

複数のビッグデータを用いた 観光者数推定手法に関する基礎的研究

西垣 友貴¹・Jan-Dirk SCHMÖCKER²・山田 忠史³・中尾 聡史⁴

¹学生会員 京都大学大学院工学研究科 (〒615-8540 京都府京都市京都大学桂)

E-mail:nishigaki@trans.kuciv.kyoto-u.ac.jp

²正会員 京都大学大学院工学研究科 (同上)

E-mail:schmoecker@trans.kuciv.kyoto-u.ac.jp

³正会員 京都大学経営管理大学院 (工学研究科併任) (同上)

E-mail: yamada.tadashi.2x@kyoto-u.ac.jp

⁴正会員 京都大学大学院工学研究科 (同上)

E-mail:nakao@trans.kuciv.kyoto-u.ac.jp

近年、観光の活発化に伴う混雑により、観光者などに不満が広がっている。しかし、経済効果を考慮すると、観光を発展させつつ問題を解決することが必要である。観光研究では、ビッグデータが注目を集めているが、その種類ごとに弱点を持つと指摘されている。そこで、本研究では、弱点を補い合う形で複数のビッグデータを活用し、観光・交通研究の基礎となる観光者数の推計を試みた。京都市を対象に、メッシュデータに基づく観光者数を、GPSデータを用いて、重回帰分析と階層線形モデル(HLM)により推計した。重回帰分析でも一定の良好な結果が得られたが、季節変動を考慮したHLMの方がより良い結果が得られた。このことから、ビッグデータを組み合わせることで、各データの弱点を補い合えること、GPSデータが推計に与える影響に季節変動があることが確認できた。

Key Words : *tourism, big data, mesh data, GPS data, hierarchical linear model*

1. 研究の背景・目的

近年、観光が活発になっており、観光客数も急増している。それに伴い、観光地における交通渋滞の深刻化などが原因で観光客のみならず、近隣住民にも不満が広がりつつあるのが現状である¹⁾。しかしながら、わが国における観光事業の経済効果は大きく、今後も観光事業を発展させていくことは重要なことである。この現状を踏まえると、観光の需要や行動をマネジメントして、上述のような不満を解消し、予防することが不可欠である。

観光の需要や行動のマネジメントに関しては、数多くの研究がなされている。この分野の課題として、非日常の行動であるがゆえに、パーソナリティ調査では扱うのが難しく、その都度アンケート調査によってデータを収集する必要がある点が挙げられる。それゆえ、大規模なサンプルの確保や多時点のデータの収集が困難となっている。また、観光行動中にアンケート調査を実施することが多く、被験者の観光行動を阻害してしまうことも課題の一つとなっている。

そこで近年、ビッグデータが注目を集めている。通信

技術の向上やスマートフォンの普及により、端末の位置情報を大規模かつ連続的に自動で収集することが可能になっている。この優位性に着目した研究も数多く実施されている。

一方で、ビッグデータにはいくつかの種類が存在しており、それぞれがメリットとデメリットを有している。例えば、株式会社ドコモ・インサイトマーケティングが提供しているモバイル空間統計では3次メッシュ内に滞在している人口を1時間ごとに把握することができる。このデータのメリットとしては、集計データとして提供されているため、個人の同意の有無に関わらず大規模なサンプルを取得できる点が挙げられる。デメリットとしては、個人を特定できないようにするために、決められた範囲内の集計データとしてのみ提供される点が挙げられる。

一方、アプリの利用者などのIDとGPSデータを、同意を得た上で収集しているものもある。このデータのメリットとしては、詳細な位置座標を把握でき、さらに経時的な行動を追跡することが可能である点が挙げられる。デメリットとしては、対象者が限定的なので、サンプル

数が比較的少なく、偏りがある点が挙げられる。

これら以外にも様々な種類のビッグデータが存在し、それぞれにメリットとデメリットが存在している。この点について、実際に生じうる課題に目を向けてみる。例えば、京都市では観光行動についてのアンケートを実施しており、その際に図-1示すような観光エリアを定義している。このように、観光需要や行動をマネジメントする場合には、エリアを定義することも想定されるが、メッシュデータだけではこのエリア定義に合致せず、思う様な使い方ができない。逆に、GPSデータだけでは、サンプル数が少なく偏りもあるため、やはり課題は残ることになる。これを踏まえると、ビッグデータが有するデメリットに対して、複数のデータを組み合わせるなどして、適切な処理を施すことが望ましい。

