

非負値行列因子分解による 都市間旅客需要データ特性の分析

菅原 優志¹・Canh Xuan Do²・塚井 誠人³

¹ 学生会員 広島大学 大学院工学研究科 (〒739-8527 東広島市鏡山 1-4-1)
E-mail:m176539@hiroshima-u.ac.jp

² Department of Informatics in Civil Engineering, Faculty of Civil Engineering,
University of Transport and Communications (UTC)
(No.3 Cau Giay Street, Lang Thuong Ward, Dong Da District, Hanoi, Vietnam)
E-mail: canh.doxuan@gmail.com

正会員 広島大学准教授 大学院工学研究科 (〒739-8527 東広島市鏡山 1-4-1)
E-mail:mtukai@hiroshima-u.ac.jp

都市間旅客需要調査に関して、従来の全国幹線旅客純流動調査 (NPTS) に代わる調査方法が検討されている。本研究では、独自に実施したインターネットを利用した Web 調査と、携帯電話情報を利用したモバイル空間統計を活用した調査 (MOBI 調査) の特徴を明らかにする。具体的には、従来の NPTS, Web 調査, および MOBI 調査のそれぞれから作成した OD 表に対して、非負値行列因子分解法 (NMF) を適用して、鉄道または航空旅客流動特性の比較を行う。NMF を用いることによって旅客需要の特徴を視覚的に示す共に、NPTS, Web 調査, MOBI 調査の違いを明らかにできた。

Key Words: inter-regional, travel demand, NMF

1. はじめに

日常生活に必要な不可欠な交通手段の確保、高齢者や障がい者などの円滑な移動、交通の利便性向上・円滑化・効率化・国際競争力の強化、地域の活力向上、及び交通に係る環境負荷の低減などの施策を行うため、都市間を結ぶ幹線交通に関しても総合的な交通体系整備の重要性が指摘されている¹⁾。そのためには、国内の幹線交通機関を使った都市間交通需要を把握し、幹線交通の整備に伴う効果を検討する必要がある。

図 1 に示すように、旅客流動は利用した各幹線交通機関を単位とする旅客総流動と、出発地 (Origin) から目的地 (Destination) までの一連を単位とする旅客純流動の 2 種類のデータフォーマットがある。このうち後者は、乗り継ぎ拠点、あるいは異なる交通機関にまたがる交通施設計画に必須の情報である。国土交通省は、1990 年以降 5 年に 1 度、旅客純流動の調査を目的として、旅行者にアンケートを配布し、その場で回答を得る方法で全国幹線旅客純流動調査 (Net Passenger Transportation Survey, 以下 NPTS) を実施してきた。

NPTS は、都市内交通需要を把握するパーソントリップ調査と同様に平日・休日の 1 日ずつだけを調査対象日としている。このため、季節変動が大きい都市間交通需

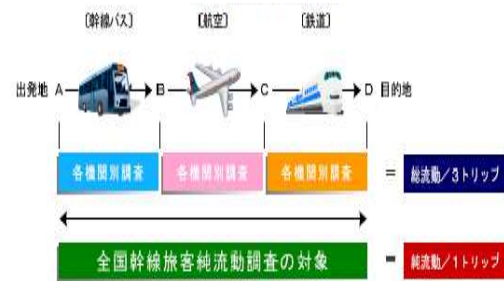


図 1 純流動と総流動

要に対する輸送計画が困難だと指摘されている¹⁾。季節変動を調査するには長期的な連続調査が有効だが、NPTS と同じ調査方法では、調査票の配布、回収、データ分析に多くの時間と費用を要するため、実施が困難である。

本研究では、NPTS とは異なる調査方法で大量のデータを短時間かつ低費用で収集するために、インターネットを利用した web 調査、携帯電話 (Mobile phone, 以下 MOBI) 情報を利用した MOBI 調査を、旅客純流動の代替調査として検討する。Web 調査と MOBI 調査は、旅客純流動を長期にわたって連続して調査できるという利点がある一方で、両調査に共通する問題点として、収集さ

れる旅客純流動の信頼性や、従来の調査との適合性が確認しにくいことがあげられる。本研究では Web 調査と MOBI 調査で収集された旅客純流動の特徴を明確にするために、NPTS, Web 調査, および MOBI 調査で収集した旅客純流動の特徴の比較を目的とする。このため、旅客純流動にもとづく O-D 行列に対して非負値行列因子分解 (Non-negative Matrix Factorization, 以下 NMF) を利用した分析を行う。

2. 既往研究

本研究では、異なる調査手段によって作成された旅客純流動の特徴を明確にするために、O-D 行列のパターンを抽出する行列分解法を用いる。確率的潜在意味解析法 (PLSA, Hofmann²⁾), 潜在ディリクレ配分法 (LDA, Blei ら³⁾) および、階層ディリクレ過程 (HDP, Teh ら⁴⁾) などの行列分解法は、文書から特徴的な単語群を見つけ出すために開発され、画像処理、顔認識、遺伝子解析などの様々な分野の研究に利用されるようになった。これらの方法は、確率的なデータ生成過程を明示したモデルによって行列分解を行う。一方で、行列の各要素を近似する典型的な方法に、非負値行列因子分解 (NMF) があげられる。

