DEIM2021 A31-3

携帯電話人口統計および施設情報の複合非負値行列因子分解に基づく都市動態の変化点検知

磯川　弘基† 豊田　正史†† 梅本　和俊†† 商 海川†,††† 是津　耕司†††

喜連川　優†,††††

† 東京大学大学院情報理工学系研究科　〒 113–8656 東京都文京区本郷 7–3–1

†† 東京大学生産技術研究所　〒 153–0041 東京都目黒区駒場 4–6–1

††† 情報通信研究機構　〒 184–8795 東京都小金井市貫井北町 4-2-1

†††† 国立情報学研究所　〒 101–8430 東京都千代田区一ツ橋 2–1–2

E-mail: †{isokawa,toyoda,umemoto,shang,kitsure}@tkl.iis.u-tokyo.ac.jp, ††zettsu@nict.go.jp

あらまし　都市空間における人の動き（都市動態）は日々様々に変化するため，それらの検知及び分析を自動化・高度化することは商業・都市計画の意思決定を行う上で重要である．都市動態の変化は周辺に存在する施設と密接に関連して発生するが，従来の人口変動に関する情報のみに基づく変化点検知では，各々の変化に関与する施設について十分に説明できない．そこで本研究では，原因を多角的に分析可能な形で都市動態の変化点を検知することを目的として，人口変動と施設情報を考慮した変化点検知に取り組む．具体的には，都市における各区画の潜在的な役割を人口変動及び施設情報を用いた同時パターン分解により抽出した上で，区画毎の複合的な都市役割の経時的変化に基づく変化点検知手法を提案する．提案手法の有効性を検証するために，人口変動に関する情報として携帯電話の GPS ログデータから得られる混雑統計データ，施設位置情報として地図データを用いた実験を行った．その結果，提案手法は人口変動単体を用いた手法と比較して，都市の潜在的な役割として汎化性能および解釈性の高いパターンを抽出可能であり，変化点検知の性能も優れていることが確認された．最後に，検出された変化点の分析を通じて，提案手法で得られる複合的なパターンが都市動態の変化原因と考えられる施設の推測に利用可能であることを確認した．

キーワード　都市動態，変化点検知，非負値行列因子分解

1 はじめに

都市において，各地域がどのように利用されているかを知ることは，地域毎の特性に応じた適切な都市計画や商業活動，災害復興を行う上で重要である．従来，地域利用のされ方に関しては交通量・通行量調査等といった人的及び時間的コストの高い方法で収集されたデータの分析がその役割を担ってきた [1]．近年では，携帯電話やカーナビをはじめとした GPS 端末の普及にともない，大量かつ広範囲に収集された人々の行動履歴を用いることで，低コスト，広範囲，長期間に渡って人々の都市活動を観測することが可能となった [2]．都市動態は主に，「各地点において，単位時間ごとにどのくらいの人や車が存在するか」という形で観測され，その値は通勤・通学，飲食，歓楽等，様々な目的を持った人々の都市活動が複合したものであると考えられる．観測された都市動態の背後で，潜在的に人々がどのような目的で各都市を利用しているのか，あるいは都市がどのような役割で利用されているのかを推定することは都市動態を理解する上で重要となる．

都市の潜在的な役割の推定に関しては多くの研究がなされており，近年の研究では，Latent Dirichlet Allocation (LDA) [3] や Non-negative Tensor/Matrix Factorization (NTF/NMF) [4,5] を用いることで，都市動態を潜在的な人口変動パターンに分解した上での分析が多くなされている [6,7]．これらの手法では，「平日昼間に人が集まる」「深夜に人が集まる」等の典型的な時系列が人口変動パターンとして抽出され，各地区の都市動態はそれら複数のパターンの重み付け和として表現される．

多くの分析において，都市動態は観測期間中変化しない，静的なものとして扱われるが，実際の都市動態は災害やイベント，施設変化等の原因により時間変化する．都市動態の変化について理解することは，人々の需要の変化を発見することに繋がるため，これらの変化を検知・分析することは商業・都市計画を行う上で極めて重要である．また，都市における施設の開閉業を速やかに捉えることは，現状人海戦術で行われる地図更新作業を効率化することにもつながる．

都市動態の局地的な変化を捉える試みとしては，潜在的な人口変動パターンに対する所属人口を，各地区で長期追跡することにより変化点を検知する手法が提案されている [8,9]．しかし，既存研究では，検知された都市動態の変化に関して，個々に周辺施設等を調べた上で，変化原因の解釈を後付け的に考察するにとどまっており，変化の原因となった施設との関連について定量的な分析はなされていない．特に，広範囲の都市動態においては日々多数の変化が検知されるため，各変化がどのよ

うな特徴を持ち，どのような施設と関連しているか，可能な限り人手をかけずに分析できることが重要である．

そこで本研究では，原因を多角的に分析可能な形で都市動態の変化点を検知することを目的として，人口変動と施設情報を考慮した解釈性の高い変化点検知手法を提案する．具体的にはまず，都市における各区画の潜在的な役割を，人口変動及び施設情報を用いた同時パターン分解により抽出する．これにより，従来の研究で抽出された人口変動パターンに加えて，対応する施設に関するパターンを同時に抽出することが可能となり，人口変動と施設との関係を明示的に理解することが可能となる．区画毎の複合的な都市役割の経時的変化に基づく変化点検知手法を提案する．その上で，人口変動パターンに対する所属人口を長期間に渡って追跡することで，都市動態に変化があった地点と時間の検知を行う．人口変動パターンは施設パターンと関連付けられているため，検知された都市動態の変化がどのような施設と関連したものなのかを分析することが容易となる．

本稿では提案手法の有効性を検証するために，人口変動に関する情報として携帯電話の GPS ログデータから得られる混雑統計データ，施設位置情報に各年度末に出版される地図データを用いた実験を行った．実験では，欠損値推定により抽出されたパターンの汎化性能を評価し，エントロピーに基づく評価からパターンの解釈性を確認した．結果として，施設情報を用いた同時パターン分解を行うことにより人口変動単体を用いたパターンと比較して優れたパターンが抽出されることを確認した．また，変化点の検知性能評価において，提案手法が従来手法に比べて優れていることを示した．最後に，人口変動が変化した点において，そこで起きた施設変化についてパターンを基にした推測が可能であることを確認した．

