

# モバイル空間統計を用いた 地方都市への来訪者分析

石橋 千尋<sup>1</sup>・塚井 誠人<sup>2</sup>

<sup>1</sup>学生会員 広島大学大学院 工学部研究科 (〒739-8527 広島県東広島市鏡山 1-4-1)

E-mail:m186495@hiroshima-u.ac.jp

<sup>2</sup>正会員 広島大学大学院准教授 工学研究院社会環境空間部門 (〒739-8527 広島県東広島市鏡山 1-4-1)

E-mail:mtukai@hiroshima-u.co.jp

交通関連のビッグデータは、データ量が膨大かつ移動目的が不明なため、そのまま集計しても有益な知見を得にくく、効果的に分析する統計的手法が必要である。本研究では、山口県萩市・島根県浜田市を対象としたモバイル空間統計を用い、どのような時間帯に、どのような地域からどのような地域へとといった来訪者の特性を把握することを目的とする。具体的には、発地・着地情報と時間情報を用い、対象エリアへの来訪者のOD×日付・時刻の交流量データを作成する。作成したデータに非負値行列因子分解を用いてパターン化することで、来訪者の特性を把握した。また、パターン毎の時間帯や発着地情報からの来訪目的の類推を行った。さらに、2市間への来訪者の違いを確認する。

**Key Words :** *mobile spatial statistics, nonnegative matrix factorization, visitors analysis, local cities*

## 1. 背景・目的

人口減少に伴う地域内消費の減少を補うため、地域外からの来訪者の消費を促す地域計画が求められている。限られた財源の下、効果的な計画を立てる必要がある。そのためには、地域の空間的特徴と来訪者特性の関係を詳細に把握が必要がある。近年、携帯電話基地局情報に基づく3次メッシュ (1km×1km) 別滞在人口データであるモバイル空間統計<sup>1)</sup>に注目が集まっている。ただし、モバイル空間統計は移動目的が不明なため、旅客の消費パターンなどの把握が難しく、そのまま集計しても有益な情報を得にくい。そのため、データ特性に即した分析手法の開発が必要である。

本研究では、モバイル空間統計に非負値行列因子分解を適用して来訪者特性の把握を行う。具体的には、同等の都市規模の2つの地方都市への来訪者を分析する。この分析を通じて、地域の空間的特徴と来訪者特性の関係や、交通アクセスに伴う来訪者の特性の把握や違いについて分析を行うことを目的とする。

本研究の構成は、以下の通りである。2.では、本研究で用いるモバイル空間統計と非負値行列因子分解の既往研究について整理する。3.では、本研究で用いるモバイル空間統計の概要について説明する。4.では、非負値行列因子分解の説明をし、データセットを分解した結果を

述べる。5.では、施設立地情報を用い、分解された結果の重回帰分析について述べる。6.では、結論を述べる。

## 2. 既往研究

本節では、モバイル空間統計を用いた研究を整理する。清家ら<sup>2)</sup>は、まちづくり分野へのモバイル空間統計の信頼性を検証し、その活用可能性について研究を行った。その結果、時間別推計・人口属性別推計・居住地属性別統計については一定の信頼性が確認されたが、人口の少ない3次メッシュレベルでは、データの信頼性が乏しいことを明らかにした。その結果を踏まえて、空間解像度については、一定以上の高い人口密度を前提にする必要があると結論付けた。室井ら<sup>3)</sup>は、モバイル空間統計と、純流動調査・パーソントリップ調査を比較して、空間の解像度・時間の解像度・速報性のメリット・デメリットをまとめた。森尾ら<sup>4)</sup>は、都市圏を対象にモバイル空間統計の特性をパーソントリップ調査と比較するとともに、モバイル空間統計の都市交通分野への適用可能性を検討した。

他のデータと統合利用した研究として、清家ら<sup>5)</sup>は、モバイル空間統計と他の人口推計を組み合わせる対象地域の特性を分析するモデルと、モバイル空間統計のみで分析するモデルを比較した。佐藤ら<sup>6)</sup>は、土砂災害

