非負値行列因子分解による AI 便乗 OD データの分析

Analysis of OD Data of Ondemand Transportation by Non-negative Matrix Factorization

野田五十樹 [1,2*]　　落合純一 [2]
Itsuki Noda[1,2]　　Junichi Ochiai[2]

[1] 産業技術総合研究所 人工知能研究センター
[1] Artificial Intelligence Research Center, AIST
[2] (株) 未来シェア
[2] Mirai Share CO. Ltd.

Abstract: We propose a method to analyze OD data of on-demand transportation services like Smart Access Vehicle Service (SAVS) and AI-bus by non-negative matrix factorization (NMF), which enables to visualize and investigate features of date and OD places for design of future transportation services. OD data represents properties of usage of transportation services, but is composed by complicated and mixtured way with many noise and uncertainly. NMF can decompose OD data into few bases, by which we can get abstracted properties of usage of transportation services. The abstracted properties provides a way to understand the difference among dates or places and to enable to plan forfuture operation of SAVS and AI bus.

1　はじめに

人々の移動を集積的に表している Origin-Destination (OD) データは様々な有用な情報を含んでいると考えられ、その情報の要素を効果的に取り出す方法の確立は重要である。人々の移動には各々なんらかの目的 (例えば通勤や買い物、医療など) が付随していることから、OD データは、単に交通機関の利用の統計データだけではなく、移動の目的である人々の活動の集積の一側面を表しているともみなすことができる。よって、OD データから移動目的に関する統計的情報を取り出すことができれば、交通機関の運行の最適化だけでなく、移動の本来の目的に対応するサービスの設計・効率化にも活用できると考えられる。

OD データは、近年、様々な形で集積されつつある。タクシーについては、従来より (精度の問題はあるにせよ) 日誌の形で乗降車地点は記録されており、近年ではその電子化も進んでいる。バス・鉄道についても、交通系カードのデータが利用可能になれば、広域の OD データとして活用できる。筆者らが関与している Smart Access Vehicle Service (SAVS) [1, 5, 3, 4] や AI 運行バスにおいても、OD データを確実に取得できており、それによるサービス改善の方法が望まれている。

一方、OD データはノイズを多く含むデータであり、かつ、多様な利用形態の混合データであるため、解析のためのモデル設計を適切に行う必要がある。そこで本稿では、人々の移動には基底となるいくつかの統計的パターンがあり、その基底パターンの混合として全体の OD が形成されていると仮定する。その上で、集計された OD データから基底パターンを抽出する方法を非負値行列因子分解の手法を用いて導出する。

本稿のこの後の構成は、2 節において OD データの構造の統計モデルを与え、3 節において、実データからその統計モデルを推定する方法を導き出す。この方法の有効性を評価した実験を 4 節で示し、最後に 5 節で評価のまとめと今後の課題について考察する。

2　混合直積分布による OD モデル

2.1　OD の基底パターン

まず、人々の移動の基底パターンとして、出発地 (origin) 確率ベクトルと目的地 (destination) 確率ベクトルの直積で OD が表されると仮定する。すなわち、ある基底パターン k の出発地確率ベクトル $\boldsymbol{u}_k$ と目的地確率ベクトル $\boldsymbol{v}_k$ が以下のように与えられたとする。

$$\boldsymbol{u}_k \;=\; \left[\; u_{ki} \;\right]$$

*連絡先：産業技術総合研究所 人工知能研究センター
〒 305-8560 つくば市梅園１－１－１ 中央第 1
i.noda@aist.go.jp

$$
\text{where} \quad \sum_i u_{ki} = 1
$$

$$
\begin{aligned}
\boldsymbol{v}_k &= \left[\ v_{kj}\ \right] \\
&\text{where} \quad \sum_j v_{kj} = 1
\end{aligned}
$$

ただし、$\boldsymbol{u}_k$ の次元数は異なる出発地の数 N、$\boldsymbol{v}_k$ の次元数は異なる目的地の数 M とする。この時、基底パターン k の OD 確率行列 $\boldsymbol{W}_k$ は $\boldsymbol{u}_k$ と $\boldsymbol{v}_k$ の直積で表される。

$$
\begin{aligned}
\boldsymbol{W}_k = \left[\ w_{kij}\ \right] &= \boldsymbol{u}_k \times \boldsymbol{v}_k \\
w_{kij} &= u_{ki} v_{kj}
\end{aligned}
$$

この基底パターンの OD の解釈は以下のとおりである。$\boldsymbol{u}_k$ は出発地を選択する確率を表し、$\boldsymbol{v}_k$ は目的地を選択する確率を表す。そして、基底パターン k では、$\boldsymbol{u}_k$ および $\boldsymbol{v}_k$ により独立して選択された出発地および目的地の組で OD が構成される。例えば、基底パターン k が買い物目的の移動を表しているすると、$\boldsymbol{u}_k$ は買い物客の居住地の分布、$\boldsymbol{v}_k$ は買い物先の分布で、買い物目的の移動需要は、その直積で表されると考える。

2.2 OD の混合パターン

全体の OD は、K 種類の基底パターン $\boldsymbol{W}_k$ の混合で表されるとする。即ち、各パターンの重みを r_k とした時、全体の OD 分布 $\boldsymbol{\mathcal{W}}$ は、

$$
\begin{aligned}
\boldsymbol{\mathcal{W}} = \left[\ w_{ij}\ \right] &= \sum_k r_k \boldsymbol{W}_k \qquad (1) \\
w_{ij} &= \sum_k r_k w_{kij} = \sum_k r_k u_{ki} v_{kj}
\end{aligned}
$$