2 関 連 研 究

本節ではまず，GPS ログデータなどから観測される都市動態から，潜在的な都市の役割を静的に同定した研究について述べる．次に都市動態の長期的な変化を捉えることを試みた研究について述べる．

2.1 潜在的な都市役割の同定

都市空間における人々の多種多様な活動の集合として現れる都市動態から，潜在的な人の活動や，土地の使われ方を抽出する取り組みは多くなされている．代表的な研究として，Toole らは携帯電話でのメッセージや通話を行った際に送信される GPS ログを時間毎にカウントしたデータを用いて，商業地や住宅地などといった都市役割のクラス分類を行った [6]．また Yuan らは，タクシーの GPS に基づく移動履歴から各地域における人流を推定し，都市役割に関する潜在パターンを LDA により抽出した上で，そのパターンを基に地域性のクラスタリングを行った [7]．これらの，静的に抽出された潜在的な都市役割に基づいて，都市動態の長期的な変化に着目した研究を

2.2 都市動態の変化に関する研究

本研究で対象とする都市動態の変化に着目した研究について述べる．Fan ら [10] は，東日本大震災前後の都市役割の変化を分析するために，NTF を用いて携帯電話人口統計データの解析を行なった．福島県全域における震災を含む 4 ヶ月間のデータに対し，1 日における日付，時刻，及び 900 m × 900 m のメッシュ [1]を単位とした場所からなる 3-mode tensor を構成し，NTF によって 9 つのパターンを抽出した．ここで各パターンは日付, 時刻, 場所の 3 つの mode に関する特徴を持つことになる．例えば，1 日の時系列を示す時刻方向には夜間に人口が増えるというような特徴がありかつ，長期的な特徴を示す日付方向には震災後に減少するといったような特徴をもつパターンが現れる．さらにその場所方向の特徴を見ることで，震災後に住宅地が減った場所を確認することが可能となる．しかし Fan らによる手法は，パターン自体がもつ日付方向の系列に注目することで長期的な変化を捉えており，都市動態の変化がパターン自体に現れる必要がある．そのため，この手法の適用範囲は大規模災害のように広範囲で同時多発的に都市動態が変化する場合に限定され，新規施設にともなう人口増加のような地域間で異なる時期に発生する局所的な変化を捉えることは難しい．

そこで Maeda ら [8] は，(| 場所 | × | 日付 |) × | 時刻 | の行列に対し NMF を行うことで，局所的な都市動態の変化を捉える手法を提案している．交通系 IC カードの利用履歴に基づく駅の人口データを用いた実験を通じて，潜在的な駅役割のパターンに基づいた変化点の検知が可能であることが確認されている．また磯川ら [9] は，携帯電話人口統計データを用いることで，特定の場所に依らず，都市役割に基づいた変化点検知を行う手法を提案している．しかし，これらの手法では，人口変動パターンと施設との関係が与えられていないため，人口変動と関連する施設や，その変化の原因となり得る施設を分析することが難しく，後付的に個別の分析を行うにとどまっている．広範囲の都市動態においては多種多様かつ多数の変化が検知されるため，それらを人手で個別に調査・分析を行うことは困難である．

そこで本研究では，従来の研究で抽出された人口変動パターンに加えて，施設に関するパターンを同時に抽出することで人の動きと施設の関係を予め明らかにした上で，人口変動に関する変化点の検知を行う．

3 提案手法：解釈性の高い都市変化点検知

都市空間においては，都市動態が日々各所で変化するため，それらを効率よく検知・分析するにあたって，求められる要件は以下の 2 つが挙げられる．

- 特定の地域に依らず，広範囲の変化点を高い性能で検知でできる
- 検知された多数の変化点に関して，個別に現地調査や文献収集といった労力をかけることなく，変化の原因の推測・分析が可能である．

特に，変化点に対する分析の自動化を考慮したとき，変化点検

1：地図を矩形分割した際に割り振られる地域メッシュコード

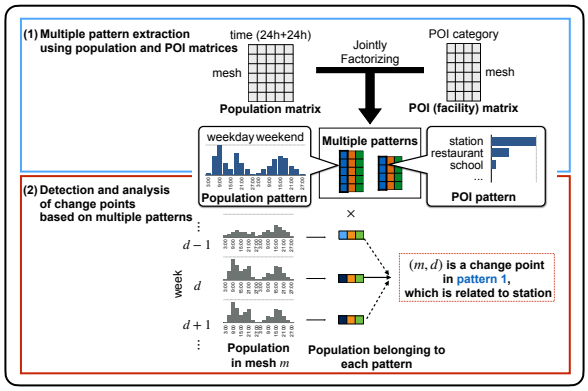

図 1: 提案手法の概要.

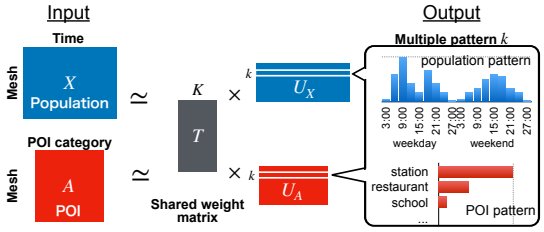

図 2: NMMF を用いた複合パターンの抽出.

知におけるモデルの解釈性の高さが求められる.

そこで本研究では，解釈性が高く，変化の原因を追求しやすい手法として，潜在的なパターンに基づく変化点検知を行う．パターンの抽出においては，検知された変化の多角的な分析を可能とすることを主な目的とし，都市動態と施設を関連付けるために，人口データと施設データの両方を用いた分解を行う．特定の地域に依存しない 2 種類のデータを組み合わせることで汎用的かつ高性能な変化点検知が期待できる.

提案手法の概要を図 1 に示す．手法は大きく 2 つのステップから構成される．まず第一に，各メッシュ，各時刻における人口を表す情報，および各場所に存在する施設（POI）に関する情報を入力として，Non-negative Multiple Matrix Factorization (NMMF) [11] を用いることで人口変動パターン，及びそれに対応する施設パターンを同時に抽出する．次に，得られた各人口変動パターンに対する所属人口を長期的に追跡することで，人口変動に変化が生じたメッシュ，週の組を検知する．検知された変化に関しては，その人口変動パターンと，それに関連した施設パターンを追跡することが可能であるため，どのような施設と関連した変化であるのかを分析することが可能となる．以下で提案手法の各要素について順に説明する.