Choo⁵⁾ らは、視覚的解析の方法として LDA と比較して NMF が優れている点に着目して分析手順を提案した。Tang and Lewis⁶⁾ は、抽出されるパターンを視覚化する NMF の結果は、同じデータセットを用いた LDA の結果に匹敵することを示した。旅客の行動の分析に NMF を適用した例として山口ら⁷⁾ は、携帯電話位置情報データから、東京都居住者の首都圏外滞在確率を求め、NMF の適用を行った。基底数 k の決定に課題が残るものの、旅行先別の滞在確率の特徴として、「平常期」、「お盆・ゴールデンウィーク期」、「年末年始期」の 3 種類の大きく異なる旅行滞在先パターンを抽出し、首都圏外滞在の特徴を明らかにした。本研究では、NMF を用いて、異なる調査から得られる O-D 表をまとめた行列のパターンを抽出して、調査手法間の需要パターン異同を明らかにすることを目的とする。

3. 旅客純流動の調査方法

本研究では、旅客純流動を NPTS, Web 調査, MOBI 調査の 3 調査法から算出する。

NPTS は、1 章で述べたように、国土交通省が 5 年に 1 度行う全国規模の旅客純流動調査である。本研究では、NPTS の需要を基準として、Web 調査, MOBI 調査の結

表 1 調査の特徴

	NPTS	WEB	MOBI
対象者	旅行者 (旅行中)	インターネット利用者 (在宅)	携帯電話利用者 (移動中)
調査日	2010 秋 (平・休日 1 日ずつ)	2015/7, 10 2016/1, 4	2015/10 (平・休日 1 日 ずつ)
旅行頻度	観測されない	同じ場所に訪 れた回数	観測されない

果を比較する。

Web 調査は、マーケティングリサーチ会社である Intage Co. Ltd 社に依頼して、埼玉県、千葉県、東京都、神奈川県 (以下、南関東) に居住する 20 歳以上の男女を対象に、南関東を出発地とする旅行を収集した独自調査である。主な調査項目は、旅行に関する項目の他、性別、年齢、年間世帯所得、家族構成などの個人属性、加えて旅行回数である。1 度の web 調査で過去 3 ヶ月間の旅行を収集することとして、2015 年 8 月から 2016 年 4 月にかけて web 調査を 4 回繰り返して実施した。得られたデータのうち 2015 年 4 月から 6 月分を春、同 7 月から 9 月分を夏、同 10 月から 12 月分を秋、2016 年 1 月から 3 月分を冬の旅客純流動データと呼ぶ。

MOBI 調査は、2013 年に株式会社ドコモ (以下、ドコモ) が事業化した、人の滞在や移動を記録したモバイル空間統計というデータベースから作成した旅客純流動調査である。モバイル空間統計は、携帯電話が周期的に位置情報を把握されていることを利用し、ある時間に基地局エリア内に存在する携帯電話データを集計する。データベースの作成に当たっては、個人を特定できない処理が行われ、同じ携帯電話が移動または滞留している情報は直接入手できない。本研究では 2015 年 10 月の平・休日それぞれ 1 日分ずつのモバイル空間統計を用いた。また 3 調査の出発地、目的地は web 調査と同一の南関東、および全国とした。3 調査の特徴を表 1 にまとめる。

4. 非負値行列因子分解 (NMF)

m 行 \times n 列サイズの非負値行列を X とする。NMF は、行列 X を $m \times k$ の非負値行列 W と $k \times n$ の非負値行列 H の積の形に近似的に分解する (図 2)。図中の k は NMF の基底数である。Paatero ら⁸⁾ は、 X を近似する行列の積 $W \times H$ の最適解を導出するために、以下の最適化問題の極小値を推定した。

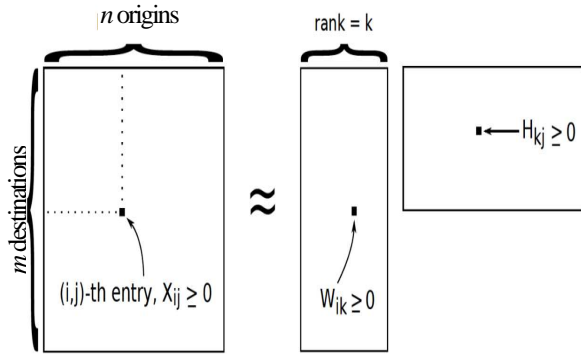


図2 非負値行列因子分解 (NMF) による行列分解

$$\min_{W, H \geq 0} [D(X, WH) + R(W, H)] \quad (1)$$

式 (1) で、D は損失関数、R は任意の正則関数である。近似精度は損失関数 D によって測定され、以下の 2 種類が広く用いられる。式 (2) は Euclid 距離に基づく距離の計算方法であり、 X_{ij} と $(W \times H)_{ij}$ の距離は、右辺から計算される左辺の期待値に対して対称の 2 乗和で定義される。式 (3) は一般化 Kullback-Leibler 距離であり、 X_{ij} と $(W \times H)_{ij}$ の距離は、期待値に対して、非対称である。

$$D(X, WH) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - (WH)_{ij})^2 \quad (2)$$

$$D(X, WH) = \sum_{i=1}^m \sum_{j=1}^n \left[X_{ij} \log \left(\frac{X_{ij}}{(WH)_{ij}} \right) - X_{ij} + (WH)_{ij} \right] \quad (3)$$

