DEIM 2020 J5-1(day2 p75)

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

磯川　弘基†　　豊田　正史††　　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
††† 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,kitsure}@tkl.iis.u-tokyo.ac.jp

あらまし　都市において，各地域がどのように使われているかを知ることで，地域毎の特性に応じた適切な都市計画や商業活動，災害復興を行うことが期待されている．そこで近年，携帯電話やカーナビをはじめとした GPS 端末の普及に伴い，大量かつ広範囲に人々の行動履歴が収集され，各地域における人々の土地利用特性をデータドリブンにモデリングする試みがなされている．一般に地域性はそこに存在する施設と密接に関わっており，施設の変化によってさまざま変化するが，施設変化と地域性の変化の関係を定量的に分析した研究は少ない．そこで本稿では施設の変化とそれに伴う地域性変化の関係を明らかにする試みとして，携帯電話の GPS ログデータに個人の非特定化処理を加えた形で提供される混雑統計データを用い，土地の潜在的な利用特性を同定した上でローカルな地域性の変化を検知する．また検知された個々の変化に対し，実際の発生した施設変化の種別と関連付けて分析を行うことで，各施設変化の種別毎に生じる地域性の変化の傾向を明らかにした．

キーワード　都市動態，変化解析，NMF

1　はじめに

都市において，各地域がどのように使われているかを知ることで，地域毎の特性に応じた適切な都市計画や商業活動，災害復興を行うことが期待され，従来から交通量・通行量調査等が行われてきた．近年では，携帯電話やカーナビをはじめとした GPS 端末の普及により大量かつ広範囲に人々の行動履歴が収集され，データドリブンに人々の都市活動をモデリングする試みがなされている．近年の研究により，「各地点において，単位時間ごとにどのくらいの人や車がいるか」といった表層の都市動態は，Latent Dirichlet Allocation (LDA) [1] や Non-negative Tensor/Matrix Factorization (NTF/NMF) [2] を用いることで潜在的な土地利用特性の和として表現可能であることが示されている．ここでは，「平日昼間に人が集まる」「深夜に人が集まる」などの典型的な時系列パタンが土地利用特性として抽出され，各地域の地域性はそれら複数の土地利用特性の重み付け和として表現される [3, 4]．例えば前者のパタンに対する重みの値が大きい地域はオフィス街，後者では住宅地に相当する，などといった分析が可能となる．これらの分析においては都市動態を静的なものとして扱っているが，実際の都市動態は災害やイベント，施設変化等さまざまな理由によって時間変化する．長期的な都市動態の時間変化に着目した研究として，Fan らは携帯電話 GPS ログを利用した混雑統計データに対し NTF を行うことで，東日本大震災を期に住宅地や商業地などといった土地利用特性の分布が広範囲で変化したことを確認した [5]．しかしこの手法の適用範囲は震災等の広範囲で同時多発的に発生する変化に限られており，個々の施設変化などといった局所的かつタイミングの異なる変化を捉えることは難しい．

局所的な地域性の変化を捉える試みとして，Maeda らは NMF により潜在的な土地利用特性を同定した上で，その重みを長期追跡することで変化を検知する手法を提案した [6]．彼らは交通系 IC カードの利用履歴データに対して手法を適用し，学期の変わり目における学校付近の駅や，付近に商業施設が開業したタイミングでの駅において，駅の利用特性に変化が発生したことを検知した．しかしここでは，データ特性の観点から検知可能な変化は駅周辺に限られ，広範囲での施設について変化を検知，分析することは難しい．またこれらの手法では，検知された個々の都市動態の変化に関して，その原因となった施設変化との関連について体系的な議論はなされておらず，人手で説明を付与するにとどまっている．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，それらの変化にどのような特徴があるのかを分析することは有用である．

そこで本研究ではまず，広範囲で取得された混雑統計データに対し Maeda らによる NMF に基づく変化点検知手法を一部修正した上で適用し，土地利用特性の同定，および変化点検知を行う．更に，各地で発生した土地利用特性の変化の原因を自動で分類することを目的とし，実際に新規大規模施設ができた場所，時間との関連付け，分析を行う．本稿ではまず，表層の都市動態として携帯電話の GPS ログデータに個人の非特定化処理を加えた形で提供される混雑統計データを用い，平休日を分割した上で NMF を行うことで，解釈可能な潜在時系列パタンの抽出が可能であることを確認した．また各地域において地域性を長期的に追跡し，統計検定を用いた変化点検知を行うことで何らかの変化があった場所，時間を抽出した．さらに，施設変化に関わる情報をウェブから収集し，都市動態の変化点と紐付けることで，施設変化がもたらす地域性の変化の様相を分