3.1 複合的な都市役割の抽出

本節では人口変動および施設情報を用いた複合的な都市役割の獲得手法について説明する．人口変動単体から抽出される都市役割は，得られた人口変動パターンの時系列から直感的にオフィスや商業地といった解釈がなされるにとどまる．またパターンを基にした変化点検知を行う際にも，関連の深い施設を各変化点に対して個別に人手で調べる必要があった．施設パターンを同時に抽出することにより，潜在的な都市役割を人口と施設の双方から解釈することが可能となる．また，後に検出される変化点に対して，個別に変化原因を調査することなく，人口と施設に関する多角的な分析が可能となる.

本研究で用いるデータに NMMF を適用する様子を図 2 に示す．ここで，入力として人口情報に関するターゲット行列 X, および施設情報に関する補助行列 A を与える．本稿において行列 X, A はいずれもメッシュを行とする．NMMF により 2 つの行列 X, A をそれぞれ，任意の要素が非負な 2 つの行列の積（$T \times U_X^{\mathrm{T}}$, $T \times U_A^{\mathrm{T}}$）に分解する．ここで，重み行列 T を共有しながら分解を行うことにより，人口パターン U_X と施設パターン U_A の各 k 行目が一対一に対応する形で抽出される．以下に各入力行列の構築手法について，ついで NMMF による複合パターン抽出手法について順に述べる.

3.1.1 人口行列 X の構築

人口情報に関する行列 X の各要素 $x_{m,t}$ はメッシュ m，時刻 t における人口を表す．時刻の粒度としては，日単位の時刻や，週単位で時刻が考えられる．一般に，人口変動は週の周期性をもっており，特に平日と休日とでは人が集まる施設は大きく異なることが予想される．日単位での時刻を用いた場合，平日と休日の区別がされず，人口変動パターンの表現力が低下する．具体的には，週単位での時刻を用いた場合，平日の昼に人が集まり休日には人が減るようなパターンを表現することが可能であるのに対し，日単位では表現することができない．そこで本研究では，平日の平均時系列と土日祝日の平均時系列を連結した平日 24 時間＋休日 24 時間の週時刻を列とする.

3.1.2 施設行列 A の構築

補助行列 A の各要素 $a_{m,e}$ はメッシュ m における施設種別 e に属する施設の数を表す．単純にはメッシュ内に存在する施設種別 e のカウント値を用いることが考えられる．しかし，一般に各地域における人口変動は近傍に存在する施設の影響を受けるのに対し，上記の方法では近隣のメッシュに存在する施設の影響を見積もることができない.

そこで，各メッシュにおける施設数を施設座標とメッシュの中心座標との間の距離に応じた値として計算する．具体的には，施設種別ごとに施設座標を入力とした二次元のカーネル密度推定（KDE）を行うことで施設の影響力をモデル化する [12–14]．ここで，今回対象とする全メッシュに接する近傍 9 メッシュまでを含めた領域内に存在する施設を用いてモデル化を行う．施設種別ごとの全施設の座標に対して，ガウシアンカーネルを用いたカーネル密度推定を行う．近傍への影響力の染み出しの大きさを調節するハイパーパラメタとしてガウシアンカーネルのバンド幅 bw rad を設定する．なお，座標上の点間の距離の算出には球面モデルによる Haversine formula 法を用い，地球を半径 6,378.137 km の真球と仮定したとき，$bw = 10^{-5}$ rad あたり約 64 m に換算されることになる．各メッシュにおける周辺メッシュからの影響を考慮した平滑化施設数は，メッシュの中心座標におけるカーネル密度の推定値に施設種別 e の領域内全体における施設数を乗じた値とする.

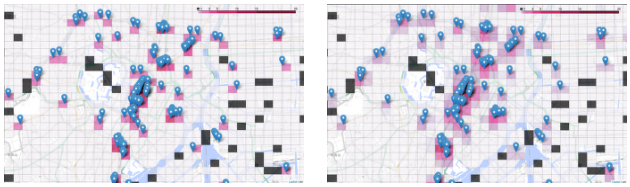

図 3: KDE を適用前後の駅の分布．左：メッシュ内の施設数，右：KDE の適用による平滑化施設数（$bw = 10^{-2}$rad）．

地理院タイル [2]（淡色地図 [3]）を加工して作成

図 3 にメッシュ内の施設数をカウントした場合（左）と KDE 適用した場合（右）の駅の施設数の分布を示す．KDE の適用により，施設の局在的な特徴を保ちながら，駅周辺かつメッシュ内には駅を含まないメッシュに対して値が染み出していることが確認できる．

3.1.3 NMMF に基づく行列の同時分解

図 2 の分解において，メッシュ方向の軸をもつ行列 T は 2 つの分解において共有されるため，入力の複数行列間のスケールは近いことが望ましい．そこで，入力行列の各メッシュにおける人口，および施設数の最大値が 1 となるよう正規化を行う．

上記のように構築した行列 X, A に対し，今回の分解において最小化すべき目的関数 D は，所定の距離関数を d を用いて次式で表される．

$$D(X, A|\hat{X}, \hat{A}; T, U^X, U^A) = \sum_{m\in M}\sum_{t\in T} d(x_{mt}|\hat{x}_{mt}) + \eta \sum_{m\in M}\sum_{e\in E} d(a_{me}|\hat{a}_{me}) \quad (1)$$

ここで，復元される行列を $\hat{X} = T \times U_X^{\mathrm{T}}$，$\hat{A} = T \times U_A^{\mathrm{T}}$ としたとき，$\hat{x}, \hat{a}$ はその要素である．また，M, T, E はそれぞれ，メッシュ，時刻，施設種別の集合である．また，η は補助行列 A の分解における重要度を定めるハイパーパラメタであり，$\eta = 0$ のとき，NMMF は NMF に一致する．