の被災地を対象に、モバイル空間統計の滞在人口データに独立成分分析を適用して、滞在人口の変動要因の把握を試みた。その結果、災害発生後の避難計画に関連する短期変動と、数か月に及ぶ長期変動を抽出できた。有村ら<sup>7)</sup>は、モバイル空間統計と都市計画基礎調査の2つのマイクロジオデータを統合的に用いて、都市内の建物用途の変更及び再配置や集積が滞在人口に与える影響を分析した。その結果、遊戯施設や宿泊施設が有意な説明変数として抽出された。山口ら<sup>8)</sup>は、モバイル空間統計の時間・空間・属性情報に非負値テンソル因子分解を適用し、北陸新幹線整備に対する感度の異なる旅行グループに分類・推定する手法を開発した。鈴木ら<sup>9)</sup>は、幹線調査とモバイル空間統計データを融合して、日別・旅行目的別の都市間旅客流動量を推計する手法を提案した。高橋ら<sup>10)</sup>は、全国総合交通分析システムとモバイル空間統計を用い、北海道新幹線開業前後の函館エリアへのアクセス所要時間の短縮、および訪問者数の変化を分析した。田中ら<sup>11)</sup>は、モバイル空間統計を用いて、主要な観光地における国内旅行者の人口動態を確認し、観光地の特性の把握を行った。

モバイル空間統計に関する既往研究では、時間や発地等の項目により、目的に応じてデータの抽出が行われてきた。しかし、これまでの分析内容は都市圏内の流動や自然災害等の突発事象・インフラ整備による滞在人口の変動の把握に限られており、日常的に来訪する交流パターン<sup>12)</sup>の把握を目的とする研究は少ない。また、地方中小都市のような人口の流動が小さい地域を対象にした研究の蓄積は少ない。

ビッグデータから統計的に特徴のあるパターンを抽出する多変量解析の方法の1つに、非負値行列因子分解(Nonnegative Matrix Factorization: NMF)がある。NMFの基礎と、適用に関する既往研究を整理する。Leeら<sup>12)</sup>は、NMFの推計について、平方ユークリッド距離の最小化と一般化KLダイバージェンスの最小化を、それぞれ補助関数法によって実現する効率的なアルゴリズムを考案した。Leeらのアルゴリズムは実装が容易なため、NMFが広く活用される契機となった。澤田<sup>13)</sup>は、NMFの適用分野として、文書データのクラスタリングと音声信号の分離を紹介した。亀岡<sup>14)</sup>は、NMFの音響信号への応用と拡張モデルについて解説を行った。山口ら<sup>15)</sup>は、携帯電話位置情報データの混雑統計を用いて、東京都居住者の都道府県別の滞在時間構成比表を作成した。その上で、NMFを用いて、空間・時間パターンに分解したところ、概ね解釈可能なパターンへ分解できた。しかし、解釈に曖昧さが残るパターンも得られた点は、課題として挙げられた。以上に見られるようにNMFには、基底数の決定に課題があるものの、パターン分類手法として

の有用さが評価され、多分野で応用が進められている。

本研究では、人口5万人程度である地方中小都市を対象地域として、モバイル空間統計を用いて、来訪者特性を把握する。またパターン数の決定問題も考慮しつつ、対象エリアへの来訪者を対象者にNMFを用い、パターン分類を行うとともに、各パターンの時間的特徴や空間的特徴を明らかにする。

### 3. モバイル空間統計による来訪者の抽出

モバイル空間統計とは、2013年に株式会社NTTドコモが事業化した、一定の空間内で一定時間内に滞在する携帯電話の台数を集計したデータベースである。調査エリアは日本国内であり、15~79歳までの男女が対象である。携帯電話の契約者の居住地は、国内では市町村単位で入手できる。時間分解能は1時間、空間分解能は都市部で4次メッシュ(500m×500m)、それ以外では3次メッシュ(1km×1km)である。