ビッグデータを活用した研究は多く存在しているが、そのほとんどの研究において、一種類のビッグデータを活用しているか、パーソントリップ調査をはじめとしたアンケート調査の結果と組み合わせている。しかし、アンケート調査と組み合わせるためには、当然ながら、アンケート調査を実施するというコストがかかり、規模を

大きくすればする程、そのコストは大きくなっていく。

そこで、本研究では、複数のビッグデータを組み合わせる結果を導出することを目的とする。具体的には、観光地ごとの時間帯別の来訪者数の推計を試みる。

観光、さらには、交通に関する研究において、各ノードの人口の把握が、最初に試みられる作業であり、その結果に基づいて、その他の研究が進められる場合が多い。それを踏まえ、本研究では、来訪者数の推計に着目する。この結果は魅力度の推計や交通量推計などにも活用することができると考えられる。また、本研究では日本でも有数の観光施設が複数存在している京都市を対象とする。また、後述するデータの性質上、日本人のみを対象としている。

2. 既往研究の整理と本研究の位置付け

観光需要や行動に関する研究は数多く存在している。本章では、扱っているデータによって分類・整理しながら、既往研究のレビューを行うとともに、本研究の位置付けを述べる。

(1) アンケート調査を用いた研究

高橋ら³⁾は、函館市が行った観光パーソントリップ調査の結果を用いて、観光客の来訪者数の算出を試みている。また、溝上ら⁴⁾は、熊本県内の主要観光施設を対象にアンケート調査を行い、観光客の来訪者数の推計に必要となる観光地の魅力度の定量化を試みている。

これらの研究が代表するように、アンケート調査を用いて来訪者数や観光地の魅力度を定量化する研究は、数多く存在している。しかしながら、先述の通り、アンケート調査を実施する場合、コストがかかってしまうため、自動的に収集できるビッグデータを活用して、同様の結果が得られる方が望ましいと考えられる。

(2) メッシュデータを用いた研究

金森ら⁵⁾は、観光が対象ではないものの、メッシュデータであるモバイル空間統計とその他のデータを組み合わせ、OD交通量を算出している。なお、ここで用いられているその他のデータとは、バスの乗降者数と小サンプルのGPSデータとなっている。このGPSデータとはビッグデータではなく、協力者を募って収集したデータである。菊池ら⁶⁾は、モバイル空間統計を用いて、パーソントリップ調査の時点更新の手法を提案している。結果として、単純に人口拡大のみで時点更新をした場合よりも、モバイル空間統計を用いて時点更新した場合の方が大幅に精度が向上することを報告している。ただし、この手法はパーソントリップ調査自体の代替になるもの

