

Kanazawa University,
Faculty of Economics and Management

Discussion Paper Series

No. 054

羽咋市におけるモバイル空間統計データ
による人口流動分析
～非負値行列因子分解法を用いて～

金沢大学大学院 人間社会環境研究科 原田 魁成
金沢大学 経済学経営学系 寒河江 雅彦

Sagae.masahiko@gmail.com

December 2020



金沢大学経済学経営学系

〒920-1192 金沢市角間町

Faculty of Economics and Management,
Kanazawa University

Kakumamachi, Kanazawa-shi, Ishikawa, 920-1192, Japan

<http://econ.w3.knazawa-u.ac.jp/DP/>

羽咋市におけるモバイル空間統計データによる人口流動分析

～非負値行列因子分解法を用いて～

金沢大学大学院 人間社会環境研究科 原田 魁成

金沢大学 経済学経営学系 寒河江 雅彦

概要

本研究は、国勢調査やパーソントリップ調査等で把握してきた人口流動に対し、携帯電話の位置情報データを用いて人の流動を季節・曜日・属性別に多面的にとらえることで、これまで把握できなかった人々の動的な移動の実態を分析することを目的とする。ビックデータである携帯電話位置情報データは広範囲かつ高頻度、高精度な構造を有していることが利点であるが、膨大なデータ故にそれを解析するための有効な分析手段が必要となる。本稿では次元縮約機能を持つ非負値行列因子分解を用いて人々の移動行動パターンを分析する。具体的には昼間人口及び夜間人口数を記録したモバイル空間統計データに、加法的な2つの構成成分に分解する非負値行列因子分解を適用することで、昼夜間の人口流動の特徴抽出を試みている。また、スパース(罰則項を付与した)非負値行列因子分解を同データに適用することで、通常非負値行列因子分解を適用する場合よりも、抽出される特徴を鮮明にできる。さらに元データに対する行列分解の残差解析をすることで、モデルで表現できなかった元データの構造を網羅的に解析する。モバイル空間統計データは石川県羽咋市居住者の、羽咋市における年代・男女別(15~19歳、20~29歳、30~39歳、40~49歳、50~59歳、60~69歳、70歳~79歳の7年代区分×2属性)1461日分(2014/3/1~2018/2/28)の昼夜間人口データを使用した。

非負値行列因子分解の適用により、基底数を3とした場合、概ね「15~19歳の男性」、「15~19歳の女性」、「20~59歳男女」の基底と解釈される特徴が抽出され、いずれも平日の流出構造は「通勤・通学」に伴うものであった。また、スパース非負値行列因子分解では「15~19歳の男性」、「15~19歳の女性」、「羽咋市民」の基底が見かり、さらに明瞭な解釈が可能となった。残差項からは日曜日に関する特徴が抽出され、特に20~39歳男女における流出構造が大きく抽出された。羽咋市民の特徴を概観すると、平日は通勤・通学に伴う市外への流出パターン、休日は20~39歳の市外への流出パターン、盆休みや年末年始などは、昼夜間人口の差が極めて小さく、同期間中は市民の多くが羽咋市内に滞在していることが示された。

キーワード 非負値行列因子分解 モバイル空間統計 スパース

Hakui population flow analysis by Mobile Spatial Data

~ Using a Non-negative Matrix Factorization ~

HARADA Kaisei

SAGAE Masahiko

We have attempted to extract useful features from Mobile spatial data. Specifically, we apply a Non-negative Matrix Factorization (NMF), which is a method of decomposing the original data into two additive matrices, to extract features of daytime and nighttime demography. By applying Sparse Non-negative Matrix Factorization to the same data, we attempt to extract features more clearly than NMF. Furthermore, we analyze the structure of the original data comprehensively by analyzing the approximate residuals of each matrix decomposition model for the original data. We used the data on the day and night population of Hakui residents in Ishikawa Prefecture for 1461 days (March 1, 2014 to February 28, 2018) by age and sex (15-19, 20-29, 30-39, 40-49, 50-59, 60-69 and 70-79 years old \times man and woman).

As a result of the analysis, when we set basis number at "3", we got the characteristics that were interpreted as the "teenage male", "teenage female" and "worker". In addition, using Sparse NMF, many of those other attributes were "0" in the "teenage male" and "teenage female", thus allowing for a clear interpretation. To summarize the characteristics of Hakui citizens, we found that there is an outflow of people who commute to work and school on weekdays, and outflow for those aged between 20 and 39 on holidays. In addition, the day and night population during the Obon holidays and New Year's holidays is very small, so many of citizens in Hakui are in their city during the same period.

Key word Non-negative Matrix Factorization Mobile Spatial Data Sparseness

1.はじめに

本研究では地域をまたいだ人口流動の実態を、携帯電話の位置情報データであるモバイル空間統計データに対して非負値行列因子分解及びスプース非負値行列因子分解を適用することで、解釈可能な行動パターンの特徴を抽出することを目的とする。

地域間移動を観測する既存の調査として、就業や通学等の近距離・中距離の移動であれば「国勢調査」、観光・出張等の都道府県をまたぐような長距離の移動であれば「パーソントリップ(PT)調査」や、「全国幹線旅客純流動調査」等がある。他にも類似した調査として「観光動態調査」もある。これらの政府統計は、基幹統計である「国勢調査」であれば全数、一般統計である「PT調査」¹⁾は都市圏単位での無作為標本調査、同じく一般統計である「全国幹線旅客純流動調査」は航空・鉄道・幹線旅客船・幹線バス・乗用車等の利用者に対する「旅客動態調査」を集計し、母集団推計した調査であり、いずれの調査も非常に大規模な調査である。そのため、調査頻度が5年に1度や10年に1度²⁾であり、調査日も国勢調査やPT調査は1日、全国幹線旅客純流動調査も平日・休日1日ずつなど、非常に限られた条件での調査となっている。そのため、調査日によるサンプルの偏り等から生じるバイアスの問題が指摘されている¹⁾。これらに対し、携帯電話の位置情報から得られるビックデータは全国から大量のサンプルデータが常時取得できることから、広範囲で高頻度、信頼性の高いデータで、人口移動の実態を把握することができる。総じて政府統計資料では市民の日常的な買い物や、余暇活動等の移動行動の把握は困難であり、我々の問題意識とする、平日・休日の活動、属性別の活動、時間帯による移動の特徴などが十分に捉えられなかった。モバイル空間統計はそれらを補うものである。携帯電話の位置情報を利用したビックデータや既存の人口流動調査に関して以下の表¹⁾でまとめる。

携帯電話の位置情報を利用したビックデータから有益な情報を抽出する手法は近年複数提案されている^{3,4)}が、特に山口・奥村(2017)⁵⁾や山口・中山(2020)⁶⁾では時系列情報を持つ集計データに対し、非負値行列(テンソル)因子分解を適用することで「旅行グループ」の特徴を抽出することに成功している。また原田・山口・寒河江(2020)⁷⁾ではモバイル空間統計を用いたCOVID-19流行前後における人口流動の変化を、非負値行列因子分解を用いて分析している。これらの研究から時系列データに対して特徴抽出を行う場合、非負値行列因子分解が有効な分析手法の1つであることが示されている。その他の類似した研究として、磯川・豊田・喜連川(2020)⁸⁾では混雑統計データに非負値行列因子分解を適用し、ショッピングモールの開業に伴う地域性の変化を明らかにしている。海外では、携帯電話の位置情報データから交通機関の利用パターンを分析した研究⁹⁾や旅行に伴う都市移動非負値テンソル因子分解を用いて3次元的に分解する研究¹⁰⁾等が行われている。本研究とは研究目的が異なるが非負値行列因子分解とその応用を用いて空間情報の特徴抽出に成功している。

他方、本研究で使用する非負値行列因子分解に関する研究について、提案者であるLee and Seung(1999)¹¹⁾は顔画像データを、目や鼻などの顔のパーツを表す特徴成分(基底行列)と正の重み成分(表現行列)の2つの行列に分解することに成功した。また近年では非負値行列因子分解に罰則項を付与することで分解される2つの行列をより解釈し

やすくするための研究がなされている。Hoyer(2002)¹²⁾は分解する2つの行列のうち、片側に不要なノイズ項を0にできる L_1 ノルムの罰則項を付与したモデルを提案している。Kim and Park(2007)¹³⁾は、分解する行列の両方に要素を縮小化できる L_2 ノルムを付与し、これらを交互に更新を行うアルゴリズムを提案している。また Cichocki, Zdunek and Amari(2007)¹⁴⁾は罰則項の付与の仕方を総括し、体系化している。その後も罰則項の付与の仕方については、上述のHoyerの提案した L_1 ノルム型とKim and parkらの提案した L_2 ノルム型を組み合わせた罰則制約(Elastic net型)¹⁵⁾や、 L_1 ノルムより制約の強い $L_{\frac{1}{2}}$ ノルムを、分解する行列の片側及び両側に罰則制約として付与する型¹⁶⁾など、様々なパターンが研究されている。本稿では、膨大な時系列情報から簡潔な特徴を得ること、またその特徴を構成する重み要素も簡潔なものにし、抽出された特徴を鮮明に解釈することを目的に、分解される基底行列と表現行列の両方の特徴から不要なノイズ項を取り除くことができる L_1 ノルムの罰則項を付与するスパース非負値行列因子分解を構築し、携帯電話の位置情報データに適用した。

以上から、本研究では非負値行列因子分解及び原田・山口・寒河江(2020)の分析アプローチの発展としてスパース非負値行列因子分解の適用により、人々の特徴的な行動パターンの抽出を試みる。携帯電話の位置情報データに非負値行列因子分解を適用する例は国内及び海外において散見されるが、スパース(罰則項を付与した)モデル、特に L_1 ノルムを付与する研究例は本研究が初となる。

