

住宅土地統計のダウンスケーリングによる 地域メッシュ単位の世帯所得データの開発

山中惇矢*・古谷貴史**・秋山祐樹***

Development of Regional Mesh-Based Household Income Data Through Downscaling of Housing and Land Survey

Junya YAMANAKA*・Takafumi FURUYA**・Yuki AKIYAMA***

Spatial statistics for small areas such as city blocks or grid cells are information useful for understanding the current situation of the area and implementing various planning. However, since many existing statistics are not available in a small area basis, methods for downscaling them into small area units are important. Therefore, in this research, Therefore, in this study, we developed a method for estimating the number of households by income group by municipality published in the Housing and Land Survey in subregions (Cho-cho-moku) and 250 m grid units by combining population census. In addition, we applied the same method to all over Japan to develop a national data set. Finally, by integrating these results with mobile phone big data, we have developed a prototype of a new spatial statistic that can estimate the average income of visitors per 500 m grid.

Keywords: 統計(statistics), ダウンスケーリング(downscaling), 住宅土地統計調査(housing and land survey), 国勢調査(population census), 小地域(small region), 地域メッシュ(regional mesh)

1. はじめに

町丁字や地域メッシュ単位の小地域を対象とした空間統計は、地域の現状を詳細に把握する上で有用な情報であり、これまでも様々な分野の数多くの研究で利用されてきた(梶田, 2008; 小西, 2010)。その一方で、近年では情報通信技術の高度化や高性能なコンピュータの普及などに伴い、建物単位や人単位といった従来の空間統計よりも時空間的に高精細な空間情報(空間ビッグデータやマイクロジオデータ)が利用可能になりつつあり(秋山, 2019)、既存の空間統計をこれらの高精細な空間情報と統合して利用する際には、その空間的粒度が十分ではなくなりつつある状況といえる。

とりわけ近年のスマートフォンの普及により、膨大な人々の移動履歴のビッグデータ(以下「モバイルビッグデータ」)が利用可能になった。また、これらの情報に空間統計を組み合わせると、各ユーザの様々な属性(年齢・性別・職業・所得など)を推定し、その結果から新しい空間統計を開発する試みも

見られるが(松島ほか, 2015; Cheng et al., 2017)、既存統計の集計単位による空間的粒度の制約が、これらの取り組みにおいて課題になっている。一方、携帯電話の契約者情報が利用できれば各ユーザの属性は容易に把握できるものの、このような個人情報に関わるデータは利用上の制約の壁が極めて高く、ほとんど利用できない状況にある。(佐藤, 2016)

そこで本研究では、以上の課題を鑑み個人情報に係わらないオープンデータを組み合わせることで、既存統計の空間的粒度を高精細化(ダウンスケーリング)する技術を提案する。本研究で対象とする統計は、市区町村単位の結果しか公開されていない住宅・土地統計調査とし、同統計を地域メッシュ単位で集計された国勢調査と組み合わせることで、住宅・土地統計調査を地域メッシュ単位にダウンスケーリングする。また同手法を日本全国に適用し、全国データの整備を行う。最後に、モバイルビッグデータと組み合わせることで、各ユーザに住宅・土地統計調査の属性を与えて、新しい空間統計を試作する。

* 学生会員 東京都市大学工学部都市工学科 (Tokyo City University)

〒158-0087 東京都世田谷区玉堤 1 丁目 2 8-1 Tel : 03-5707-0104 E-mail : g1818100@tcu.ac.jp

** 学生会員 東京都市大学大学院総合理工学研究科 (Tokyo City University)

*** 正会員 東京都市大学建築都市デザイン学部都市工学科 (Tokyo City University)

2. 使用する統計

本研究では住宅・土地統計調査を地域メッシュ集計（5次メッシュ集計）された国勢調査と組み合わせることで、地域メッシュ単位にダウンスケーリングする。以下にそのために使用する統計を示す。なお、本研究では住宅・土地統計に収録された統計のうち、世帯の年間収入に関する情報をダウンスケーリングするものとする。