本研究の分析対象期間は、2015年5・8・11月である。GWやお盆・三連休などの来訪が増えるイベントを含む期間を選定した。対象地域は、島根県浜田市と山口県萩市の3次メッシュである。この2市は、人口・面積ともに同規模である。本研究で用いる携帯電話の契約地の単位は、市町村単位である。集計の結果、対象期間中に滞在人口がカウントされた地点数は、275メッシュとなった。本研究では、携帯電話保有者の出発地は、その電話の契約地とする。日常的な生活交通を除いて分析するため、対象エリア内(各市内)を契約地とする滞在人口を除外する。以下、これらの条件を満たすデータを来訪者と呼ぶ。なお、市町村境界線上のメッシュでは、対象市町村以外の契約地をもつ滞在人口が定常的に存在していた。市町村境界線上のメッシュの住民は、対象市町村以外の隣接する市町村の居住者の可能性があり、その場合はたとえば、そこが自宅ならば夜間でも常に一定数の滞在をカウントするという問題が生じる。そこで境界メッシュ上では、隣接市町村を契約地にもつ滞在人口を分析対象外にした。

### 4. 来訪者のパターン分類

#### (1) 非負値行列因子分解(NMF)

NMFはデータセットのもつ情報量の圧縮やパターン抽出のための多変量解析の一種である。NMFでは、 $m \times n$ サイズの非負値の観測行列 $X$ を、 $m \times k$ 非負値行列 $T$ と $k \times n$ 非負値行列 $V$ の積の形に分解する。この分解は、 $X$ と $TV$ の近似の程度を距離 $D(X,TV)$ と定義し、距離 $D(X,TV)$ を評価する損失関数を非負制約の下で最小化する数値最

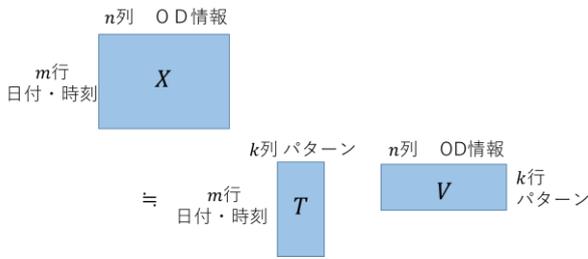


図-1 NMFのイメージ図

適化問題として定式化される。NMFは、非負のデータを非負の成分に分解できるため、応用範囲が広い。

本研究では、モバイル空間統計を用いてOD表を作成する。その手順は以下の通りである。携帯電話所有者の居住地をO（交流先）、着地をDと仮定して、モバイル空間統計の時間的情報である日付・時刻を行側に、期間中に滞在人口がカウントされたOD情報を列側とする。OD表の大きさは、672（日付・時刻）×1917（OD情報）である。日付・時刻に関しては、各月での比較を行うため、大きさを揃えた。具体的には、日曜日から土曜日までの1週間単位を各月ともに第4週分を抜き出した。そのため、5月データでは5月3日～5月30日、8月データでは8月2日～8月29日、11月データでは11月1日～11月28日の0～23時を対象とした。2市による来訪者の特性を把握するために、2市のOD情報を1つのデータセットでまとめ、3月分のデータセットを作成する。モバイル空間統計を用いて作成したOD表のデータを基に、時間的情報とODの空間的情報に関してパターン分類を行う。これにより、来訪者の特性把握を行う。

今回は、目的関数DとしてEuclid距離を採用した。また、NMFを適用する際にKは、データから取り出すパターン数であり、その値は、分析者が設定しなくてはならない。本研究では、様々なKで非負値解釈が可能なパターンのみが出るパターン数を選択する。また、パターンの時間的特徴を示す行列Tのコサイン類似度を算出し、類似度が低いパターン数を採用した。

$$\min D_c(X, TV) = \min \sum_{m=1}^M \sum_{n=1}^N d_c(x_{mn}, t_m^T v_n) \quad (1)$$

$$d_{EU}(x_{mn}, t_m^T v_n) = (x_{mn} - t_m^T v_n)^2 \quad (2)$$

$$X \cong TV \quad (3)$$

(2) NMFの結果

本紙では、紙面の都合上、5月データの考察を行う。5

月データは、パターン数5を選択した。図-2に、NMFを適用する前の実測値であるOD交通量と、適用後の予測値である行列T×Vの再現OD交通量の散布図を示した。

図-2より、少数のOD交通量の大きい値に引っ張られて行列分解が行われていることがわかる。そのため、OD交通量の大きいODの説明の精度が高く、小さいと精度が低い傾向が確認できた。次に、行列Tを用いて、パターンの時間的特徴分布を見ることが出来る。行列Tの各列の和は1で正規化されているため、パターン間の比較ができる。図-3,4,5,6に各パターンの日付・時刻毎の行列Tの値を示した。紙面の都合上、5/3～5/16の半月分を示している。図-3,4,5,6を参考にし、各パターンを、“平日昼間パターン”・“休日昼間強めパターン”・“GW昼間パターン”・“夜間パターン”と名づけた。