ただし、r_k の総和は 1 である。

$$
\sum_k r_k = 1
$$

つまり、全体の OD は、目的別など異なる OD 基底パターンが一定の割合で混合して形成されるとみなす。

3 基底パターンの最尤推定

本節では、OD の実データから各基底パターン及びその混合割合を最尤推定で求める方法を導出する。

まず、OD の実データとして、出発地 i から目的地 j へ移動した頻度 (回数)$\acute{c}_{ij}$ が各 i, j の組に与えられているとする。このデータから、以下のように実データの OD 確率行列 $\acute{\boldsymbol{W}}$ を求めておく。(以下、実データに基づく値には $\acute{\bigcirc}$ を、推定して求める値には $\hat{\bigcirc}$ をアクセントとして着けて区別する。)

$$
\begin{aligned}
\acute{\boldsymbol{W}} = \left[\ \acute{w}_{ij}\ \right] &= \frac{1}{\acute{C}} \left[\ \acute{c}_{ij}\ \right] \qquad (2) \\
\acute{C} &= \sum_{ij} \acute{c}_{ij}
\end{aligned}
$$

また、最尤推定のために、以下のような対数尤度 $\log \mathcal{L}$ を定義しておく。

$$
\log \mathcal{L} = \log \mathcal{L}\left(\hat{\boldsymbol{\mathcal{W}}} | \acute{\boldsymbol{W}}\right) = \sum_{ij} \acute{w}_{ij} \log \hat{w}_{ij} \qquad (3)
$$

さらに、以下のような負担率 γ_{kij} を定義しておく。

$$
\gamma_{kij} = \frac{\hat{r}_k \hat{w}_{kij}}{\hat{w}_{ij}} \qquad (4)
$$

3.1 混合分布の重みの最尤推定

対数尤度 $\log \mathcal{L}$ を最大化する重みを求めるため、ラグランジェ乗数 λ_r を導入して、以下の評価関数 E を考える。

$$
E = \log \mathcal{L} + \lambda_r \left(1 - \sum_k \hat{r}_k\right)
$$

E を、$\hat{r}_k$ および λ_r で偏微分して各々 0 と等しいすれば、以下が求まる。

$$
\hat{r}_k = \sum_{ij} \acute{w}_{ij} \gamma_{kij} \qquad (5)
$$

3.2 基底パターンの基底ベクトルの最尤推定

対数尤度 $\log \mathcal{L}$ を最大化する出発地・目的地ベクトルを求めるため、ラグランジェ乗数 λ_{uk}、 λ_{vk} を導入して、以下の評価関数 F を考える。

$$
\begin{aligned}
F = \ & \log \mathcal{L} + \\
& \sum_k \left[\lambda_{uk}\left(1 - \sum_i \hat{u}_{ki}\right) + \lambda_{vk}\left(1 - \sum_j \hat{v}_{kj}\right)\right]
\end{aligned}
$$

F を、$\hat{u}_{ki}, \hat{v}_{kj}$ および $\lambda_{uk}, \lambda_{vk}$ で偏微分して各々 0 と等しいすれば、以下を導出できる。

$$
\begin{aligned}
\lambda_{uk} = \lambda_{vk} &= \hat{r}_k \\
\hat{u}_{ki} &= \frac{1}{\lambda_{uk}} \sum_j \acute{w}_{ij} \gamma_{kij} \qquad (6) \\
\hat{v}_{kj} &= \frac{1}{\lambda_{vk}} \sum_i \acute{w}_{ij} \gamma_{kij} \qquad (7)
\end{aligned}
$$

Procedure ODD-NMF

given: $\acute{c}_{ij}$, K
(2) 式 により、$\acute{\boldsymbol{W}}$ を求める。
基底ベクトル $\hat{\boldsymbol{u}}_k$、$\hat{\boldsymbol{v}}_k$, $\hat{r}_k$ を初期化 (initFactors)。
repeat
  (4) 式により負担率 γ_{kij} を求める。
  (5) 式により $\hat{r}_k$ を更新。
  (6) 式、(7) 式により $\hat{\boldsymbol{u}}_k$、$\hat{\boldsymbol{v}}_k$ を更新
until ($\hat{\boldsymbol{u}}_k$、$\hat{\boldsymbol{v}}_k$, $\hat{r}_k$ が収束)

図 1: 非負行列因子近似による OD 基底分解 (ODD-NMF)

3.3 非負値行列因子分解としての最尤推定

上記で導出した最尤推定は、以下のように、非負値行列因子分解 (Non-negative Matrix Factorization) [2] とみなすことができる。

各要素が非負であり、各行の総和が 1 である 2 つの行列、$\boldsymbol{U} = \left[\ u_{ki}\ \right]$、$\boldsymbol{V} = \left[\ v_{kj}\ \right]$ を考える。ただし、$\boldsymbol{U}$ は K 行 N 列、$\boldsymbol{V}$ は K 行 M 列とする。また、対角要素がすべて非負で、総和が 1 となる K 行 K 列の対角行列 $\boldsymbol{R} = \mathrm{diag}[r_k]$ を考える。この時、OD 混合パターン $\boldsymbol{\mathcal{W}}$ は、

$$\boldsymbol{\mathcal{W}} \quad = \quad {}^{\mathrm{T}}\boldsymbol{URV}$$

となる。よって、実データ $\acute{\boldsymbol{W}}$ の最尤推定 $\hat{\boldsymbol{\mathcal{W}}}$ を求めることは、$\acute{\boldsymbol{W}}$ を最尤の基準で ${}^{\mathrm{T}}\hat{\boldsymbol{U}}, \hat{\boldsymbol{R}}, \hat{\boldsymbol{V}}$ の 3 つの非負値行列に近似分解することに対応する。