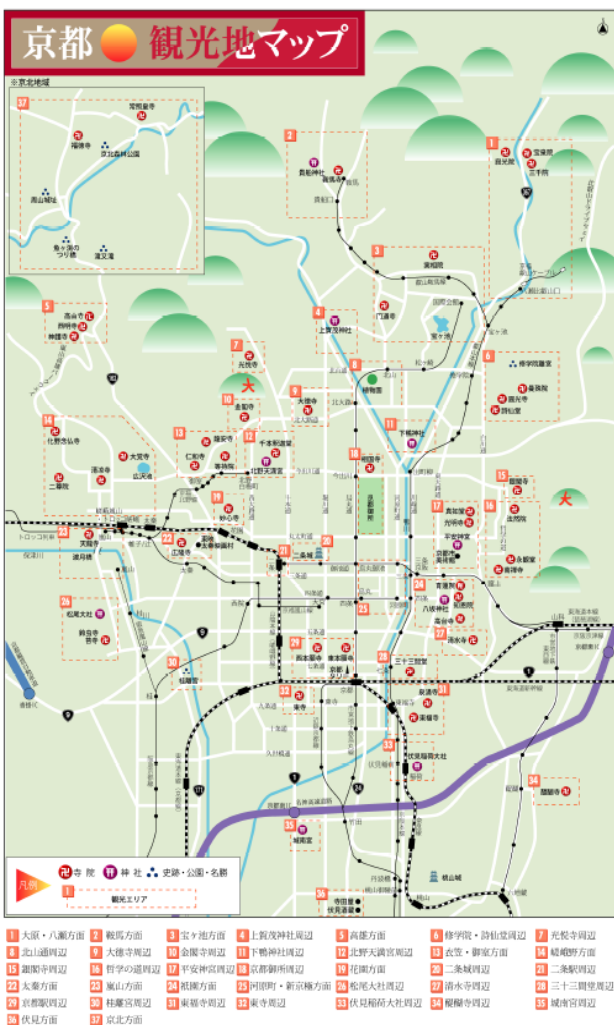


図-1 京都市観光エリアマップ (出典：京都市)

ではなく、依然として大きなコストをかけて、パーソントリップ調査を続けていくことが前提となっている。川上ら⁷⁾は、モバイル空間統計に基づいて算出された人口流動統計を用いて、観光地間のOD交通量の算出を行っている。ここでは、モバイル空間統計のデータを別の方法で集計したデータを用いているが、集計方法の性質上、大きな誤差が発生し得ることを指摘している。

これらの研究が代表するようにメッシュデータを用いた観光や交通の研究は数多く存在しているが、集計データであるがゆえのデメリットの存在も指摘されている。一方で、他のデータと組み合わせることを試みた研究も存在しているが、ビッグデータと組み合わせている例は確認できない。

(3) GPSデータを用いた研究

生形ら⁸⁾は、ゼンリンデータコム社が提供する混雑統計データを用いて、石川県域および石川県加賀市域への観光客の来訪頻度の分析や、滞在期間・時間、主要観光スポットの月別来訪客数、周遊行動の実態などの分析を行っており、概ね良好な結果が得られたと結論付けている。しかしながら、研究の課題として、GPSデータが有する属性の偏りについて指摘している。井坪ら⁹⁾は、31人の被験者に対し、webダイアリー調査を行うとともに、携帯電話を所持してもらうことにより、GPSの追跡を実施し、従来の紙媒体でのアンケート調査との比較を行った。その結果、紙媒体でのアンケート調査よりも様々なメリットが確認できたと結論付けている。

このように、GPSデータに関しては、ビッグデータのみならず、研究者が必要な数の被験者に協力を依頼してデータを収集することが容易であることも、一つの特徴であると考えられる。また、サンプル数が膨大に確保できた場合においては、来訪頻度や滞在時間の分析など、様々な方法で活用できることが確認できたが、やはりデータの偏りが課題となっており、この点に対して何らかの対処が必要であることも確認できた。

(4) 本研究の位置付け

(1)~(3)から、観光だけでなく様々な交通行動の推計に関する研究があることが確認でき、また、そのデータソースとしては、パーソントリップ調査をはじめとしたアンケート調査やGPSデータ、メッシュデータなどのビッグデータ用いられていることが分かる。特に、近年の研究においては、何らかのビッグデータや、サンプル数が小さい同じ性質のデータを用いている場合が多く見受けられた。しかしながら、それぞれの研究において、用いているデータ特有の課題を抱えており、そのことを研究の課題として挙げているものが多く見受けられ、複数のビッグデータを、それぞれのデメリットを補い合う形で

活用している事例は見られなかった。これらのことから、本研究では複数のビッグデータを用いて、様々な研究の基礎となる来訪者数の推定を試みることにする。

3. 利用データ

本章では、本研究で用いる各種データ（メッシュデータ、GPSデータ、公共交通データ）について概説する。

(1) メッシュデータ

本研究では、株式会社ドコモ・インサイトマーケティングが提供しているモバイル空間統計というデータを用いる。このデータは、ドコモの基地局エリア内に存在する携帯端末の台数を集計し、その集計値をドコモの普及率に基づいて滞在人口へと拡大したものである。その後、メッシュごとの人口を再集計する。また、端末保有者が居住する都道府県や市町村ごとの集計も可能である。本研究では、この都道府県ごとの集計結果を用いており、観光客を抽出するという目的の下、京都府在住者のデータを除外している。また、本研究では、観光客が対象であるため、対象の時間帯を8時から23時としている。