本研究では、式 (1) を極小化する行列 W, H を、式 (3) の目的関数の下で最小化する。以下の式 (4)、(5) で表される更新式に基づいて推定する。

$$H_{kj} \leftarrow H_{kj} \frac{\left(\sum_l \frac{W_{lk} X_{lj}}{(WH)_{lj}} \right)}{\sum_l W_{lk}} \quad (4)$$

$$W_{ik} \leftarrow W_{ik} \frac{\left(\sum_l \frac{H_{kl} X_{il}}{(WH)_{lj}} \right)}{\sum_l H_{kl}} \quad (5)$$

NMF では基底数 k を行列 W, H の推定に先立って与えなくてはならない。基底数 k の決定には、数種類の異なる k を設定して得られる結果から、最も有用な行列分解を与える k を選択する経験的手法が用いられる。k の評価尺度をいくつか示す。

i. 共役相関係数

k によって分類される集団の安定性の尺度として Brunet ら⁹⁾ が提案した。

ii. RSS

式 (6) で示される観測行列 V と推定行列 V' との残差平方和である。

$$RSS = \sum_{ij} (V_{ij} - V'_{ij})^2 \quad (6)$$

i. evar

式 (7) で示される。1 に近いほどよい。

$$evar = 1 - \frac{RSS}{\sum_{i,j} V_{ij}^2} \quad (7)$$

ii. 因子分解の希薄性

iii. silhouette 値

-1 から +1 の範囲をとり、値が大きいほどパターンわけが適切になされている。

以上の値に基づいて、k の最適値を選択するアプローチがいくつか提案されている。Brunet らは、共役相関係数が減少し始める k の最小値を選択することを提案した。一方、Hutchins ら¹⁰⁾ は RSS 曲線が変曲点を示す k の最小値を選択することを提案した。

本研究で NMF を行う対象とする行列は、旅客純流動調査から得られる O-D 行列を 3 種類の調査別の非負値行列 X として、行方向に m 個の目的地、列方向に n 個の出発地、要素として旅客純流動量をとる。基底数 k は旅客純流動のパターン数を表す。

NMF によって推定される、行列 W の第 k 列は、旅客純流動のパターン k の目的地選択の特徴を表す。一方で、行列 H の第 k 行は、旅客純流動のパターン k の出発地別の重みを表す。

5. NMF の結果

(1) データセット

本研究では、航空もしくは鉄道を主に利用する旅行について分析を行う。MOBI 調査では利用交通機関に関するデータが得られないので、MOBI 調査の交通機関シェアを NPTS の交通機関シェアと同じと仮定する。その結果、航空を主に利用する旅客純流動は、NPTS で約 62 万回、Web 調査で約 14 万回、MOBI 調査で約 58 万回記録され、鉄道を主に利用する旅行は、NPTS では約 120 万回、Web 調査では約 16 万回、MOBI 調査では約 140 万回記録されていた。

3 調査 : NPTS, Web, MOBI それぞれから得られる旅客純流動について、行方向に共通の目的地 D を、列方向に 3 調査の出発地 O をそれぞれ並べて、データセットを作成した。なお O, D の集計単位は都道府県である。O-D 行列のサイズを合わせるために、出発地 O は Web 調査で対象とした南関東に限定した。また目的地 D から NPTS で対象とされない南関東と鉄道による移動ができない沖縄県を除いた。また目的地 D のうち北海道を道北、道東、道央、道南の 4 つに分割する。以上から目的地を 45 府県地域、出発地を 4 都県とする 3 つの O-D