1) 住宅土地統計調査（平成30年）

「第44-4表 世帯の年間収入階級（9区分）・世帯の種類・住宅の所有の関係」

2) 国勢調査（平成27年）

・小地域集計 第7表 住居の種類・住宅の所有の関係（6分類）

・町丁・字等別境界データ

・国勢調査に関する地域メッシュ統計（250mメッシュ）

・世界測地系（250mメッシュ）その1人口等基本集計に関する事項

・都道府県・市区町村別主要統計表

なお、住宅・土地統計調査は、市、区及び人口1万5千人以上の町村を対象としているため、それ以外の地域は未収録となるが、平成30年の場合、日本全国の1,741自治体のうち、1,068自治体（61.3%）、世帯数にすると全国約5,400万世帯のうち、約5,200万世帯（96.5%）の範囲の情報を収録しており、日本全国の世帯年間収入の空間分布を網羅的に把握するのに十分な量の統計を含んでいるといえる。

3. データ開発の流れ

本研究では、5次メッシュごとの世帯・個人所得を算出する。図1にデータ開発の流れを示す。以下に図1中①～③の処理方法について説明する。

3.1 町丁字毎の年間収入階級別世帯数の推定

まず、住宅・土地統計調査第44-4表「世帯の年間収入階級（9区分）・世帯の種類・住宅の所有の関係」から、住宅の所有関係別の世帯の年間収入階級別の世帯数（市区町村別）を取得する。次に国勢調査小地域集計第7表「住居の種類・住宅の所有の関係（6

分類）」から、住宅所有形態別の世帯数（町丁字別）を取得する。これらの情報に基づき、小地域（町丁字）毎の年間収入階級別世帯割合を推定する。市区町村*i*の町丁字*j*の住宅所有形態*k*の年間収入階級*l*の世帯数は式1で定義する。

$$N_{ijkl} = \frac{h_{ikl}}{\sum h_{ik}} H_{ijk} \quad (1)$$

$\sum h_{ijk}$ は住宅・土地統計から得られる市区町村*i*の住宅所有形態*k*の世帯総数、 h_{ikl} は住宅・土地統計から得られる市区町村*i*の住宅所有形態*k*の年間収入階級*l*の世帯数、 H_{ijk} は国勢調査から得られる市区町村*i*の町丁字*j*の住宅所有形態*k*の世帯数、そして N_{ijkl} は市区町村*i*の町丁字*j*の住宅所有形態*k*の年間収入階級*l*の世帯数である。

以上の計算を住宅・土地統計調査に収録された日本全国の全ての市区町村に適用することで、町丁字毎の年間収入階級別世帯数のデータを整備することが実現した。

3.2 5次メッシュ毎の年間収入階級別世帯数の推定

まず、国勢調査町丁・字等別境界データと、国勢調査に関する地域メッシュ統計（5次メッシュ）から、町丁字と5次メッシュのポリゴンデータを抽出する。続いて、町丁字のポリゴンデータと5次メッシュのポリゴンデータを空間結合（ユニオン）し、



図1 データ開発手順

生成された領域の面積を算出する。続いて、面積按分を行い、町丁字別の年間収入階級別世帯数を 250m メッシュ別に世帯数割合に変換する。例えば図 2 のように、メッシュ m 内に 3 つの町丁字集計された住宅・土地統計調査の結果が分布した場合を考えてみる。町丁字 A, B, C の境界とメッシュで分割された領域の面積がそれぞれ S_A , S_B , S_C であり、メッシュの面積が S_m だとする。また、町丁字 A, B, C の収入階級 l の世帯割合がそれぞれ RI_A , RI_B , RI_C であるとする。以上を用いてメッシュ m の収入階級 l の世帯の割合 RI_m は式 2 で与えられる。

$$RI_m = \frac{S_A}{S_m} RI_A + \frac{S_B}{S_m} RI_B + \frac{S_C}{S_m} RI_C \quad (2)$$

