非負値行列因子分解を用いた観光地来訪人口の時空間分布特性の分析

沼田　祥太朗[1]・奥村　誠[2]

[1]非会員　東北大学大学院　工学研究科（〒980-0845 仙台市青葉区荒巻字青葉 468-1, S502b）
E-mail: shotaro.numada.r5@dc.tohoku.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, S502b）
E-mail: mokmr@tohoku.ac.jp (Corresponding Author)

旅行者は目的に応じて活動を実行する時間を決めるが，その時間分布は自然（暦や天候）によって影響する分布が含まれると考えられる．本研究ではNTTドコモが提供するモバイル空間統計から，対象地域内の旅行者人口分布を抽出し，非負値行列因子分解（NMF）を用いて活動に対応する時間分布パターンを導出する．導出パターンの特性を分析し，自然影響パターンが存在することを確認する．

Key Words: *Non-Negative Matrix Factorization, tourism, distribution of population*

1. はじめに

(1) 観光地における活動時間分布

観光地において，旅行者は複数の異なる活動を行うことが多い．限られた滞在時間の中で，それぞれの活動をその内容にあったような時間帯に実行できるように，1日のスケジュールを計画すると考えられる．ただし施設の営業時間や混雑の影響で，その活動に最も適した時間帯で実行することが難しい場合には，その前後にずらすという調整が行われることもある．

例えば，「お昼に繁華街でランチを食べたい」といった活動を計画する場合，「ランチを食べる」という活動内容に適していると思われる正午付近の時間帯に実行するだろう．しかし「この超有名店で食べたい」というように場所を重視している場合には，混雑が激しい正午付近の時間帯を避けて空いている時間帯にずらす，また「どうしてもお昼にこの超有名店でランチを食べたい」であれば，食べるまでの行列にも時間を割り振るというように，状況に応じて時間の使い方を調整することになる．このような本来の希望活動時刻からの調整は，その施設が混雑する場合ほど行われやすく，混雑が激しくない日には旅行者はほぼ希望活動時刻に活動することができるであろう．したがって「ランチ」のような一つの活動が実行される時間分布は，本来の希望時刻にピークを持ち，混雑などの影響を受ける日にはその前後の時間帯に広がるという分布を持つと考えられる．

「活動」をどの程度の細かさで分類するかに依るものの，活動内容が大きく異なれば，その内容に適した希望時間帯は異なり，実際に活動が行われる時間分布も異なると考えられる．逆に，異なる活動と考えられていた2つの活動が同じ時間帯に行われているのであれば，一つの活動施設が休業や混雑に見舞われた場合に，他方の活動で代替される可能性もある．もともとは和食の店で海鮮ランチを食べたいと考えていたが，近くの中華レストランでランチを済ませて，前後のスケジュールを変更せずに別の活動を楽しむ，という例がこれにあたる．

(2) 本研究の長期的な目標

コロナ禍の収束による観光客の回復が見込まれる一方で，一旦解雇した労働者を再雇用することが容易ではなく，以前よりも混雑が激しくなっている観光地も少なくない．このような観光地で，観光客への対応を誤ると，インターネット情報などを通して失望感が拡散し，長期的な観光地の停滞につながる危険性もある．観光地としては，自らの観光地においてどのようなサービスが提供できており，それは観光客のニーズに十分応えられているのか，さらに混雑などを原因とする失望感を与えている危険性がないかについて，注意を払うことが求められる．

本研究では，観光地において行われている様々な活動を，時間分布の共通性に基づいて分類，把握し，当該観光地内で充分な種類の活動機会が提供されているのか，それらは混雑などの問題を発生させていないかを，特別な能動的な調査をせずに，近年容易に使用できるように

なった携帯電話位置情報ビッグデータを活用して分析する方法の提案を長期的な目標としている．ただし，現段階では具体的な研究成果が得られている段階には至っておらず，まずは実際のデータを扱いながら，その特徴を把握するにとどまっている．

(3) 自然時間に依存する活動時間分布

上述したように，旅行者が行う活動には，その内容に対応した希望時間帯が存在すると考えられるが，その中には特定の時刻に固定的に行われるものだけでなく，自然（暦や天候）の持つ時間サイクルに依存する活動も存在すると考える．例えば，「水平線に沈む夕日を見る」という活動は，季節によって好適な時間は異なる．また，牧場や遊園地など屋外施設は，降雨時には魅力が減少し，実際に活動を行う人は少ないと考えられる．

