

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

山口 裕通¹・奥村 誠²

¹正会員 金沢大学特任助教 自然科学研究科 (〒 920-1192 金沢市角間町)

E-mail: hyamaguchi@se.kanazawa-u.ac.jp

²正会員 東北大学教授 災害科学国際研究所 (〒 980-0845 仙台市青葉区荒巻字青葉 468-1 S-502W)

E-mail: mokmr@m.tohoku.ac.jp

近年、携帯電話位置情報データの活用が始まったことにより、これまで観測が困難であった都市間旅行の長期にわたる時間的な変動情報を得ることができるようになった。本研究では、混雑統計^⑧から得られたデータを対象に、長距離旅行行動の年周期変動パターンの分析方法を実施しその有用性を検討する。具体的には、東京都居住者の都道府県別の滞在時間構成比表を空間・時間パターンごとに分解し、その特徴を整理した。分解においては、まず各変数の時間変動パターンから、年(365日)周期の年周期パターンを抽出し、さらにそれを非負値行列因子分解にかけることによって、3種類の時空間パターンに分解した。上述の方法で抽出された時空間パターンは、観光/業務/私用・帰省といった旅行目的別と考えられるものに近い特徴を持つことが分かった。これは、多時点データの時間変動情報を用いることで、携帯電話位置情報には含まれない旅行目的などに近い情報を抽出できていることを示唆している。

Key Words: mobile phone GPS data, long-distance travel, year periodric pattern, non-negative matrix factorization

1. はじめに

都道府県をまたぐような長距離旅行行動の調査については、行動が各個人にとって低頻度かつ、その旅行頻度の個人差が大きといった特徴から、全容の把握が困難であることが指摘されてきた¹⁾。そのため、我が国全体の都道府県間流動を把握するために実施されてきた全国幹線旅客純流動調査は、旅行者を対象とした非常に大規模な調査となり、5年おきの秋期の平休日1日ずつの情報しか得ることができない。さらに、拡大面における課題から、バイアスが存在するという課題も報告されている²⁾。一方で、携帯電話位置情報をはじめとするパッシブ型の位置情報ビッグデータは、大量のサンプル(携帯電話ユーザ)の位置情報を、広範囲かつ高頻度で取得した信頼性の高いデータであり、国レベルでの長距離旅行の実態を俯瞰的かつ高頻度に把握することが可能である。すでに、Ahas et al. (2007, 2008)^{3) 4)} や室井ら (2015)⁵⁾, Janzen et al. (2016)⁶⁾ をはじめとして、観光旅行・長距離旅行の行動分析に活用され始めている。

これまでの研究では、従来調査を代替に向けた方法論が主に検討されてきた。これらは、携帯電話位置情報の「空間範囲の広さと空間解像度の高さ」に着目して、旅行目的などの情報を付与しながら^{6) 7)}、従来型

の長距離旅行分析・需要予測に必要なデータを取得しようとするものである。我が国においても、パーソントリップ調査と対応するデータを抽出する方法⁸⁾などの検討が進められている。一方で、携帯電話位置情報の特徴として、面的な移動情報を高頻度に取得しているという、「時間的に高い解像度」という特徴がある。全国幹線旅客純流動調査などの従来調査では、コスト面から調査頻度を上げることは非常に困難であったため、都道府県間 OD 表のような面的な移動情報の時系列変化を理解し、予測に適用するようなアプローチはほとんど検討されてこなかった。

もともと、長距離旅行行動については、お盆・年末年始などの特定のピークに需要が集中するなど、大きな季節変動(年周期変動)を持つという特徴がある。そのため、ピーク時の需要を分散化させた混雑緩和・利便性向上などの施策を検討するためには、長距離旅行の時間変動情報とそれを議論できるようなモデルが必要不可欠となる。さらに、新幹線や空港などの長距離旅行向けのインフラ整備についても、我が国の長距離旅行の年周期変動パターンと交通サービスがそのパターンに与える影響を明らかにすることで、より効率的・効果的なインフラ整備計画の検討につながることを期待できる。