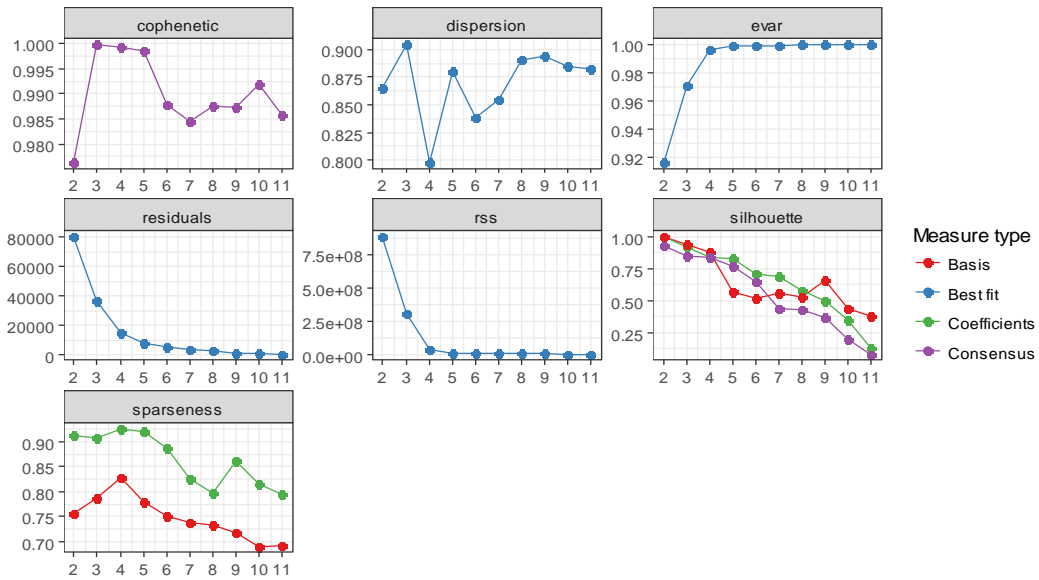


図 3 基底数 k における、航空旅行の NMF 分析が 100 回計算された尺度

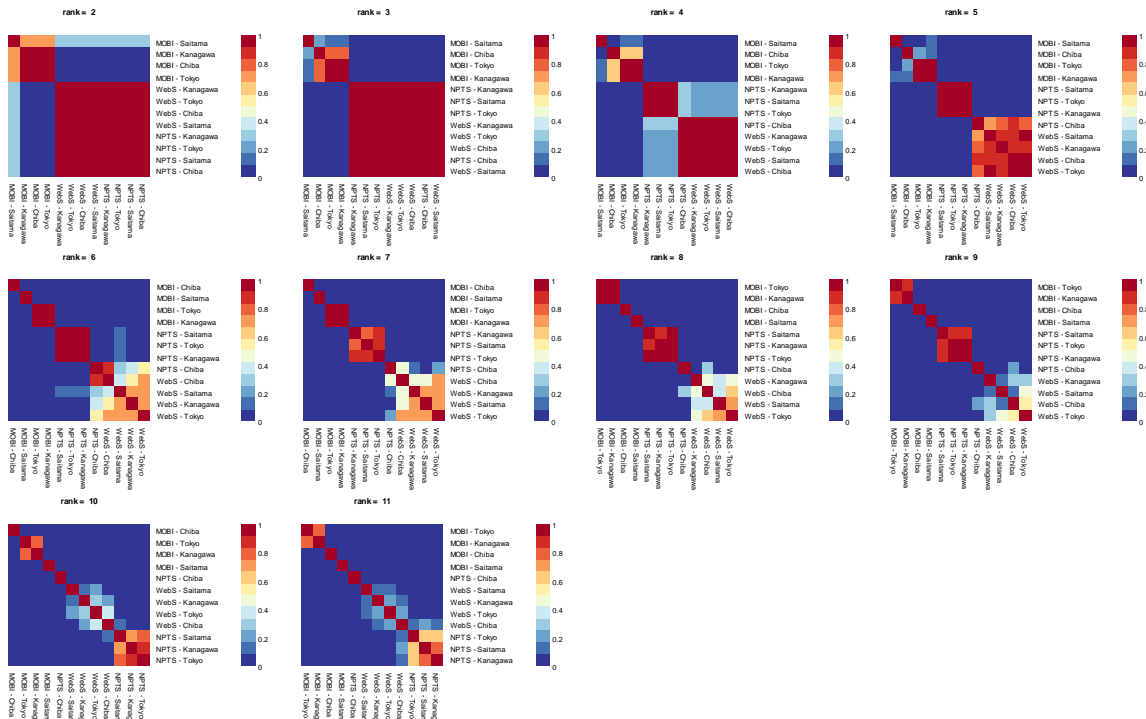


図 4 基底数 k における、集団分類後の一致度

行列を列側に並べて、結合した非負値行列を NMF への入力 X とする。

NMF では、カウント数が大きい要素で大きな誤差を生じないように、その要素に対する当てはまりを良くするくせがある。しかし先述のように、3 調査の観測できた数には大きな違いがあるため、トリップ数そのものを入力すると、最も規模の小さな web 調査の結果がパターン抽出に反映されるおそれがある。そこで、Web 調査の O-D 行列、MOBI 調査の O-D 行列の要素総数を基準とした NPTS の O-D 行列の要素総数に合わせる。このため、以下の式で定義する重みを与える。

$$W_{NPTS-Web} = \frac{N_i^{NPTS}}{N_i^{Web}} \quad (8)$$

$$W_{NPTS-MOBI} = \frac{N_i^{NPTS}}{N_i^{MOBI}} \quad (9)$$

ここで、 $W_{NPTS-Web}$ 、 $W_{NPTS-MOBI}$ はそれぞれ Web 行列、MOBI 行列の各要素にそれぞれ与えられる重みであり、 N_i^{NPTS} 、 N_i^{Web} 、 N_i^{MOBI} はそれぞれ NPTS 行列、Web 行列、MOBI 行列の要素総数である。