一方、国勢調査の「世界測地系 (250mメッシュ) その 1 人口等基本集計に関する事項」から、5 次メッシュごとの世帯数 H_m を把握することができるため、メッシュ m の収入階級 l の世帯数 HI_m は式 3 で与えられる。

$$HI_m = RI_m H_m \quad (3)$$

以上の処理を一般化すると、メッシュ m の収入階級 l の世帯数 HI_m は式 4 で与えられる。

$$HI_m = \left(\sum_{k=1}^n \frac{S_{m_k}}{S_m} R_{ml_k} \right) H_m \quad (4)$$

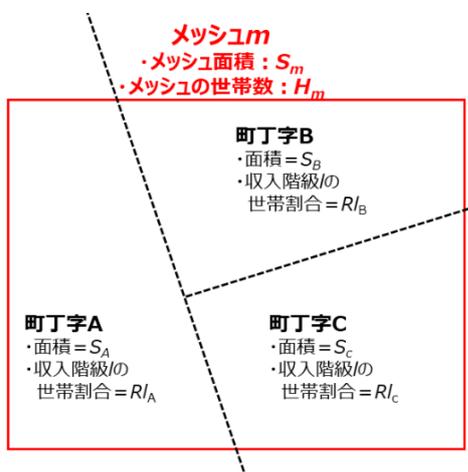


図 2 同一メッシュ内に複数の町丁字が分布する例

3.3 5 次メッシュ毎の平均世帯収入および 1 人あたり平均収入の推定

最後に、5 次メッシュごとの平均世帯収入と 1 人あたり平均収入を推定する。住宅・土地統計調査から得られる収入階級は表 1 の通りである。メッシュ m における平均収入 \bar{I}_m は式 5 で与えられる。

$$\bar{I}_m = \sum_{k=1}^n \bar{I}_{m_k} r_{mk} \quad (5)$$

\bar{I}_{m_k} はメッシュ m の収入階級 k の平均収入、 r_{mk} はメッシュ m の収入階級 k の世帯割合である。なお、収入階級が 1500 万円以上の世帯は国民生活基礎調査を参考に、 \bar{I}_{m_k} を 1804.545 万円とした。

最後に 1 人あたりの平均収入も算出する。先の計算で 1 世帯あたりの平均収入は明らかになっており、また国勢調査の「世界測地系 (250mメッシュ) その 1 人口等基本集計に関する事項」から、メッシュごとの世帯数および人口も明らかなので、メッシュ m の 1 人あたりの平均収入 \bar{i}_m は式 6 で与えられる。

$$\bar{i}_m = \bar{I}_m \frac{P_m}{H_m} \quad (6)$$

表 1 世帯の年間収入階級区分と世帯数および世帯割合 (日本全国 2018 年)

年間収入階級	世帯数 [万世帯]	世帯割合 [%]
100 万円未満	192.4	3.7
100~200 万円	369.2	7.1
200~300 万円	473.2	9.1
300~400 万円	535.6	10.3
400~500 万円	572.0	11.0
500~700 万円	1097.2	21.1
700~1000 万円	1081.6	20.8
1000~1500 万円	650	12.5
1500 万円以上	228.8	4.4
合計	5200.0	100.0

H_m はメッシュ m の世帯数, P_m はメッシュ m の人口である. なお, 以上の計算を町丁字単位で推計した結果に適用することで, 町丁字単位の平均世帯収入および1人あたり平均収入の推定も可能である.

以上の処理により, 町丁字単位および5次メッシュ単位の平均世帯収入および1人あたり平均収入の推定が実現した. 図3および4は本研究の手法で推定した町丁字ごとの平均世帯収入の推定結果と, 5次メッシュ別の居住者1人あたり平均収入の推定結果を示す. 図3より, 東京都心とその近郊に高収入の世帯が数多く存在し, 東京都心から離れるにつれて概ね世帯収入が低くなるのが分かった. また, 図4が示すように世帯当たり収入を1人あたり収入に, 町丁字毎を5次メッシュ毎に細分化することで, 一部の高収入者の空間分布を詳細に把握できるようになった.