析した．結果として，ショッピングモールといった大規模施設ができることによる地域性の変化の傾向を定量的に確認した．

2 関 連 研 究

データドリブンな都市計画や出店計画の意思決定に向け，都市動態に関わる様々なデータから，都市における人々の動きや土地の使われ方をモデル化する試みがなされている．以下ではまず，GPS ログデータなどといった表層の都市動態から，潜在的な都市の役割や，人々の行動パタンの同定を行った研究を示し，次にそ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 潜在的な地域の役割の同定

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．教師あり学習による手法として，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，住宅地などといった土地の利用のされ方のクラス分類を行った [3]．また教師なしによる手法として 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，土地の利用のされ方に関する潜在パタンを LDA により抽出した上で，そのパタンを元に地域性のクラスタリングを行った [4]．ここで土地の利用のされ方は，いくつかの基本的なパタンの組み合わせであるという前提の下での解析がなされている．さらに，付加的な情報を組み合せた研究として，Sun らは地図検索の際に送信されるクエリのログを大規模に収集し，解析を行った [7]．検索クエリには，検索された行き先情報の他に，検索を行なった場所，および時間の情報が含まれる．行き先に関する情報から，地域ごとの人口を推定し，それぞれの地域において，いつ，どこから人が移動してくるのかに関する潜在パタンを抽出するといった，より詳細な分析を行なっている．

2.2 都市空間の変化分析

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 [5] は，2011 年東日本大震災前後の都市の利用特性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，一日の時刻 (time)，日付 (day)，及び 900 m× 900 mの空間メッシュを単位とした場所の系列 (area) からなる 3-mode tensor を構成し，NTF によって 9 つのパタン抽出している．ここで各パタンは time, day, area の 3 つの mode に関する特徴をもつ．例えば，一日の時系列を示す time 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 day 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ンが現れる．さらにその area 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．これらのパタンを定性的に分析することで，震災の前後で住宅地や商業地域に対応する場所が減少した地域，および仮設住宅等により人口が増えたと考えられる地域を特定した．しかし Fan らによる手法は，パタン自体がもつ日付 (day) 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ン自体に現れる必要がある．そのため適用範囲は広範囲で同時多発的に地域の使われ方が変化した場合に限定され，局所的に様々発生する地域性の変化を捉えることは難しい．

そこで Maeda ら [6] は，一日の時刻 × (場所，日付）の系列の 2-mode tensor，つまり行列の形で NMF を行うことで，時空間に対して局所的に生じる地域特性の変化を捉える手法を提案した．交通系 IC カードを用いて収集された駅改札の出入場記録データに対して提案手法を適用し潜在的な駅利用特性パタンを抽出した上で，平日午前の時間帯に出入場のピークが発生するパタンに対する重みの値が特定の駅において変化したことなどを検知している．この変化に関する分析として，付近に大学があること，さらに変化の時期が学期の始業や終業のタイミングに一致していることから，この駅利用特性の変化が学期の切り替わりによって引き起こされたものであると考察している．

しかしここでは，データ特性の観点から検知可能な変化は駅周辺に限られ，広範囲での施設に関する変化を検知，分析することは難しい．またこれらの手法では，検知された個々の都市動態の変化について，その原因となった施設変化との関連について体系的な議論はなされておらず，人手で説明を付与するにとどまっている．特に GPS ログデータを用いた広範囲の都市動態においては多数の多種多様な変化が検知されるため，それらの変化について人手で解釈を行うことは困難であると考えられる．そこで本研究ではまず，広範囲で取得された混雑統計データに対し Maeda らによる NMF に基づく変化点検知手法を一部修正した上で適用し，土地利用特性の同定，および変化点検知を行う．更に，各地で発生した土地利用特性の変化の原因を自動で分類することを目的とし，実際に新規大規模施設ができた場所，時間に関するデータを web から収集した上で施設変化と都市動態変化との関係を分析する．

3 局所的な地域性の変化解析手法

分析は，大きく 3 つの連続するステップからなる．まず第一に，表層の都市動態を表すログデータから NMF を用いて複数の典型的な時系列パタン，及びその空間的な分布を得る手法を説明する．次に，その地域性を長期的に追跡することで，各地で発生する変化を検知する．最後に検知された変化について，ウェブ上から収集した実際の施設変化情報と組み合わせることで定量的に分析する．

3.1 パタン抽出

本節では，Non-negative matrix Factorization(NMF) による地域特性の抽出手法について説明する．NMF は，何らかの行列 X が与えられたとき，これを要素すべてが非負な 2 つの行列の積として表現する手法である．Fig. 1 に今回適用する NMF の概略図を示す．ここで週 t における各エリア及び一週間内の各時刻における推定人口を要素とした行列 X_t を与える．なお一般に都市動態は平休日で大きく異なるため，後の分析は別個