そこで、本研究では、長距離旅行行動の年周期変動

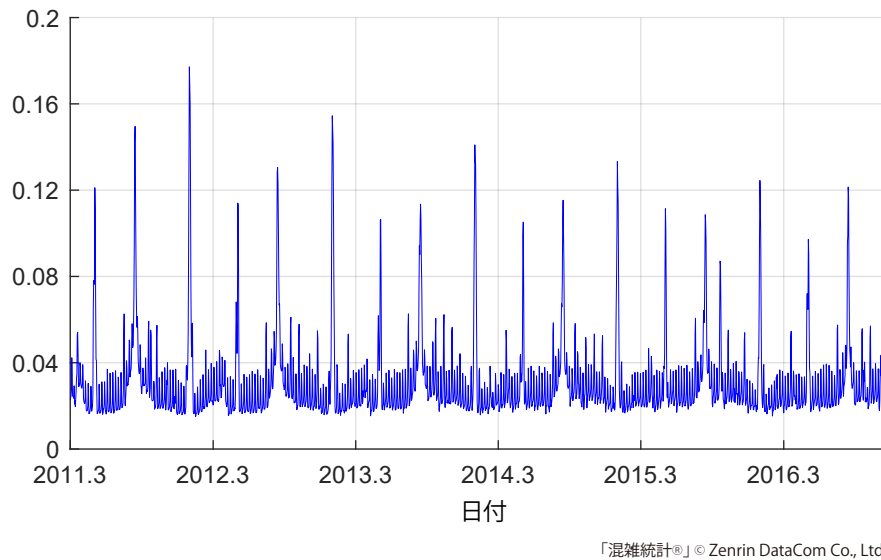


図-1 東京都居住者の首都圏外滞在確率の日変動

に着目し、その変動パターンの特徴を整理しつつ、それらを抽出・理解する方法を検討する。具体的には、携帯電話位置情報データの一つである混雑統計[®]から得られた、東京都居住者の都道府県別の滞在時間構成比表を空間・時間パターンごとに分解することでその特徴を整理した。まず、各変数の時間変動パターンに着目し、年（365日）周期と週（7日）周期の時間変動パターンを抽出した。そのうえで、43道府県での滞在確率の年周期パターンを、非負値行列因子分解にかけることで、3種類の特徴的な時空間パターンに分解した。抽出された時空間パターンは、観光/業務/私用・帰省といった旅行目的別と想定されるものが抽出されることを明らかにした。

本論文の構成は、以下のとおりである。2.では、本研究で用いる混雑統計[®]データの概要を説明する。3.では、長距離旅行の時系列データの特徴を整理しつつ、年・週周期の変動成分を抽出する方法を説明する。4.では、非負値行列因子分解を説明したうえで、年周期の変動成分を分解した結果を述べる。5.は本論文の結論である。

2. 混雑統計[®]による東京居住者の滞在時間データ

本研究では、株式会社ゼンリンデータコムが販売している混雑統計[®]データを利用して分析を行う。

この「混雑統計[®]」データは、NTTドコモが提供する「ドコモ地図ナビ」サービスのオートGPS機能⁹⁾利用者の中から、許諾を得た上で送信される携帯電話の位置情報をもとに作成される集計データである。元となるオートGPS機能利用者から得られる情報は、通

信が不可能な時などを除き、最短5分間隔で送信されるGPSによる緯度経度情報である（サンプル数は約50-70万人）。なお、この情報には性別・年齢などの個人属性に関する情報は含まれない。そして、携帯電話利用ユーザーごとの識別IDがついた位置情報を、NTTドコモが総体的かつ統計的に加工を行ったものであり、個人を特定できないよう十分に処理を施したうえで販売・提供される。

上述のようなスキーム上、提供されるデータは集計データのみであるが、その集計方法は用途に応じて細かく指定することが可能である。本研究では、居住地が東京と判定されたユーザーの2011年3月1日～2017年11月30日の2102日間のある一日 $(y, d) \in D$ における、滞在都道府県 $j \in Z$ ごとの集計滞在時間 $t_{j,(y,d)}$ を用いる。ただし、2102日の期間中に、混雑統計[®]データのユニーク・ユーザー数は時々刻々と変化している。そのため、本研究では次式から算出される滞在構成比 $p_{j,(y,d)}$ を分析に用いる。

$$p_{j,(y,d)} = \frac{t_{j,(y,d)}}{\sum_{j \in Z} t_{j,(y,d)}} \quad \forall ((y, d) \in D) \quad (1)$$

なお、集計滞在時間 $t_{j,(y,d)}$ を算出するために必要となる、携帯電話のGPS位置情報データにおける滞在/移動中・自宅地の判定には、山口ら¹⁰⁾で示されているものと同じアルゴリズムを用いた。