2. 使用したモバイル空間統計データ

本研究の使用データとして石川県羽咋市³⁾居住者の、年代・男女別(15~19歳、20~29歳、30~39歳、40~49歳、50~59歳、60~69歳、70歳~79歳の7年代区分×2属性)1461日分(2014/3/1~2018/2/28)の昼夜間人口データを用いた。ここで羽咋市居住者とは、羽咋市内にて契約を行った携帯電話を有する者である。また本データは、羽咋市居住者の動態のみに着目したものであり、羽咋市居住者以外の他市への流出及び他市からの流入は含まない。

本研究では羽咋市民の他市への流出パターンを特徴抽出することを目的に、属性別に夜間人口から昼間人口の差分を取り、さらにデータを基準化するため、夜間人口の平均で除した、夜間人口1人当たりの「昼間流出率」を算出し、これを使用データとした⁴⁾。このデータの特徴として、値が0(0%)に近いほど昼夜間の人口差が小さいことを示し、1(100%)に近いほど昼間人口が他市へ流出していることを示す。使用データとして例えば、図1は2017/3~2018/2における性別・年代別昼間流出率を可視化したものであり、図1a)は15~19歳男女、図1b)は20~79歳男女に着目した図である。図1の横軸は上記期間の時系列、縦軸は昼間流出率であり、図1a)と図1b)でスケールを統一している。図1a)において、図中の数字は該当区間中の男女昼間流出率の平均値である。また図1b)では年代別に色分けがなされており、色が薄いほど若年層、色が濃いほど高齢層を表す。

表1 人口流動調査まとめ

項目		携帯電話 ビックデータ	国勢調査	PT(パーソン トリップ)調査	全国幹線旅客 純流動調査
量	調査対象	利用者による 大量サンプル	全数	居住者への無 作為標本抽出	都道府県を超えた 通勤・通学以外
	調査地域	全国	全国	都市圏単位	全国(機内・車内)
質	属性	△(年齢:10歳区 分、性別)	○(年齢・性別・職 業・住所)	○(年齢・性別・ 職業・住所)	○(年齢・性別・職 業・住所)
	移動目的	×	△(通勤・通学)	○(施設・目的・ 時間等)	○(目的・出発・目 的地)
	交通手段	×(複数交通手 段は×)	△(10年に1度)	○(22分類)	○(目的地までの 交通手段)
頻度	調査日	24時間365日	1日(10/1)	ある1日	秋期の平日・休日1 日ずつ
	調査頻度	常時取得	5年に1度	概ね10年に 1度	5年に1度

(国土交通省都市局作成「総合都市交通体系調査におけるビックデータ活用の手引き」²⁾を参考に筆者加筆)

図1a)より、15～19歳男性と15～19歳女性の昼間流出率はおおよそ類似した特徴であり、3月や8月、年末年始等は昼間流出率が、それぞれ24%、26%、18%と、他の区間と比べて低い。その他の区間として4月～7月や9月～12月下旬、1月上旬～2月等は相対的に昼間流出率が高く、それぞれ41%、37%、29%である。これらから15～19歳男女は主に通学で市外に流出していると推察され、学校の開校時と閉校時に関して周期性が見られる。他方、図1b)より20～79歳男女は5月上旬(ゴールデンウィーク)、8月中旬(盆期間)、年末年始等の部分的に昼間流出構造が普段と異なり、市内に滞在している様子が伺えるが、その他期間は昼間流出率が安定的に推移している。また流出の構造は若年層ほどその傾向が高く、高齢層ほど昼間時の流出が少ないことがわかる。また図1a)と図1b)より、15～19歳男女は周期性を持っているが故に、他の属性と流出構造が異なっていることが言える。

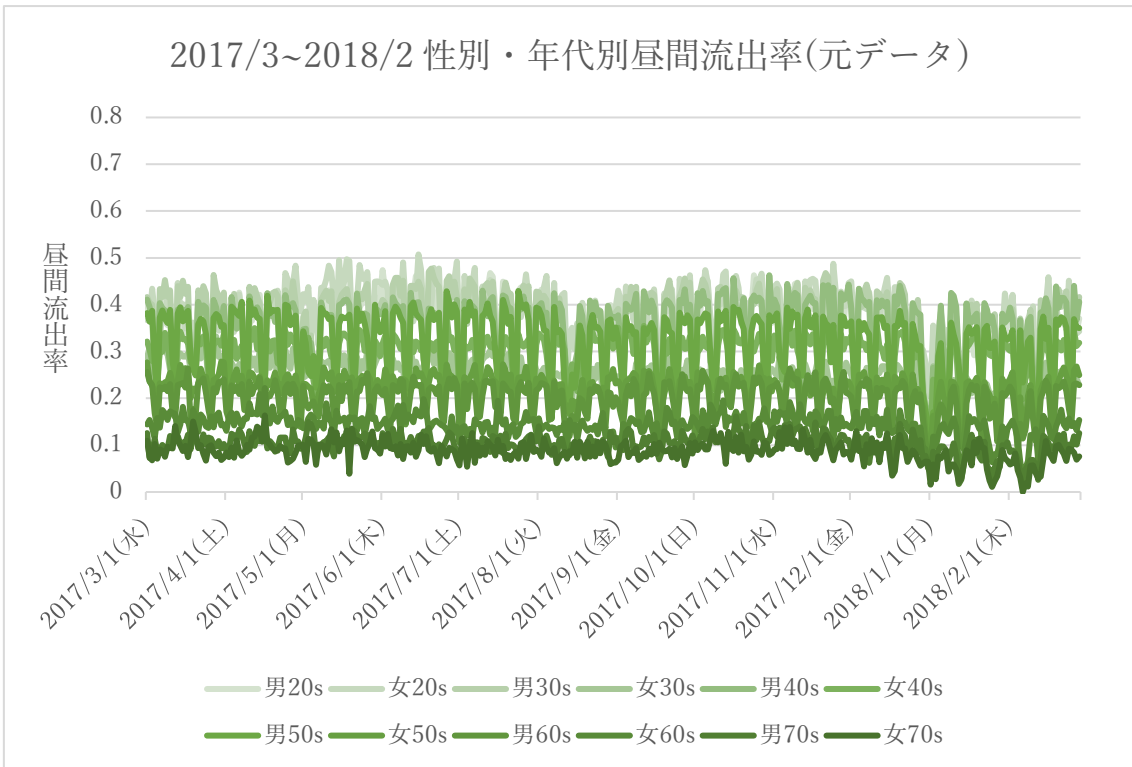
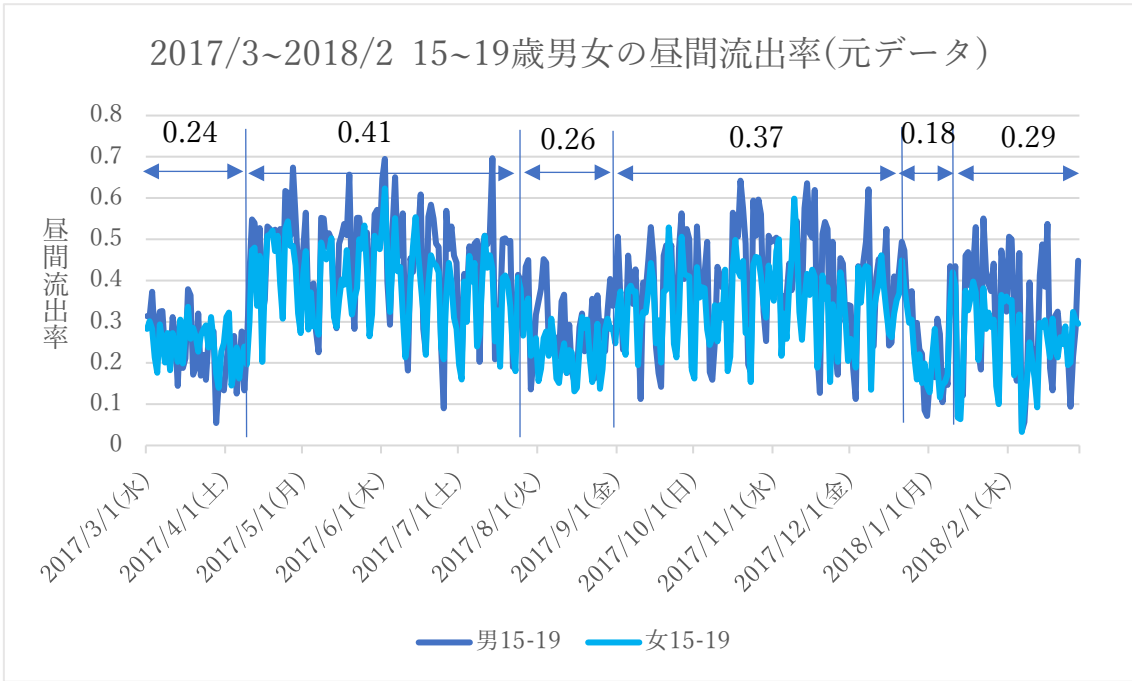


図1 使用データに基づく羽咋市民の2017/3~2018/2期の昼間流出率
 (上(a):15~19歳男女の昼間流出率、下(b):20~79歳男女の昼間流出率)

3.石川県羽咋市の基本情報(H27 国勢調査)