図-7に、行列Vが得られるパターン毎のOD重みの2市別の構成比率を図-7に示した。図-7をみると、5/27夜間のパターンに関しては、他のパターンに比べOD重みが小さいことに加え、特異的な特徴であることがわかる。よって、以下では5/27夜間パターンについて深く考察は行わない。

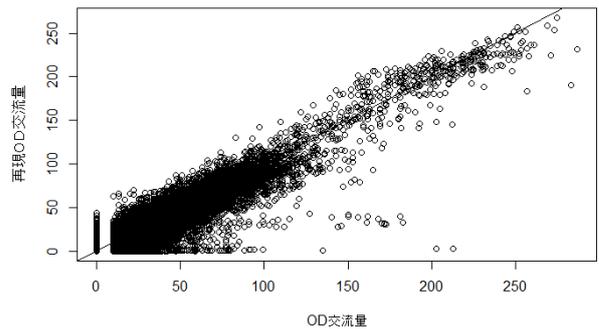


図-2 NMF適用前後のOD交通量の散布図

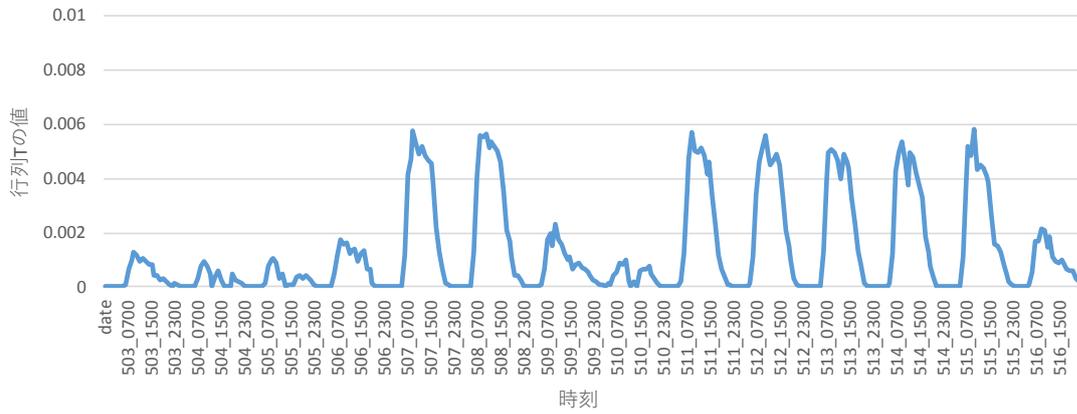


図-3 平日昼間パターンの日付・時刻毎の行列Tの値 (5/3-5/16)

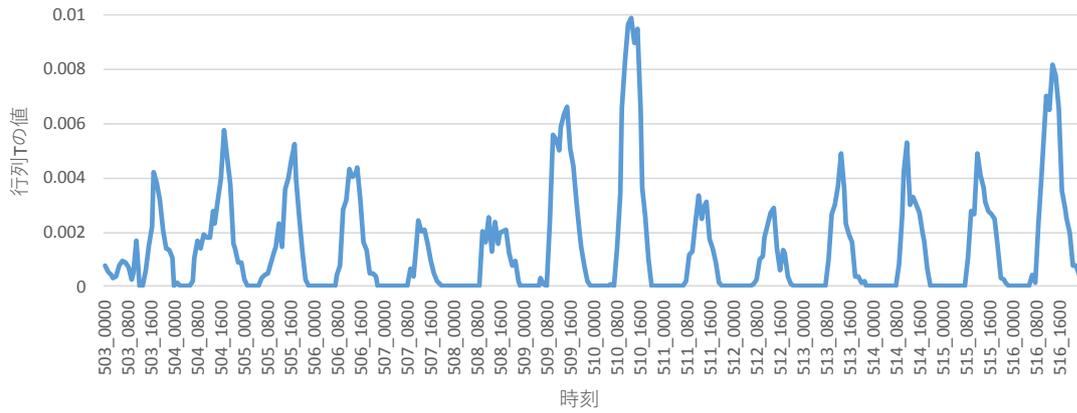


図-4 休日昼間強めパターンの日付・時刻毎の行列Tの値 (5/3-5/16)