3. 時間変動データからの年・週周期変動成分の抽出

(1) 東京都居住者の首都圏外滞在確率の日変動

まず、都市間旅行の基本的な日変動を、東京居住者の首都圏外滞在確率から見ていこう。首都圏外滞在確

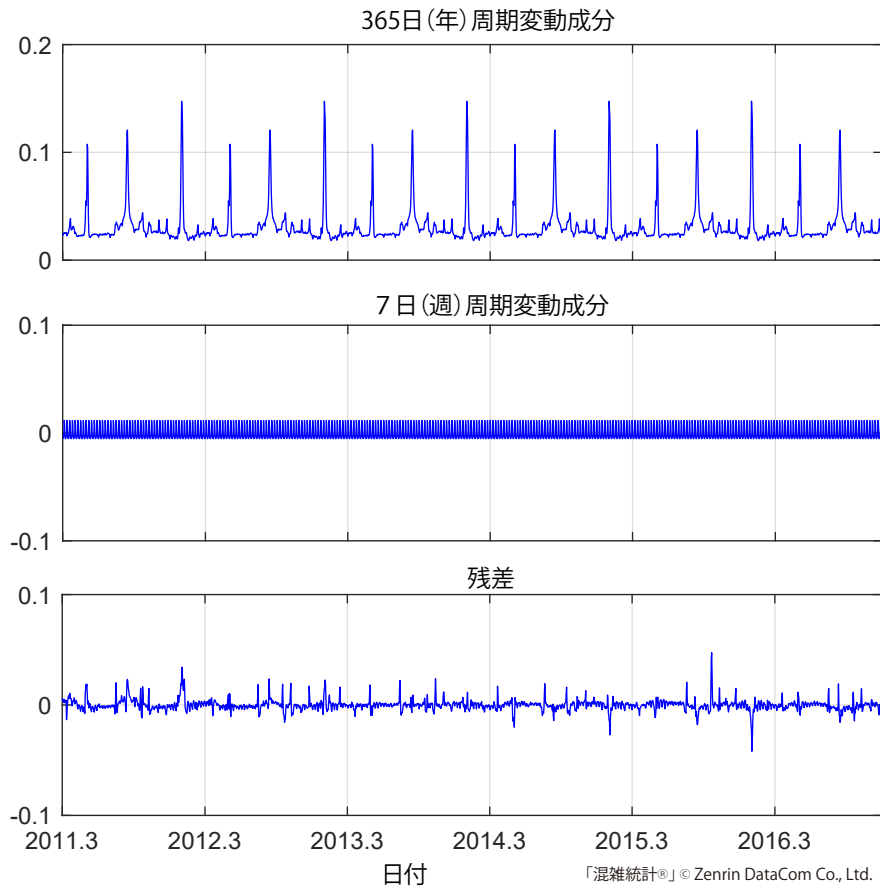


図-2 東京都居住者の首都圏外滞在確率の時間変動分解

率 $p_{\text{trav.},(y,d)}$ は、1都3県以外の道府県 $j \in Z'$ に滞在した時間の構成比として、次式から算出できる。

$$p_{\text{trav.},(y,d)} = \sum_{j \in Z'} t_{j,(y,d)} \quad \forall ((y,d) \in D) \quad (2)$$

図-1 は、この首都圏外滞在確率の日推移を図示したものである。この図から、東京居住者が首都圏外に滞在する確率の日変動の特徴として、以下の2点があることが読み取れる。(1) 平常時は約2-3%程度で変動している。(2) ゴールデンウィーク、お盆、年末年始の年に3回、大きなピークがあり、この期間においては10%以上(平常時の約5倍以上)が首都圏外に滞在している。

(2) 最小二乗法による周期変動情報の抽出

図-1の首都圏外滞在確率から、年(365日)周期と週(7日)周期の変動を抽出する。ここでは、混雑統計データから得られた首都圏外滞在確率 $p_{\text{trav.},(y,d)}$ は、次式のような二種類の周期変動と誤差の和として得られたものと考える。

$$p_{\text{trav.},(y,d)} = x_{\text{trav.},\text{year},d} + x_{\text{trav.},\text{week},w(y,d)} + \epsilon_{\text{trav.},(y,d)} \quad \forall ((y,d) \in D) \quad (3)$$

このうち、 $x_{\text{trav.},\text{year},d}$ は年周期の変動、 $x_{\text{trav.},\text{week},w(y,d)}$ は週周期の変動を示し、 $w(y,d)$ は年月日と曜日の関係を示す関数である。