ここでは国勢調査を用いて、羽咋市における静的な流動の実態を確認する。本研究の対象地域とした石川県羽咋市は人口 21,729 人の都市であり、金本・徳岡(2002)¹⁷⁾の都市雇用圏において中心都市と設定される金沢市と七尾市の間に位置する(ただしいずれの郊外都市にも該当しない)。15 歳未満の年少人口、15 歳以上 65 歳未満の生産年齢人口、65 歳以上の老年人口数及び比率はそれぞれ、2,272 人(10.5%)、11,340 人(52.2%)、7,922 人(36.5%)である(ただし年齢不詳者 195 人を除く)。また 65 歳以上 80 歳未満の人口は 5,330 人(24.5%)であることから、本データでは 15 歳以上 80 歳未満の 16,670 人(76.7%)分の行動パターンを分析していることに対応する。

羽咋市民の性別・年代別の流出率⁵(図 2)では、男性では 20~59 歳において市外への流出率が約 50%と高く、男性 15~19 歳も次いで 42.1%と高い。女性では 15~19 歳が 49.2%と最も高く、次いで 20~29 歳が 43.3%である。これらより、羽咋市は特に男性就労者や 15~19 歳男女の通勤・通学に伴う流出が多く、一方で 30 歳以上の女性や高齢の男女は同流出が少ない特徴がある地域である。

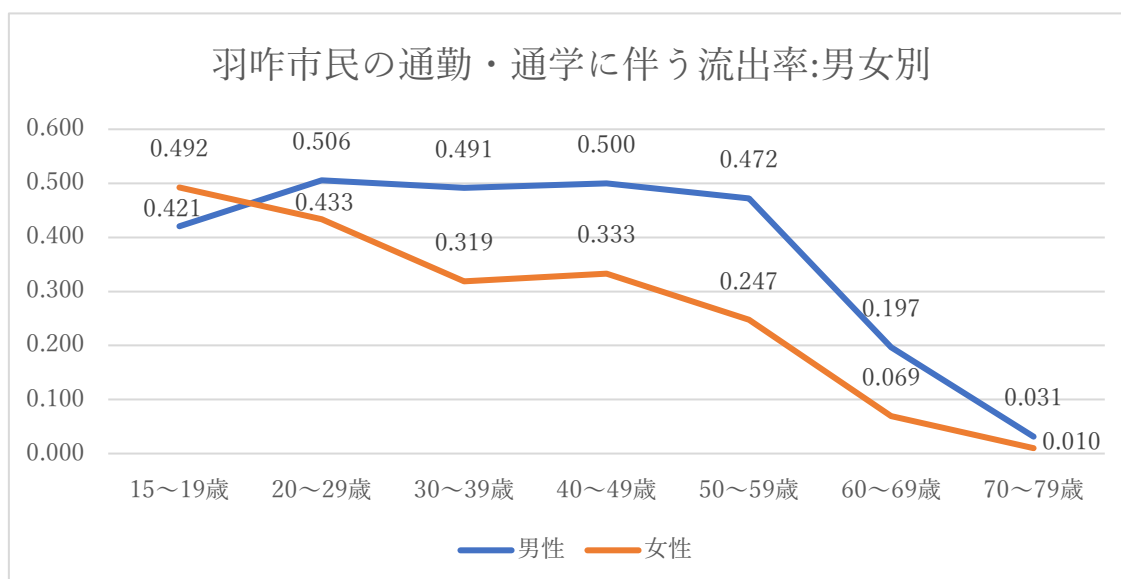


図 2 羽咋市民の通勤・通学に伴う性別・年代別流出率、国勢調査より筆者作成

4.分析方法

4.1 非負値行列因子分解

非負値行列因子分解は、観測された非負値の行列データを加法的な 2 つの行列、基底行列 H と表現行列 U に分解する分析手法である。以降では本データの構造に合わせて、基底行列 H を時系列パターン、表現行列 U を属性パターンと呼ぶ。すなわち、観測データを $Y \in \mathbb{R}_+^{n \times m}$ 、分解される行列をそれぞれ、時系列パターン $H \in \mathbb{R}_+^{n \times k}$ 、属性パターン $U \in \mathbb{R}_+^{k \times m}$ とすると

$$Y \sim HU$$

で表される。ここで k は観測データを分解する基底数を表す。これより二乗誤差基準に基づく非負値行列因子分解モデルは以下で表される。

$$L = \|Y - HU\|_F^2 \quad H_{i,j} \geq 0, U_{i,j} \geq 0$$

また、上記式を最小化する行列 H と U の最適化アルゴリズムは以下のように表される。

$$H_{i,j} \leftarrow H_{i,j} \cdot \frac{(YU)_{i,j}^T}{(HUU^T)_{i,j}}$$

$$U_{i,j} \leftarrow U_{i,j} \cdot \frac{(H^T Y)_{i,j}}{(H^T H U)_{i,j}}$$

ただし、上記式は要素ごとの演算である。

4.2 スパース非負値行列因子分解

スパース非負値行列因子分解について、本研究では Cichocki, Zdunek and Amari (2007) における最適化アルゴリズムを参考に構成した。すなわち二乗誤差基準に基づくスパース非負値行列因子分解モデルは以下で表される。

$$L = \|Y - HU\|_F^2 + \mu \|H\|_1 + \lambda \|U\|_1$$

$$H_{i,j} \geq 0, U_{i,j} \geq 0, \mu \geq 0, \lambda \geq 0$$

ここで右下の添え字 F はフロベニウスノルム、同添え字 1 は L_1 ノルムを表し、 μ と λ はハイパーパラメータである。また上記式を最小化する行列 H と U の更新式は以下のように表される。

$$H_{i,j} \leftarrow H_{i,j} \cdot \frac{(YU)_{i,j}^T}{(HUU^T)_{i,j} + \mu(\mathbf{1}_n \mathbf{1}_k^T)_{i,j}}$$

$$U_{i,j} \leftarrow U_{i,j} \cdot \frac{(H^T Y)_{i,j}}{(H^T H U)_{i,j} + \lambda(\mathbf{1}_k \mathbf{1}_m^T)_{i,j}}$$

ただし $\mathbf{1}_n$ は要素が全て1の $n \times 1$ ベクトルである。

4.1 と 4.2 に共通して、最適アルゴリズムにおける初期値は一様分布に基づくランダムな値とした。またアルゴリズムの収束条件は更新前後の変化率が H と U ともに 10^{-3} を下回った場合と設定した。基底数 k の設定について、本研究では $k = 3$ とした。基底数の定量的な決定方法として、統計モデルの良さを評価する指標である情報量基準のうち、AIC や MDL 等を使用する場合があるが、本研究例の場合、尤度関数に対応する二乗誤差量のオーダー数と、罰則項の値のオーダー数との間に大きな乖離があり、基底数増加に伴う二乗誤差量の減少よりも罰則項の線形増加の影響が強大なため、これらの情報量基準が有効に機能しなかった。そのため、特徴の解釈を優先し、一般的に利用される都市間旅行データにおける旅行目的分類が業務、観光、私用・帰省の3つであることから、基底数を3と設定した。また、スパース非負値行列因子分解におけるハイパーパラメータについては、現在有効なパラメータ数の設定

方法が論証されていない。そのため本研究では分解される行列のうち、不要な要素を0にすることを目的に、 $\mu|H|_1$ と $\lambda|U|_1$ のそれぞれにおいて二乗誤差量のオーダー数と概ね等しい値となるように $\mu = 10^{-3}$ 、 $\lambda = 10^1$ と設定した。

5. 分析結果

5.1 非負値行列因子分解による分析結果

(1)3つの基底

基底数を3とした時の非負値行列因子分解による分析結果をそれぞれの基底ごとについて詳細に論述する。属性パターン U (図3~5、図7~9のa))は横軸が性別・年代別の属性であり、縦軸は総和が1の重みを表している。 H は時系列パターンと曜日周期に分けられ、時系列パターン(図3~5、図7~9のc))では横軸が1年分の時系列、縦軸が昼間流出率を表し、7日間移動平均をとった4年分の時系列パターンをそれぞれ図中にプロットしている。またその時系列パターンのうち、曜日に関して平均を算出したものが曜日周期のパターン(図3~5、図7~9のb))であり、横軸は月曜日から日曜日までの曜日周期、縦軸は曜日ごとの昼間流出率を表している。

1つ目の特徴として、「15~19歳男性」が抽出された。図3a)を見ると、15~19歳男性の特徴が0.22と最も高く抽出されており、20~59歳男性や20~29歳女性など、多くの属性の特徴が同時に抽出されている。他方、図3b)をみると時系列パターンでは、3月や8月、年末年始などに他市への流出が少なく、その他の期間は流出が多い。また、図3b)をみると曜日周期パターンでは、平日の方が土日より流出が多い。これらを総じて、様々な属性が混在する中でも、概ね「15~19歳男性」の特徴と解釈され、3月や8月、年末年始等はそれぞれ春休み、夏休み、冬休みに該当し、平日は市外の学校や会社に通勤・通学し、土日は市内に滞留している特徴がある。

2つ目の特徴として、「15~19歳女性」が抽出された。図4a)を見ると、15~19歳女性の特徴が0.41と最も高く抽出されている。他方、図4b)をみる時系列パターンでは、3月や7月中旬~8月、年末年始などに他市への流出が少なく、その他の期間は流出が多い。また、図4c)をみると曜日周期パターンでは、平日の方が土日より流出が多い。これらを総じて、概ね「15~19歳女性」の特徴と解釈され、「15~19歳男性」と類似した特徴を示している。すなわち3月や7月中旬~8月、年末年始はそれぞれ春休み、夏休み、冬休みに該当し、平日は市外の学校や会社に通勤・通学し、土日は市内に滞留している特徴がある。また、「15~19歳女性」の方が、U側で特徴が突出しているため、より「学生らしい」行動パターンとなっている。特に「夏休み」について石川県の公立高校は7月中旬頃から8月末日までを長期休暇とする特徴があり、その影響で学校が休みとなり、羽咋市内に滞留する時系列パターンとなっていることが推察される。