以上の例は，活動の目的と自然が直接的に関係している例だが，より間接的に自然の時間が旅行者の時間分布に影響を与える可能性がある．例えば，「夏は明るい時間が長いので，ホテルにいる時間が短くなる」「高級ディナーを食べる時間は暗い方を好む」など，無意識のうちに選ぶ活動も自然が関わっている可能性がある． すなわち，意識的かどうかを問わず，自然が時間分布に影響を与えているような活動は少なくないと考える．

このような観光地での自然（とりわけ天候）と旅行者の行動との関係の分析 [1]は多くおこなわれている．また，旅行者自身では変えられない自然の要因によって，活動時間を希望時間から移動させたり，活動を取りやめざるを得なくなると，結果として旅行者の満足度が大きく低下することを示した研究もある [2]．観光地を運営する側としては，自然が影響する活動の時間分布を把握し，悪天候などの影響が観光地の魅力や評判の低下につながらないように注意することが必要であろう．

本論文では，携帯電話位置情報ビッグデータを活用して，旅行者の滞在人数の時間分布の中から，自然の時間（暦や天候）が強く影響しているような活動の存在を確認することを当面の目的とする．

具体的には，代表的な携帯電話位置情報データである NTT ドコモのモバイル空間統計から旅行者の時空間別人口分布データを抽出し，非負値行列因子分解を用いてパターン分解を行い，その中に自然の時間に依存する活動が含まれていることを示す．

以下本論文は 4 章で構成する．2 章では使用データ概要と分解手法について述べる．3 章では観光地を対象に 2 章の分解手法を適用し，自然が時間分布に与える影響を確認する． 4 章は結論であり，本研究の成果をまとめるとともに，今後の研究方向を展望する．

2. 分析の方法

(1) 使用データ

本研究では，旅行者の時間分布を把握するために，NTT ドコモが提供するモバイル空間統計の人口分布データを用いる．モバイル空間統計は各基地局のエリアごとに所在する携帯電話の台数を集計し，ドコモの普及率を加味して対象区域の人口を推計したデータで，性別，年代，居住地（都道府県・市区町村）といった属性データも併せて推計されている．なお，個人の特定を防ぐため，運用データに非識別化処理，集計処理，秘匿処理を行う．人口分布データは単位区域の 1 時間ごとの人口を上記方法で推計した集計データである．1 日の中の 1 時間単位の人口分布を長期間取得可能であるため，自然の時間の影響を量的に捉えることが可能である．

本研究の分析では，単位区域を 500m メッシュとし，分析期間は COVID-19 発生前である 2019 年の全休日とした．分析地域は観光地によって異なるが，500ｍメッシュが数百～千個程度の範囲とした．旅行者が観光地を 1 日で回る範囲として妥当であると考える．また，旅行者人口の定義として，居住地が対象観光地の所属都道府県以外の人口と置いた．

(2) 非負値行列因子分解（ＮＭＦ）

人口分布データを類型化して分解する手法として，非負値行列因子分解(NMF; Non-negative Matrix Factorization)を用いる．NMF とは，非負値のデータを加法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である．各観測ベクトルがＫ個の基底ベクトルの重み付き和によって表されたものとみなし，全ての観測ベクトルを最もよく説明するＫ個の基底ベクトル及び重み係数を推定する．

NMF は，主成分分析や因子分析などの手法と比較して，データを分解した結果が非負値であるため，値の大小のみ考慮すればよく，結果の解釈が行いやすい．また，要素の表現は足し算のみであるため，分解した行列の情報量が増え，特徴的な情報を取り出すことが可能である．

観測行列 X（サイズ M×N）を２つの非負値行列 W（サイズＭ×Ｋ），H（サイズＫ×Ｎ）で近似する．この時，基底数Ｋは(2)式を満たす必要がある．

$$X \simeq W \cdot H \tag{1}$$

$$K \leq \min(M, N) \tag{2}$$

基底数Ｋは，事前に外生的に与える必要がある．様々な決定方法があるが，AIC や BIC などの情報量基準を利用した方法が頻繁に使われている．本研究では安川(2015)[4]に従い損失関数をユークリッド距離，情報量基準

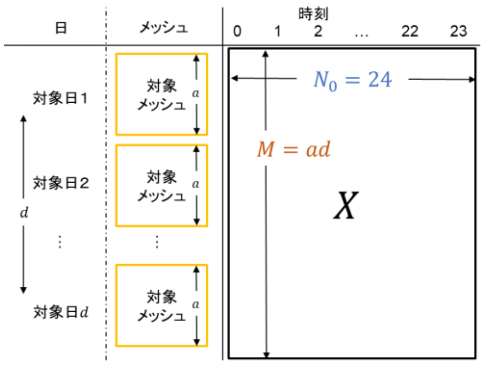


図-1　人口分布データ

を(Owen&Perry2009)の BIC の定義を採用した.

$$D_k = \sum\sum (x_{mn} - \mu_{mn})^2 \tag{3}$$

x_{mn} ：観測行列Xの要素，μ_{mn} ：NMF によって近似した行列の要素

$$BIC = \log(D_k) - K\frac{m+n}{mn}\log\frac{mn}{m+n} \tag{4}$$

(3) 本研究のＮＭＦの特徴

モバイル空間統計人口分布データから得られたデータを，対象メッシュ数をa，対象日数をdとし，図1のように定義する．分解した行列 W,H は，1 日の時間分布に着目すると，それぞれ係数行列，基底ベクトル行列と捉えることができる.

