、欠損値補間なしに一定のパターンが推定できたNMFとの比較を行う。各基底で、左側はNMFによって抽出された基底行列 H の分布である。色が濃いほど値が大きいことを示す。欠損値補間を行わないNMFの結果では、実際に訪問が記録された数箇所のみで値が大きくなっている。これに対し、右側の拡張可変基底NMFの結果では、訪問可能性がある多くの場所が薄く検出されている。欠損箇所の補間を行った上で、同時に複数の基底がアクティベートされずに、基底毎の成分が独立になるよう、アクティベーションのスパース化を行うことにより、図8の下部に示されているような行動パターンを抽出することが可能になった。加えて、薄く色がついている訪問場所には、実際に訪問した場所だけでなく、訪問可能性が高い場所群が表現されており、各状態が、場所群への「訪問気分」に該当すると期待される。こういった「訪問気分」が獲得できれば、情報配信の際に、より多くの候補を用意して情報の確度を上げることに貢献できる可能性がある。

図 6 拡張可変基底 NMF (提案法) の $U_{k,t}$ の出力図 7 周期性付与を行わず欠損値補間のみを行った可変基底 NMF の $U_{k,t}$ の出力図 8 NMF と拡張可変基底 NMF の H_{ω,t,z_t^k} の出力比較

5. 考察

4 章では, ある特定のユーザのログを用いて, 1 週間周期・24 時間周期の訪問系列パターンが抽出できたことを示した. 本章では, 発見されたパターンの妥当性を, 正当性・網羅性の 2 つの観点から評価する.

5.1 パターンの正当性

各パターンの各状態 (s) で $H_{\omega,k,s}$ の値が大きく算出された場所が, 本当に観察期間を通して想定した周期で繰り返し訪問されているかを検証する. 各場所 ω に対する, 曜日または時間帯別訪問回数を集計したものを, それぞれ $C_{\omega,d}, C_{\omega,h}$ とする. これを用いて, パターン k ・状態 s 毎に, $H_{\omega,k,s}$ の値で重みづけした場所 ω 別訪問時間分布の

和を求めたものを図 9 に示す. すなわち, 図 9 (左) のパターン 1 (1 週間周期) では, $\sum_{\omega} H_{\omega,k,s} C_{\omega,d}$ の値が, 図 9 (右) のパターン 2 (24 時間周期) では, $\sum_{\omega} H_{\omega,k,s} C_{\omega,h}$ の値がヒートマップ形式で描画されている. 赤色が濃いほど値が大きいことを示す.

パターン 1 の状態 1 と 2 に含まれる場所は毎週土曜日に訪問される傾向が強かったこと, 状態 3 に含まれる場所は月曜日から金曜日に訪問される傾向が強かったことが分かる. また, 状態 2 は月曜日に, 状態 3 は土日に訪問される場合も見られ, これが 4 章で述べた状態 2 から状態 3 へと続く行動パターンに該当すると考えられる. 同一曜日内に繰り返し起こる訪問系列を状態 1 と 2 に抽出することができたのは, 3.4 節で述べたように付与周期を考慮したデータの再生成を行わなかったことの効果であると言える. パターン 2 では, 状態 1 に含まれる場所は朝 8 時台と夜に, 状態 2 に含まれる場所は夜のみに, 状態 3 に含まれる場所は朝 8 時台のみに訪問されやすいことが分かる. 朝→夜→朝の 24 時間周期のパターンを発見した上で, 朝だけ, 朝と昼の両方, 夜だけに訪問されやすい場所を分類できた. 以上から, 提案手法が観察期間を通して特定の曜日または時間帯で訪問されやすい場所群とその訪問系列を抽出できたことが確認できた.

5.2 パターンの網羅性

次に, 提案手法が, 特定周期で連続して訪問される場所を網羅的に抽出できているかどうかを検証する. 図 10 は, 本論文で利用した特定のユーザの観察期間中の訪問履歴のうち, 訪問回数の上位 15 箇所について, パターン 1 では曜日別, パターン 2 では時間帯別の訪問回数の標準偏差 σ_d, σ_h と $\hat{H}_{\omega,k,s}$ の値を示したものである. ここで,

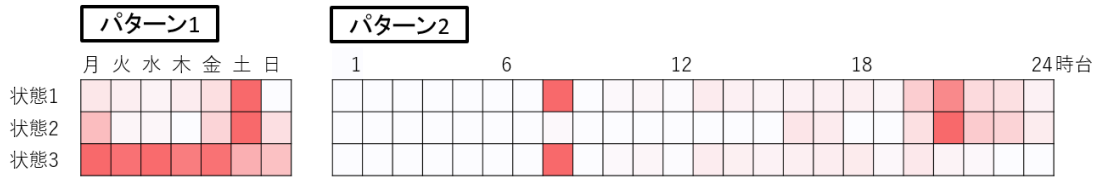


図9 パターンの正当性評価. 各パターン・状態の時間別訪問傾向.