3.4 ODD-NMF アルゴリズム

以上で導出してきた OD の基底分解をアルゴリズムの形にまとめると、図 1 となる。このアルゴリズムを、非負行列因子近似による OD 基底分解 (OD Decomposition by Non-negative Matrix Factorization、ODD-NMF) と呼ぶ。

このアルゴリズムのうち、初期化 (initFactors) については、以下のように行う。本アルゴリズムは漸進的に最尤推定を行うため、収束は初期値に依存する。この収束を安定化させるため、0 番目の基底には元データの平均を用いた値を与え、残りの基底には、その平均からの残渣を順に解消するような初期値を与えるものとする。具体的には図 2 に示す手順で初期化を行う。

また、各基底を更新するループの終了条件については、以下の実験では簡単のため、十分に収束したとみなせる 100 回に繰り返しを固定している。

Procedure initFactors

given: $\acute{\boldsymbol{W}}$, K
$\hat{u}_{0i} = \sum_j \acute{w}_{ij}$、$\hat{v}_{0j} = \sum_i \acute{w}_{ij}$ を求める。
$\hat{\boldsymbol{u}}_0 = \left[\ \hat{u}_{0i}\ \right]$、$\hat{\boldsymbol{v}}_0 = \left[\ \hat{v}_{0j}\ \right]$ とする。
for $k = 1$ to $K - 1$ **do**
  残渣 $\acute{\boldsymbol{W}} - \hat{\boldsymbol{u}}_0 \times \hat{\boldsymbol{v}}_0$ のうち、k 番目に大きな値を持つ要素の index を (i', j') とする。
  $\hat{u}_{ki} = \begin{cases} 1.0 & ;\text{where } i = i' \\ (0, 0.1) \text{ 区間の一様乱数} & ;\text{otherwise} \end{cases}$
  $\hat{v}_{kj} = \begin{cases} 1.0 & ;\text{where } j = j' \\ (0, 0.1) \text{ 区間の一様乱数} & ;\text{otherwise} \end{cases}$
  $\hat{\boldsymbol{u}}_k = \left[\ \hat{u}_{ki}\ \right]$、$\hat{\boldsymbol{v}}_k = \left[\ \hat{v}_{kj}\ \right]$ とし、各々総和が 1 となるように正規化。
end for
$\forall k : \hat{r}_k = 1/K$

図 2: 基底パターンの初期化

3.5 基底数 K の決定

ODD-NMF では基底数 K は与えられることを仮定している。これについて、実際の OD データに前提知識として K が仮定できる場合はそれを用いることに成る。次節の実験 1 ではそれを仮定した評価を行っている。

前提知識が無い場合には何らかの情報量基準、例えばベイズ情報量基準 (BIC) などにより K を決める必要がある。実験 2 においてはそのような方法で K を求めている。幸い、本手法の計算量は乗降地点の数の自乗程度のオーダであり、多数の K で BIC を求めるのはそれほど困難ではない。

4 実験

4.1 実験 1:トイプロブレム

3 節 で導出した ODD-NMF の動作を確認するために、以下のような小規模な実験を行った。

まず、推定のターゲットとなる OD データモデルとして、次のようなものを考える。

$$\acute{\boldsymbol{W}} \quad = \quad (1-\rho)(\acute{r}_0 \acute{\boldsymbol{W}}_0 + \acute{r}_1 \acute{\boldsymbol{W}}_1) + \rho\boldsymbol{\epsilon} \qquad (8)$$

where

$$\begin{aligned} \acute{r}_0 &= \acute{r}_1 = 0.5 \\ \acute{\boldsymbol{W}}_0 &= \acute{\boldsymbol{u}}_0 \times \acute{\boldsymbol{v}}_0 \\ \acute{\boldsymbol{W}}_1 &= \acute{\boldsymbol{u}}_1 \times \acute{\boldsymbol{v}}_1 \\ \acute{\boldsymbol{u}}_0 &= {}^{\mathrm{T}}[.5, .5, .0, .0, .0] \end{aligned}$$

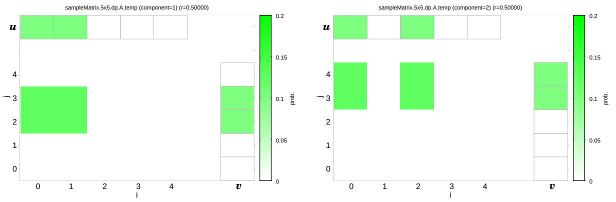


図 3: ターゲットの OD 分布の構成基底 (左:$\acute{\boldsymbol{W}}_0 = \acute{\boldsymbol{u}}_0 \times \acute{\boldsymbol{v}}_0$、右:$\acute{\boldsymbol{W}}_1 = \acute{\boldsymbol{u}}_1 \times \acute{\boldsymbol{v}}_1$)

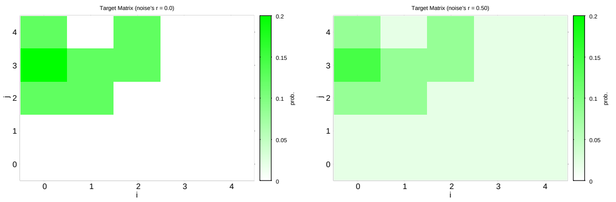


図 4: ターゲットの OD 分布 (左:ノイズレベル $\rho = 0.0$、右:$\rho = 0.5$)

$$\begin{aligned} \acute{\boldsymbol{v}}_0 &= {}^{\mathrm{T}}[.0, .0, .5, .5, .0] \\ \acute{\boldsymbol{u}}_1 &= {}^{\mathrm{T}}[.5, .0, .5, .0, .0] \\ \acute{\boldsymbol{v}}_1 &= {}^{\mathrm{T}}[.0, .0, .0, .5, .5] \end{aligned}$$