そしてこの周期変動成分 $\mathbf{x}_{\text{trav.}}$ を、以下の最小二乗法を用いて推定する。

$$(\mathbf{x}_{\text{trav.},\text{year}}, \mathbf{x}_{\text{trav.},\text{week}}) = \underset{\mathbf{x}_{\text{trav.},\text{year}}, \mathbf{x}_{\text{trav.},\text{week}}}{\text{argmin}} \left(\sum_{(y,d) \in D} \epsilon_{\text{trav.},(y,d)}^2 \right) \quad \forall ((y,d) \in D) \quad (4)$$

(3) 東京都居住者の首都圏外滞在確率の分解結果

図-1の首都圏外滞在確率の日変動を、式(4)を用いて2つの周期成分と残差に分解した結果が、図-2である。この図から、平常時に高頻度で起こる1%程度の変動は週周期変動として、年間に3回あるピークは年周期変動として抽出されることが分かる。その結果として、残差成分はそれぞれの変動と比較して小さく、都市間旅行の日変動は年周期・週周期の二種類の変動要素が大きく効いていることが確認できる。一方で、不規則であるが大きな残差がある日も存在する。この残差には、観測誤差に加え、日付が固定されていない祝日の影響や、災害・交通機関の運行停止などの突発事象の影響が含まれている。

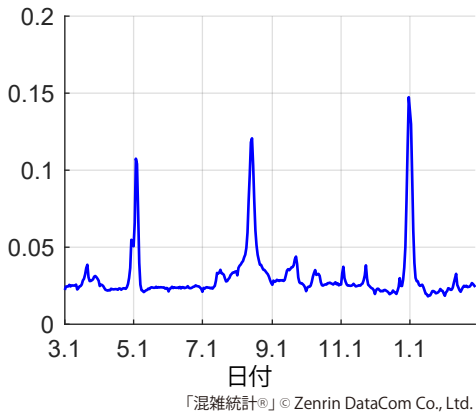


図-3 東京都居住者の首都圏外滞在確率の時間変動分解（年周期変動）

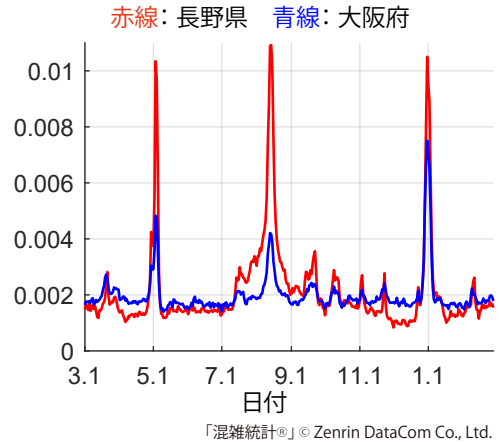


図-5 長野県・大阪府での滞在確率の年周期変動

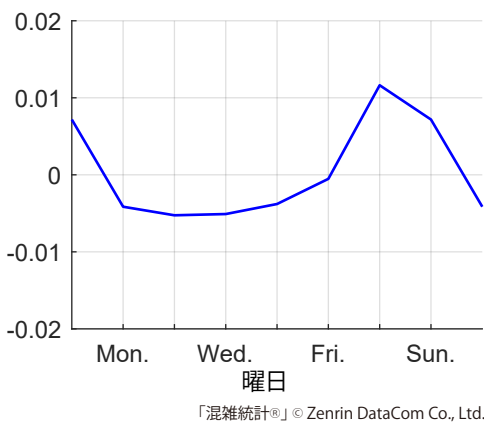


図-4 東京都居住者の首都圏外滞在確率の時間変動分解（週周期変動）

つぎに、推定した各周期成分の特徴を見ていこう。まず、図-3は、年周期成分を拡大したものである。この図からわかる特徴としては、以下の3点が読み取れる。(1) 年間に3回の強いピークが存在している。それぞれ5月初旬のゴールデンウィーク、8月中旬のお盆、1/1前後の年末年始であり、この期間は平常時の4倍以上の人が首都圏外に滞在している。(2) 9月から11月にかけて、祝日に一致して1%程度首都圏外に滞在する量が増加している。(3) 一方で、全く祝日のない6月にはほとんど年周期の変動が見られない。