順位	パターン1($k = 3$: 1週間周期)				パターン2($k = 1$: 24時間周期)			
	σ_d	$\hat{H}_{\omega,3,1}$	$\hat{H}_{\omega,3,2}$	$\hat{H}_{\omega,3,3}$	σ_h	$\hat{H}_{\omega,1,1}$	$\hat{H}_{\omega,1,2}$	$\hat{H}_{\omega,1,3}$
1	1.59	0.00E+00	0.00E+00	0.00E+00	9.65	8.28E-02	9.40E-04	5.69E-02
2	0.83	0.00E+00	0.00E+00	0.00E+00	12.01	4.49E-02	8.71E-04	5.69E-02
3	0.83	0.00E+00	0.00E+00	0.00E+00	12.01	4.49E-02	8.71E-04	5.69E-02
4	2.32	0.00E+00	0.00E+00	0.00E+00	12.26	0.00E+00	0.00E+00	0.00E+00
5	2.26	3.12E-02	2.93E-02	9.04E-02	9.23	0.00E+00	0.00E+00	0.00E+00
6	2.86	3.12E-02	2.78E-02	7.90E-02	9.45	0.00E+00	0.00E+00	0.00E+00
7	5.99	0.00E+00	0.00E+00	0.00E+00	12.58	1.74E-07	4.04E-07	2.85E-07
8	6.01	0.00E+00	0.00E+00	0.00E+00	11.58	9.90E-08	4.86E-07	2.89E-07
9	5.52	0.00E+00	0.00E+00	0.00E+00	11.38	2.85E-08	5.17E-07	3.62E-07
10	3.52	0.00E+00	0.00E+00	0.00E+00	5.48	2.53E-01	1.72E-01	2.01E-04
11	2.13	3.71E-02	2.07E-02	6.67E-02	5.49	0.00E+00	0.00E+00	0.00E+00
12	3.80	6.14E-04	3.20E-04	1.08E-01	7.78	4.42E-08	3.79E-07	1.57E-07
13	3.77	6.14E-04	3.18E-04	1.07E-01	7.58	1.29E-07	9.47E-08	1.53E-07
14	3.84	9.40E-08	1.02E-07	2.96E-07	6.78	1.26E-07	3.85E-07	2.97E-07
15	1.05	0.00E+00	0.00E+00	0.00E+00	1.42	0.00E+00	0.00E+00	0.00E+00

図10 パターンの網羅性評価. 訪問回数が特に多い場所における曜日・時間帯別の訪問回数の偏りの強さ(標準偏差)と $\hat{H}_{\omega,k,s}$ の値.

$\hat{H}_{\omega,k,s}$ は $H_{\omega,k,s}$ の値を全場所の値の合計 $\sum_{\omega} H_{\omega,k,s} = 1$ となるように正規化したものである. セルの色は, 上位15箇所だけでなく全場所を通じて, 値が大きいほど濃い色になるように色づけされている. また, $H_{\omega,k,s}$ の値が0となっていた際には青色で表記した. 上位15箇所を選んだのは, 訪問回数が多い場所ほど, 周期的に訪問される可能性が高いと考えたためである. なお, 個人情報の保護のため, 上位15箇所の具体的な訪問回数の記載は避けるが, 該当ユーザの場所別の訪問回数分布は図11のようになっている. 図11で灰色の帯に含まれている箇所が, 図10で抽出されている箇所に該当する. 16位以下の場所の訪問回数が15以上に比べ極端に少ないため, 上位15箇所を抽出した.