データの収集期間など、その他のデータ概要を表-1に示している。また、京都市内の集計結果を図-2に示す。なお、本研究では、対象を日本人に限定しているが、これは、このデータが集計データであることに起因する。個人の追跡を不可能にするために、メッシュ内の人口が10人未満であれば0に修正される。すなわち、集計対象となる外国人が元々少なく、日本人と同じように集計す

表-1 メッシュデータの概要

取得期間	2018.10.17, 20, 21, 24, 27, 28 2018.11.14, 17, 18, 21, 23, 24, 25 2018.12.5, 8, 9, 12, 15, 16, 19, 22, 23, 24 2019.1.9, 12, 13, 14, 16, 19, 20
データ内容	日時、メッシュコード、居住地（都道府県）、メッシュ人口
対象者	京都府以外に在住している日本人

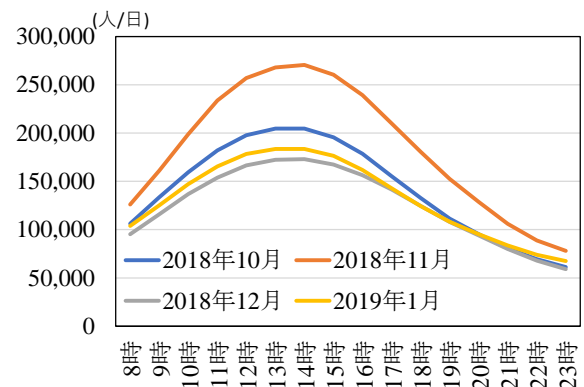


図-2 メッシュデータに基づく京都市内観光者数

ると、秘匿される情報が多くなるため、集計方法が異なっている。このことから、日本人と外国人では異なるアプローチが必要となるため、本研究では日本人のみを対象としている。

(2) GPSデータ

本研究では、経路検索システムである、「歩くまち京都」アプリの利用者のGPSデータを用いる。このアプリを起動した時と、その後のバックグラウンド処理中にデータが取得される。取得されるデータは、時刻、座標、IDなどである。データの収集期間など、その他のデータの概要を表-2に示している。表-2の対象者については、次の手順で分類を行っている。個人IDを基にして、個人ごとに集計を行い、京都市内に訪れた日数と、図-1に示す観光エリアのうち、訪れているエリアの数を算出したのち、エリア数を日数で除する。その値が1.0以上であるデータを観光者として、本研究の対象として扱うこととする。このような手順で分類した対象者のデータを用いて、各時間帯における、各メッシュ内のユニークユーザー数を集計したものが、図-3である。なお、比較のため、メッシュデータと同じメッシュで集計を行っている。

(3) 公共交通データ

本研究では、これらのビッグデータだけでなく、公共交通のデータも併せて用いる。具体的には、電車とバスの運行データに基づく一般化費用を活用する。

$$C_{lt} = \min_k \left[\frac{d_l}{v^k} + \frac{60}{2f_{lt}^k} + \frac{F_l^k}{\alpha} \right] \quad (1)$$

表-2 GPSデータの概要

取得期間	2018.10.1~2019.1.31
データ内容	個人ID, 日時, 緯度, 経度, 使用言語
対象者	「歩くまち京都」アプリの利用者のうち、使用言語が日本語で条件*を満たす者 ※基準については本文中にて説明

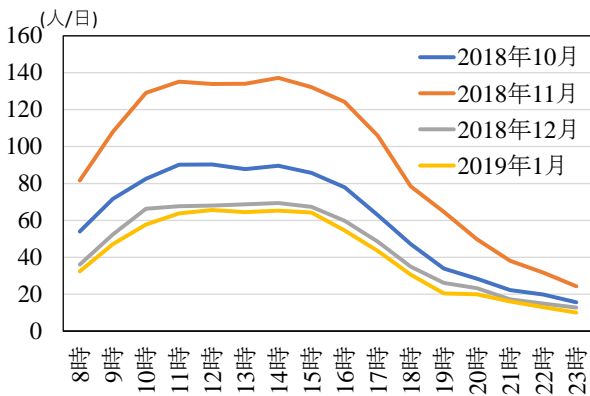


図-3 GPSデータによる京都市内のユニークユーザー数

$$C_{ijt} = \min_n [\sum_l \delta_l^{ijtn} C_{lt}] \quad (2)$$

$$C'_{at} = \frac{C_{at}}{C_{ijt}} \quad (3)$$

ここに、

C_{lt} : 時間帯 t におけるリンク l の最小一般化費用[min]

d_l : リンク l の距離[m]

k : 交通手段 (徒歩, バス, 電車)

v^k : 交通手段 k の移動速度[m/min]

f_{lt}^k : リンク l , 時間帯 t における交通手段 k の運行頻度 [本/hour]