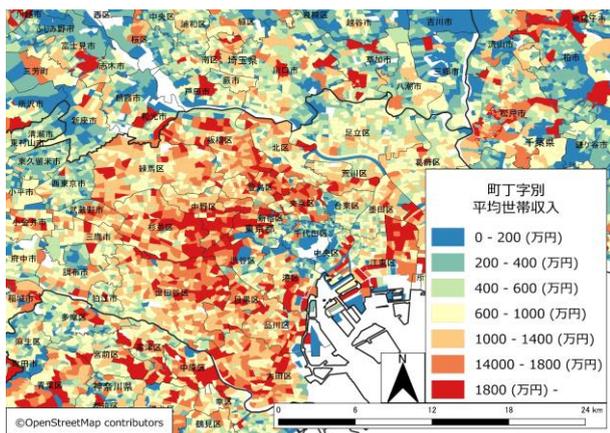


図3 平均世帯収入の推定結果
(東京23区とその周辺・町丁字集計)

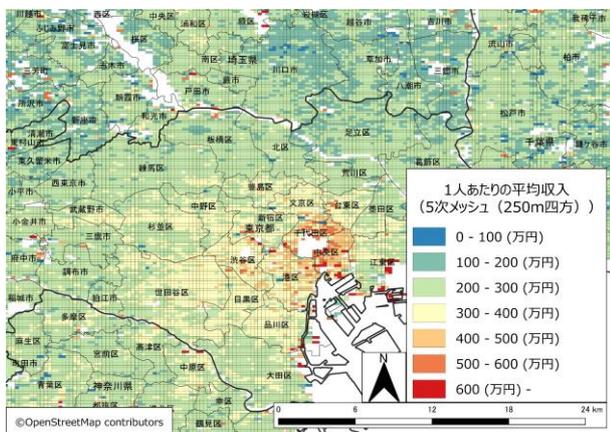


図4 居住者1人あたり平均収入の推定結果
(東京23区とその周辺・5次メッシュ集計)

4. 新しい空間統計の提案

1章においてモバイルビッグデータを用いた新しい空間統計を開発しようとする試みについて, その際に利用可能な空間統計の空間的粒度がその障害になっていることを紹介したが, 本研究で開発した高精細な年収推計データを用いることで, この障害を突破し新しい空間統計を提案することも可能である.

そこで, 本研究では2019年の東京都全域の「ポイント型流動人口データ」(株式会社Agoop)と, 本研究で開発した5次メッシュごとに推計した1人あたりの推定収入データを組み合わせることで, 任意の時間における来訪者の平均収入分布を推定する方法を提案する. 同結果は出店計画や店舗における在庫管理, 広告の計画などのマーケティング戦略上重要な情報となるだけでなく, 将来的にはこれらの情報を時空間的にクラスタリングすることで, 新しいジオデモグラフィクスの開発にも繋がる可能性がある (Dalton and Thatcher, 2015). なお, 本研究ではまず本研究で提案する手法により以下の一連のデータ処理が可能であることを検証するために, 2019年10月7日~10月11日の土日祝日を含まない月~金曜日のデータのうち, 東京都が居住地であると考えられるユーザのデータを用いてデータの開発を行った.

4.1 モバイルビッグデータの処理

まず, モバイルビッグデータから得られる各ユーザの移動履歴の情報を Akiyama et al. (2016)の手法で処理し, 各ユーザの居住地を推定する. この処理により各ユーザの居住地が経度緯度で推定されるため, 本研究で開発した5次メッシュ集計の1人あたりの収入情報を各ユーザに付与することができる. また, 各ユーザの居住地を5次メッシュと対応付けることができるため, メッシュの人口をそのメッシュに居住地が分布すると推定されたユーザ数で除すことで, 各ユーザが何人の人口を代表しているか(拡大係数)を明らかにすることも可能である.