1 章で，旅行者の時間分布は，通年を通して一定である通常パターンと，自然に応じて変動するパターンが加算されて出現すると仮定を立てた．自然パターンの影響を解釈しやすいよう，時間分布データに自然由来のデータを合わせて分解を行う．分解は図 2 のように表現できる.

行列 Y は自然由来のデータを格納する行列であり，サイズを $M \times N_1$ とし，対象日・メッシュに対応するデータを与える．X と Y を横向きに並べたデータに対して NMF を行えば，通常パターンの係数 W0・基底ベクトル H0 に加え，自然パターンの係数 W1・基底ベクトル H1 に分解できる．F0, F1 はダミーデータと自然パターンの共起係数行列を表す.

なお，結果の解釈を行いやすくするため，分解行列に対して標準化を行い，基底ベクトルの最大値を 1 とした．標準化行列の内積と，元の分解行列の内積は等しい.

$$\mathrm{H}'_{\mathrm{k,n}} = \frac{\mathrm{H_{k,n}}}{\max\limits_{\mathrm{k}}(H_n)} \tag{5}$$

$$\mathrm{W}'_{\mathrm{m,k}} = W_{m,k} \cdot \max_{\mathrm{k}}(H_n) \tag{6}$$

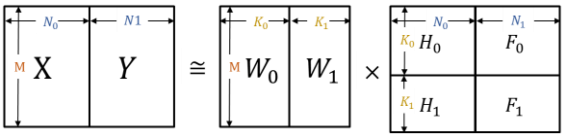


図-2　分解手順

3. パターン分解

(1) 追加自然データ

通年を通して時間分布が一定でないパターンを考慮する．本節では日の出日の入りを対象とし，対象地域の日の出日の入り時刻をそれぞれ読み取り，$\pm 1h$ の人口を合計したデータを日の出人口，日の入り人口とし，自然データY_2（サイズ M×2）とした．日の出日の入り時刻は国立天文台を参照し，人口は人口分布データの 1 時間単位の人口を按分し求めた.

$$Q_d = (60 - y_d)q_{x_d-1} + q_{x_d} + y_d q_{x_d+1} \tag{7}$$

Q_d ：日付 d の日の出日の入り人口（各メッシュの添え字は省略） x_d, y_d ：日付 d の日の出日の入り時刻,(x時y分) q_n ：n 時の人口

(2) 季節間比較

対象地域として湘南鎌倉地区（鎌倉市，藤沢市，茅ヶ崎市，逗子市，葉山町）を選定した.

まず，データ X をそのまま分解した結果と，自然データを追加し分解した結果を比較し，分解における自然データの影響を見た（図 3）．抽出パターン数は前節の情

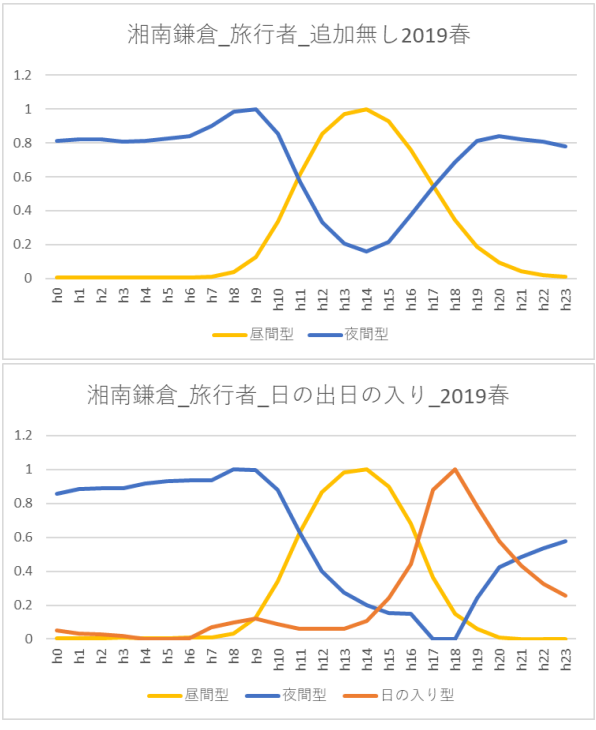


図-3　自然追加 NMF



図-4　パターン分布

報量基準を基に決定する．対象期間は 2019 年春（3 月～5 月）の休日である．前者の分解は，昼間型と夜間型の 2 パターンが抽出された．後者の分解は，昼間型・夜間型に加え，18 時付近にピークがある日の入り型のパターンが抽出された．