F_l^k : リンク l における交通手段 k の運賃[円]

α : 時間価値[$\text{円}/\text{min}$]

C_{ijt} : 時間帯 t における二地点 ij 間の最小一般化費用 [min]

δ_l^{ijtn} : 時間帯 t において ij 間の n 番目のルートにリンク l が含まれていれば1, そうでなければ0となるダミー変数

C'_{at} : 時間帯 t における地点 a から全ての駅やバス停への移動コストの平均[min]

式(1)は、存在し得る全てのリンクの一般化費用を算出する式である。 k は交通手段を表しており、同一のリンク l に複数の手段が存在する場合、そのうち最もコストが小さいものを選択することを示している。例えば、バス停 a からバス停 b に移動する場合、徒歩で移動することも可能であるため、徒歩での一般化費用とバスでの一般化費用の小さい方を選択することになる。

式(2)は、二地点間を結ぶ全てのルートのうち、一般化費用が最小になるルートを選択するための式であり、ダイクストラ法を用いて最小値の算出を行う。式(3)は、駅(バス停) i から全ての駅(バス停)への移動コストの平均と、メッシュ a の中心から駅(バス停) i への移動コストとの和の平均を算出する式である。各変数に代入する具体的な数値を表-3に示している。

運行頻度や運賃に関しては、関連サイトから手作業でデータ収集を行った。運行頻度に関しては、存在する全ての組合せを正確に収集するには、膨大な時間を要するため、簡略化したものを用いる。各ルートの始発駅(もしくは、バス停)のみについて、一時間ごとの運行本数をカウントして、それをその路線を代表する運行頻度としている。しかし、始発駅だけを考慮しているため例えば、始発駅を22時59分に出発し、それが最終電車であった場合、そのルートのリンクは、23時には存在しないことになり、現実とそぐわない。また、9時ちょうどに出発し、8時台に電車が一本もない場合、8時台においては、

表-3 各変数の値

v^{walk}	70[m/min]	v^{bus}	300[m/min]
v^{train}	700[m/min]	f_{lt}^{walk}	∞
$f_{lt}^{bus}, f_{lt}^{train}$	路線ごとに設定		
F_l^k	路線ごとに設定		
α	29.8[円/min]		

深夜のような全く電車が運行していない時間帯と同様の扱いになってしまうため、これもまた現実とそぐわない。そこで、本研究では、前後1時間を含めた3時間ごとの運行頻度の平均を考慮することで、上記の課題に対処する。また、平日、土曜日、日曜日の三パターンの頻度を収集しており、それらの代表する値として、一週間の平均値を使用した。

運賃に関しても、存在する全ての組合せを正確に収集するには膨大な時間を要するため、簡略化したものを用いている。バスに関しては、京都市内の一部範囲内では一律230円となっているため、本研究では、京都市内全域においてこの値段としている。範囲外では230円よりも安い区間もあれば高い区間もあるが、総じて平均的な値段になっていると考えられる。電車に関しては、運営会社ごとに基準が異なり、それを公表していない会社がほとんどである。しかし、京都市営地下鉄に関しては、明確な基準を公表しており、会社ごとに極端な差が存在しているわけではない。それゆえ、簡単のため、全てのルートで京都市営地下鉄の基準に当てはめて運賃を算出している。なお、時間価値に関しては、加藤ら¹⁰が報告している値を引用している。

4. 観光地ごとの来訪者数の推計

(1) 重回帰分析

本章では、第3章で概説したデータを用いて、来訪者数の推計を試みる。具体的には、サンプル数が多く概ね真値であると考えられるドコモデータに基づくメッシュ人口を被説明変数とし、その他変数を説明変数として重回帰分析を行った。その結果を表-4に、変数の説明を表-5にそれぞれ示す。なお、重回帰分析に投入する変数については、あらかじめ相関係数を求めておき、相関係数の絶対値が0.3を超えない変数のみを投入している。