ここで、$\boldsymbol{\epsilon}$ はノイズ成分で、一様分布 (全ての OD の組合せが均等に生じる) としておく。また、$\rho \in [0, 1]$ はノイズレベルである。

(8) 式によるモデルは、図 3 に示すような 2 つの基底を均等に混合し、それにノイズを重畳したものになっている。図 4 にそのモデルの確率行列を図にしたものを示す。これらの図の左右の各グラフのうち、右端の黒線で囲った縦長帯はカラーバー、その左に OD 確率行列を出発地確率ベクトル $\boldsymbol{u}$ と目的地確率ベクトル $\boldsymbol{v}$ とともに示している。上部に灰色で囲ったものが $\boldsymbol{u}$、右部に灰色で囲ったものが $\boldsymbol{v}$ である。左下の残りの部分が行列要素 (5×5) である。各要素は、行列・ベクトルの値により白 (値が 0) から緑 (値が 0.2 以上) で色付けされている。以降の同様の図に於いても、ベクトル $\boldsymbol{u}$、$\boldsymbol{v}$ の表示が無い場合を含めて、同じ形式で示している。

実際に ODD-NMF に与えるデータのモデルとしては、図 4 のようなものとなる。このうち、左側がノイズのない ($\rho = 0$)2 基底だけの混合であり、右側は半分 ($\rho = 0.5$) ノイズを重畳したものである（行列の右や下の部分に薄っすらと色がついている)。

なお、この実験では簡単のため、(8) 式の OD データモデルからのサンプリングデータ $\acute{c}_{ij}$ を取るのではなく、(8) 式そのものをサンプリングから求めた実データの確率行列とみなして、アルゴリズムの評価を行っている。

まず、$K = 2$ として ODD-NMF により基底分解を行った結果を図 5 に示す。この図は、ノイズレベル ρ の値を 0.0, 0.5, 0.75, 0.9 とした時について、基底分解を行った結果の混合確率行列 $\hat{\mathcal{W}}$ の値を示している。これを見るとわかるように、ノイズレベルが 0.5 までは、図 4 に示したもとのデータの特徴 ((2,2) 要素や (1,4) 要素の部分が欠けた形になっている) をうまく復元できているのに対し、ノイズレベル 0.75 以上ではその特徴が潰れてしまっている。

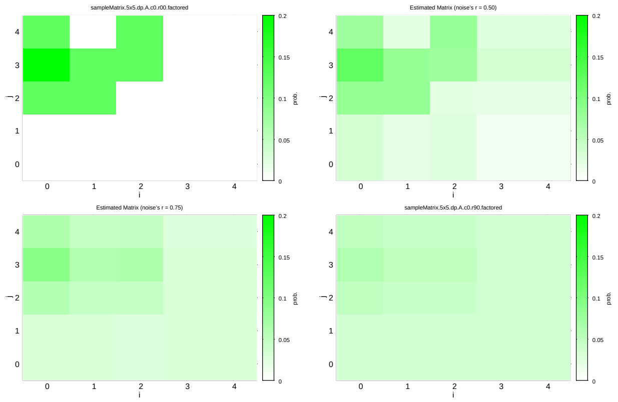


図 5: ODD-NMF で推定された OD 分布 (左上:ノイズレベル $\rho = 0.0$、右上:$\rho = 0.5$、左下:$\rho = 0.75$、右下:$\rho = 0.9$)

獲得された基底の詳細を見てみると図 6 のようになる。この図の最上段 ($\rho = 0.0$) の左右 2 つを見ると、図 3 に示したもとの基底とほぼ完全に一致していることがわかる。2 段目 ($\rho = 0.5$) においても、一様分布するノイズ成分を取り込みながらも、図 3 に近い基底を獲得できていることがわかる。一方、3 段目以下 ($\rho = 0.75, 0.9$) では第 1 基底がノイズ成分を表しており、第 2 基底のみで図 3 の特徴を表そうとしていることがわかる。このため、(2,2) 要素や (1,4) 要素がないといった特徴が潰れてしまう結果になっている。

続いて $K = 3$ とした場合の基底分析結果を図 7 に示す。このうち、最上段のノイズなしの場合は、第 2 基底が $\acute{\boldsymbol{W}}_0$ に対応していることはわかるものの、$\acute{\boldsymbol{W}}_1$ の成分が第 1 基底 (左) と第 3 基底 (右) がに分割されて表現されてしまっていることがわかる。つまり、基底の数がもとのデータのものより多いため、獲得された表現が冗長になっている。一方、$\rho = 0.5$ の場合は、ノイズが重畳しているものの、第 1 成分が $\acute{\boldsymbol{W}}_0$、第 2 成分が $\acute{\boldsymbol{W}}_1$ という対応を取りながら基底分解できている。$\rho = 0.75, 0.9$ の場合においても、第 2 成分が $\acute{\boldsymbol{W}}_1$、第 3 成分が $\acute{\boldsymbol{W}}_0$ という対応が見て取れる。

以上をまとめると、ODD-NMF は、K が妥当な値になっている場合には適切な基底分解が可能であり、ノイズに対してもある程度頑健 (本実験では $\rho = 0.5$ 程度) であることが示された。