3つ目の特徴として、「20~59歳男女」が抽出された。図5a)を見ると20~29歳女性が0.12であることを筆頭に20~59歳男性や20~59歳女性の特徴が高く抽出されている。他方、図5b)をみると時系列パターンでは、2月末~3月末、7月中旬~8月末に昼間流出率が高い。その他期間では、同比率は概ね安定的に推移している。また、図5c)

をみると曜日周期パターンでは、日曜日の昼間流出率は相対的に低いが土曜日は高い。これらを総じて、概ね「20~59歳男女」の特徴と解釈される。平日は通勤パターンであるが、2月末~3月末や7月中旬~8月末あたりに昼間流出率が高いことや、土曜日も同比率が高いことから、市外での宿泊を含む旅行に関する流出パターンも推察される。また4月はこれらの属性において転出が発生していると考えられる。例えば今まで羽咋市内に在住しており、就労で市外に流出していた者が転出を行うと、羽咋市内の昼間人口は変化しないが、夜間人口は減少し、それによって昼夜間の人口差が減少するため昼間流出率は減少することとなる。そのため4月は他の期間よりも昼間流出率が低いことが推察される。

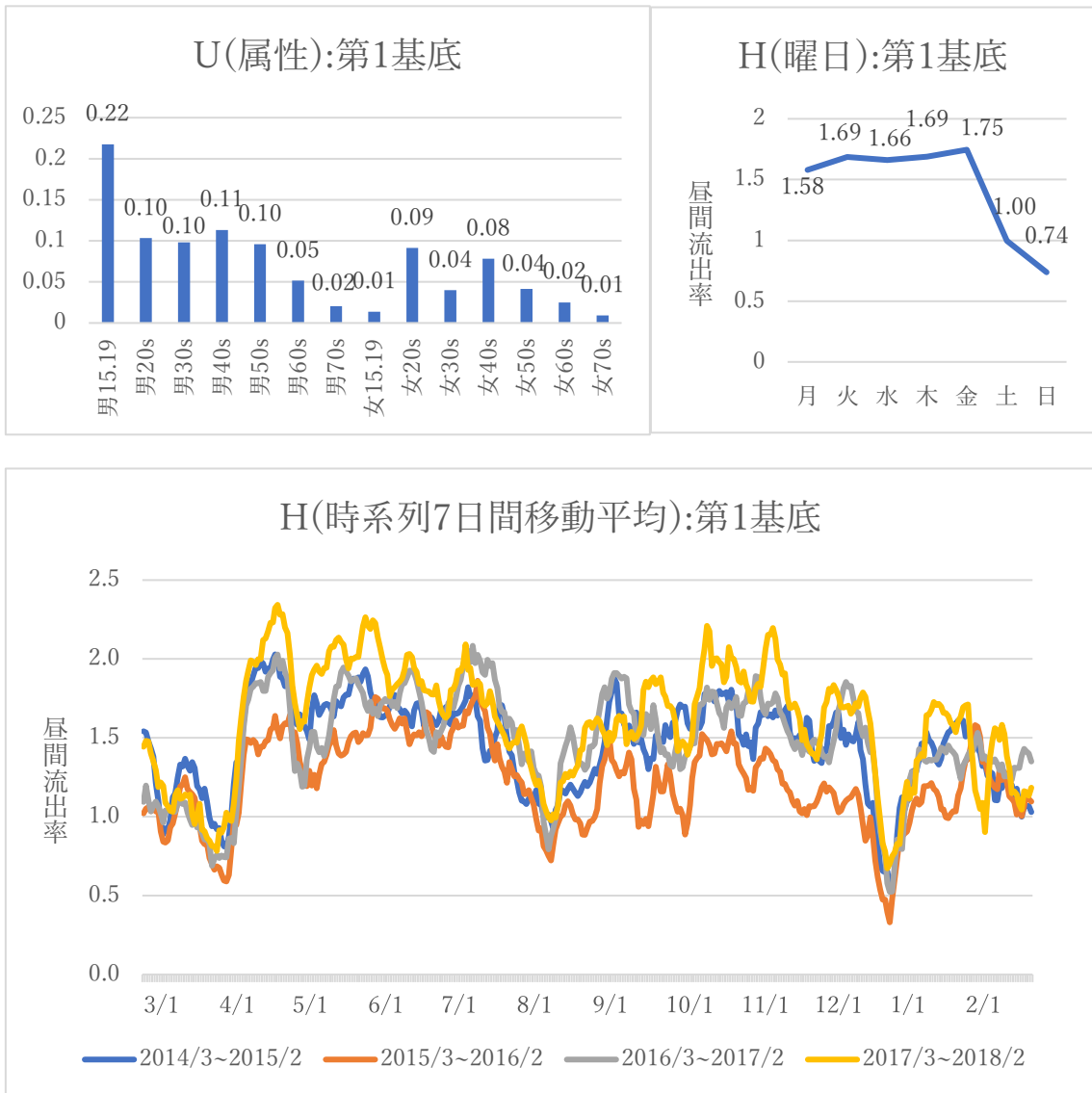


図3 非負値行列因子分解による行列分解 第1基底
 (左上 a):U 側属性、下 b):H 側時系列周期、右上 c):H 側曜日周期)
 (※ a 図における「男(女)15.19」は「15~19歳」を表す)

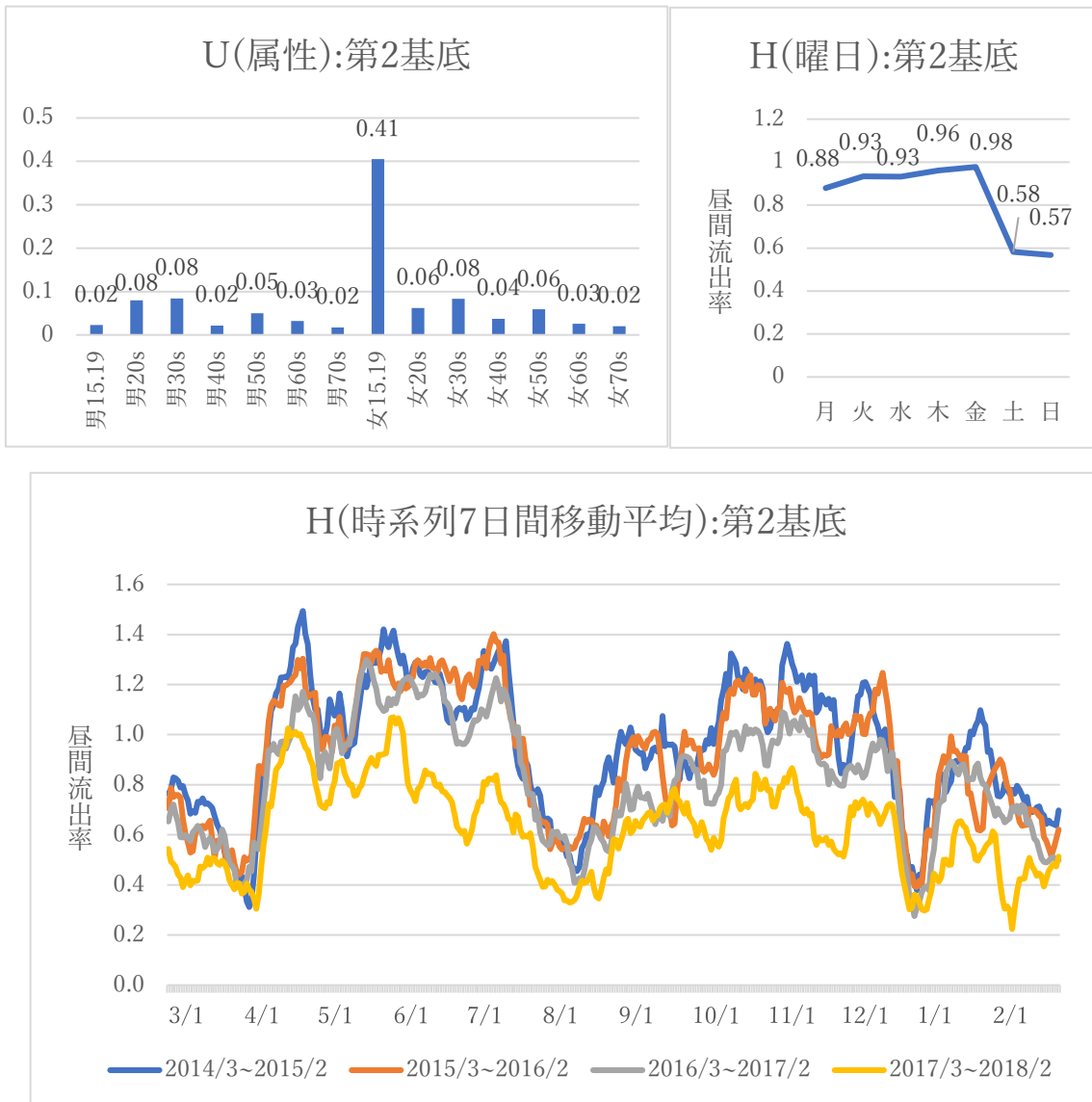


図4 非負値行列因子分解による行列分解 第2基底
 (左上 a):U側属性、下 b):H側時系列周期、右上 c):H側曜日周期)
 (※ a図における「男(女)15.19」は「15~19歳」を表す)