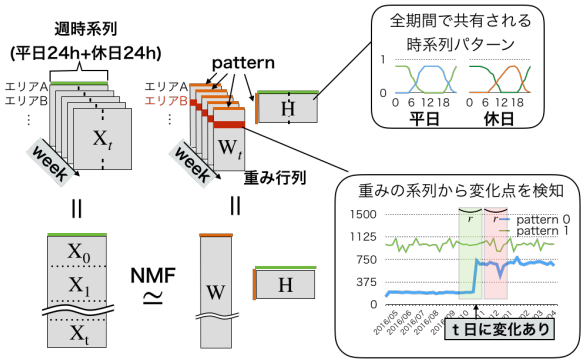

Fig. 1: 本手法で用いる NMF の概略図

でなされることが望ましい．そこで一週間の人口は平休日で平均値により集約し，平休日合わせて 48 時間の列とした．また，対象とするエリア数を a, 対象とする期間の週数を b とすると，行列 X の行数 n はその組み合せ数 $n = ab$ となる．このように定義した $m \times n$ 行列 X に対し，NMF を行うことで Fig. 1 に示す形で各要素が非負である 2 つの行列 W, H が得られる．ここで，W の列数，H の行数 K はハイパーパラメタとして予め設定される．

NMF は以下の損失関数を最小化することにより求められる．

$$D(X, WH) = \sum_{i=1}^{I} \sum_{j=1}^{J} d(X_{i,j}, \mathbf{t}_i^T \mathbf{v}_j)$$

損失関数には 2 乗誤差の他，一般化 KL-divergence 等が用いられ，本稿では後者を用いる．なお最適化は NP 困難 [8] であることが知られており，大域解を求めることは難しい．最適化のアルゴリズムはさまざま提案されており，乗法的更新アルゴリズムによる手法が広く用いられる [9,10].

NMF によって得られる行列 H はデータ X から抽出されたパタン数 K 個の典型的な時系列パタンを表す．また W は各 (area, week) の組に対して与えられる，それぞれの時系列パタンに対する重みと解釈することができる．

3.2 都市変化検知

行列 W に着目することで，NMF 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．Fig. 9 に，特定の area における各パタンへの重みを週系列にプロットした例を示す．ここから週系列に変化点，つまり特定の週を境に統計的な性質が有意に変わるような週があったか，ある場合それがどの週なのか検出する，変化点検知の問題を考える．異常の集合としての変化点を検知することは容易ではなく，これまでにさまざまな手法が提案されている [11–13]．ここでは Maeda ら [6] と同様，各点の局所的な前後の分布を比較することで変化点を検知する手法を用いる．

まず NMF の結果得られる行列 W から特定の area において，特定のパタン k に関する重みの週系列 $\mathbf{w}_k = \{w_1, \ldots, w_j, \ldots, w_n\}$ を取り出す．さらにこの系列からある週 i の前後 r 週を抜き出し，$P_i = \{w_{i-r}, \ldots, w_{i-1}\}$, $Q_i = \{w_i, \ldots, w_{i-1+r}\}$ を得る．この P_i と Q_i との間の距離 $D(P||Q)$ を週 i における変化度 S_i と定義する．変化度を各週について計算することで，特定のエリア，パタンに対する変化度の週系列が得られる．なお，window size r はハイパーパラメタであり，r 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．

さらにこの変化度に関して，式 (1) により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求めることができる．ここで，S_i^* を週 i における change score と定義する．

$$S_i^* = \begin{cases} S_i & (\text{if } S_i = \max\limits_{i-r \leqq j \leqq i+r} S_j) \\ 0 & (\text{otherwise}) \end{cases} \tag{1}$$

変化度における距離関数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は，P, Q それぞれに正規分布を仮定した上で Kullback-Leibrer divergence（式 (2)) や Jensen-Shannon divergence(式 (3)) が用いられることが多い [6]．しかし，本稿で対象とする混雑統計データでは，メッシュ数が大きいために人の出入りが少なく分散が 0 となるメッシュが大量に現れる．これに対し上のパラメトリックな距離関数を用いると普段人がいない地域に偶然一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という問題がある．

$$D_{KL}(P||Q) = \log \frac{\sigma_q}{\sigma_p} + \frac{\sigma_p^2 + (\mu_p - \mu_q)^2}{2\sigma_q^2} - \frac{1}{2} \tag{2}$$

$$D_{JS}(P||Q) = \frac{1}{2} \log \frac{\sigma_p^2 + \sigma_q^2}{2\sigma_p \sigma_q} + \frac{(\mu_p - \mu_q)^2}{4(\sigma_p^2 + \sigma_q^2)} \tag{3}$$