式 1 の最適化手法について説明する．目的関数は非凸であるため，解析的に大域最適解を求めることができない．局所最適解を求めるために交互最適化に基づく手法が提案されており，各因子に関する更新式は以下のように導かれる．

$$t_{mk}^{(\mathrm{new})} = t_{mk} \frac{\sum_{t\in T}\left[\frac{x_{mt}}{\hat{x}_{mt}} u_{tk}^X\right] + \eta \sum_{e\in E}\left[\frac{a_{me}}{\hat{a}_{me}} u_{ek}^A\right]}{\sum_{t\in T} u_{tk}^X + \eta \sum_{e\in E} u_{ek}^A}, \quad (2)$$

$$u_{tk}^{X(\mathrm{new})} = u_{tk}^X \frac{\sum_{m\in M}\left[\frac{x_{mt}}{\hat{x}_{mt}} t_{mk}\right]}{\sum_{m\in M} t_{mk}}, \quad (3)$$

$$u_{ek}^{A(\mathrm{new})} = u_{ek}^A \frac{\sum_{m\in M}\left[\frac{a_{me}}{\hat{a}_{me}} t_{mk}\right]}{\sum_{m\in M} t_{mk}}. \quad (4)$$

ここで，更新式の各項はすべて正であり，初期値を正とする限りにおいて非負性は保たれる．NMMF によって得られる行列 U^X は行列 X から抽出されたパターン数 K 個の人口変動パターンを表す．また T はそれぞれの人口変動パターンに対する各メッシュにおける重みと解釈することができる．

分解後の行列は，パターンごとに定数倍に関して自由度を持つ．そこで，変化点検知に利用する人口変動パターンを基準とした各行列の正規化を行う．ここで，人口変動パターンは平日と休日の時刻を持つことを考慮して合計が 2 となるよう正規化する．

3.2 都市変化点検知

3.1 節で獲得した人口変動パターン U_X を用いて変化点を検出する．変化点検知では，まず，各人口変動パターンに対する所属人口を各メッシュ m，各週 d において推定する．さらに，所属人口を特徴量とし，その週系列から変化点の検出を行う．

3.2.1 パターン所属人口の推定

変化点の候補である週 d における人口行列 X_d を考える．変化点検知における入力人口行列 X_d は複合パターン抽出時とは異なり，各パターンに所属する人数の増減を捉えるためにスケールを保持した行列とする．この行列に対して，$\|X_d - W_d U_X^T\|_2$ を最小化するような重み行列 W_d を求める．この行列の要素 $w_{m,k}$ は，週 d において，メッシュ m，パターン k に所属すると考えられる人口に相当する．本稿では NMMF と同様の交互最適化アルゴリズムによって推定する．

3.2.2 パターン所属人口を用いた変化点検知

週毎に計算される行列 W_d $(d = 1, 2, ...)$ の系列に着目することで，NMMF によって得られた各地域の地域性の長期的な変化をみる．変化点検知の手法は多数存在するが，ここでは Maeda ら [8] と同様にして各点の局所的な前後の分布を比較することにより検知を行う．

まず行列 W_d からメッシュ m においてパターン k に所属する人口の週系列 $\mathbf{w}_{m,k} = \{w_{m,d,k}|d = 1, \dots\}$ を取り出す．この系列から，ある週 d の前後 s 週を抜き出し，それを $P_{m,d,k} = \{w_{m,d-s,k}, \dots, w_{m,d-1,k}\}$，$Q_{m,d,k} = \{w_{m,d,k}, \dots, w_{m,d-1+s,k}\}$ とする．この $P_{m,d,k}$ と $Q_{m,d,k}$ との間の距離 $D_{\mathrm{CPD}}(P_{m,d,k}||Q_{m,d,k})$ を変化度 $S_{m,d,k}$ と定義する．なお，窓幅 s はハイパーパラメタであり，s を大きくするほど単発的な異常の影響を小さく見積もることができる．しかし，大きすぎる場合には検知が遅くなることに加えて感度が鈍くなる可能性がある．

変化度における距離関数 D_{CPD} はデータや目的に適したものに設定する必要がある．分布間の差を求めるにあたって Maeda らは，P, Q それぞれに正規分布を仮定した上で次式で表される Jensen-Shannon divergence を用いた [8]．

$$D_{JS}(P||Q) = \frac{1}{2}\log\frac{\sigma_p^2 + \sigma_q^2}{2\sigma_p\sigma_q} + \frac{(\mu_p - \mu_q)^2}{4(\sigma_p^2 + \sigma_q^2)} \quad (5)$$

ここで，$\mu_p, \mu_q, \sigma_p, \sigma_q$ はそれぞれ，集合 P, Q の平均，標準偏差である．

しかし，本稿で対象とする携帯電話人口統計データでは，以下の理由により，零点が連続して現れる系列を考慮する必要がある．

- 個人情報保護の観点から，一定値以下の人口は秘匿処理

2 : https://maps.gsi.go.jp/development/ichiran.html

3 : Shoreline data is derived from: United States. National Imagery and Mapping Agency. "Vector Map Level 0 (VMAP0)." Bethesda, MD: Denver, CO: The Agency; USGS Information Services, 1997.

されることが多い [15]

- 人口規模の小さい地域を扱う必要がある．
- 非負値行列因子分解はスパースなパターンや重みが得られやすいことが知られている．

このような点群に対し，上記のパラメトリックな距離関数を用いた変化度を定義した場合，人口が少ない地域に偶然 1 人が通りかかった場合に変化度が無限大に発散してしまうなどといった問題が生じる．そこで本稿では式 6 で示す平均値の差によって変化度を定義する．

$$S_{m,d,k} = D_{\mathrm{CPD}}(P_{m,d,k}||Q_{m,d,k}) = \frac{1}{s}\sum_{s'=1}^{s} w_{m,d-s',k} - \frac{1}{s}\sum_{s'=1}^{s} w_{m,d+s'-1,k} \quad (6)$$

3.2.3 検知性能向上のための追加処理

前述の方法で，零点が連続するような今回のデータに対しても変化点の検知が可能となる．しかし，都市空間における変化点の数は非変化点の数に比べて圧倒的に少数であるため，注目したい少数の変化点が多くの非変化点に埋もれてしまうという問題が存在する．そこで，注目対象の点における変化度を強調するための追加処理を 2 点提案する．