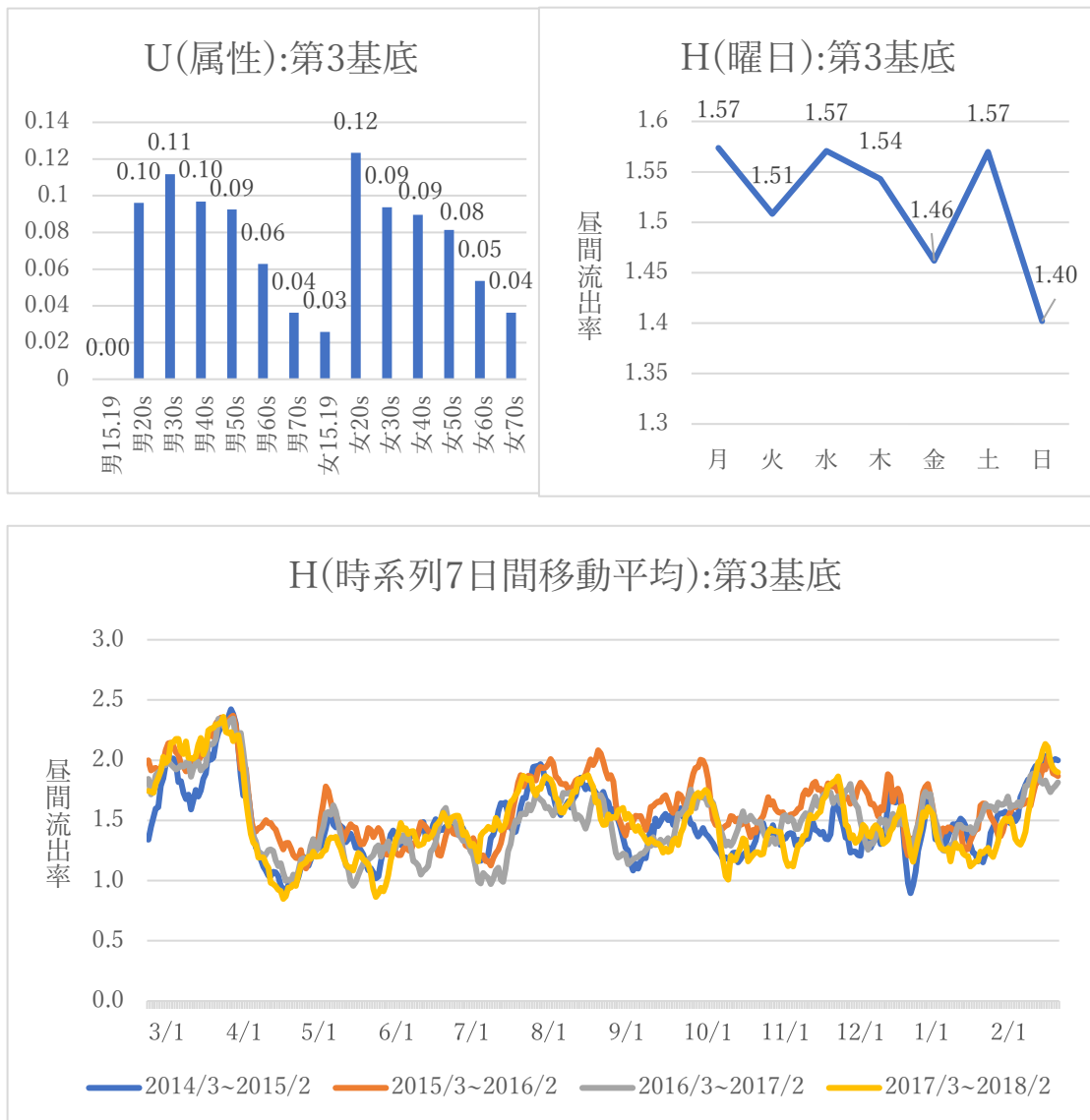


図5 非負値行列因子分解による行列分解 第3基底
 (左上 a):U側属性、下 b):H側時系列周期、右上 c):H側曜日周期)
 (※ a図における「男(女)15.19」は「15~19歳」を表す)

(2)非負値行列因子分解近似の残差項 ($Y - HU$)

非負値行列因子分解を適用することで、(1)で示された3つの特徴が抽出されたが、他方で元データを3次元に低次元近似した際のデータの損失分が残差項として存在する。これは、上述の3つの特徴パターンの増減だけでは表せない特徴であり、この損失分を分析することで元データに対して網羅的に解析することが可能である。

図6はそれぞれ元データに対する非負値行列因子分解モデルでの近似残差の2乗構造を表しており、上図(a)は時系列ごと(行方向)に総和をとった「時系列残差二乗ベクトル」、下図(b)は属性パターンごと(列方向)に曜日別で平

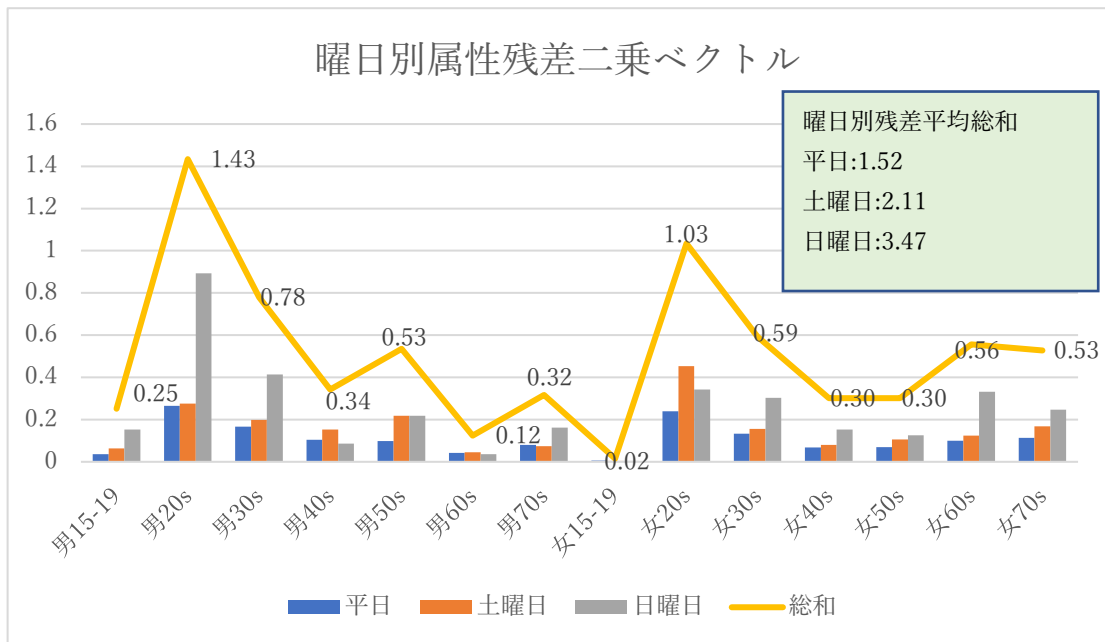
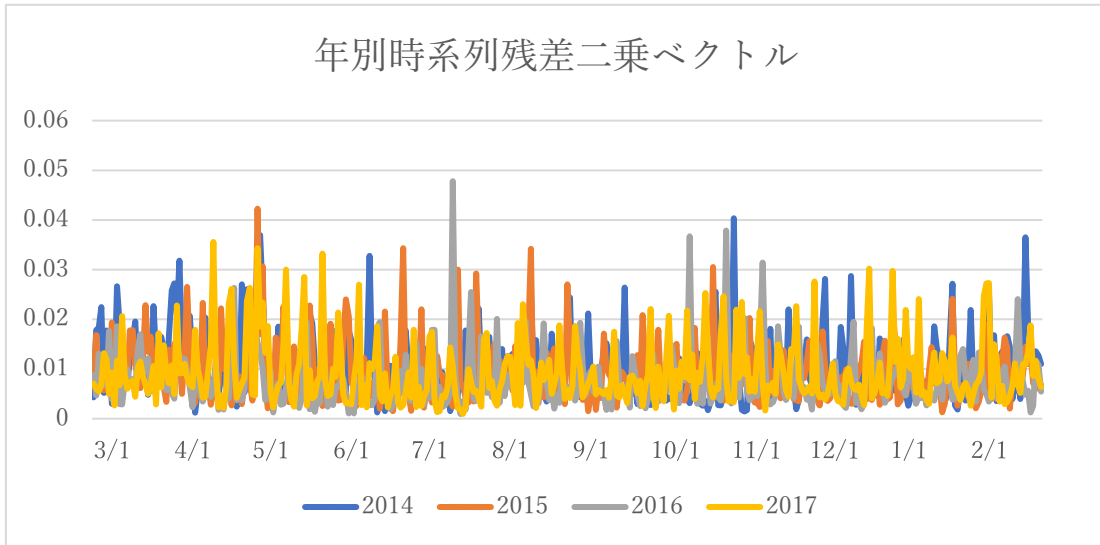


図6 元データに対する非負値行列因子分解近似の残差項

(上図 a):年別時系列残差二乗ベクトル、下図 b):曜日別属性残差二乗ベクトル)

均をとった「属性残差二乗ベクトル」である。いずれも縦軸は残差を表しており、縦軸の値が大きいほど、モデルと元データの残差が大きいことを示す。

図(a)において、普段の行動パターンとは極端に異なるような突出した残差項はここでは見られないが、残差の大きい値としてスパイクする日は一定の間隔を保っており、周期性があることが推察される。そこで(b)を確認すると、曜日別残差平均の総和では日曜日の残差が3.47と最も大きく、次いで土曜日が2.11、平日が1.52であることから(a)の周期性は概ね日曜日に関するものであることがわかる。また(b)より属性別の残差二乗ベクトル総和は、20～