図-4は週周期成分を拡大したものである。この図から、土日 > 平日であり、土曜日が最も首都圏外での滞在確率が高いことが読み取れる。

(4) 旅行先別滞在確率データの年周期変動

ここまで見てきた、首都圏外滞在の日変動について、旅行先別の滞在確率 $p_{j,(y,d)}$ を用いてさらに詳しく見ていこう。ここでは、旅行先 j ごとに以下の式から

週周期・年周期成分を算出する。

$$(\mathbf{x}_{j,\text{year}}, \mathbf{x}_{j,\text{week}}) = \operatorname{argmin} \left(\sum_{(y,d) \in D} \epsilon_{j,(y,d)}^2 \right) \quad (5)$$

$$\forall (j \in Z', (y,d) \in D)$$

図-5は、長野県と大阪府の滞在確率の年周期変動推定結果 $(\mathbf{x}_{\text{Nagano,year}}, \mathbf{x}_{\text{Osaka,year}})$ を示したものである。この図から、旅行先別に滞在確率の年間変動は大きく異なることが示唆される。長野県と大阪府の違いとしては、以下の2点を読み取ることができる：(1) ゴールデンウィーク、お盆、9月～11月の祝日におけるピークが、長野県と比較して大阪府は非常に小さい。(2) 一方で、年末年始には大阪でも強いピークが見られる。これは、旅行目的の差異などに起因すると考えられる。大阪府と比較して長野県に滞在する行動は、観光目的の旅行が多いためにゴールデンウィークやお盆などの行楽シーズンに大きなピークがあり、一方で、大阪府では平常時に実施される業務目的が占める割合が多いために、行楽シーズン時と平常時との差異が小さいと推察できる。

4. 旅行先別の年周期滞在確率変動の非負値行列因子分解

(1) 非負値行列因子分解

以降では、首都圏外の43道府県すべての旅行先を対象とした年間変動の特徴を見ていこう。本研究では、43変数という多変量の時間変動を理解するために、非負値行列因子分解を用いる。非負値行列因子分解とは、「非負値のデータを下方的な構成成分に分解することを目的とした多変量解析手法」¹¹⁾であり、画像解釈・分解¹²⁾をはじめとして、テキスト分解など非負値となるデータへの適用が進みつつある。

本研究では、式(5)で算出した、首都圏を除く43都

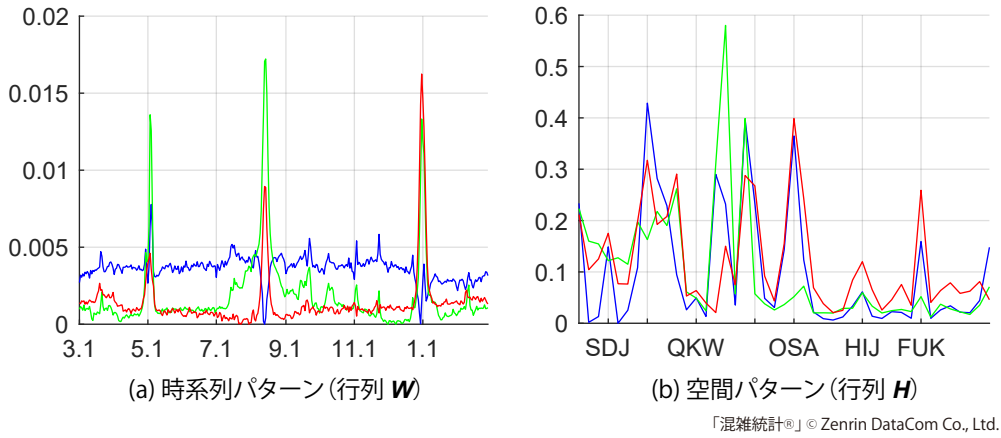


図-6 年周期滞在確率変動の非負値行列因子分解結果

道府県 × 365 日分の年周期変動成分から構成される行列 X_{year} を、非負値のみで構成される二つの行列 $W \in \mathbb{R}^{365 \times k}$, $H \in \mathbb{R}^{k \times 43}$ の内積として以下のように近似する：

$$\begin{aligned}
 X_{\text{year}} &\sim WH \\
 &= \begin{pmatrix} w_{1,1} & \dots & w_{1,k} \\ \vdots & \vdots & \vdots \\ w_{365,1} & \dots & w_{365,k} \end{pmatrix} \\
 &\quad \times \begin{pmatrix} h_{1,1} & \dots & h_{1,43} \\ \vdots & \vdots & \vdots \\ h_{k,1} & \dots & h_{k,43} \end{pmatrix} \quad (6) \\
 w_{d,l} &\geq 0 \quad \forall d \in D, \forall l \in [1, \dots, k] \\
 h_{l,j} &\geq 0 \quad \forall j \in Z', \forall l \in [1, \dots, k] \\
 \sum_{j \in Z'} h_{l,j} &= 0
 \end{aligned}$$