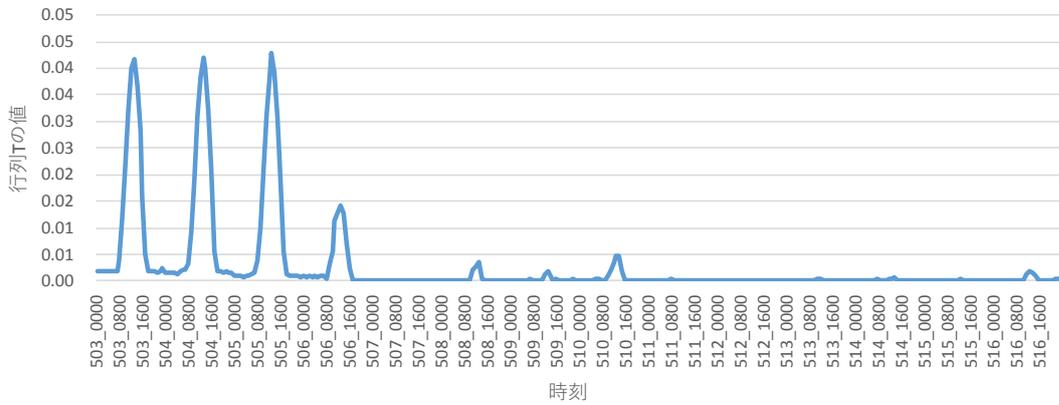


図-5 GW昼間パターンの日付・時刻毎の行列Tの値 (5/3-5/16)

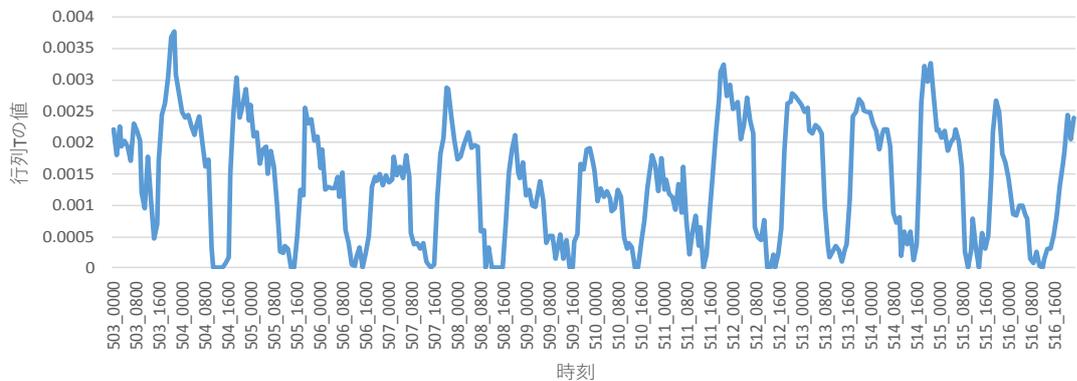


図-6 夜間パターンの日付・時刻毎の行列Tの値 (5/3-5/16)