そこで本稿ではまず，二群の差の検定によりある週の前後で統計的に有意な増加が見られる時空間メッシュを選択する．検定には，人口の増加に伴い分散が増加することを考慮し，二群に等分散性を仮定しない Welch の t 検定，およびノンパラメトリックな手法として，母集団の分布に正規分布を仮定せず，群内における値の順位のみを用いる Brunner-Munzel 検定 [14] の二つの手法を検討した．また，本稿では新規施設の出現によりいずれかのパタンで人口に増加が現れることを検知するために各パタン毎に片側検定を行い，いずれかのパタンにおいて棄却された時空間メッシュを選択する．

このようにして得られたメッシュの前後における各パタンの重みの平均値の差によって変化度を式 (4) により定義し，式 (1) により change score を計算する．なおここで，各場所において計算される週 i の変化度，change score はともにパタン数 K 次元のベクトルである．

$$\begin{aligned} D(P||Q) &= \mu_q - \mu_p \\ S_i &= D(P_i||Q_i) \end{aligned} \tag{4}$$

次に，各時空間メッシュに関して得られる 10 次元の change score を用いてランキングを行う．ランキングにはそれぞれのパタンごとに標準化を行なった上での最大の要素を代表値として用いた (式 (5))．

$$S_{\max}^* = \max \left(\frac{S_r^* - \mu_r}{\sigma_r} \right) \tag{5}$$

なお，μ_r, σ_r はそれぞれ，パタン r の change score の平均，標準偏差である．

3.3 変化の識別

変化検知においては様々な特徴をもつ変化が同時に検出される．それがどのような変化であるのかを NMF により抽出されたパタンを元に分析，分類することを目標とする．本項では変化点における変化の原因を分類する試みとしてまず，K 次元の change score から分散分析により変化の種類ごとに現れるパタン変化の特徴を分析する．なおここでは，ウェブ上から特定の新規施設に関する情報を収集し，変化点に対するラベルデータとする．

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 ログを元に，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 [1]を用いた．用いた期間は 2014 年 12 月から 2018 年 11 月の 4 年間である．

また変化検知，分析において用いる，実際に POI に変化があった事例として，当該期間内に新規にオープンしたショッピングモール及び宿泊施設に関するデータをウェブ上 [2][3]から収集した．収集したデータには施設名，及びオープンの日付が含まれており，施設名をもとに位置情報を取得した．

4.2 NMF によるパタン抽出

「混雑統計®」データから，週毎に平日の平均時系列，及び土日祝日の平均時系列を計算し，それらを連結した平日 24 時間＋休日 24 時間の系列を一週間の時系列とする．また，推定人口の期間を通した平均値が一定以下の空間メッシュを除き，22979 メッシュを対象とした．つまり，全期間 208 週を考えると，NMF を行う行列は (22979×208) ×48 の行列である．なお本稿では以下，22979 地点からなる各 area を「250 m メッシュ」，あるいは「空間メッシュ」と呼び，area と週から定義される要素を「時空間メッシュ」と呼ぶ．

パタン数 K=10 として NMF をした結果得られる時系列パタンに関する行列 H を Fig. 2 に示す．ここで，それぞれの時系列パタン k は行列 H の k 行目に対応する．この時系列を見ることで，解釈が可能な時系列パタンが抽出できていることが定性的に確認できる．例として，パタン 5 は平日の昼間に人口が増えるため，オフィス街を表すようなパタンであるといった

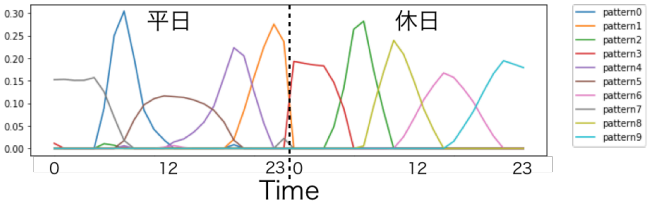

Fig. 2: NMF により抽出された 10 個の時系列パタン
「混雑統計®」© ZENRIN DataCom CO., LTD.

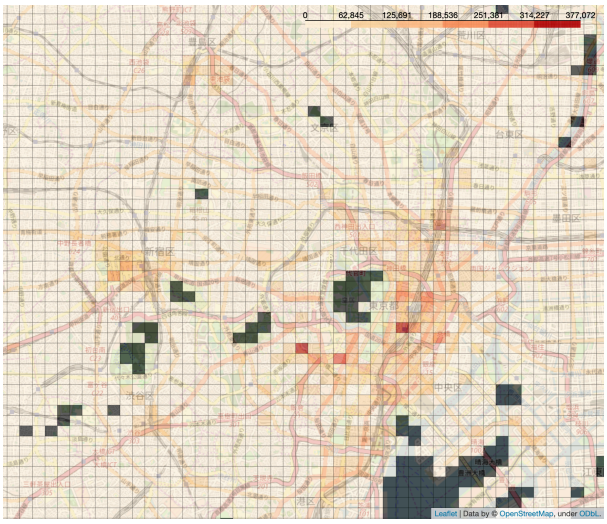

Fig. 3: pattern 5 の地理分布（2014 年 12 月初週）
「混雑統計®」© ZENRIN DataCom CO., LTD.