本研究では、matlab の ‘nnmf’ に実装された、ランダムな初期値からの交互最小二乗アルゴリズムによる方法で W, H を推定した。

なお、本研究では、 $k = 3$ として非負値行列因子分解を実施した結果を示す。 k の値を変えると、結果も大きく異なる。また、本研究で適用したデータ X_{year} はスパースなデータではないため、一意な分解結果を得ることができない¹³⁾。以降で示す結果は、複数ある解の一つでしかない。そのため、「 k にどのような値を設定すべきか？」と、「一意な解をえるためにどのような制約条件を付与すべきか？」の二点については今後検討を進める必要がある。

(2) 非負値行列因子分解の結果

図-6 は、非負値行列因子分解の結果得られた 2 つの行列の成分を示したものである。そして、図-7-9 は、それぞれの基底成分を 1 種類ずつ抽出したものである。なお、

図-7-9 の (a) では時点 d において基底成分 k が占める割合 $w_{d,k} / \sum_{k \in K} w_{d,k}$ を、(b) では基底 k の空間成分について全基底の平均値からの差分 $h_{k,j} - (\sum_{k \in K} h_{k,d} / 3)$ を示している。

まず、図-7 の 1 つ目の成分を見ていこう。この成分の時系列構成比図-7 (a) をみると、以下の 2 点の特徴が読み取れる。(1-1) 平常時には、おおよそ 60% 程度を占める、(1-2) ゴールデンウィーク、お盆、年末年始には、極端にその構成比が低くなる。そして、空間的な特徴を示す図-7 (b) を見ると、(1-3) 北海道・宮城県・近隣県（北関東 3 県と山梨県・静岡県）・愛知県・大阪府・福岡県・沖縄県といった場所で大きい。このような時間的・空間的な特徴から、平常時の「業務旅行」に近い成分であろうと推察される。

つぎに、図-8 の 2 つ目の成分を見ていこう。この成分の時系列構成比図-8 (a) をみると、以下の 2 点の特徴が読み取れる。(2-1) ゴールデンウィーク、お盆においては過半数をしめ、年末年始にも 40% 程度をしめる、(2-2) お盆のピークにつて裾野が広く、7 月後半から 8 月全体にかけて（学校等の夏休み期間で）構成比が高い。そして、空間的な特徴を示す図-8 (b) を見ると、以下の特徴が読み取れる：(2-3) 宮城県を除く東北地方、新潟県、山梨県・長野県といった、鉄道で 1-2 時間程度でアクセスできる場所で大きい値をとる。(2-4) 一方で、北関東 3 県・愛知県・大阪府・福岡県などでは小さい値をとる。このような時間的・空間的な特徴から、「観光・行楽旅行」に近い成分であろうと推察される。

最後に、図-9 の 3 つ目の成分を見ていこう。この成分の時系列構成比図-9 (a) をみると、(3-1) とくに年末年始で 60% と占める割合が高く、お盆においても比較的大きい値をしめる特徴が読み取れる。そして、空間的な特徴を示す図-9 (b) をみると、(3-2) 北関東 3 県・山梨県・長野県で小さい特徴がある。このような時間的・空間的な特徴から、「私用・帰省」に近い成分であ