(2) 航空利用者の純流動パターン

図 3 に航空を利用する旅行データから作成した行列 X

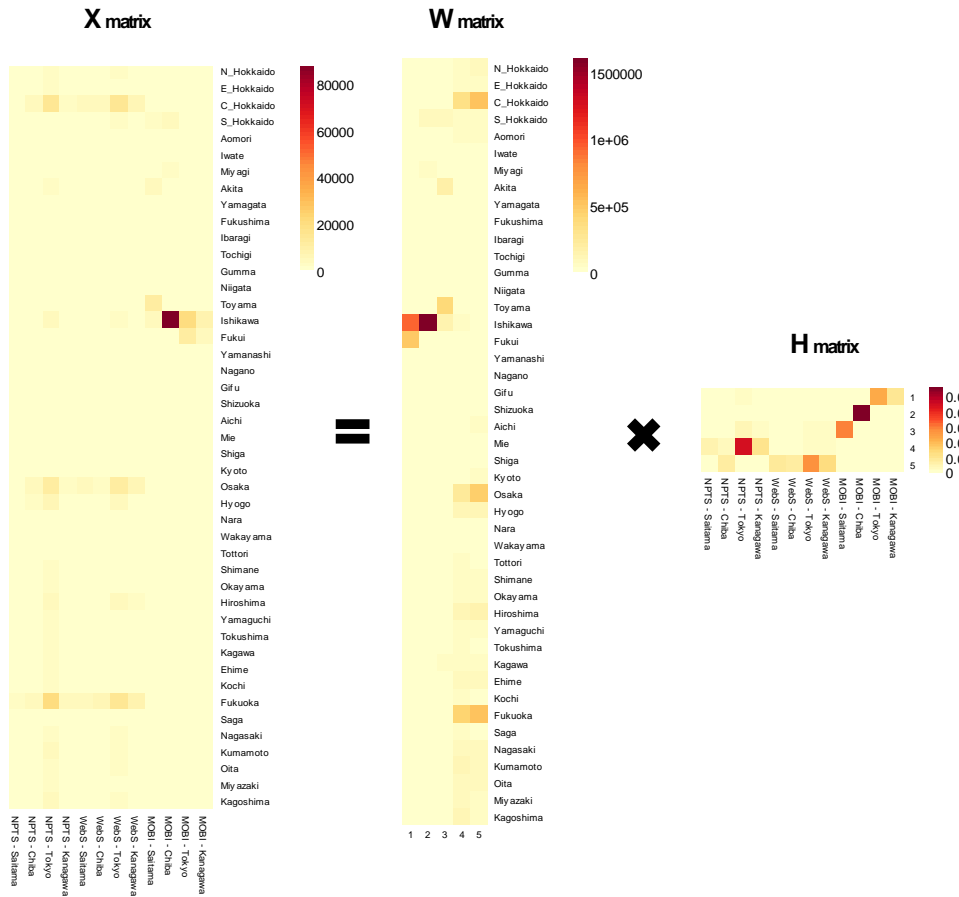


図5 基底数（パターン）5における，航空旅行のNMF

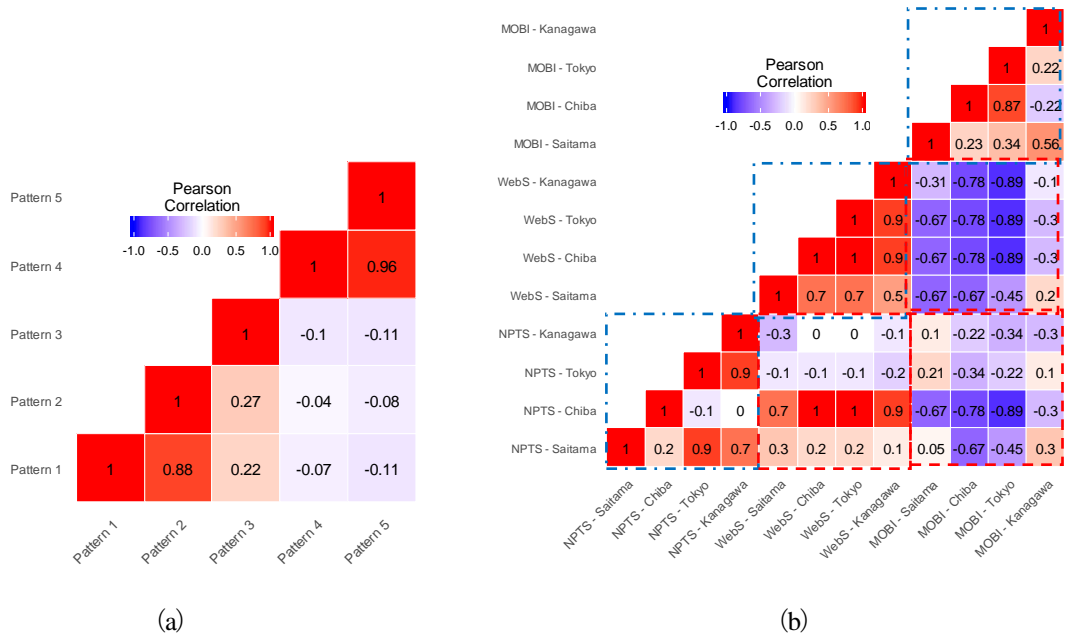


図6 目的地分布 (a) と重み (b) のピアソン相関試験結果

の NMF 分析の基底数 k を定める指標を示す。図 3 の共役相関係数 (cophenetic) から $k=5$ および 10 が候補となる。さらに sparseness と silhouette 値を参照して $k=5$ を選択する。図 4 に示すように、 $k=5$ のときの分類は最良であり以下の分析では $k=5$ とした。

図 5 の行列 H から、Web と NPTS はそれぞれパターン

4, 5 の特徴を持ち、MOBI はパターン 1, 2, 3 の特徴を持つ。行列 W から、パターン 4, 5 は、北海道、大阪、福岡への流動が大きい。しかし、パターン 1-3 は石川、富山、福井への流動が比較的大きい。これは、MOBI 調査のデータ収集時にこれらの地域でのイベントが開催されていた可能性があるためと考えられる。