続いて, 以上の結果を任意の時間・空間で集計することで, その時間の任意の空間における来訪者の平均収入を明らかにすることができる. 時間 t における空間 s の平均収入 I_{ts} は式7で与えられる.

$$I_{ts} = \frac{\sum_k M_{tsk} i_{tsk}}{\sum_k M_{tsk}} \quad (7)$$

M_{tsk} と i_{tsk} はそれぞれ時間 t の空間 s に来訪した人 k の拡大係数と収入である。

以上の方法で東京都における 12 時、19 時、23 時の東京都民の来訪による来訪者の平均所得の変化を 4 次メッシュ (500m メッシュ) で表現した結果を図 5、図 6、図 7 に示す。

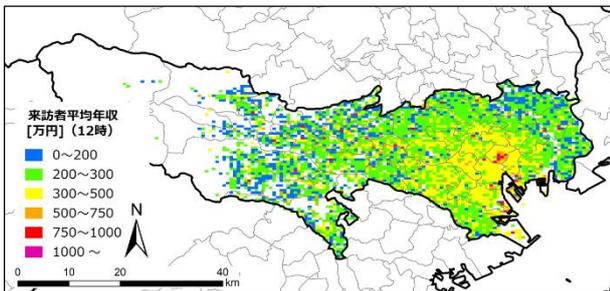


図 5 12 時における東京都民の来訪者による平均所得の空間分布 (2019 年)

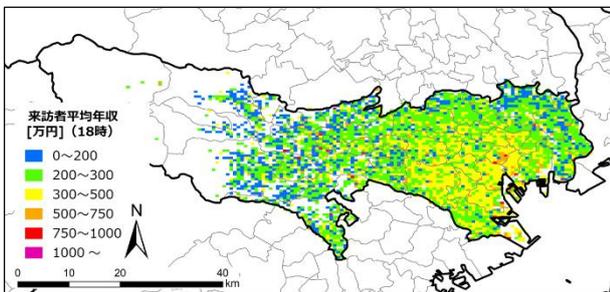


図 6 19 時における東京都民の来訪者による平均所得の空間分布 (2019 年)

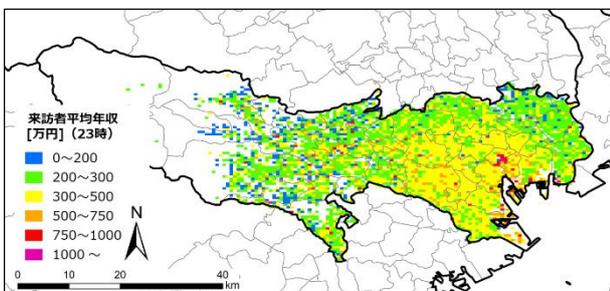


図 7 23 時における東京都民の来訪者による平均所得の空間分布 (2019 年)

12 時は多くの人々が勤務地に滞在していることから、概ねその勤務地付近の企業に務める人の平均的な給与が推定できる。また、19 時は帰宅途中であると同時に夕食の時間帯でもあり、23 時はナイトライフに伴うアクティビティが活発な時間帯であることから、これらの活動に伴い、特にどの地域に高額所得者が集積し、活発な消費活動を行っているのかということを推定することができる。

まず、図 5 より各年収区分が空間的に似通った地域に分布していることが確認できることから、同程度の収入を持つ人々 (あるいは同程度の収入を給与として与える企業) が空間的に集積することが分かる。また、図 6 からは、各年収帯の分布が散らばっている時間帯であることから、多くの場所で帰宅に伴う移動が行われていることが推測される。さらに、図 7 をみると図 5 と同様に類似する収入区分の人々の空間的集積が確認できる。このことから図 7 の時間帯においても、類似する収入を持つ人々が時空間的に集積することが明らかとなった。

以上より、時間帯に応じて同じ地域 (メッシュ) でも時間により来訪する人々の平均年収は大きく変化することが明らかになった。これらの結果から時間毎・時間帯毎や曜日・季節の違いなどを細分化して分析することにより、メッシュ単位というミクロな領域において来訪者の収入の時系列的な変化をミクロにモニターすることが可能である。これらの結果は例えば出店計画、広告戦略、在庫管理などのマーケティング関連分野はもちろんのこと、他にも交通、防災、都市地域計画などにも応用ができる可能性があるといえる。