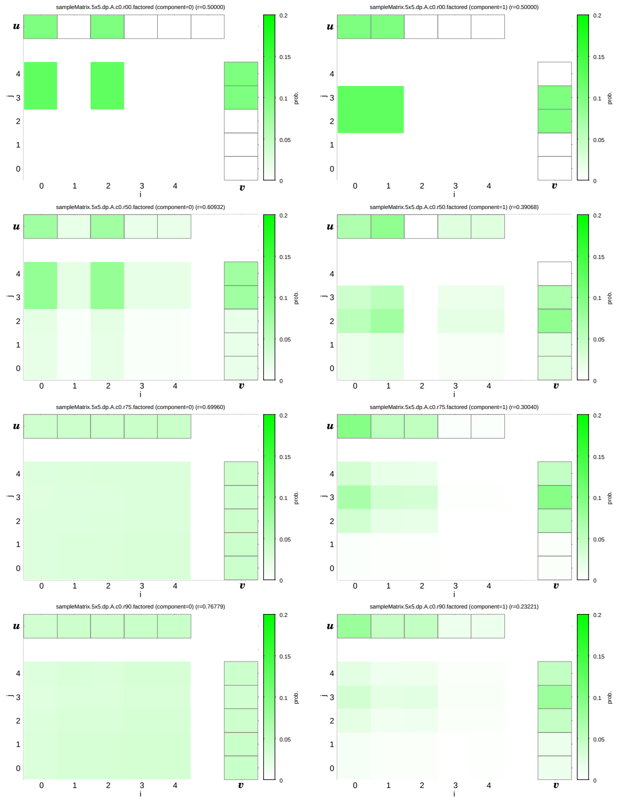


図 6: OD 分布の第 1、2 基底の分布 (上より、ノイズレベル $\rho = 0.0, 0.5, 0.75, 0.9$)

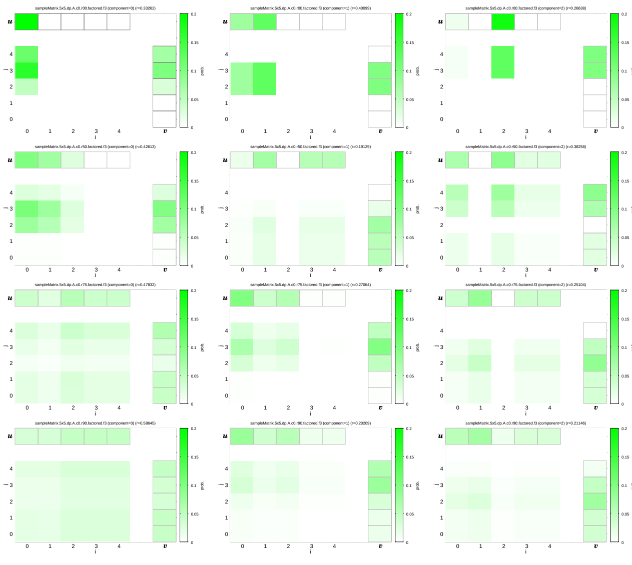


図 7: 3 基底で分解した場合の第 1、2、3 基底の分布 (上より、ノイズレベル $\rho = 0.0, 0.5, 0.75, 0.9$)

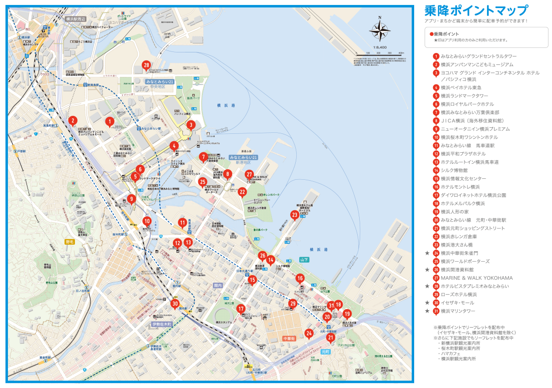


図 8: 横浜における AI 運行バスのサービスエリア

4.2 実験 2:AI 運行バスの実 OD データの分析

4.2.1 AI 運行バス実証実験

ODD-NMF の OD の実データ分析の例として、横浜市関内・みなとみらい 21 エリア実証実験で運営した AI 運行バスで取得した OD を用いた分析結果を示す。

本実証実験は、横浜市都心臨海地区において 2018 年 10 月 5 日から 12 月 10 日に渡って行われた。この AI 運行バスは仮想バス停版 SAVS というべきものであり、乗降場所は予め決められたバス停のみという制約の他は、乗降場所の組合せおよび時刻は SAVS と同じように自由に指定できるようになっている。図 8 は本実験の運行エリアを示しており、図中にある赤い丸が仮想バス停となっている。各バス停の詳細は表 1 に示すとおりである。

4.2.2 ODD-NMF による分析

ODD-NMF による解析では、本実証実験で取得・処理された全件、28,300 件のデマンドを用いる。これを確率分布 $\dot{\boldsymbol{W}}$ として整理したものが図 9 に示す行列である。なお、これら確率行列を示す図の読み方は、次元数が 31 となる他は 4.1 節のものと同じである。

まず解析に当たっては、適切な基底の数 K を定めるために、BIC を用いた解析を行った。即ち、様々な K の値に対して各々ODD-NMF を行い、BIC を求めてその値が最小となる K を求めた。図 10 が各 K に対する BIC の値の変化である。この図より、$K = 5$ において BIC が最小となることがわかる。

この $K = 5$ の時に求められた各基底の確率行列 $\hat{W}_k$ を示したものが図 11 である。すなわち実証実験で得られた OD は、この 5 つの基底パターンの混合で表されると考えることができる。