まず、Case1として、GPSデータのみを用いて推計を行った。その結果、決定係数が0.773であり、良好な結果が得られた。しかし、結果の誤差について確認すると、京都駅を含むメッシュで、大きな誤差が生じていた。京都駅は、多くの観光者にとっての観光の起終点になり得る場所であり、他のメッシュとは異なる性質を有することが原因であると考えられる。また、本研究で使用したGPSデータの特長として、経路検索アプリを開いたタイミングで情報が収集される。このことも踏まえると、京都駅に着いて、その後の交通手段の確認をするという行動も十分想定されるため、このような結果になっていると考えられる。つぎに、Case2として、公共交通データを追加して推計を行った。公共交通データとして、式(3)で定義した C_{at}^i をそのまま用いた。

その結果、決定係数が0.774となり、わずかではあるが改善が見られた。このことは、京都駅を含むメッシュのアクセシビリティが、周囲よりもかなり大きくなることに起因すると考えられる。一方、わずかしか改善しなかった一因として、観光施設に接続しない路線も考慮していることが考えられる。公共交通のデータのみを活用するのではなく、観光施設へのアクセシビリティを算出して用いることで、改善が見られるものと推察できる。また、Case2においても、京都駅を含むメッシュの誤差が大きくなっていることが確認できた。そこで、Case3として、表-4に示す通り x_4^{imt} を追加して推計を実施した。その結果、決定係数が0.798となり、さらなる改善が見られた。

これらの結果から、GPSデータを用いると、メッシュ人口を高精度で推計できることや、京都駅のような多くの観光者にとっての起終点となり、他のメッシュと大きく異なる性質を持つ場所については、それを考慮できる変数を加えることが望ましいものと推察される。

表-4 重回帰分析の結果

	Case1			Case2			Case3		
	β	S.E.	t	β	S.E.	t	β	S.E.	t
定数項	-	1.66	24.5**	-	6.73	25.0**	-	6.37	30.6**
x_1^{imt}	0.879	4.14	440**	0.871	4.20	430**	0.822	4.16	408**
x_2^{it}	-	-	-	-3.96E-2	1.85E-2	-19.5**	-4.64E-2	1.76E-2	-24.2**
x_4^{imt}	-	-	-	-	-	-	0.162	21.6	82.3**
R^2	0.773			0.774			0.798		

*:p<0.05, **:p<0.01

表-5 重回帰分析で用いた変数の説明

x_1^{imt}	GPSデータから算出されるメッシュ <i>i</i> , 月 <i>m</i> , 時間帯 <i>t</i> の一日あたりの人口
x_2^{it}	公共交通データから算出されるエリア <i>i</i> , 時間帯 <i>t</i> の移動コストに関する変数($x_2^{it} = C_{at}^i$)
x_3^{imt}	x_1^{imt} からメッシュごとの平均値を引いた値
δ^i	エリア <i>i</i> に京都駅が含まれていれば1そうでなければ0のダミー変数
x_4^{imt}	x_3^{imt} のうち京都駅を含むメッシュ以外の値を0にしたもの($x_4^{imt} = x_3^{imt} \times \delta^i$)

表-6 階層線形モデルの結果

		Case3			Case4		
		β	S.E.	t	β /分散	S.E.	t
固定 効果	定数項	-	6.37	30.6**	-	12.9	12.7**
	x_1^{imt}	0.822	4.16	408**	0.905	161	11.6**
	x_2^{it}	-4.64E-2	1.76E-2	-24.2**	-3.91E-2	1.67E-2	-21.4**
	x_4^{imt}	0.162	21.6	82.3**	0.151	308	5.36**
変量 効果	定数項	-	-	-	6.62E+2	-	-
	x_1^{imt}	-	-	-	1.04E+5	-	-
	x_4^{imt}	-	-	-	3.76E+5	-	-
AIC		834573.9			828673.7		
RMSE		364.56			351.97		

*: $p < 0.05$, **: $p < 0.01$

(2) 階層線形モデル

重回帰分析を用いた推計では、概ね良好な結果が得られた。しかしながら、京都市の観光の実態を考慮すると、季節によって、人気のある観光地や行動が変化し得る可能性がある。すなわち、重回帰分析で得られる回帰係数が季節によって異なる可能性が考えられる。そこで、季節による回帰係数の差を考慮するために、変数を2018年10月、11月、12月、2019年1月の4つの集団に分割して、階層線形モデルを用いた推計を試みる。