パターンごとの標準化．日中に人が集まるようなパターンに所属する人口は，夜間のそれに比べて人口規模が大きいと考えられる．したがって，前者のようなパターンにおける変動が，後者に比べて優先される可能性がある．このようなパターン間の不平等性を解消し，人口の差異が小さくとも特異なパターンに対する所属人口の変化を検知するために，パターンごとに値の標準化を行う (式 7)．

$$S^{*}_{m,d,k} = \frac{S_{m,d,k} - \mu_k}{\sigma_k} \quad (7)$$

$$\mu_k = \frac{1}{N}\sum_{(m,d)\in CP} S_{m,d,k}, \quad \sigma_k = \sqrt{\frac{1}{N}\sum_{m,d\in CP}(S_{m,d,k}-\mu_k)^2}$$

ここで，CP は全ての変化点候補 (m,d) の集合であり N はその総数である．

週近傍のピーク抽出処理．標準変化度 $S^{*}_{m,d,k}$ に関して，週近傍でのピークを取ることで変化のあったタイミングを強調する．考慮する週近傍を窓幅 sp としたとき，ピーク処理後の変化度 $S^{**}_{m,d,k}$ は以下の式で表される．ピーク処理後の変化度 $S^{*}_{m,d,k}$ をメッシュ m, 週 d, パターン k を以下の式により求める．

$$S^{**}_{m,d,k} = \begin{cases} S^{*}_{m,d,k} & (\text{if } S^{*}_{m,d,k} = \max\limits_{d-sp\leqq d'\leqq d+sp-1} S^{*}_{m,d',k}) \\ 0 & (\text{otherwise}) \end{cases} \quad (8)$$

上記で算出された変化度から，最終的に計算すべき変化に関する値は，メッシュ m，週 d に関するスカラー値である．そこで，各パターンに関する変化度を集約することで，点 (m,d) についての代表値を求める．これを変化スコア $\tilde{S}_{m,d}$ と定義する．代表値の計算方法は平均値やユークリッドノルム等，さまざま考えられる．ここでは，パターンに対する変化の大小を直接的に考慮するために最大値を用いることとする．したがって，点 (m,d) の分類に用いる最終的な変化スコアは以下の式となる．

$$\tilde{S}_{m,d} = \max_{k} S^{**}_{m,d,k} \quad (9)$$

4 実　　験

4.1 データセット

都市動態に関するデータセットとして，携帯電話から収集された GPS ログを基に，東京都，神奈川県全域における 250 メートル四方メッシュ内の 1 時間ごとの人口の推定を行った「混雑統計®」データ [4]を用いた．対象期間は 2014 年 12 月から 2019 年 11 月の 5 年間である．

複合土地役割抽出における施設情報は，拡張版全国デジタル道路地図データベース（2016 年度，2018 年度，2019 年度，2020 年度）から抽出した．本データベースには施設名，施設種別，緯度，経度が記述されている．施設種別は，粒度を揃えるためにデータベースに登録されているカテゴリを一部削除，統合した上で 41 種別を用いた．

また変化点検知・分析においては，実際に施設に変化があった正解事例としてそのメッシュと時期に関する情報が必要である．上記地図データベースには変化の時期に関する詳細な情報が含まれないため，ウェブ上 [5][6]，および上記の拡張版全国デジタル道路地図データベースを基に手動で，新規に開業した施設の名称，位置情報，及び開業日を収集した．結果として得られた，当該期間内に新規にオープンしたショッピングモール（62 件），宿泊施設（189 件）のデータを正解事例とした．

4.2 都市役割に関する複合パターン抽出

4.2.1 実験設定

「混雑統計®」データから，解釈性の高いパターンを獲得するために，以下の 2 つの条件を同時に満たす 5,460 メッシュを，複合パターンの抽出対象として選択した．

- 2014 年 12 月から 2015 年 11 月における 1 時間あたりの平均人口が 1,000 人より多い
- 1 つ以上の施設を有する

これらの各メッシュにおいて，2014 年 12 月から 2015 年 11 月の 1 年間における平日の平均時系列，及び土日祝日の平均時系列を連結することでターゲット行列 X を構築した．

4.2.2 抽出された複合パターンの評価

NMMF による複合パターン抽出はハイパーパラメタに依存して結果が大きく変わる．そこで，パラメタ選択を行うために得られたパターンの定量評価を行った．

モデルの汎化性能評価

NMMF によって得られたモデルの汎化性能を欠損値補完問題とすることで評価した [16]．具体的には，行列 X の要素の

4：「混雑統計®」データとは，NTT ドコモが提供するアプリケーション（※）の利用者より，許諾を得た上で送信される携帯電話の位置情報を，NTT ドコモが総体的かつ統計的に加工を行ったデータである．位置情報は最短 5 分毎に測位される GPS データ（緯度経度情報）であり，個人を特定する情報は含まれない．またデータの加工には「非特定化」「集計処理」「秘匿処理」がなされており個人が特定されることはない．※ドコモ地図ナビサービス（地図アプリ・ご当地ガイド）等の一部のアプリ．

5：http://www.jcsc.or.jp/sc_data/sc_open/2019opensc

6：https://www.traveltowns.jp/hotels/japan-new-hotels/

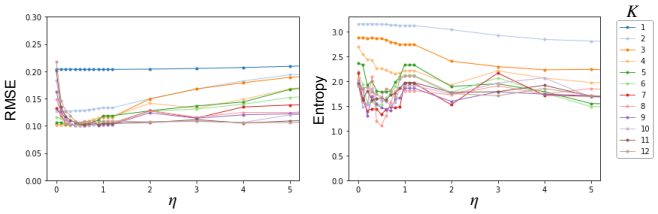

図 4: 施設行列に KDE による平滑化を行った際の複合パターン抽出評価．（左）人口行列 X の欠損値復元性能．（右）エントロピーによる施設パターンの解釈性評価．
「混雑統計®」© ZENRIN DataCom CO., LTD.

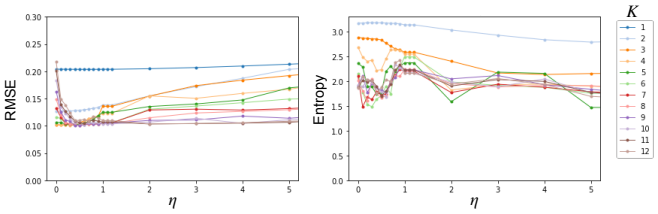

図 5: 施設行列に KDE による平滑化を行わない場合の複合パターン抽出評価．（左）人口行列 X の欠損値復元性能．（右）エントロピーによる施設パターンの解釈性評価．
「混雑統計®」© ZENRIN DataCom CO., LTD.