29歳男性の残差が1.43と最も高く、次いで20～29歳女性が1.03、30～39歳男性が0.78、30～39歳女性が0.59と続いている。その一方で、第1基底（図3）や第2基底（図4）で抽出された、それぞれ15～19歳男性、15～19歳女性の残差は小さく、モデル内で特徴が抽出出来ていることがわかる。20～39歳男女の属性は第3基底（図5）でも大きく抽出されていたが、図5(b)より、それは平日や土曜日に関する特徴であった。対して残差においてこれらの属性は、特に日曜日に関して大きな値をとっている。3つの基底では日曜日に関する特徴を大きく抽出されていないことを踏まえると、非負値行列因子分解モデルでは日曜日に関する近似を過小評価していることが推察される。これらを総括して、羽咋市の20～39歳男女は(1)(2)合わせて平日においても、また土日においても他の属性より大きな流出構造を持っていることが推察される。

5.2 スパース非負値行列因子分解による分析結果

(1)3つの基底

1つ目の特徴として、「15～19歳男性」が抽出された。図7a)を見ると、15～19歳男性の特徴のみが抽出されている。図7c)をみると時系列パターンでは、3月や8月、年末年始等に他市への流出が少なく、その他の期間は流出が多い。図7b)をみると曜日周期パターンでは、平日の方が土日より流出が多い。これらを総じて、「15～19歳男性」の特徴と解釈され、3月や8月、年末年始はそれぞれ春休み、夏休み、冬休みに該当し、平日は市外の学校や会社に通勤・通学し、土日は市内に滞留している特徴がある。罰則項の影響により、U側の主要な要素以外の特徴を0にすることができ、属性がほぼ15～19歳のみとなった。したがってH側の時系列パターンも15～19歳男性の行動パターンを抽出したものであると推察できる。

2つ目の特徴として、「15～19歳女性」が抽出された。図8a)を見ると、15～19歳女性の特徴が0.98と極めて高く抽出されている。他方、図8c)をみると時系列パターンでは、3月や7月中旬～8月、年末年始などに他市への流出が少なく、その他の期間は流出が多い。また、図8b)をみると曜日周期パターンでは、平日の方が土日より流出が多い。これらを総じて、概ね「15～19歳女性」の特徴と解釈される。図7の「15～19歳男性」と同様に罰則項の影響によりほぼ15～19歳女性以外の特徴が0となったため、同属性の特徴が際立ったものとなっており、それに伴って時系列パターンの解釈を容易なものとしている。

3つ目の特徴として、「羽咋市民全体」が抽出された。図9a)を見ると、20～29歳女性や20～39歳男性の特徴が0.11ずつ占めていることを筆頭に全ての属性で特徴が抽出されている。図9c)をみると時系列パターンでは、GW、盆休み、年末年始に昼間流出率が低くなっており、その他は安定的に推移している。また、図9b)をみると曜日周期パターンでは、平日の方が土日より流出が多い。これらを総じて、概ね「羽咋市民全体」の特徴と解釈される。羽咋市民の基本的な行動パターンとして、平日の方が土日よりも昼間流出が多く、安定した周期的行動をとり、またGWや盆休み、年末年始は市内に滞留している特徴があることが推察される。

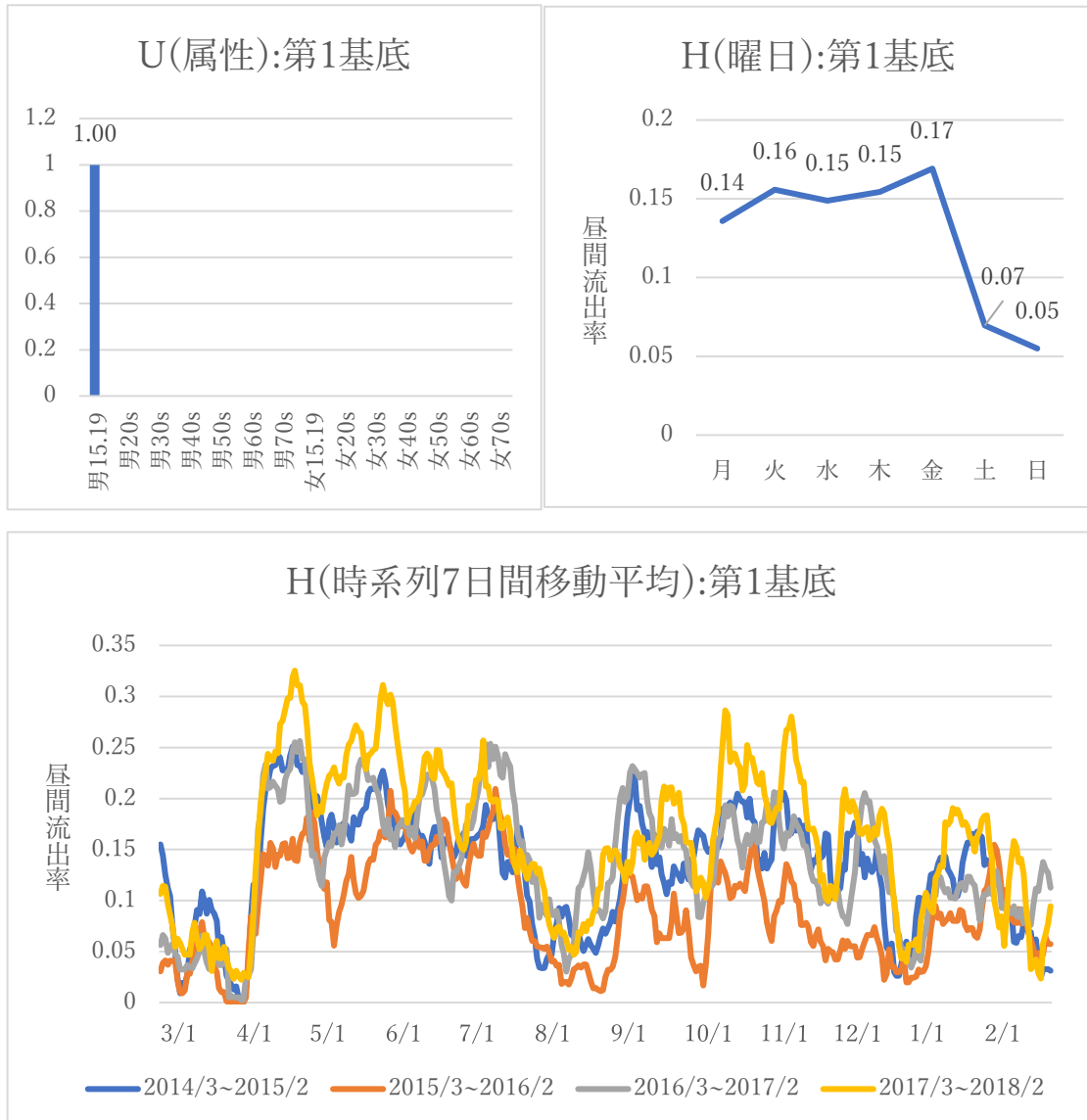


図7 スパース非負値行列因子分解による行列分解 第1基底
(左上 a):U 側属性、下 b):H 側時系列周期、右上 c):H 側曜日周期)