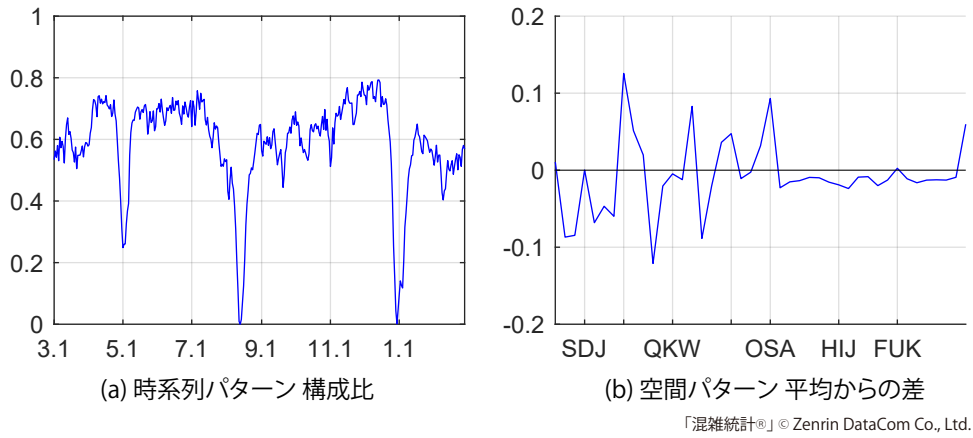


図-7 非負値行列因子分解結果（第一成分）

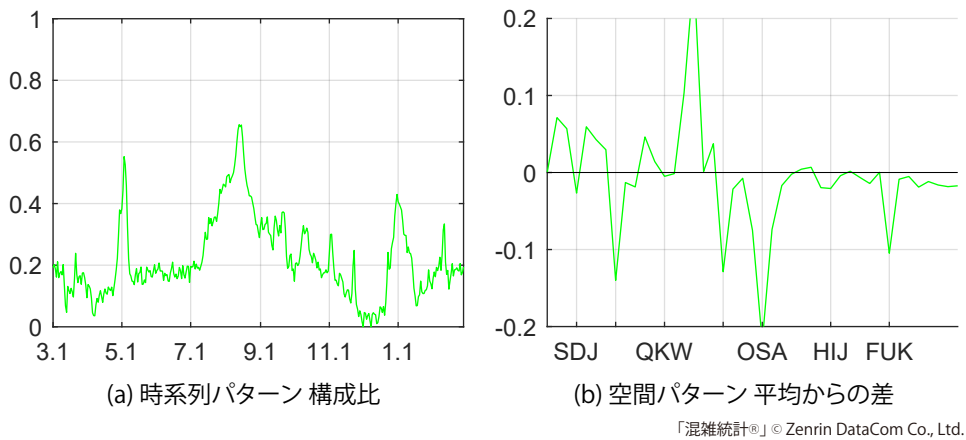


図-8 非負値行列因子分解結果（第二成分）

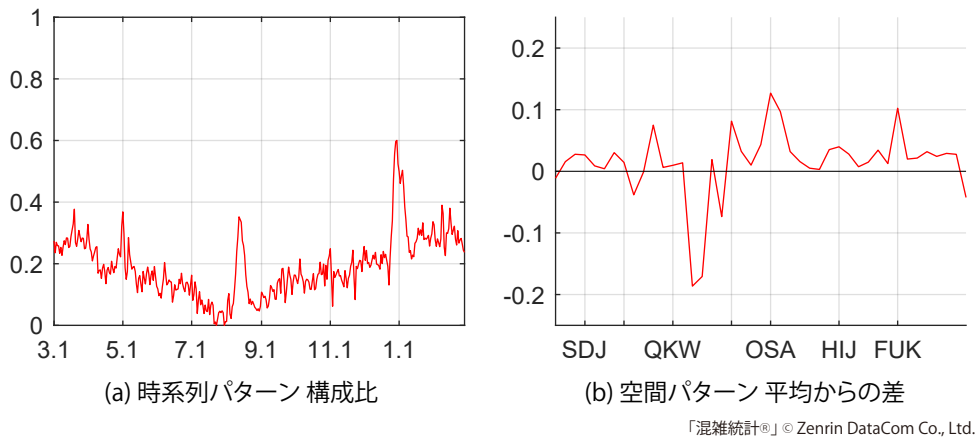


図-9 非負値行列因子分解結果（第三成分）

ろうと推察される。

以上のような分解結果が得られたことは、旅行先別の滞在時間確率の季節変動の特徴として、「平常時」、「お盆（+夏季）・ゴールデンウィーク」、「年末年始」の3種類で、それぞれ旅行先構成のパターンが大きく異なるという特徴があることを示している。

5. おわりに

本研究では、長距離旅行行動の年周期変動に着目し、混雑統計[®]から得られた、東京都居住者の都道府県別の滞在時間構成比表を空間・時間パターンごとに分解し、その特徴を整理した。まず、各変数の時間変動パター

ンから、年（365 日）周期の年周期パターンを抽出し、さらにそれを非負値行列因子分解にかけることで、3 種類の時空間パターンに分解した。このような方法で抽出された時空間パターンは、観光/業務/私用・帰省といった旅行目的別と考えられるものに近い特徴を持つことが分かった。