うち，1% をランダムに欠損させた上で分解を行い，欠損値の補完を行った．評価指標には RMSE を用いた．欠損位置は 5 通りに変化させ，評価値はそれらの平均を用いた．

複合パターンの解釈性評価

複合パターン抽出において，得られたパターンの解釈性を直接的かつ定量的に評価することは難しい．そこで，各パターンは施設パターンにおける 1 つ以上の少数の施設が大きな値をもつことが望ましいという仮定に基づき，得られたパターンの解釈性を評価した．具体的には，施設パターンを確率分布として扱い，エントロピーを計算することで解釈性の評価指標とした．ここで，エントロピーの値が小さいほど解釈性の高いパターンであると見なすことができる．ただし，施設パターンを確率分布として扱うにあたり，微小値での差分が大きく見積もられることを避けるため，施設パターンの全ての値が一定の微小値以下であるものに対しては，施設種別数を C としたとき，$1/C$ の値をもつ一様な分布として扱った．各パターンのエントロピーを平均したものを分解結果の評価値とした．

4.2.3 複合パターン抽出結果

異なるハイパーパラメータを用いた際の NMMF の欠損値推定性能を図 4 に示す．ここで，η は NMMF の分解において，補助行列 A をどの程度考慮するかを調節するパラメタであり，大きいほど施設情報の分解を重視することになる．なお，$\eta = 0$ は施設情報を用いず，人口行列を単独で NMF することに対応する．この結果から，以下の 2 点が確認された．

施設情報を用いることによる補完性能向上: パターン数 $K = 1$ を除く全てのパターン数において，人口行列のみを用いた分解 ($\eta = 0$) と比較して，NMMF を用いて施設情報を一定程度 ($\eta = 0.5$ 付近）考慮することにより，人口行列の欠損値補完性能が向上していることがわかる．このことから，人口変動を表現するパターンの汎化性能において，NMMF を用いた分解が，従来の人口変動単体を用いた分解を上回ることが確認された．

施設影響力の周辺メッシュへ染み出しを考慮したことによる分解パターンの解釈性能向上: 図 4 右のエントロピーに基づく施設パターンの解釈性能において，KDE（$bw = 10^{-5}$ rad）を用いることにより，$\eta = 0.5$ 付近において，図 5 右の KDE による処理を行わなかった場合に比べて同等以上に良いことが確認される．以降の変化点検知では，KDE による施設の平滑化を行った上でのパターンを用いる．ここで，具体的に抽出された複合パターンについて分析する．図 6 に $\eta = 0.5, K = 8$ としたときの複合パターンを示す．例として，パターン 6 が平日の昼や夕方，休日の昼に人が集まり，施設はレストランやレジャーに関わるようなパターンとして抽出されており，飲食・娯楽・ショッピングに関わるような都市の役割を表していると解釈できる．このような分析から，NMMF によって，定性的にも解釈可能な複合パターンが獲得できているといえる．

4.3 変化点検知

4.3.1 実験設定

変化点検知においては広範囲のメッシュの長期間の人口変動を用いる．期間は，2014 年 12 月から 2019 年 11 月までの 260 週を用いた．実用上，一定程度人が集まる場所での変化に興味があることが多いため，上記期間中の 1 週以上における 1 時間あたりの平均人口が 1,000 人より多く，かつ既知の大規模なイベント会場を除く 12,972 メッシュを対象とした．分解におけるパターン数は，汎化性能評価において最も良い値を示した $K = 8$ とし，変化度の算出における窓幅は，少数サンプルを用いた予備実験において良好な感度が確認された $s = 8, sp = 4$ に固定した．

4.3.2 変化点検知評価

新規にショッピングモールあるいはホテルが開業したメッシュとその週をどの程度捉えられているかを評価する．各メッシュと週の組 (m, d) に対して算出される変化スコアを用いた分類を行った際の分類性能を AUC (Area Under the ROC Curve) [17] により評価した．ここで，評価は近傍 9 メッシュ，および前後 1 週間のずれを許容した．具体的には，正解となる点 $(\tilde{m}, \tilde{d})$ の周辺の点 $\{(\tilde{m}-1, \tilde{d}-1), \ldots, (\tilde{m}+1, \tilde{d}+1)\}$ の変化スコアの最大値を $(\tilde{m}, \tilde{d})$ における変化スコアとした．また，正解予測の重複を防ぐために，$(\tilde{m}, \tilde{d})$ の周辺の点は評価から除外した．比較対象として以下のベースライン手法を用いた．

Maeda らの手法 [8]：（週×メッシュ）×（週時系列）を入力とした NMF により，人口変動パターンとそれに対する重み行列を獲得する．分解時のパターン数 K は同数比較のため 8 とした．さらに，重み行列を用い，式 5 により（メッシュ，週，パターン）の組に対する変化度を計算する．元論文では（メッシュ，週）に対する変化度の代表値の算出手法に関して記述されていない．そこで，実験では提案手法と同様の処理により代表値を計算した．

weekly summation：パターンに対する所属人口を特徴量として用いることの有効性を示すためのベースラインである．本手法は各メッシュ，週における特徴量として，各パターンに