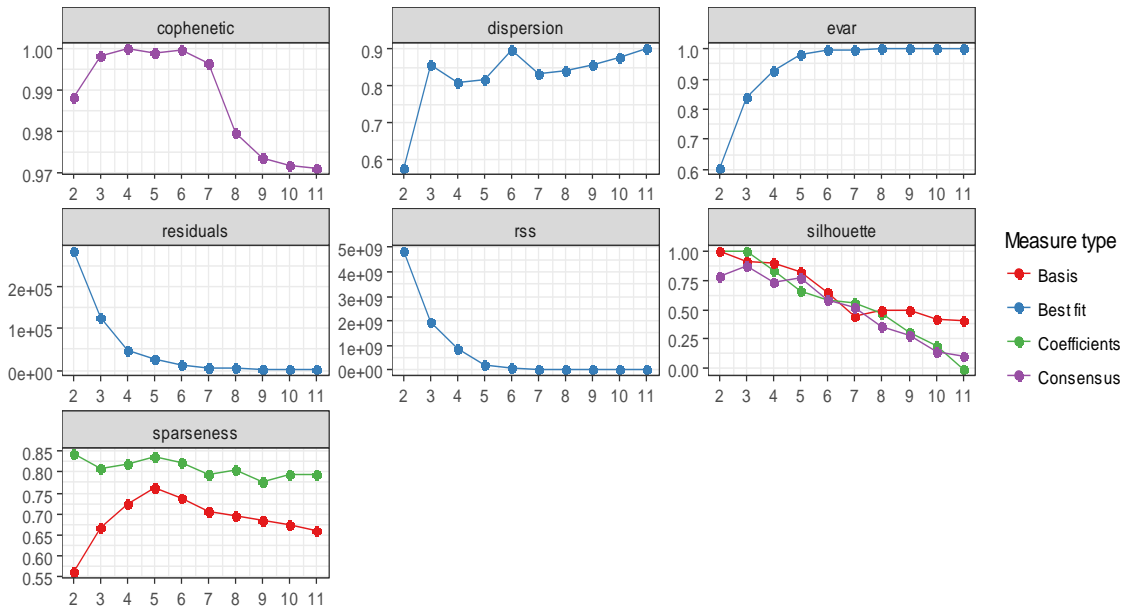


図 7 基底数 k における、鉄道旅行の NMF 分析が 100 回計算された尺度

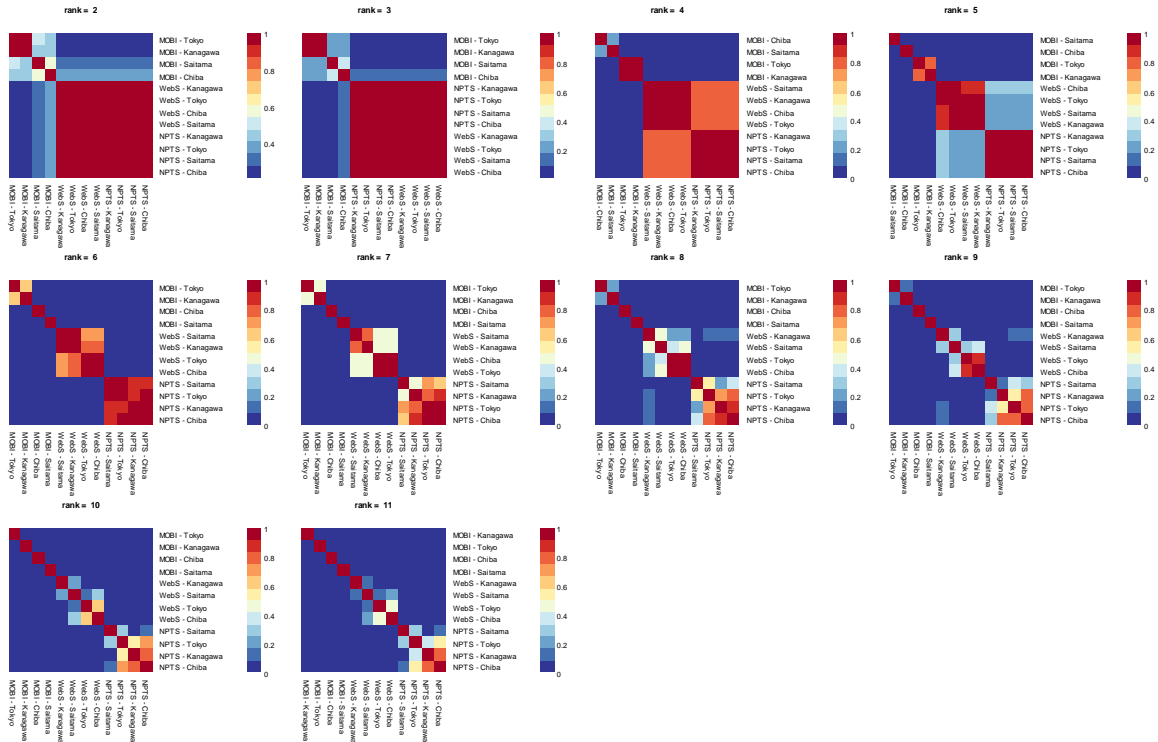


図 8 基底数 k における、集団分類後の一致度

図 6 にピアソン相関による、純流動パターンの目的地分布の類似性 (W 列方向成分: a) と純流動パターンへの出発地重みの類似性 (H 列方向成分: b) をそれぞれ示す。図 6 は、NPTS と Web 調査には類似性がみられるものの、MOBI 調査は他の調査との間に類似性が乏しい。