解釈が可能である．また，特定の週における各 area での重みの値を可視化することでパタンごとの空間的な分布を得ることができる．Fig. 3 に，2014 年 12 月初週における pattern 5 の分布を地図上にヒートマップで表す．なお，地図タイルには OpenStreetMap を用いた．ここから，このパタンが東京駅や新宿といったオフィス街に分布していることが確認できる．次節では，このようなパタンの重みに関して長期的に追跡した際に，何らかの持続的な変化が起きた場所，週を特定することを考える．

4.3 変化点検知

特定のメッシュに対して 10 パタンそれぞれ Brunner-Munzel 検定，または Welch の t 検定を行い，いずれかのパタンにおいて $p < 0.01$ に含まれる週を統計的に有意な変化があったタイミングとして抽出する．なおいずれも片側検定とし，変化点検知においてすべての window size r は 8 とした．抽出された時空間メッシュに対して，change score を計算し，ランキングを行なった．なお，東京ドームや東京競馬場などといった既知の大規模なイベント会場を含む空間メッシュについては，ランキングを行う対象から省いている．

4.3.1 変化検知の結果

ウェブから収集したショッピングセンター 46 件，宿泊施設 97 件のオープンした場所，及びタイミングが捉えられているかを評価した．ここで，新規施設のランキングは前後一週間のずれを許容した値を用いた．Fig. 4, Fig. 5 に，change score に

1：「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．

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

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

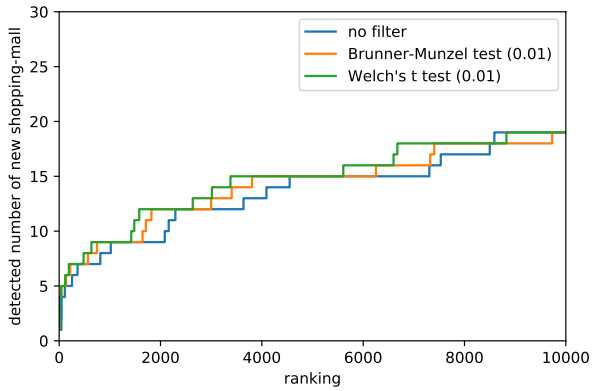

Fig. 4: ショッピングモールのランキング「混雑統計®」© ZENRIN DataCom CO., LTD.

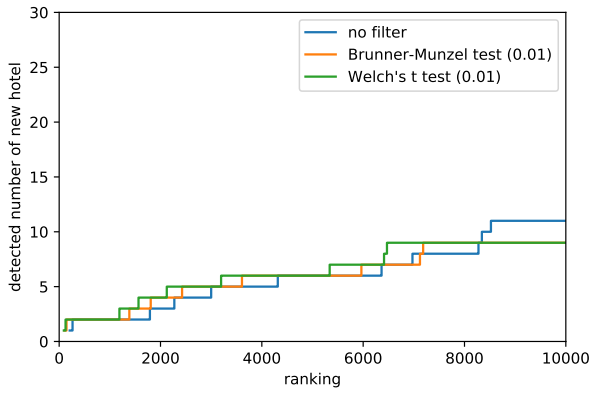

Fig. 5: 宿泊施設のランキング「混雑統計®」© ZENRIN DataCom CO., LTD.

よるランキング上位 10000 までの時空間メッシュを取ったときのショッピングモール，及び宿泊施設の検知数の推移をそれぞれ示す．ここで，検定による時空間メッシュの選択をせず，すべての点に対しランキングを行なった場合 (no filter)，および Brunner-Munzel 検定, Welch の t 検定それぞれによって統計的有意な増加があったと考えられる点のみを選択した上でランキングした場合とを比較した．この結果から，ショッピングモール，宿泊施設の双方において二群の検定によりランキングの対象となる時空間メッシュ数を減らすことで，既知の新規出店があった時空間メッシュが上位にランキングされるようになっており，検定にって時空間メッシュを絞ることの有効性が確認された．なお以後の事例分析においては，Welch の t 検定により時空間メッシュを選択した上での change score を用いたランキング値を利用した．

4.3.2 検知成功事例

検知が成功している例として，2015 年 10 月 29 日にグランドオープンした「三井ショッピングパークららぽーと海老名」における，NMF によって得られたパタンごとの重みの週系列を Fig. 9 に示し，そこから得られる change score の系列を Fig. 10 に示す．ここで青の点線はオープンした週である．この時空間メッシュにおけるランキングはすべての時空間メッシュの中で 10 位であった．この結果から，当施設オープンに伴うパタン変化が捉えられていることを定性的に確認できる．