係数行列 W の値について GIS 上に表示したものを図 4 に示す．日の入り型の分布は藤沢駅，鎌倉駅，江ノ島駅周辺が高い値を取った．駅周辺の日帰り客の帰宅行動を表すと考えられる．

続いて，季節間の比較を行った．2019 年夏（6 月~8 月）の休日の分析との比較を図 5 に示す．昼間型はほぼ変わらないが，日の入り型のピークに着目すると，春は 17~18 時なのに対し，夏は 18~19 時であり，日の入り型は時期に応じて変動する結果が得られた．また，夏の日の入り型の空間分布は，春とほぼ同じく駅周辺で高い値を示す分布となった．夏の方が遅くまで観光地に滞在していることを示唆している．

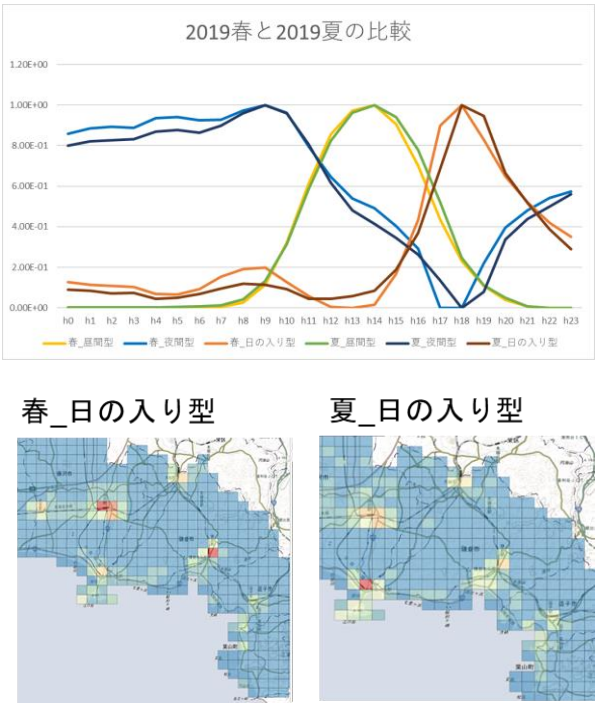


図-5　季節間パターン分布比較

4.　おわりに

(1) 本研究の成果

本研究では人口分布データを類型化し，旅行者の時間分布に自然の影響を受けるパターンが含まれることを確認した．自然パターンとして，期間に応じて時間分布が変動するパターンを抽出できた．

(2) 今後の課題・展望

降水量や日照時間など，1 日の中で時間に応じて変動する要素が時間分布に与える影響についても分析を進めたい．また，観光地の活動機会の多様性の把握を目標に，時間分布・空間分布特性について，より詳しい分析を続けたい．

参考文献

1) M. Belen Gomez-Martin：Weather, climate and tourism, A geographical perspective, Annals of Tourism Research, 32(3),pp.571-591, 2005.
2) 山田雄一：旅行先での天候が顧客満足およびロイヤルティ意識に及ぼす影響，日本国際観光学 Vol.19 pp.47-52，2012. [Yamada,Y.: Effects of weather conditions over formation of CS and loyalty at a destination, Jonan Foundation for International Tourism, Vol.19, pp.47-52, 2012.]
3) 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. [Yasukawa,T.: Text data analysis using Nonnegative Matrix Factorization, Bulletin of the Computational Ststistics of Japan, Vol.28, No.1，pp.41-55，2015.]

(Received October 13, 2023)
(Accepted ?)

Analysis of Spatio-Temporal Distribution Characteristics of the Population Visiting Tourist Attractions Using Non-Negative Matrix Factorization

Shotaro NUMADA, Makoto OKUMURA

Travelers decide the time to perform an activity according to their objectives, and the time distribution is expected to include distributions affected by nature (calendar and weather). In this study, we extract the distribution of the traveler population in the target area from the Mobile Spatial Statistics Data provided by NTT DOCOMO, and derive the time distribution pattern corresponding to the activity using non-negative matrix factorization (NMF). The characteristics of the derived patterns are analyzed to confirm the existence of natural influence in time distribution.