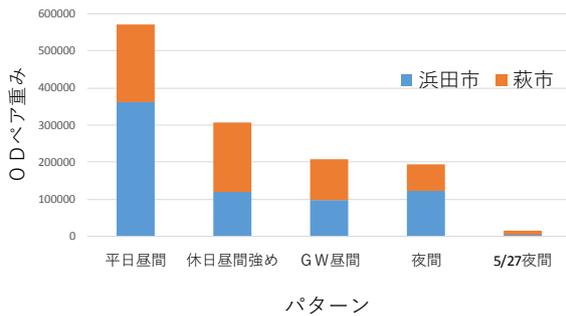


図-7 パターン毎のOD重み構成比率

## 5. パターン別来訪目的の推定

4章までに日付・時刻の時間的特徴を用いて、5月データのパターンの解釈を行った。次に、地理情報等の空間的特徴とパターン毎のOD交流量の対応関係に基づいて来訪目的を推定するため、重回帰分析を行った。目的変数として、行列Vから得られるパターン別OD交流量のうち、発地Oの情報を集約して着地D毎のパターン毎のOD重み値を用いた。この値は、パターンが表現している重みの推定値に相当する。説明変数として、国土交通省が提供している国土数値情報より、医療機関・文化施設（美術館・資料館・図書館・水族館・動植物園）・スポーツ施設・観光資源を用いる。また、株式会社ゼンリンが提供している座標付きデータベースであるテレポイントPack!より、事業所数を用いた。また、株式会社旅行出版社が提供している宿泊施設データを用いた。それぞれ、各メッシュに対しての施設数を算出し、説明変数として用いる。なおサンプル数は両市をあわせて275である。モデルの解釈を容易にするため、切片＝定数項は推定対象外とした。

各パターンの重回帰分析の結果を表-1に示す。平日昼間パターンでは、事業所・観光資源・文化施設・スポーツ施設・医療機関の影響が大きいことがわかる。平日の昼間パターンの来訪目的は、ビジネス・観光・私用といった様々な目的が考えられるが、両市については文化施設への来訪が多く、次に医療、観光施設となっている。休日昼間強めパターンでは、事業所・宿泊施設・観光資源・文化施設の影響が大きい。平日の昼間パターンに比べ、宿泊施設の影響が大きい一方、医療機関の影響が小さい。すなわち、主に観光目的の来訪者が多いと考えられる。GW昼間パターンに関しては、決定係数が低い結果となった。ただし、GW昼間という時間的特徴から、観光目的と推測できる。そのため、集客力のある文化施設の影響が大きくなった

と考えられる。夜間パターンに関しては、宿泊施設の影響が大きく宿泊目的で来訪していることが分かる。ただデータの性質上、観光目的かビジネス目的の宿泊かは分からない。

以上のように、重回帰分析によって各パターンと施設立地数との関係を通じて、来訪目的を推定できた。ただし一部の推定の結果、たとえば、夜間パターンに対してスポーツ施設の影響があるなど、解釈が困難な結果もあった。この理由として、地方中小都市では、町自体がコンパクトであって、宿泊施設とスポーツ施設が近接していることが考えられる。なお、これは統計的には多重共線性の問題だが、VIFなどの統計指標は異常を示していないことは確認している。

表-1 重回帰分析の結果

平日昼間パターン				
説明変数	回帰係数	t 値	p 値	判定
事業所数	16.122	4.821	2.4E-06	***
宿泊施設数	572.112	1.168	0.243805	
観光資源数	1141.129	3.517	0.000512	***
文化施設数	4249.445	5.832	1.57E-08	***
スポーツ施設数	965.4	3.899	0.000122	***
医療機関数	1751.653	4.71	4.13E-06	***
定数項	0	決定係数	0.6643	
休日昼間強めパターン				
説明変数	回帰係数	t 値	p 値	判定
事業所数	12.821	1.825	1.75E-11	***
宿泊施設数	1004.438	267.396	0.000211	***
観光資源数	980.847	177.029	7.18E-08	***
文化施設数	1780.352	397.594	1.12E-05	***
スポーツ施設数	184.6	135.124	0.173031	
医療機関数	438.164	203.302	0.03203	*
定数項	0	決定係数	0.6647	
GW昼間パターン				
説明変数	回帰係数	t 値	p 値	判定
事業所数	6.138	2.577	0.0179	*
宿泊施設数	-25.308	377.534	0.9466	
観光資源数	340.604	249.945	0.1741	
文化施設数	3449.508	561.36	2.87E-09	***
スポーツ施設数	94.38	190.78	0.6212	
医療機関数	25.621	287.04	0.6212	
定数項	0	決定係数	0.9289	
夜間パターン				
説明変数	回帰係数	t 値	p 値	判定
事業所数	3.674	2.157	0.0319	*
宿泊施設数	2416.475	9.683	2E-16	***
観光資源数	-372.341	-2.254	0.025	*
文化施設数	228.388	0.615	0.5388	
スポーツ施設数	932.439	7.394	1.81E-12	***
医療機関数	276.308	1.456	0.1465	
定数項	0	決定係数	0.6536	