図10を見ると, σ_h の値が特に大きく, σ_d の値はあまり大きくない順位1,2,3の場所は, パターン1には含まれず, パターン2のみに含まれている. 一方, σ_d, σ_h のどちらも相対的に大きな値となっている順位5~14の場所は, 大きさの大小や他に各パターンで抽出された周期性との近さによってパターン1と2の両方もしくは片方に含まれている. σ_d, σ_h の値がどちらも小さい15位の場所は, パターン1にも2にも含まれていない. また, 時間帯毎の訪問数の偏りが大きいにもかかわらず, パターン2に含まれなかった場所4は, パターン2で抽出された朝8時ではなく, 毎朝9時頃に多く訪問される場所であり, 他に1週間周期を与えた基底 $k=4$ のパターンで

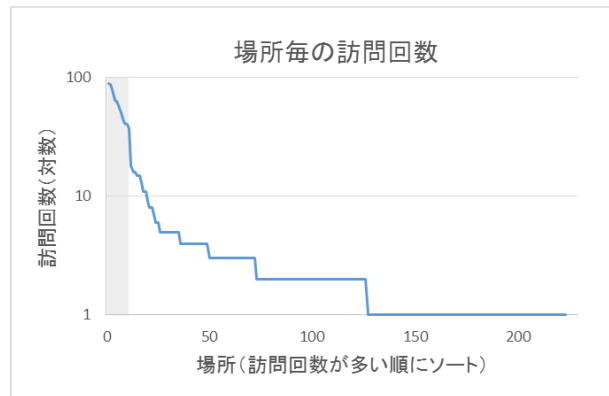


図11 該当ユーザの場所ごとの訪問回数分布

$H_{\omega,k,s}$ の値が大きくなっていた. 以上から, 訪問回数が上位15のうち, 1週間周期・24時間周期の繰り返し訪問が多かった場所のほとんどが, 各周期を仮定したパターンのいずれかには含まれており, 抽出されたパターンの網羅性は高いと考えられる.

加えて, 提案手法では行動を複数の周期を持つ訪問系列の重みづけ和で近似することを目的としていたが, 図10を見ると, 例えば順位12,13,14の場所など, σ_d, σ_h の値がどちらも同じくらい大きい場合には2つのパターンに含まれていることから, 複数周期の重ね合わせで行動が近似できていることも確認できる. その一方で, 青色で示されているように周期成分が十分に強くない場合や,

抽出された周期性と異なる場合には、 $H_{\omega,k,s}$ の値は 0 となり、各基底成分の独立性も保たれている。これは、提案手法で基底の独立性を保つために $U_{k,t}$ のスパース化制約を考慮したことの効果であると考えられる。

6. ま と め

本稿では、可変基底 NMF を位置情報ログの欠損や周期性を考慮して拡張し、位置情報ログに周期的に出現する、階層的な周期をもつ行動パターンを抽出する手法を提案し、抽出されたパターンの妥当性を評価した。提案手法では、欠損の多いデータにおいても、各時刻で訪問可能性の高い場所群を予測できると期待される。また、場所に応じた訪問周期、滞在時間、次に訪問する場所の傾向など、複数のユーザに共通する行動のパターンを把握して、都市計画などに役立てたり、ユーザ毎に特有な行動パターンを抽出することにより、情報配信のタイミングや内容に関する満足度を向上させたりといった効果が期待される。今後は、より大規模なデータへの適用を試み、より柔軟かつ多様な周期性を持つ階層的な行動パターンの抽出を試みたい。また、各行動が習慣に沿っているかどうかを数値化する習慣度算出手法 [Hayashi 14] や、NMF において U に周期性を仮定して、その周期を自動的に推定する手法 [林 16b] との融合を行いたい。