4.3.3 評価対象とはしていないが上位にランキングされている事例

ランキングの上位 2 つを Fig. 11, Fig. 12 に示す．一位の東京駅，および二位の溜池山王における変化については，変化の理由が判明していない．しかし，東京駅においては，平日昼に人が増える pattern 5, 平日夕方の pattern 4，休日夕方の pattern 6 など多くのパタンに増加が見られれていることから，駅内部の動線に変化があったなどといった推測が立てられる．また，溜池山王においては，平日の昼，夕方のパタンに増加が見られることから，オフィスの転入があったといったことが変化の原因として推測される．

4.3.4 検知失敗事例の分析

今回収集したショッピングモール，宿泊施設に関して，変化点の検知に失敗した例を示す．Fig. 13 は 2015 年 11 月 19 日にオープンした渋谷モディにおける pattern ごとの重みの週系列である．この時空間メッシュは，施設に変化があったにも関わらずパタンの変化が見られず，前後の分布に有意な増加がないとする検定で棄却されなかったためにランキングの対象とはならなかった．渋谷モディのある地域は商業施設が密集しており，当該施設が都市動態に大きな影響を与えていなかったと考えられる．

4.4 変化の識別

Welch の t 検定によるメッシュの選択を行なった上で，ランキング上位 5000 に含まれた時空間メッシュを用いて，新規ショッピングモール，および宿泊施設のオープンに伴ってどのパタンに変化が見られるかを分析する．ここでまず，上位の時空間メッシュにおける 10 次元の change score を t-sne [15] を用いて二次元に次元削減した結果を Fig. 6 に示す．新規ショッピングモールがオープンした時空間メッシュに対応する点は互いに近くにプロットされており，同一のクラスタを形成していることが確認できる．これに対し，宿泊施設 (hotel) については検知点数が少ない上に，それぞれが離れて分布しており，同一のクラスタを形成しているとは言えないことが分かる．本稿ではさらに，change score の内，どの要素が分類に寄与するのかを示す指標として，各パタンについて分散分析を行なった．ここで「ショッピングモール」と「ショッピングモール以外」の二群について，パタンごとに分散分析を行なったときの F 値を Fig. 7 に示す．なおここで，自由度 (1, 4998) の F 分布より，1 % 棄却域は 6.64 である．この結果から，新規ショッピングモールがオープンした時空間メッシュでは，他の時空間メッシュと比較して pattern 6，4，8 の値に有意差があるといえる．ここから，ショッピングモールができた時空間メッシュでは，休日夕方に人口が集まるという定性的な判断とも一致する．ここで，10 次元の change score のうち，pattern 6 の値のみでランキングを行なった際には上位 10 位中 6 点が新規ショッピングモールのオープンに対応する時空間メッシュであり，有効な特徴量であることが確認された．

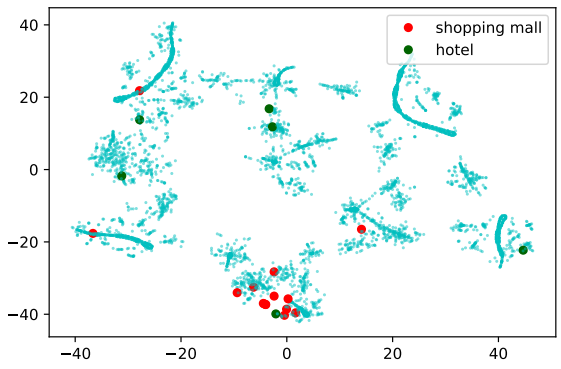

Fig. 6: 上位 5000 の変化点における change score を t-sne [15] により可視化した結果「混雑統計®」© ZENRIN DataCom CO., LTD.

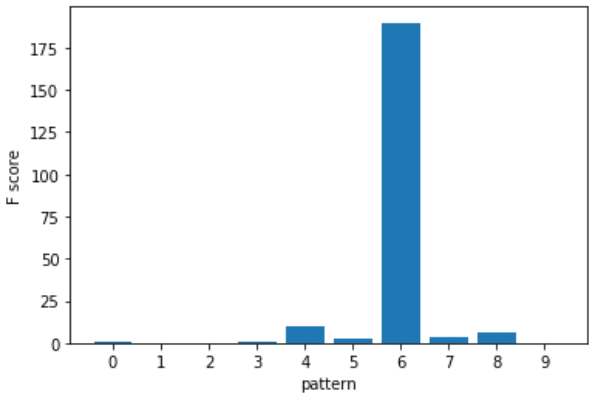

Fig. 7: 「ショッピングモール」と「ショッピングモール以外」の二群に対してパタンごとに分散分析を行なった結果「混雑統計®」© ZENRIN DataCom CO., LTD.