表 1: AI 運行バスの仮想バス停

i/j	**バス停 (乗降ポイント) 名**
0	みなとみらいグランドセントラルタワー
1	横浜アンパンマンこどもミュージアム
2	ヨコハマ グランド インターコンチネンタル ホテル
3	横浜ベイホテル東急
4	横浜ランドマークタワー
5	横浜ロイヤルパークホテル
6	横浜みなとみらい万葉倶楽部
7	ＪＩＣＡ横浜 （海外移住資料館）
8	ニューオータニイン横浜プレミアム
9	横浜桜木町ワシントンホテル
10	みなとみらい線　馬車道駅
11	横浜平和プラザホテル
12	ホテルルートイン横浜馬車道
13	シルク博物館
14	横浜情報文化センター
15	ホテルモントレ横浜
16	ダイワロイネットホテル横浜公園
17	ホテルメルパルク横浜
18	横浜人形の家
19	みなとみらい線　元町・中華街駅
20	横浜元町ショッピングストリート
21	横浜赤レンガ倉庫
22	横浜港大さん橋国際客船ターミナル
23	横浜中華街朱雀門
24	横浜ワールドポーターズ
25	横浜開港資料館
26	MARINE & WALK YOKOHAMA
27	ホテルビスタプレミオみなとみらい
28	ローズホテル横浜（横浜中華街内）
29	イセザキ・モール
30	横浜マリンタワー

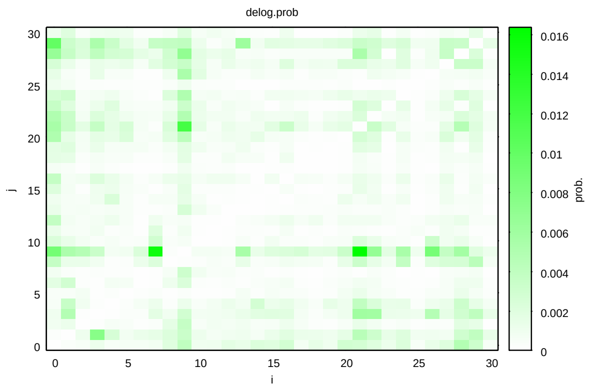


図 9: 実証実験の OD 分布

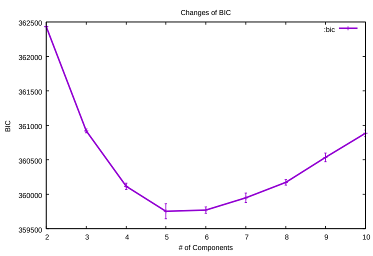


図 10: BIC の変化

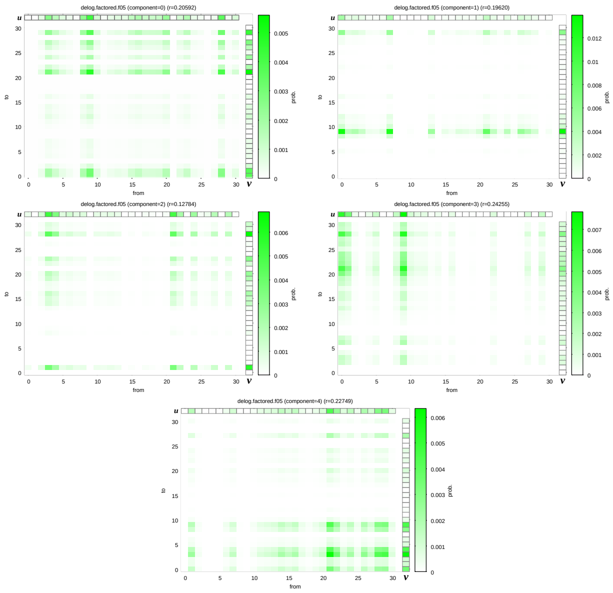


図 11: 獲得された 5 つの基底 (左上:$k=0$、右上:$k=1$)、右中:$k=2$)、右中:$k=3$)、下:$k=4$)

4.2.3 基底の機能分析

各基底が表している OD の特徴を洗い出すために、休日及び平日で分類した OD データの解析を行った。図 9 で示した全 OD データを休日分と平日分に分けてまとめると、図 12 のようなものになる。この 2 つの OD 行列に対して、5 つの基底の $\hat{\boldsymbol{u}}_k$、$\hat{\boldsymbol{v}}_k$ を図 11 のものに固定し、$\hat{r}_k$ のみを調整して別々に最尤推定を行う。すなわち、図 1 に示したアルゴリズムのうち、3 行目の初期化と 7 行目の $\hat{\boldsymbol{u}}_k$、$\hat{\boldsymbol{v}}_k$ の更新をスキップして実行する。

このようにして求められた $\hat{r}_k$ を、全 OD データ、休日分 OD データ、平日分 OD データ別に示したのが図 13 である。このグラフのうち、青色のバーが全データに対する $\hat{r}_k$ の値、橙色のバーが休日分データ、黄色のバーが平日分データに対する値であり、横に並んだ 5 つのグループは図 11 の各基底に対応している。このグラフから、最初の基底 ($k=0$) は休日分に強く寄与しており、2 番めの基底 ($k=1$) は平日分に寄与していることがわかる。一方、$k=2,3,4$ の基底については、平日・休日の間で大きな差は見られない。

この休日と平日を特徴づけている $k=0$ および $k=1$ の基底の特徴を明らかにするために、図 11 に示している $k=0$ および $k=1$ の基底の $\hat{\boldsymbol{u}}_k$、$\hat{\boldsymbol{v}}_k$ に着目する。これらのベクトルのうち大きな成分となっているところをのバス停を取り出すと、以下のようになる。