◇ 参 考 文 献 ◇

- [Ashbrook 03] Ashbrook, D. and Starner, T.: Using GPS to learn significant locations and predict movement across multiple users, *Personal and Ubiquitous Computing*, Vol. 7, No. 5, pp. 275–286 (2003)
- [Csiszár 75] Csiszár, I.: I-divergence geometry of probability distributions and minimization problems, *The Annals of Probability*, pp. 146–158 (1975)
- [Eddy 96] Eddy, S. R.: Hidden Markov models, *Current Opinion in Structural Biology*, Vol. 6, No. 3, pp. 361–365 (1996)
- [Fan 14] Fan, Z., Song, X., and Shibasaki, R.: CitySpectrum: A non-negative tensor factorization approach, in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 213–223 (2014)
- [Ghahramani 97] Ghahramani, Z. and Jordan, M. I.: Factorial hidden Markov models, *Machine Learning*, Vol. 29, No. 2-3, pp. 245–273 (1997)
- [Hayashi 14] Hayashi, A., Matsubayashi, T., and Sawada, H.: Regular behavior measure for location based services, in *Proceedings of the 2014 ACM Conference on Web Science*, pp. 299–300 (2014)
- [林 16a] 林 亜紀, 亀岡 弘和, 松林 達史, 澤田 宏: 位置情報履歴の欠損と周期性を考慮したパターン抽出手法, 第 7 回 WII2 研究会予稿集 (2016)
- [林 16b] 林 亜紀, 亀岡 弘和, 松林 達史, 澤田 宏: 非負値周期成分分析手法による音楽音響信号の音源分離, 日本音響学会 2016 年春季研究発表会講演論文集, pp. 639–642 (2016)
- [Hyvärinen 04] Hyvärinen, A., Karhunen, J., and Oja, E.: *Independent Component Analysis*, John Wiley & Sons (2004)
- [Lee 01] Lee, D. D. and Seung, H. S.: Algorithms for non-negative matrix factorization, in *Advances in Neural Information Processing Systems*, pp. 556–562 (2001)
- [Nakano 10] Nakano, M., Le Roux, J., Kameoka, H., Kitano, Y., Ono, N., and Sagayama, S.: Nonnegative matrix factorization with Markov-chained bases for modeling time-varying patterns in music spectrograms, in *Latent Variable Analysis and Signal Separation*, pp. 149–156, Springer (2010)
- [Ozerov 09] Ozerov, A., Févotte, C., and Charbit, M.: Factorial scaled hidden Markov model for polyphonic audio representation and source separation, in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA-2009)*, pp. 121–124 (2009)
- [澤田 12] 澤田 宏: 非負値行列因子分解 NMF の基礎とデータ/信号解析への応用, 電子情報通信学会誌, Vol. 95, No. 9, pp. 829–833 (2012)
- [Shashua 05] Shashua, A. and Hazan, T.: Non-negative tensor factorization with applications to statistics and computer vision, in *Proceedings of the 22nd International Conference on Machine Learning*, pp. 792–799 (2005)
- [Wang 14] Wang, Y., Zheng, Y., and Xue, Y.: Travel time estimation of a path using sparse trajectories, in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 25–34 (2014)

[担当委員: 奥 健太]

2016 年 5 月 9 日 受理

—— 著 者 紹 介 ——



林 亜紀(正会員)

2010 年, お茶の水女子大学理学部情報科学科卒業。2012 年, 同大学院人間文化創成科学研究科博士前期課程修了。同年, 日本電信電話株式会社入社。NTT サービスエボリューション研究所にて, 可視化およびデータマイニングの研究開発に従事。2015 年, お茶の水女子大学大学院人間文化創成科学研究科博士後期課程修了, 博士(理学)。2016 年より株式会社 NTT ドコモサービスデザイン部にて, 音声エージェントサービスの開発に従事。日本データベース学会, 日本音響学会各会員。第 7 回 WII2 研究会優秀研究賞等受賞多数。



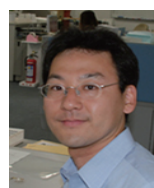
亀岡 弘和(正会員)

2002 年, 東京大学工学部計数工学科卒業。2004 年, 同大学院情報理工学系研究科修士課程修了。2007 年, 同大学院情報理工学系研究科博士課程修了。同年, 日本電信電話株式会社入社。NTT コミュニケーション科学基礎研究所配属。2011 年, 東京大学大学院情報理工学系研究科客員准教授。2016 年, 国立情報学研究所客員准教授。音声・音楽を対象とした音響信号処理・機械学習の研究に従事。日本音響学会, 情報処理学会, IEEE 各会員。博士(情報理工学)。IEEE Signal Processing Society 2008 SPS Young Author Best Paper Award 等受賞多数。



松林 達史(正会員)

2000 年, 京都大学理学部物理学科卒業。2002 年 10 月より 2 年半, 理化学研究所非常勤研究員。2005 年, 東京工業大学大学院理工学研究科地球惑星科学専攻博士課程修了。同年, 日本電信電話株式会社入社。以来, NTT コミュニケーション科学基礎研究所, NTT サービスエボリューション研究所にて, グラフ可視化および機械学習の研究開発に従事。博士(理学)。



澤田 宏(正会員)

1991 年, 京都大学工学部情報工学科卒業。1993 年, 同大学院情報理工学系研究科修士課程修了。同年, 日本電信電話株式会社入社。以来, 同社コミュニケーション科学基礎研究所にて, VLSI 向け CAD および計算機アーキテクチャの研究に従事。2000 年より, 信号処理, 特にブラインド音源分離の研究に従事。2009 年より, 知能創発環境研究グループリーダー。2013 年より, サービスエボリューション研究所ヒューマンアナリシスプロジェクトグループリーダー。2016 年よりコミュニケーション科学基礎研究所協創情報研究部部長, 現在に至る。2001 年, 京都大学博士(情報学)。電子情報通信学会, 日本音響学会, IEEE 各会員。