(3) 鉄道利用者の純流動パターン

図 7 に鉄道を利用する旅行データから作成した行列 X の NMF 分析の基底数 k を定める尺度を示す。図 7 の共役相関係数 (cophenetic) から k=6 および 7 が候補となる。さらに sparseness と silhouette 値を参照して k=6 を選択す

る。さらに、図 8 に示すように、k=6 のときの分類が最良であり以下の分析では k=6 とした、図 8 に結果を示す。

図 8 の行列 H から、Web はパターン 2, 4 を持ち、NPTS はパターン 2 を、MOBI はパターン 1, 3, 5, 6 の特徴を、それぞれ持つ。行列 W から、パターン 2, 4 は大阪、静岡、愛知への流動が比較的多い。一方で、パターン 1, 3, 5, 6 は大阪、静岡、愛知への流動が少なく、新潟、富山、石川、福井など北陸地方への流動が比較的多いという特徴を持つ。この傾向は、航空の場合にも現れた傾向である。図 9 にピアソン相関の計算結果を示す。

図 9 から、鉄道では航空よりも、MOBI 調査と他の調

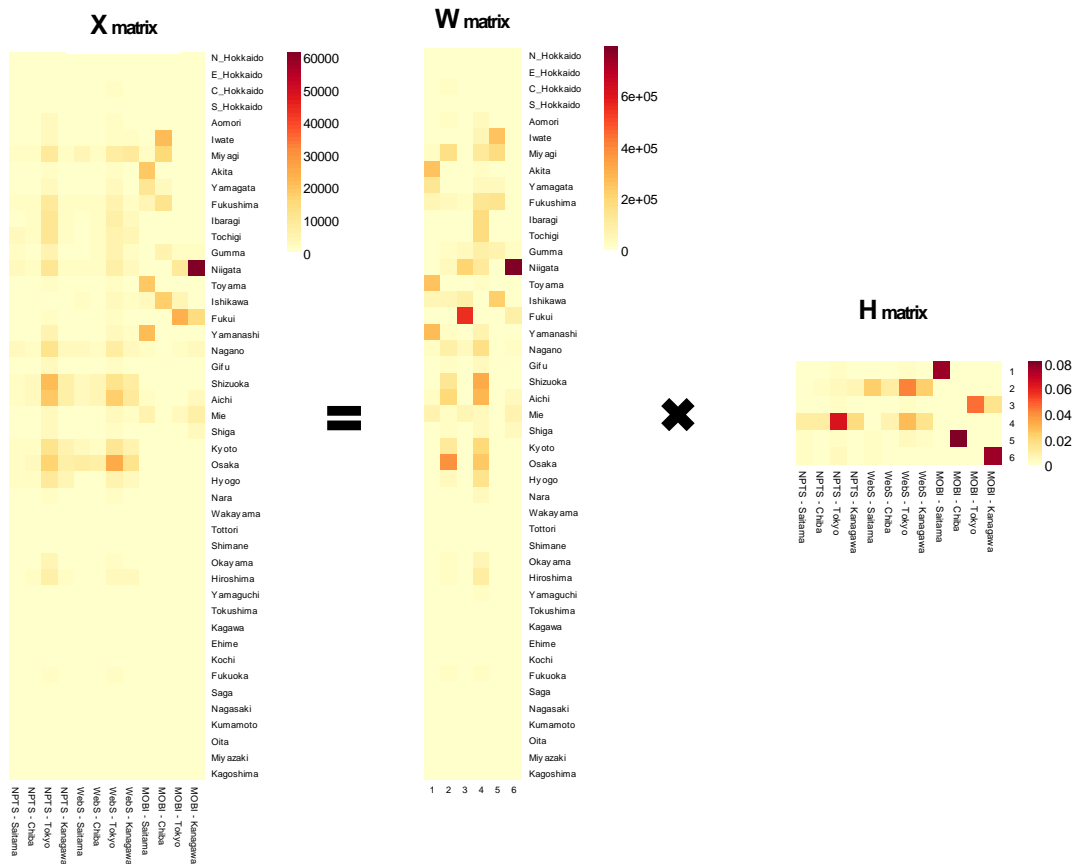


図 9 基底数（パターン）6における，鉄道旅行の NMF

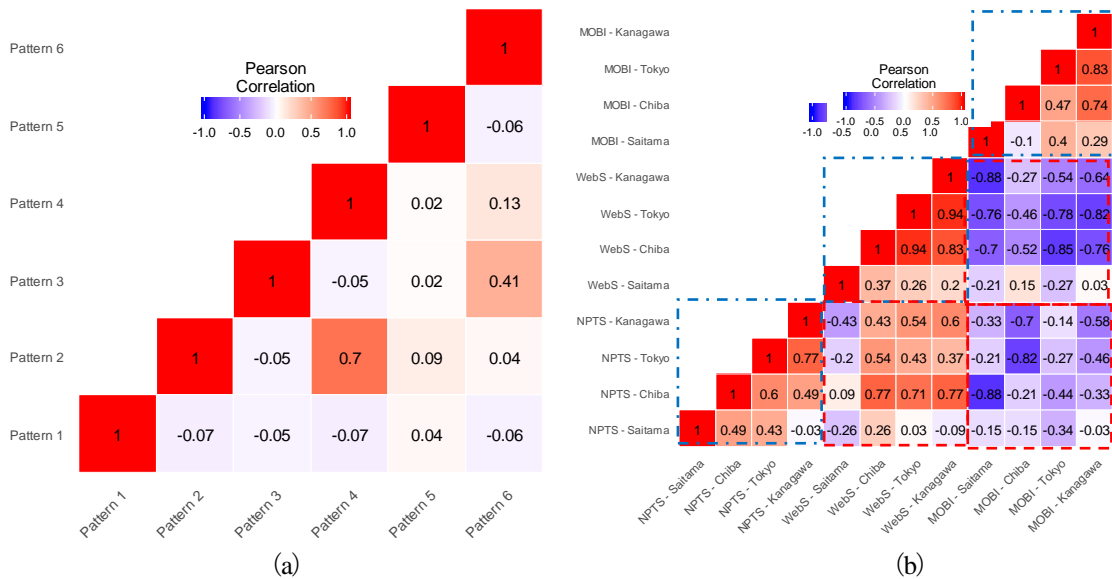


図 10 目的地分布 (a) と重み (b) のピアソン相関試験結果

査に共通するパターンが多いことが明らかとなった。

6. まとめ

本研究では，Web 調査，MOBI 調査で収集された旅客純流動の特徴を NPTS の結果との比較に基づいて明らか

にするために，航空と鉄道を利用する純流動について NMF を適用した。

分析の結果，MOBI 調査で収集される旅客純流動の特徴は，NPTS および Web 調査で収集された旅客純流動の特徴と有意に異なった。さらに，MOBI 調査の旅客純流動は，出発地が異なる場合に目的地の分布も異なるという特徴が明らかとなった。NPTS と Web 調査の旅客純流

動には、航空では千葉県を出発地とする純流動に、鉄道では千葉県もしくは神奈川県を出発地とする純流動の類似性が高いことがあきらかとなった。

なお本研究から得られた旅客純流動の特徴は、3 調査それぞれの回答者の個人属性の違いに対処していない。今後は個人属性の違いを除く方法として、傾向スコア法の導入を検討する。

参考文献

- 1) 奥村誠：全国幹線旅客純流動調査の意義と課題，運輸と経済，第 72 巻，pp.21-30，2012
- 2) Hofmann, T. (1999) . Probabilistic latent semantic analysis, Paper presented at the 15th Conference on Uncertainty in Artificial Intelligence. Stockholm, Sweden, 1999: Morgan Kaufmann Publishers Inc.
- 3) Blei, D.M., Ng, A.Y., Jordan, M.I. (2003) . Latent dirichlet allocation. Journal of machine Learning research, 3 (Jan) , 993-1022.
- 4) Teh, Y.W., Jordan, M.I., Beal, M.J., Blei, D.M. (2005) . Sharing clusters among related groups: Hierarchical Dirichlet processes, Paper presented at the Advances in Neural Information Processing Systems 17 (NIPS 2004) . Vancouver, Canada, 2005.