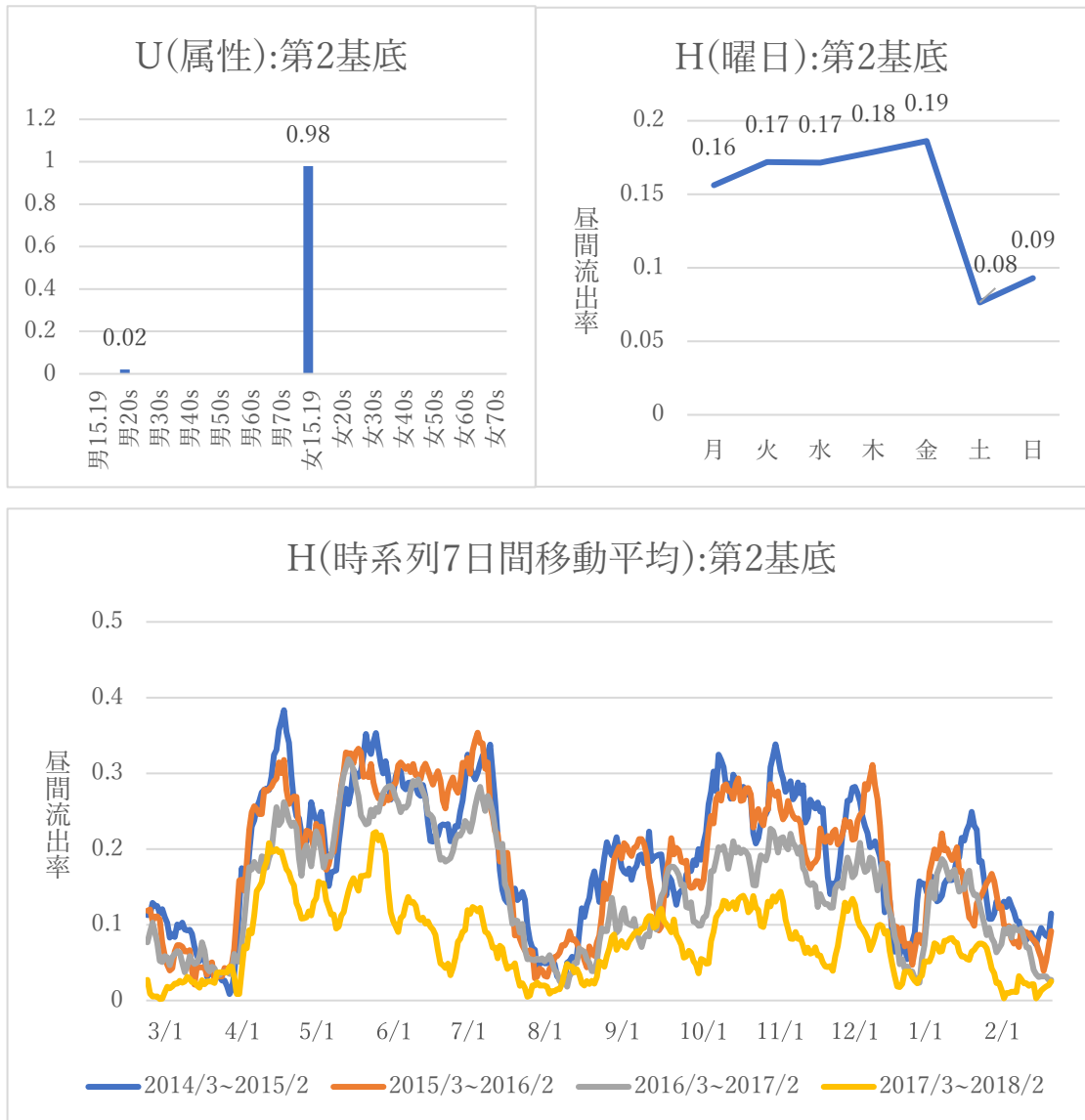


図8 スパース非負値行列因子分解による行列分解 第2基底
 (左上 a):U 側属性、下 b):H 側時系列周期、右上 c):H 側曜日周期)

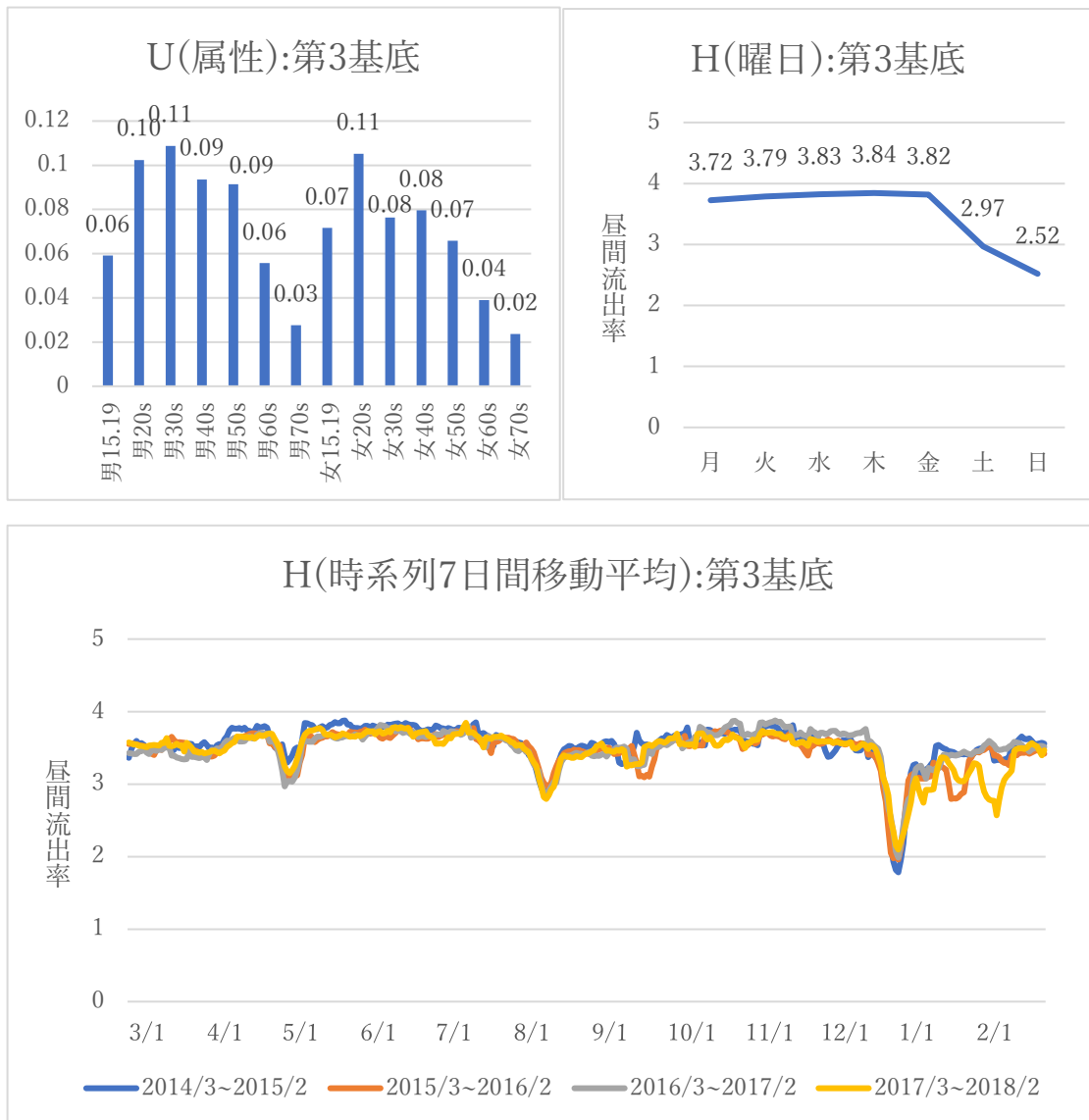


図9 スパース非負値行列因子分解による行列分解 第3基底
(左上 a):U 側属性、下 b):H 側時系列周期、右上 c):H 側曜日周期)

(2) スパース非負値行列因子近似の残差項 ($Y - HU$)

5.1.(2) と同様の分析をスパース非負値行列因子分解モデルにおいても行う。図 10(a)において、図 6(a)の特徴とおおよそ類似しており、普段の行動パターンとは極端に異なるような突出した残差項はここでも見られず、残差の大きい値としてスパイクする日は周期性があることが推察される。図 10(b)を確認すると、曜日別残差平均の総和では日曜日の残差が 4.47 と最も大きく、次いで土曜日が 2.61、平日が 1.72 であることから(a)の周期性は特に日曜日に関するものであったことがわかる。また(b)より属性別の残差二乗ベクトル総和を見ると、20~29 歳男性の残差が 1.43 と最も高く、次いで 20~29 歳女性が 1.31、30~39 歳女性が 0.92、60~69 歳女性が 0.78 と続いている。

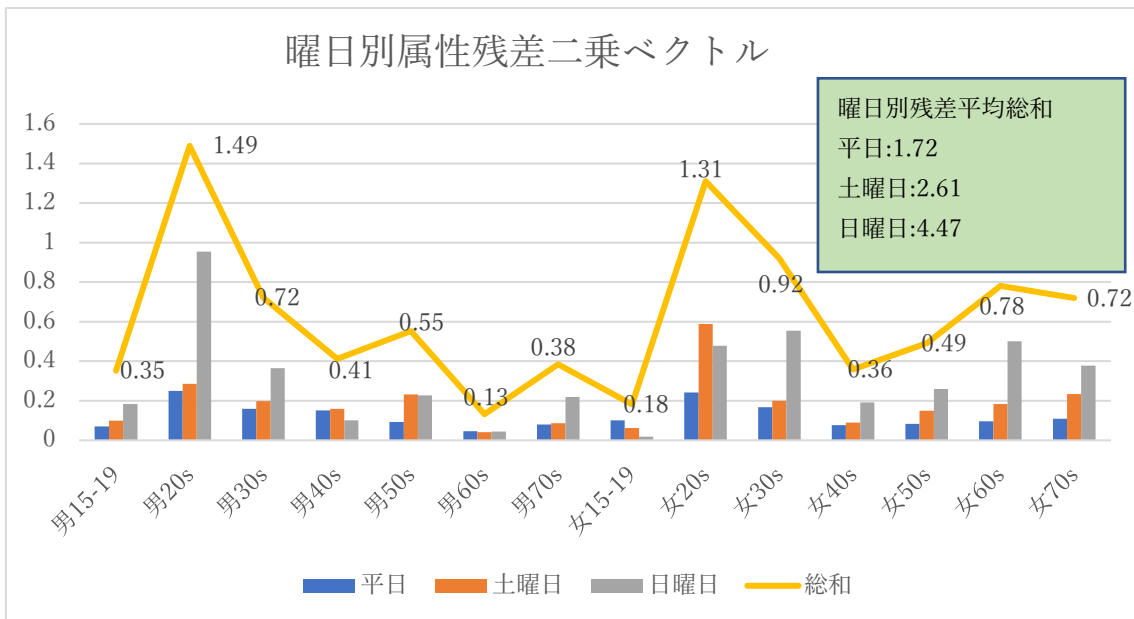
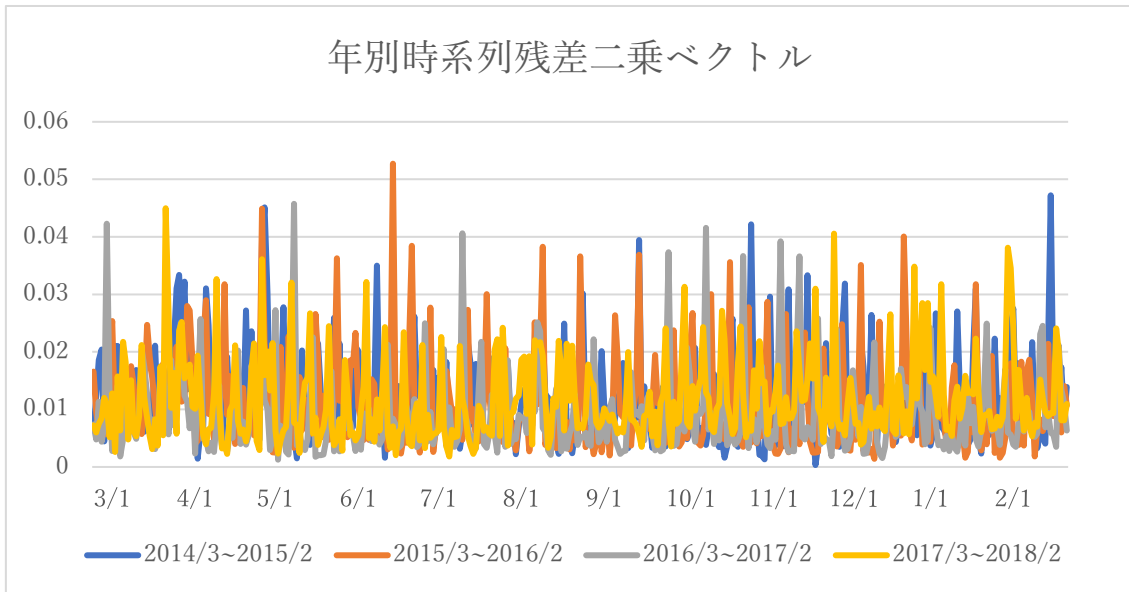