5. おわりに

本研究では、既存統計等のオープンデータを組み合わせることで、既存統計の空間的粒度を高精細化 (ダウンスケーリング) する技術を提案した。即ち、市区町村単位の結果しか公開されていない住宅・土地統計調査を、地域メッシュ単位で集計されている国勢調査と組み合わせることで、地域メッシュ単位にダウンスケーリングする技術を開発した。また同手法を日本全国に適用し、全国データの整備を行っ

た。さらに、この成果をモバイルビッグデータと組み合わせることで、各ユーザに住宅・土地統計調査の属性を与えて、4次メッシュ単位で集計することで任意の時間におけるメッシュごとの来訪者の平均収入を推定できる新しい空間統計を試作した。

本研究で提案・開発した既存統計のダウンスケーリング技術は、様々な空間統計や空間情報のダウンスケーリングに応用が可能であり、またこのような高精細な空間情報は様々な分野の分析に汎用的に利用可能である(上杉ほか, 2010)。また、所得に関する空間情報は、地域の特徴を分析する上で重要な指標の一つであることは既存研究でも指摘されており(豊田, 2021)本研究で開発したデータのさらなる発展は学術的にもその需要は高いものと期待される。さらに、本研究のように5次メッシュという高い空間的粒度で既存統計をダウンスケーリングすることで、本研究で試作したように、空間的に非常に高精細なモバイルビッグデータと結合することも実現する。この延長には来訪者の属性に応じた時空間的なクラスタリングによる新しいジオデモグラフィクスの実現の可能性もある(宮本ほか, 2018)。

本稿で紹介した内容は前述の通り依然として検討すべき課題は数多いが、これらの課題を順次解決していくことで、信頼性の高い統計データのダウンスケーリング技術の実現と、モバイルビッグデータと組み合わせた新しいダイナミック(時系列)データの提案を進めていきたい。

謝辞

本研究は科研費・挑戦的研究(萌芽)JSPS(19K21660)の助成を受けて実施した研究成果である。また本研究で使用したモバイルビッグデータ「ポイント型流動人口データ」を株式会社Agoopより提供を受けた。本以上、ここに記して謝意を表したい。

参考文献

秋山祐樹(2019)ビッグデータは何を語るか?〜地理学×ビッグデータの可能性とその将来展望〜, 「地理空間」, 12(3), 159-178.
上杉昌也・浅見泰司(2010)小地域の所得分布推計

および転居の所得要因分析, 東京大学空間情報科学研究センターディスカッションペーパー, 103.
梶田真(2008)国勢調査における小地域統計の整備過程とその利用可能性, 「東京大学人文地理学研究」, 19, 31-43.
小西純(2010)現状把握のための小地域統計データの利用と共有-情報共有媒体としての地方公共団体統計ホームページ-, 「日本統計研究所」, 33-48.
佐藤一郎(2016)ビッグデータと個人情報保護法 データシェアリングにおけるパーソナルデータの取り扱い, 「情報管理」, 58(11), 828-835.
豊田哲也(2021)市町村別データを用いた住宅地地価の形成要因分析 所得の地域格差と世帯密度の影響に注目して, 「日本地理学会発表要旨集」, 2021s(0), 123.
松島敏和・橋本浩良・高宮進(2015)移動履歴データと外生データを組み合わせた移動目的判別システムの提案, 「第52回土木計画学研究発表会・講演集」, 52, CD-ROM.
宮本旺周・仙石裕明・清水千弘(2018)電話帳データを用いた全国の地域クラスタリング, 「第27回地理情報システム学会講演論文集」, C46.
Akiyama, Y., Ueyama, S., Shibasaki, R. and Adachi, R. (2016) Event Detection Using Mobile phone Mass GPS Data and Their Reliability Verification by DMSP/OLS Night light Image, *ISPRS Annals of the photogrammetry, Remote Sensing and Spatial Information Sciences*, III-2, 77-84.
Cheng, X, Fang, L, Hong, X and Yang, L. (2017) Exploiting Mobile Big Data : Sources, Features, and Applications, *IEEE Network*, 31(1), 72-79.
Dalton, C. M and Thatcher, J. (2015) Inflated granularity: Spatial “Big Data” and geodemographics, *Big Data & Society*, 2(2), doi.org/10.1177/2053951715601144.