今後の課題としては、以下の 2 点があげられる。第一に、非負値行列因子分解で抽出された、各成分と旅行目的別の旅行先構成比との関係を明らかにすることがあげられる。本論文では、「旅行目的別に近いという推察」にとどまっているが、全国幹線旅客純流動調査や旅行観光消費動向調査などのデータと比較することで、旅行目的別の特徴との関係を明らかにすることができる。あるいは、別データから「旅行目的別に近い成分を抽出する」という制約条件を付与して、分解するといったアプローチも考えられる。

第二に、図-2 の残差成分についても精査とさらなる分解が必要である。例えば、「日付の変わる祝日」などは、年周期変動・週周期変動でも表現することはできないが、カレンダー上の効果として同様に扱う方が望ましいと考えられる。このような「予測できる」変動成分を残差からさらに分離することで、将来時点での長距離旅行量の予測精度向上が期待できる。さらに、それ以外の要素は、「交通サービスレベルの変化」や災害などの「突発的な異常事態」に相当するものが抽出されることが考えられ、この部分についても今後の都市間旅客交通サービスの向上に資する重要な情報があると考えられる。

謝辞：本研究は、科学研究費補助金・若手研究 (B) 「ビッグデータの時間情報が拓く、新たな都市間旅行需要予測手法の開発」(代表：山口裕通，課題番号：17K14736)，ならびに基盤研究 (B) 「在宅 Web 調査をフル活用したデータフュージョン型都市間旅客需要調査手法の開発」(代表：塚井誠人，課題番号：15H04060) からの支援を受けて行われた。また、本研究を遂行するにあたって、東京工業大学の福田大輔准教授，中西航助教から多くのアドバイスを受けた，ここに記して感謝の意を表します。

参考文献

- 1) Axhausen, K. W.: *Capturing long-distance travel*, Research Studies Press, pp. 342, 2003.
- 2) 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, pp.911-918, 2011.
- 3) Ahas, R., Aasa, A., Mark, Ü., Pae, T. and Kull, A.: Seasonal tourism spaces in Estonia: Case study with mobile positioning data, *Tourism Management*, Vol. 28, No. 3, pp.898-910, 2007.
- 4) Ahas, R., Aasa, A., Roose, A., Mark, Ü. and Silm, S.: Evaluating passive mobile positioning data for tourism surveys: An Estonian case study, *Tourism Management*, Vol. 29, No.3, pp. 469-486, 2008.
- 5) 室井寿明, 磯野文暁, 鈴木俊博: モバイル・ビッグデータを用いた都市間旅客交通への活用に関する研究, 土木計画学研究・講演集, Vol. 51, 2015.
- 6) Janzen, M., Vanhoof, M., Axhausen, K. and Smoreda, Z.: Estimating Long-Distance Travel Demand with Mobile Phone Billing Data, *In 16th Swiss Transport Research Conference*, 2016.
- 7) Alexander, L., Jiang, S., Murga, M. and 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, pp. 240-250, 2015.
- 8) 森尾淳, 牧村和彦, 山口高康, 池田大造, 西野仁, 藤岡啓太郎, 今井龍一: 東京都市圏におけるモバイル空間統計とパーソントリップ調査の比較調査-都市交通分野への適用に向けて-, 土木計画学研究発表会・講演集, Vol.52, pp.882-889, 2013.
- 9) NTT ドコモ: オート GPS, (http://www.nttdocomo.co.jp/service/auto_gps/index.html, last access: 2017/7/25).
- 10) 山口裕通, 奥村誠, 金田穂高, 土生恭祐: 携帯電話 GPS 情報から分かる 熊本地震による行動パターンの被災・回復過程, 土木計画学・研究講演集 Vol.54(CD-ROM), 2016.
- 11) 亀岡弘和: 非負値行列因子分解, 計測と制御 Vol.51, No.9, pp.835-844, 2012.
- 12) Lee, D. D. and Seung H. S.: Learning the parts of objects by non-negative matrix factorization, *NATURE*, Vol.401, pp.788-791, 1999.
- 13) 安川武彦: 非負値行列因子分解を用いたテキストデータ解析, 計算機統計学, Vol.28, No.1, pp.41-55, 2015.

(2017 年 7 月 31 日 受付)

Analysis of Japanese year periodric inter-prefectural travel volume change
by non-negative matrix factorization

Hiromichi YAMAGUCHI and Makoto OKUMURA