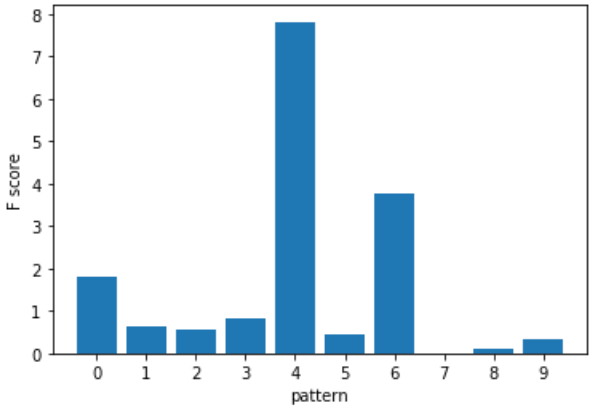

Fig. 8:「宿泊施設」と「宿泊施設以外」の二群に対してパタンごとに分散分析を行なった結果「混雑統計®」© ZENRIN DataCom CO., LTD.

同様に，「宿泊施設」と「宿泊施設以外」の二群について分散分析を行なった結果を Fig. 8 に示す．この結果から，新規宿泊施設のオープンに伴って，平日午後に人口が増加することに対応する，pattern 4 に有意な変化が現れることが分かる．pattern 4 についてのランキング上位を見ると，「ホテル グレイスリー新宿」の開業が 19 位に現れている．

5 おわりに

本論文では，携帯電話人口統計データによる表層の人口動態から，NMF により複数の潜在パタンを抽出し，地域性を同定した．さらに，パタンによって定義される地域性に対し，その長期的な変化を追跡，検知した．次に，実際に人口動態に変化を与える可能性のある大規模新規施設に関するデータを取得し，検知ができているかを定量評価した．最後に，変化に対する説明を加える手法として，特定の大規模新規施設ができた時空間メッシュとそれ以外の変化点を二群とした統計分析を行った．結果として，ショッピングモールといった大規模施設ができることによる地域性の変化の様相を定量的に確認した．

しかし現状として，検知された変化点のうち，web 上からその原因が特定できている点は少ない．そこで今後は，それらの変化理由の特定が難しい変化点に対する情報を複数の情報源から収集することを検討する．またさらに，検知された変化点からその原因を自動で分類する手法についても検討していきたい．

謝　　辞

本研究を進めるにあたり共同研究を通じて支援，コメントを頂いた国立研究開発法人情報通信研究機構の是津様，梅本様に感謝いたします．また，本研究の一部は，JST，CREST，JPMJCR19A4 の支援を受けたものです．

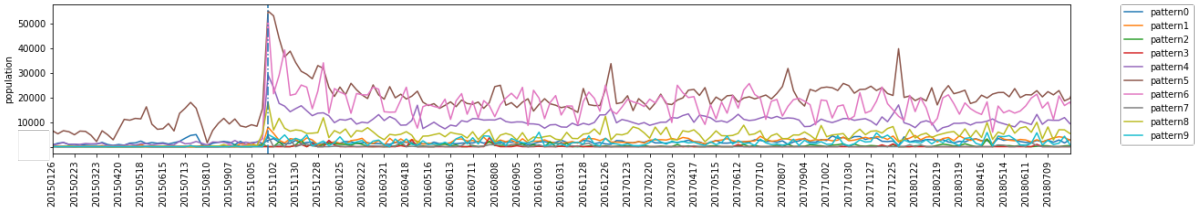

Fig. 9: 「三井ショッピングパークららぽーと海老名」における NMF により抽出されたパタンの週系列
「混雑統計®」© ZENRIN DataCom CO., LTD.

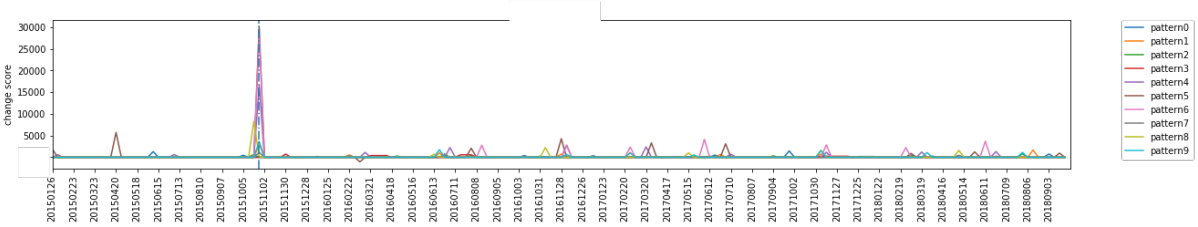

Fig. 10: 「三井ショッピングパークららぽーと海老名」における NMF により抽出されたパタンに対する change score の週系列
「混雑統計®」© ZENRIN DataCom CO., LTD.