- $k=0$
  - from($\hat{\boldsymbol{u}}_0$) : 9(横浜桜木町ワシントンホテル), 28(ローズホテル横浜), 3(横浜ベイホテル東急)
  - to($\hat{\boldsymbol{v}}_0$) : 21(横浜赤レンガ倉庫), 29(イセザキ・モール), 0(みなとみらいグランドセントラルタワー), 1(横浜アンパンマンこどもミュージアム)

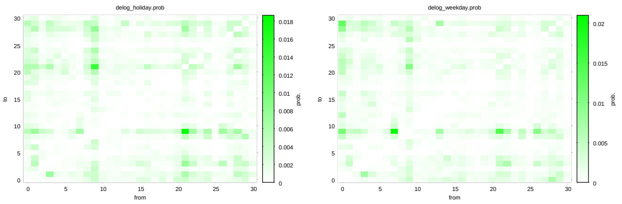


図 12: 休日および平日の OD 分布 (左：休日、右：平日)

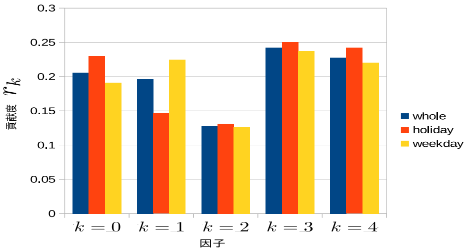


図 13: 休日/週日別 OD 分布への 5 基底の貢献度

- $k=1$
  - from($\hat{\boldsymbol{u}}_1$)：7(ＪＩＣＡ横浜), 0(みなとみらいグランドセントラルタワー),
  - to($\hat{\boldsymbol{v}}_1$)：9(横浜桜木町ワシントンホテル), 29(イセザキ・モール)

これを見ると、$k=0$ の基底の方は、ホテルなどから観光地などへの流れであり、休日分に大きく寄与している裏付けとみなせる。一方、$k=1$ の基底については、to に対応するワシントンホテルとイセザキ・モールは、各々、桜木町駅と関内駅の最寄りバス停となっており、オフィス街から駅へ向かう流れとして平日に強く現れることが妥当と考えることができる。

4.2.4 負担率の活用

ODD-NMF のアルゴリズム中で、図 4 で導入した負担率 γ_{Kij} は事後確率であることから、ある OD が発生した時、その OD の (i,j) の組に対して、どの基底 k より発生したものかを表しているといえる。実際に上記の分析で実際に得られた最終的な γ_{Kij} の値を示したものが図 14 である。

これらの図の中で、$k=2$ に対応するグラフでは (0,1) 要素や (21,26) 要素が強い (γ の値が 1 に近い) ことが見て取れる。つまり (0,1) や (21,26) に対応する OD のデマンドは、この基底に属する可能性が高い。これらの地点を見ると、みなとみらいグランドセントラルタワーから横浜アンパンマンこどもミュージアムへ、横浜赤レンガ倉庫から MARINE & WALK YOKOHAMA へとなっている。このことから、$k=2$ は観光地のはしごのような移動を表した基底とみなすことができるかもしれない。

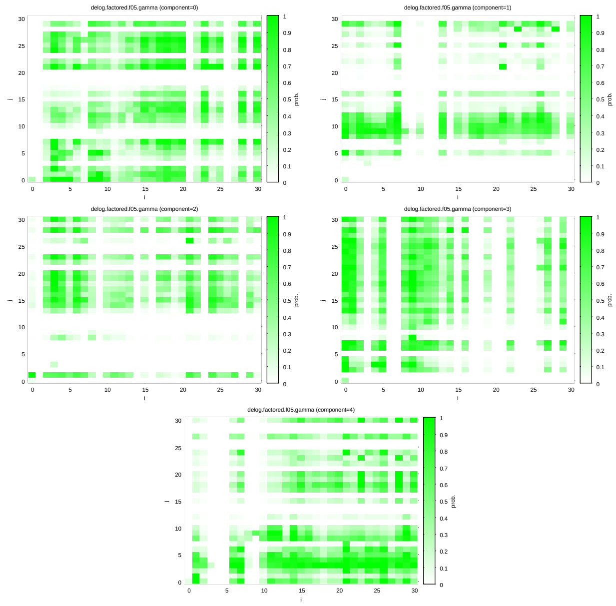


図 14: 各 OD ペアにおける各基底の負担率

このように、負担率と OD の各地点の性質を用いて基底の属性解析や実際の OD の目的推定が行えると考えられる。

4.2.5 各バス停の特徴分析

各基底内のベクトル $\hat{\boldsymbol{u}}_k$、$\hat{\boldsymbol{v}}_k$ の各要素の値は対応するバス停 (place) の属性とみなすことができる。そこで、各バス停毎に $\hat{u}_{ki}$、$\hat{v}_{kj}$ を取り出して並べたベクトルをそのバス停の特徴ベクトルとしてまとめたのが図 15 である。この特徴ベクトルを用いて各バス停の特徴空間上での距離を求めてみたのが図 16 の上図である。ただし、距離としては、特徴ベクトルの方向余弦の値を 1 から引いた値とした。

$$
\begin{aligned}
\text{距離}\ (i,j) &= (1 - \frac{\boldsymbol{f}_i \cdot \boldsymbol{f}_j}{|\boldsymbol{f}_i| \cdot |\boldsymbol{f}_j|}) \\
\boldsymbol{f}_i &= \left[\ \hat{u}_{0i}, \hat{v}_{0i}, \hat{u}_{1i}, \hat{v}_{1i}, \cdots\ \right]
\end{aligned}
$$