図10 元データに対するスパース非負値行列因子分解近似の残差項
 (上図 a):年別時系列残差二乗ベクトル、下図 b):曜日別属性残差二乗ベクトル)

これらの属性は第3基底(図9)でも大きく抽出されていたが、図9(b)より、それは平日に関する特徴であった。加えて第1基底(図7)、第2基底(図8)の基底がそれぞれ15~19歳男性、15~19歳女性を説明する基底であったことから、第1~第3基底までで大きく抽出されなかった土日に関する特徴が残差として表れ、非負値行列因子分解モデルの残差(図6b)より大きいことが推察される。

6. おわりに

本研究では羽咋市における昼間人口及び夜間人口数を計測したモバイル空間統計データに、非負値行列因子分解及びスパース非負値行列因子分解を適用し、ビックデータから羽咋市民の行動パターンの特徴を抽出することを試みた。分析結果として非負値行列因子分解では「15～19 歳男性」「15～19 歳女性」「20～59 歳男女」、スパース非負値行列因子分解では「20～59 歳男女」に代わり「羽咋市民全体」の特徴を抽出できた。そしてこれらは図1や図2と比較しても十分に特徴抽出できていることが伺える。元データの昼間人口率に基づいた図1では、15～19 歳男女の行動パターン周期的であったことや他のパターンとは異なった推移をしていたことが示されていた。国勢調査から算出した昼間流出率に基づいた図2と比較しても、特にスパース非負値行列因子分解では15～19 歳男女の昼間流出率が高い特徴だけでなく、男性の方が女性より昼間流出率が高いこと特徴も第3基底で抽出できている。

また、元データとモデルとの近似誤差に対応する残差項を確認すると、第1～第3基底までに大きく抽出されてこなかった日曜日に関する残差が大きく表れており、特に20～39 歳男女においてその特徴が大きかった。20～39 歳男女の行動パターンは平日の通勤・通学等の流出だけでなく、休日においても流出構造を持っていることが推察された。

今後の発展として、本稿において人口移動データに対してスパース非負値行列因子分解が有効に機能することが示せたが、厳密には昼間人口及び夜間人口データのみを使用していたため「流出先」に関しては明確ではなかった。そのため、OD(Origin destination)に対応するデータを使用し、「流出先」を明確にすることで、移動目的に関しても例えば、買い物目的の他市への流出や、旅行目的の流出など行動パターンについて深く考察できるであろう。

本研究より携帯電話の位置情報から得られるビックデータを解析する上で、非負値行列因子分解や罰スパース非負値行列因子分解は有効な分析方法の1つであると言えよう。特にこれらの客観的な分析アプローチから有益な特徴を抽出できたこと、またそれが図1のようにある属性、時系列に絞って定点観測した分析結果と類似した特徴パターンが抽出できたことは統計的に価値のある事であろう。

注釈

¹ PT 調査には都市圏PT 調査と全国PT 調査があるが本稿のPT 調査は前者のことを指す。後者は5年に1度の更新、小規模都市も対象とする等のメリットがあるが、移動の始点、終点の総量を表すOD(Origin destination)量が調査の対象外であり、地域の移動に関して議論する本研究の趣旨とは合わない。

² PT 調査は、三大都市圏及び地方中枢都市圏以外の地域は連続して実施されない場合があり、過去6回の調査のうち、実施回数が1回の都市圏で41.5%(27都市圏/65都市圏)を占めている。

³ 研究対象を羽咋市とした理由として、2017年に金沢大学と羽咋市(及びNECの3者間)においてデータに基づいた政策立案のための連携協定を締結したことが挙げられる。政策立案において人口動態の実態把握は重要な要素である

が、国勢調査等の静的な統計資料から解析できない日常的な市民の動態がモバイル空間統計の使用によって解析可能であり、この解析結果を羽咋市の連携協定プロジェクトにおける政策立案に役立てるために、同市と研究対象として

いる。
4 本データは羽咋市居住者以外の流入・流出は含まないため、基本的に夜間人口と昼間人口の差分が負にならない。しかし使用データ中に3要素分のみ、負値をとるものが存在し、2016/3/6(日)15~19歳男性が「-8人」、2016/5/15(日)15~19歳男性が「-1人」、2018/2/6(火)70~79歳女性が「-3人」であった。負値を取りうる原因として、市外での宿泊を伴う大人数(修学旅行など)での外出などが挙げられる。市外で宿泊するため、翌日の夜間人口(3時)としてはカウントされないが、翌日の13時までに市内に帰宅すると昼間人口としてカウントされる。今回生じた負値は、15~19歳男性の夜間人口平均が478人、70~79歳女性の同平均が1682人であることから、これらの他に及ぼす影響は微小なものと判断し、それらの値を「0」とした。

5 流出人口÷夜間人口で定義した。

参考文献

- 1)奥村誠, 山口裕通, 大窪和明, 「全国幹線旅客純流動調査の鉄道サンプル拡大方法に関する研究」, 土木学会論文集 D3, Vol.67, No.5, 2011, pp.911-918
- 2) 国土交通省 都市交通調査・都市計画調査 パーソントリップ調査に関連する手引き「総合都市交通体系調査におけるビックデータ活用の手引き」https://www.mlit.go.jp/toshi/tosiko/toshi_tosiko_tk_000031.html 2020年9月11日閲覧
- 3)Alexander, L., Jiang, S., Murga, M., González, M.C.,
“Origin-destination trips by purpose and time of day inferred from mobile phone data” Transportation Research Part C, Emerging Technologies, Vol. 58, 2015, pp.240-250
- 4)Alsger, A., Tavassoli, A., Mesbah, M., Ferreira, L., Hickman, M., “Public transport trip purpose inference using smart card fare data” Transportation Research Part C, Emerging Technologies, Vol. 87, 2018, pp.123-137
- 5)山口裕通, 奥村誠「非負値行列因子分解による都道府県間滞在分布の年周期変動の分析」, 第56回土木計画学研究発表会・講演集, 2017, pp.23-29
- 6) Yamaguchi, H., Nakayama, S.: Detection of base travel groups with different sensitivities to new high-speed rail services: Non-negative tensor decomposition approach, Transport policy, Vol. 97, 2020, pp.37-46
- 7)原田魁成, 山口裕通, 寒河江雅彦「スパース非負値行列因子分解を用いた COVID-19 流行期の県間旅行行動の変容分析」土木学会論文集, 2020, 投稿中

- 8) 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」
DEIM, 2020
- 9) Graells, Garrido, E., Caro, E., Parra, D., "Inferring modes of transportation using mobile phone data." EPJ Data Sci. 2018, 7, 49
- 10) Dianhai, Wang., Zhengyi, Cai., Yanlei, Cui. Xiqun, (Michael), Chen., "Nonnegative tensor decomposition for urban mobility analysis and applications with mobile phone data" *Transportmetrica A: Transport Science*, 2019
- 11) D, Lee. H, S, Seung., "Learning the parts of objects by non-negative matrix factorization", *Nature* 401, 1999, pp.788-791
- 12) P, O, Hoyer. , "Non-negative sparse coding", In *Neural Networks for Signal Processing XII (Proc. IEEE Workshop on Neural Networks for Signal Processing)*, 2002, pp.557-565
- 13) H, kim. H, park. , "Sparse non-negative matrix factorization via alternating non-negative-constrained least squares for microarray data analysis", *Bioinformatics*, 23(12), 2007, pp.1495-1502
- 14) Cichocki, A. Zdunek, R. Amari, S. , "Hierarchical ALS Algorithms for Nonnegative Matrix and 3D Tensor Factorization", *Independent Component Analysis and Signal Separation. Lecture Notes in Computer Science*, vol 4666. Springer, Berlin, Heidelberg , 2007 , pp.169-176
- 15) Y, Qian. S, Jia. J, Zhou. A, Robles-Kelly. , "Hyperspectral unmixing via $L_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol.49, no.11, 2011, pp. 4282-4297
- 16) Peter, Ballen. Sudipto, Guha. , "Elastic Nonnegative Matrix Factorization" *IEEE International Conference on Data Mining Workshops (ICDMW)*, 2018
- 17) 金本・徳岡 (2002) 「日本の都市圏設定基準」 *応用地域学研究* No.7, 1-15