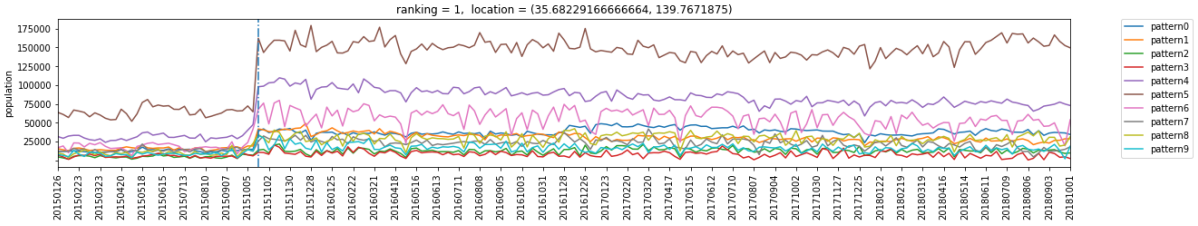

Fig. 11: 一位　東京駅「混雑統計®」© ZENRIN DataCom CO., LTD.

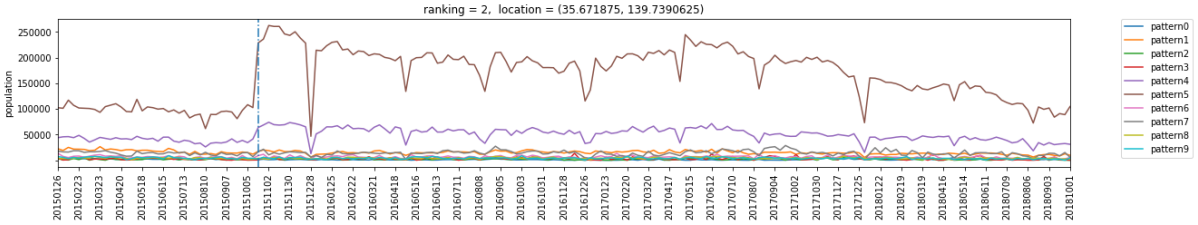

Fig. 12: 二位　溜池山王「混雑統計®」© ZENRIN DataCom CO., LTD.

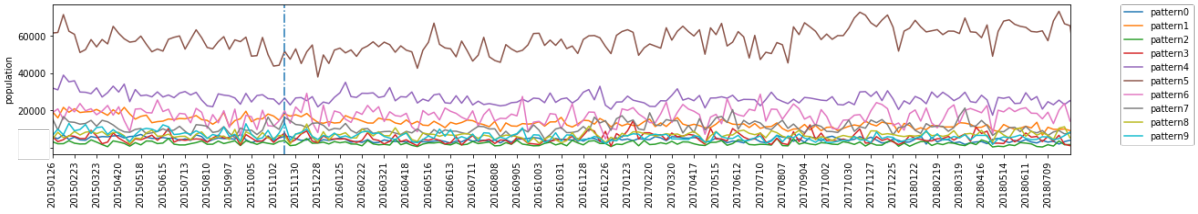

Fig. 13: 新規ショッピングモールができたが変化点の検知ができなかった例：渋谷モディ「混雑統計®」© ZENRIN DataCom CO., LTD.

文　　献

[1] 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.

[2] 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.

[3] Jameson L Toole, Michael Ulm, Marta C González, and Dietmar Bauer. Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD international workshop on urban computing*, pp. 1–8. ACM, 2012.

[4] Jing Yuan, Yu Zheng, and Xing Xie. Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 186–194. ACM, 2012.

[5] Zipei Fan, Xuan Song, and Ryosuke Shibasaki. Cityspectrum: a non-negative tensor factorization approach. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 213–223. ACM, 2014.

[6] 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.

[7] Ying Sun, Hengshu Zhu, Fuzhen Zhuang, Jingjing Gu, and Qing He. Exploring the urban region-of-interest through the analysis of online map search queries. pp. 2269–2278, 07 2018.

[8] Stephen A Vavasis. On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization*, Vol. 20, No. 3, pp. 1364–1377, 2009.

[9] 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.

[10] Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pp. 556–562, 2001.

[11] Jun-ichi Takeuchi and Kenji Yamanishi. A unifying framework for detecting outliers and change points from time series. *IEEE transactions on Knowledge and Data Engineering*, Vol. 18, No. 4, pp. 482–492, 2006.

[12] Yoshinobu Kawahara and Masashi Sugiyama. Sequential change-point detection based on direct density-ratio estimation. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, Vol. 5, No. 2, pp. 114–127, 2012.

[13] David S Matteson and Nicholas A James. A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, Vol. 109, No. 505, pp. 334–345, 2014.

[14] Edgar Brunner and Ullrich Munzel. The nonparametric behrens-fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, Vol. 42, No. 1, pp. 17–25, 2000.

[15] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, Vol. 9, No. Nov, pp. 2579–2605, 2008.