まずは、階層線形モデルを適用することが妥当かどうかの確認を行う。これには、清水¹⁾がまとめている基準を参考にする。被説明変数の集団間の誤差を全体の誤差で除することで算出できる級内相関 (ICC) が0.1を超えている場合や、ICCに基づいて算出されるデザインエフェクト (DE) が2.0を超えている場合には、階層線形モデルの適用は妥当と判断される。これらの数値を算出した結果、ICCは9.32E-4、デザインエフェクトは14.31となった。ICCは基準を大きく下回る結果となっているが、DEは基準を大きく上回っている。これらの基準は、どちらかを満たしていれば、階層線形モデルの適用が妥当であると言えるので、本研究においても適用可能と判断した。

階層線形モデルを適用する際は、各変数の回帰係数が集団ごとに異なるかどうか、すなわち、変量効果を仮定するかどうかを設定する必要がある。そこで、先述のCase3で用いた変数に関して、変量効果を仮定する場合としない場合で考えられる全ての組み合わせに関して、推計を行った。そのうちAICが最小であった組み合わせをCase4として、推計結果を表-6に示している。なお、階層線形モデルを扱う際には、説明変数をそのまま用いるのではなく、集団ごとの平均値を減じて、集団中心化を行った変数を用いるのが望ましいとされているため、本研究でも、集団中心化を行った変数を投入している。

表-6を見ると、移動コストに関する変数以外に関しては、変量効果が認められており、その結果、重回帰分析の結果よりも、AICとRMSEともに小さな値となってい

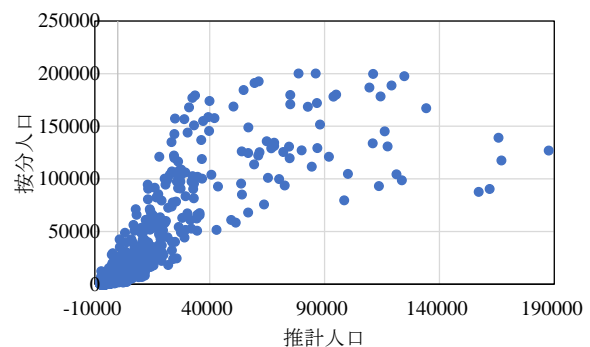


図-4 按分人口と推計人口の比較

る。このことから、移動コストが人口推計に与える影響に季節変動はないが、GPSデータが人口推計に与える影響には季節変動が存在しており、季節変動を考慮したモデルを構築することで、より良い結果が得られることが確認できた。

また、Case4で得られたモデルを用いて、図-1で示したエリアごとの人口推計を行った。推計結果と比較するために、メッシュデータの人口をGPSデータ分布に従って按分し、エリアの人口を算出した。その結果を図-4に示す。図-4より、大きく外れた推計結果ではないことが見受けられる。基本的に按分した人口の方が大きい値を示しているが、これについては、単純に按分した値と推計した値を比較していることから、総数が一致しないためであると考えられる。この差を考慮したとしても、大きく値が異なっているエリアが存在している。また、推計値が負値となってしまっているエリアも見受けられる。これらについては、モデル式の構築の際に、1km四方のメッシュだけを考慮しているが、推計には様々な面積のエリアが存在していることが一因と考えられる。改善の余地が見られるものの、全体の傾向としては概ね良好な結果が示唆されたものと考えられる。

5. 結論と課題

本研究では、観光需要や行動のマネジメントに向けて注目を集めるビッグデータに関して、複数のビッグデー

タを用いることで、それぞれのデメリットを補い合うことができることを示しつつ、観光者数の推計を試みた。メッシュデータとGPSデータを用いたが、メッシュデータには、メッシュ単位でしか人口を把握できないというデメリットがあり、GPSデータには、サンプル数が比較的少なく、偏りがあるというデメリットがある。そこで、GPSデータを用いてメッシュデータを推計するモデルを構築し、そのモデルを活用することで、メッシュデータと同規模の任意のエリアの人口が把握できる可能性を示した。また、モデルを構築するにあたって、京都駅のような、他の場所とは異なる性質を持った場所に対しては、その点を考慮できる変数を導入することで、推計結果を向上させることができた。観光者数の推計に際しては、GPSデータの影響が季節により変動しており、階層線形モデルを用いて、その変動を考慮することで、より良い推計結果が得られることを示唆した。