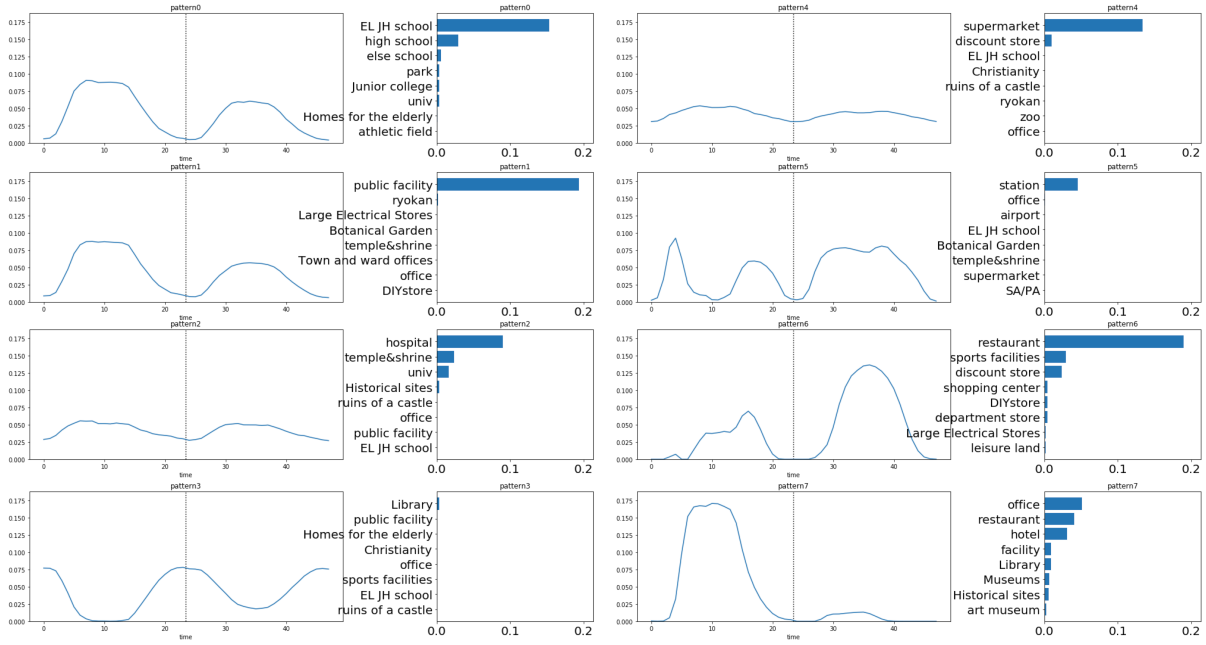

図 6: NMMF により抽出された複合パターン．「混雑統計®」© ZENRIN DataCom CO., LTD.

表 1: ショッピングモールおよびホテルの開業に対し，各変化点検知手法を適用したときの AUC による検知性能比較．

ラベル	スコア処理	**NMMF**	NMF	Maeda	sum
mall	なし	0.87714	0.90036	0.85434	0.90135
	標準化	0.90011	0.89660	0.85455	0.90135
	ピーク	0.94635	0.92812	0.87748	0.88599
	標準化＋ピーク	**0.95458**	**0.94355**	**0.88724**	**0.91295**
hotel	なし	0.89549	0.90429	0.81693	0.89896
	標準化	0.90391	0.90538	0.83106	0.89896
	ピーク	0.94935	0.93268	0.86365	0.80717
	標準化＋ピーク	**0.95200**	**0.95056**	**0.87896**	**0.83022**

「混雑統計®」© ZENRIN DataCom CO., LTD.

対する所属人口の代わりに，総人口を用いた．特徴量を基にした変化度の算出は提案手法と同様とした．

NMF：施設情報を用いた複合パターンを抽出することの変化点検知性能への影響を調べるために，人口行列のみから NMF を用いることでパターン抽出を行った結果として得られる人口変動パターンを用いた実験を行う．なお，NMF により抽出されるパターンは，NMMF における $\eta = 0$ とした場合の結果を用いた．また，抽出パターン数 K は，同数比較のため 8 とし，変化度の算出は提案手法と同様とした．

4.3.3 変化点検知結果

表 1 にショッピングモール，および宿泊施設の検知に関する実験結果を示す．この結果から，以下の 4 点が確認できる．1 点目は，NMMF と NMF の比較についてである．NMMF と NMF は，人口変動パターンが施設情報を加味しながら抽出されたものであるかにおいて違いがある．提案手法である NMMF は，4.2 節で示されたように，解釈性の高い施設情報に関するパターンを獲得しつつ，優れた検知性能を示すことが確認された．2 点目は，NMMF と週の合計人口を用いた手法との比較についてである．これらは異なる特徴量に対して同様の変化点検知手法を適用したものである．この比較から，特徴量として合計値よりも提案手法の人口変動パターンに対する所属人口を特徴量とした場合がよりよい性能を示すことが確認された．3 点目は，NMMF と Maeda らの手法の比較についてである．これらはいずれも人口変動パターンに対する所属人口を特徴量として用いているという点で共通しているが，変化度の計算手法が異なる．この結果から提案手法における変化点検知手法が，既存手法を上回る性能を示したことが確認された．4 点目は，変化点検知において行った変化度の標準化処理，週近傍でのピーク抽出処理の有効性についてである．ショッピングモールおよびホテルのいずれの開業の検知に関しても，標準化処理，および週近傍でのピーク処理がいずれも検知の性能向上に寄与しており，両者を組み合せることでさらに性能が向上することが確認された．

4.4 変化点に対する分析

変化点の例として，ららぽーと湘南平塚（2016 年 10 月 6 日開業）における人口変動パターンの変化と施設の関係を分析した．開業前後における生の時系列を図 7 に示す．また，各複合パターン（図 6）に対する所属する人口の系列を図 8 に示す．ここから算出される変化度ベクトルを基に，変化した施設種別を推定した結果を図 9 に示す．推定値 (pred) は，変化点 (m, d) における変化度ベクトル $(S_{m,d,1}, \cdots, S_{m,d,K})$ と施設パターンの積について，和が 1 となるよう正規化を行った値を用いた．参考値（map）は変化週の前後（2016 年度，2018 年度）における施設データを用い，該当メッシュに実際に存在する施設数の差分を同様に正規化した値である．両者を比較した結果，いずれも restaurant, および discount store が上位の施設として現れており，施設パターンが施設分布の変化の推定に利用可能であることが確認される．

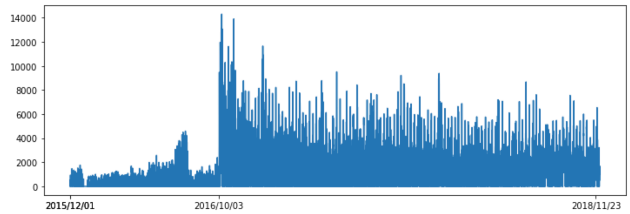

図 7: ららぽーと湘南平塚における生時系列．「混雑統計®」© ZENRIN DataCom CO., LTD.