## 6. おわりに

本研究では、モバイル空間統計を用いて来訪者の時間的特徴と空間的特徴に分解し、来訪者の特性について考察を行った。その結果、時間的特徴から概ね解釈が出来るパターン分類が行えた。さらに、重回帰分析を行うことで、各パターンの来訪する目的を明らかにできた。

本研究では、モバイル空間統計が持つ性別や年代といった属性を分類せずに滞在人口をカウントしたデータを用いた。これは、都心や巨大な観光地に比べ来訪者の少ない地方中小都市を対象としているため、属性で分類したデータを用いると、モバイル空間統計が行う秘匿処理によって、データの多くが消去される可能性が高いためである、このため、推測される来訪目的の精度については限界があるという課題が残った。

また、今回発地Oに関する考察を行えていない。メッシュへの所要時間や発地Oの人口を踏まえ、来訪者の特性や消費に繋がるような情報の確認は、発表時に示したい。

### 参考文献

- 1) NTT ドコモ：モバイル空間統計に関する情報，[https://www.nttdocomo.co.jp/corporate/disclosure/mobile\\_spatial\\_statistics/](https://www.nttdocomo.co.jp/corporate/disclosure/mobile_spatial_statistics/).
- 2) 清家剛，三牧浩也，原祐介，小田原亨，永田智大，寺田雅之：まちづくり分野におけるモバイル空間統計の活用可能性に係る研究，日本都市計画学会・都市計画論文集，Vol.46，No.3，pp.451-pp.456，2011.
- 3) 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，第 51 回土木計画学研究発表会・講演集，2015.
- 4) 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：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分析ー都市交通分野への適用に向けてー，第 52 回土木計画学研究発表会・講演集，2015.
- 5) 清家剛，三牧浩也，森田祥子：モバイル空間統計を活用した都市拠点地区の人口特性分析に係る研究
- 6) 昼夜間を通じて都市の実態人口分布を捉える新たな統計手法，日本建築学会計画系論文集，Vol.80，No.713，pp.1625-1633，2015.
- 7) 佐藤忠典，力石真，藤原章正：モバイル空間統計を用いた復旧・復興過程の独立成分分析：2014 年 8 月広島土砂災害を例に，第 55 回土木計画学研究発表会・講演集，2017.
- 8) 有村幹治，鎌田周，浅田拓海：マイクロジオデータの統合化による建物用途別メッシュ入込人口の推計，土木学会論文集，Vol.72，No.5，pp.515-522，2016.
- 9) 山口裕通，三宅寛之，中山晶一郎：非負値テンソル因子分解を用いた旅行需要の北陸新幹線開業に対する感度の計測，第 59 回土木計画学研究発表会・講演集，2019.
- 10) 鈴木新，山口裕通，福田大輔：データ融合による日別・旅行目的別都市間旅客流動量推計，運輸政策研究，21 巻，2019.
- 11) 高橋央亘，浅田拓海，有村幹治：北海道新幹線開業前後における函館エリア来訪者数の比較分析，土木学会論文集 D3 (土木計画学) ,VOL.74,No.5 (土木計画学研究・論文集第 35 巻) I\_827-I\_835,2018.
- 12) 田中敦士，岡本直久，鈴木俊博，浅野礼子，白川洋司：人口分布統計データを活用した観光地の特性把握，第 54 回土木計画学研究発表会・講演集，2016.
- 13) Lee, D. D. and Seung, H. S. : Algorithms for nonnegative matrix factorization, Advances in Neural Information Processing Systems 13 (NIPS 2000), pp.556-562,2000.
- 14) 澤田宏：非負値行列因子分解 NMF の基礎とデータ/信号解析への応用，第 13 回 ITS シンポジウム電子情報通信学会誌 Vol.95，No.9，pp.829-833，2015.
- 15) 亀岡弘和：非負値行列因子分解とその音響信号処理への応用，日本統計学会誌 Vol.44，No.9，pp.383-pp.407，2015.
- 16) 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

(?受付)

## ANALYSIS OF VISITORS TO LOCAL CITIES USING MOBILE SPATIAL STATISTICS

Chihiro ISHIBASHI and Makoto TSUKAI