これらの結果は、観光交通計画などにおいて、今後ビッグデータをさらに活用していくうえで、有効な知見となると考えられる。また、このモデルを活用して、任意のエリアごとの観光者数が推計できれば、エリアごとの魅力度の算出や交通量の算出などにも寄与できる可能性がある。

今後の課題として、まずは、メッシュデータを真値として扱っているが、その検証が十分にできていないことが挙げられる。メッシュデータを被説明変数として投入する前に施すべき処理があるのか、ある場合どのような処理を行えば良いのかについて、さらなる検討を加えていくことが必要である。また、公共交通のデータを推計に用いたが、精度をわずかに向上させるに留まった。公共交通のデータだけを活用するのではなく、観光施設との接続を考慮したデータを用いることで、より精緻に推計できる可能性がある。按分人口と比較した際に、大きく相違したエリアが存在しており、按分人口と推計人口のどちらに原因があるかは判断できないものの、いずれにしても改善の余地がある。さらに、推計人口が負値となる場合があることについても、改善の余地がある。本

研究では、1kmメッシュの人口のみを推計しており、エリアごとの人口を推計する場合には、より大きな面積のエリアの人口を推計している。面積の違いが推計結果にどの程度の影響をもたらすのかについても検証する必要がある。

参考文献

- 1) 宮崎耕輔, 谷本圭志, 森山昌幸: 観光客の過剰な流入による地域住民の外出控えに関する一考察-離島におけるケーススタディ, 農村計画学会誌, 特別号, Vol.34, pp.231-236, 2015.
- 2) 今泉博国: 持続可能な地域づくりと観光—プロプラー・ツーリズムの視点からオーバーツーリズムを考える—, 経済学論纂, Vol.60, No.5-6, pp.91-106, 2020.
- 3) 高橋清, 五十嵐日出夫: 観光スポットの魅力度を考慮した観光行動分析と入込み客数の予測, 土木計画学研究・論文集, Vol.8, pp.233-240, 1990.
- 4) 溝上章志, 森杉壽芳, 藤田素弘: 観光地域魅力度と観光周遊行動のモデル化に関する研究, 都市計画論文集, Vol.27, pp.517-522, 1992.
- 5) 金森亮, 水野敬太, 野田五木樹, 中島秀之: 滞留人口データを利用した交通手段別 OD 交通量推計手法の提案, 情報処理学会研究報告, Vol.8, pp.1-6, 2015.
- 6) 菊池雅彦, 岩館慶多, 羽藤英二, 茂木渉, 加藤昌樹: 交通ビッグデータによる実用的な都市圏 PT 調査マスターデータの時点更新, 土木学会論文集, Vol.74, No.5, pp.667-676, 2018.
- 7) 川上陸, Jan-Dirk Schmöcker, 宇野伸宏, 中村俊之: モバイル空間統計のデータ特性を考慮した OD 推計手法: 京都観光地間流動におけるケーススタディ: 土木学会論文集, Vol.75, No.6, pp.379-391, 2020.
- 8) 生形嘉良, 関本義秀, Teerayut Horanont: 大規模・長期間の GPS データによる観光統計調査の活用可能性~ 石川県を事例に~, 土木学会論文集, Vol.69, No.5, pp.345-352, 2013.
- 9) 井坪慎二, 羽藤英二, 中嶋康博: 情報技術の活用による交通行動調査の効率化・高度化に関する研究, 第 31 回土木計画学研究・講演集, CD-ROM, 2005.
- 10) 加藤浩徳, 橋元稔典. "我が国の旅客交通時間価値に関するメタ分析." 第 38 回土木計画学研究・講演集 CDROM, 2008.
- 11) 清水裕士: 個人と集団のマルチレベル分析, ナカニシヤ出版, 2014.

A NEW APPROACH TO ESTIMATE THE NUMBER OF TOURISTS WITH SEVERAL BIG DATA

Tomoki NISHIGAKI, Jan-Dirk SCHMÖCKER, Tadashi YAMADA, Satoshi NAKAO