- 5) Choo, J., Lee, C., Reddy, C.K., Park, H. (2013) . Utopian: User-driven topic modeling based on interactive nonnegative matrix factorization. IEEE transactions on visualization and computer graphics, 19 (12) , 1992-2001.
- 6) Tang, J., Lewis, P.H. (2008) . Non-negative matrix factorisation for object class discovery and image auto-annotation, Paper presented at the International Conference on Content-based Image and Video Retrieval. Niagara Falls, Canada, 2008.
- 7) 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
- 8) Paatero, P., Tapper, U. (1994) . Positive matrix factorization: A non - negative factor model with optimal utilization of error estimates of data values. Environmetrics, 5 (2) , 111-126.
- 9) Brunet, J.-P., Tamayo, P., Golub, T.R., Mesirov, J.P. (2004) . Metagenes and molecular pattern discovery using matrix factorization. Proceedings of the national academy of sciences, 101 (12) , 4164-4169.
- 10) Hutchins, L.N., Murphy, S.M., Singh, P., Graber, J.H. (2008) . Supplement Text and Figures: Position-Dependent Motif Characterization Using Nonnegative Matrix Factorization.

ANALYSIS ON INTER-REGIONAL PASSENGER DEMAND CHARACTERISTICS ON THE BASIS OF NON-NEGATIVE MATRIX DECOMPOSITION Yushi SUGAHARA, Canh Xuan DO and Makoto TSUKAI

In Japan, the Ministry of Land, Infrastructure, Transport and Tourism has conducted the Net Passenger Transportation Survey (NPTS) since 1990 in every five year, which gives valuable information about inter-regional passenger flows. However, NPTS has been criticized in several limitations such as cost, long time to publish and no observation for seasonal demand change. Therefore, this study conducts new survey using web survey and mobile phone data. This study purposes is to compare NPTS and new surveys, in railway and airplane travel, based on non-negative matrix factorization (NMF) . NMF was useful to clarify the characteristics of inter-regional passenger flows.