また、グラフでは、距離が小さいほど色が濃くなるように表示している。

このバス停について、その類似度がより明確になるように、この距離を用いてクラスタリングを行い、その順に並べ替えて距離行列を書き直したものが図 16 の下図である。なお、並べ替えの順番は、

29, 11, 12, 8, 9, 10, 1, 0, 5, 3, 4, 17, 18, 15, 20, 30, 28, 19, 23, 14, 16, 26, 7, 13, 2, 27, 6, 25, 22, 21, 24

となっている。このグラフを見ると、各バス停がクラスタを形成していることが見てとれる。即ち、各バス

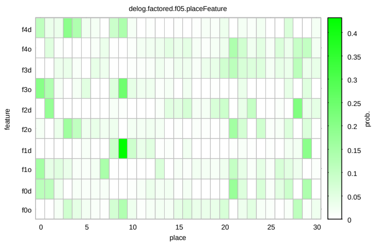
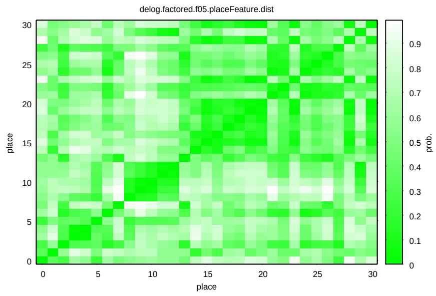
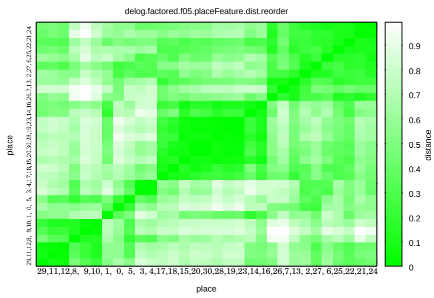


図 15: 各基底の u, v の値による各バス停の特徴量



(整列前)



(クラスタリングによる整列後)

図 16: 各バス停の特徴量間距離行列

停をクラスタ毎に性格づけすることが可能となり、利用者への各種レコメンデーションなどの根拠として活用できると考えられる。

5 おわりに

本稿では、交通移動の OD データを、出発地ベクトルと目的地ベクトルの直積の混合で表されると仮定し、最尤推定で非負値行列因子分解を行う方法である ODD-NMF を提案し、トイプロブレムによりその機能・性質の検証を行った。さらに、実データに ODD-NMF を適用することで、実データの特徴を複数の側面から分析することを試みた。トイプロブレムによる検証では、基底の数を適切に設定することで提案手法はノイズに頑健にもとの基底を再現できることが確認された。また、実データに適用した分析では、抽出された各基底を目的別に解釈できること、OD の場所を特性に応じてクラスタリングできる可能性が確認できた。

交通サービス利用の OD の実データから (1) 式のような形で基底分解できることで、よりリアリティのある OD のバリエーションが構成できると考えられる。例えば、獲得された $\hat{\boldsymbol{u}}_k$、$\hat{\boldsymbol{v}}_k$ はそのままに、$\hat{r}_k$ を変更することで、基底パターンの混合度合いが異なる OD データを構成することができる。このような OD パターンのバリエーションは、交通サービスをシミュレーションで評価していく際に欠かせないものである。

本研究では、またいくつかの課題が残されている。まず、各基底の機能のうち、$k = 2, 3, 4$ についてはまだ分析できていない。休日・平日の別だけでなく、時間帯別でも同様の分析を行うなど、さらなる解析方法の確立が必要である。

また、抽出された要素パターンが、おのおの、解釈可能な目的行動 (例えば買い物) の分類になっているという保証はない。出発地・目的地確率ベクトルの形によっては明確に分割できないか、あるいは過剰に分割されてしまう可能性もある。今後は、この方法の限界を確認していく必要がある。

この他、移動サービスの OD データの場合、出発地と目的地が同じ場合に潜在的な移動要求がデータに現れない可能性がある。このようなデータが本手法でどのように抽出できるかも今後の検討の課題である。

謝辞

この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務の結果得られたものです。

参考文献

[1] Hideyuki Nakashima, Hitoshi Matsubara, Keiji Hirata, Yoh Shiraishi, Shoji Sano, Ryo Kanamori, Itsuki Noda, Tomohisa Yamashita, and Hitoshi Koshiba. Design of the smart access vehicle system with large scale ma simulation. In *Proceedings of the 1st International Workshop on Multiagent-based Societal Systems (MASS2013)*, May 2013.

[2] 亀岡弘和. 非負値行列因子分解. **計測と制御**, 51(9):835–844, 9 月 2012.

[3] 中島秀之, 小柴等, 佐野渉二, 落合純一, 白石陽, 平田圭二, 野田五十樹, and 松原仁. Smart access vehicle system : フルデマンド型公共交通配車システムの実装と評価. **情報処理学会論文誌**, 57(4):1290–1302, 4 月 2016.

[4] 中島秀之, 松原仁, and 田柳恵美子, editors. **スマートモビリティ革命 — 未来型** *AI* **公共交通サービス** *SAVS* —. 公立はこだて未来大学出版会, 3 月 2019.

[5] 中島秀之, 野田五十樹, 松原仁, 平田圭二, 田柳恵美子, 白石陽, 佐野渉二, 小柴等, and 金森亮. バスとタクシーを融合した新しい公共交通サービスの概念とシステムの実装. **土木学会論文集** *D3* (**土木計画学**), 71(5):I_875–I_888, 12 月 2015.