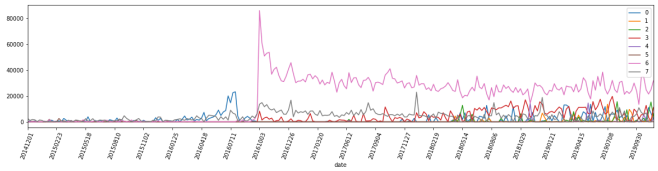

図 8: ららぽーと湘南平塚における人口変動パターンに対する所属人口の推移．「混雑統計®」© ZENRIN DataCom CO., LTD.

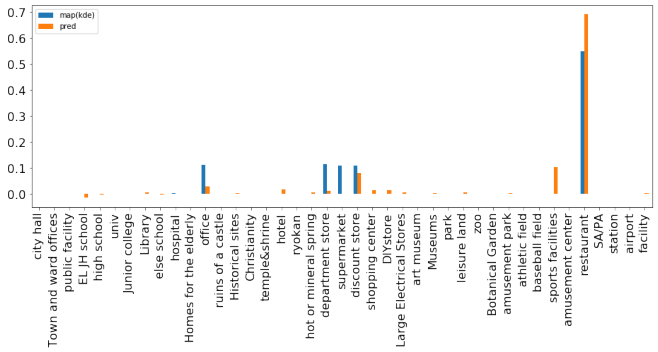

図 9: ららぽーと湘南平塚における施設パターンを用いた変化施設の推定．「混雑統計®」© ZENRIN DataCom CO., LTD.

5 おわりに

本論文では，携帯電話人口統計データおよび地図データから，施設の開閉業等にともなう人口動態の変化を検知する手法を提案した．提案手法はまず，NMMF を用いた同時因子分解により，人口と施設の複合パターンを抽出する．次に，得られたパターンに基づいて，人口変動に変化の生じたメッシュと週の組を検知する．大型ショッピングモールや宿泊施設の開業情報を用いた評価実験により，提案手法の変化検知性能は人口データのみを利用するベースライン手法を上回ることが示された．また，変化のあった人口変動パターンに対応する施設パターンを観察することで，人口変動パターンの変化と関連の深い施設の推定に活用できることを確認した．

本稿における提案手法では，人口変動の変化を検知に 8 週間程度の遅延を必要とする．高度な商業計画等に活用することを検討する場合には，さらに早期の検知が望まれる．そのための拡張として，提案手法の複合パターン抽出と，オンライン型の変化点検知手法を組み合わせることが考えられる．

また，本稿では検知する施設変化はデータ収集の観点からショッピングモールとホテルの開業に限定されていた．しかし実際の変化にはより多くのバリエーションが存在する．そこで，検知された変化点に対しては，原因となった施設の変化を自動で多クラスに分類することが望ましい．このような変化原因の追求の高度化に向けて，さらなるラベルデータの収集および，施設変化の推測・自動分類手法についても検討していきたい．

謝　辞

本研究の一部は，JST，CREST，JPMJCR19A4 の支援を受けたものです．

文　献

[1] Susan Hanson and Perry Hanson. Gender and urban activity patterns in uppsala, sweden. *Geographical Review*, pp. 291–299, 1980.

[2] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. Urban computing: concepts, methodologies, and applications. *ACM TIST*, Vol. 5, No. 3, pp. 1–55, 2014.

[3] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, Vol. 3, No. Jan, pp. 993–1022, 2003.

[4] Pentti Paatero and Unto Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, Vol. 5, No. 2, pp. 111–126, 1994.

[5] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, Vol. 401, No. 6755, p. 788, 1999.

[6] Jameson L Toole, Michael Ulm, Marta C González, and Dietmar Bauer. Inferring land use from mobile phone activity. In *UrbComp*, pp. 1–8, 2012.

[7] Jing Yuan, Yu Zheng, and Xing Xie. Discovering regions of different functions in a city using human mobility and pois. In *KDD*, pp. 186–194, 2012.

[8] Takashi Nicholas Maeda, Narushige Shiode, Chen Zhong, Junichiro Mori, and Tetsuo Sakimoto. Detecting and understanding urban changes through decomposing the numbers of visitors' arrivals using human mobility data. *Journal of Big Data*, Vol. 6, No. 1, p. 4, 2019.

[9] 磯川弘基, 豊田正史, 喜連川優. 携帯電話人口統計データを用いた新規施設に関わる都市動態の変化解析. In *DEIM*, 2020.

[10] Zipei Fan, Xuan Song, and Ryosuke Shibasaki. Cityspectrum: a non-negative tensor factorization approach. In *UbiComp*, pp. 213–223, 2014.

[11] Koh Takeuchi, Katsuhiko Ishiguro, Akisato Kimura, and Hiroshi Sawada. Non-negative multiple matrix factorization. In *Twenty-Third International Joint Conference on Artificial Intelligence*, 2013.

[12] Bin Liu, Hui Xiong, Spiros Papadimitriou, Yanjie Fu, and Zijun Yao. A general geographical probabilistic factor model for point of interest recommendation. *IEEE TKDE*, Vol. 27, No. 5, pp. 1167–1179, 2014.

[13] Jia-Dong Zhang and Chi-Yin Chow. Core: Exploiting the personalized influence of two-dimensional geographic coordinates for location recommendations. *Information Sciences*, Vol. 293, pp. 163–181, 2015.

[14] Ling Cai, Jun Xu, Ju Liu, and Tao Pei. Integrating spatial and temporal contexts into a factorization model for poi recommendation. *International Journal of Geographical Information Science*, Vol. 32, No. 3, pp. 524–546, 2018.

[15] Yasunori Akagi, Takuya Nishimura, Takeshi Kurashima, and Hiroyuki Toda. A fast and accurate method for estimating people flow from spatiotemporal population data. In *IJCAI*, pp. 3293–3300, 2018.

[16] Koh Takeuchi, Ryota Tomioka, Katsuhiko Ishiguro, Akisato Kimura, and Hiroshi Sawada. Non-negative multiple tensor factorization. In *ICDM*, pp. 1199–1204. IEEE, 2013.

[17] Andrew P Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, Vol. 30, No. 7, pp. 1145–1159, 1997.