

複数のデータベースの組み合わせによる 人口統計データの高度化について

大佛 俊泰

近年、携帯電話の位置情報に基づく人口統計データがさまざまな分野で活用されている。ただし、プライバシー保護の観点から詳細な属性情報などは秘匿されており、また、滞留中の建物用途や滞留目的などの情報も所与のデータから直接知ることはできない。本稿では、現在利用可能な人口統計データに備わる特徴を整理し、各データの長所を活かし短所を補いながら統合することで、滞留中の建物用途、滞留目的、性別・年齢、居住地などの情報を人口統計データに付与する方法について論述する。

キーワード：人口統計、滞留人口、建物用途、滞留目的、居住地、時空間分布

1. はじめに

1.1 地理空間情報の醍醐味

多くの地理情報システム (GIS) では、リレーショナルデータベース (RDB) と呼ばれる、複数の表形式のデータベースを組み合わせて活用するデータモデルが採用されている。たとえば、ベクトルデータの場合、地物の空間情報 (建物の位置や形状など) と属性情報 (建物の用途、階数、面積など) は別々のデータベースとして管理され、両者は地物に固有の識別子 (キー) を用いて統合され利用される。一般に属性情報は多数のデータベースに分けて収集・更新されることが多いことから、RDB はデータベース管理の容易さにおいて優れていると考えられている。

一方、GIS において利用可能な情報は、データベースとして整備された所与の情報だけではない。たとえば、住宅の位置情報と店舗の位置情報を用いれば、各住宅にとって最も近い店舗を検出することができる。このとき、住宅と店舗は、両者を直接対応付ける固有の「キー」ではなく、位置情報 (空間情報) に基づき対応付けられる。すなわち、データベースに共通する「キー」が備わっていなくても、空間情報を活用することで複数のデータベースを関連付けることができる。ここに地理空間情報の醍醐味がある。

本稿では、空間情報のみならず、時間情報も活用して、複数のデータベースを統合し、データベースの価値を高める方法について概説する。具体的な分析方法

や分析結果の詳細については文献 [1] を参照されたい。

1.2 人口統計データの特徴と限界

都市計画や地域防災計画などについて検討するうえで、都市内滞留人口の時空間分布に関する情報は、基礎データの一つとして非常に重要である。しかし、滞留人口の空間分布は時々刻々と変動しており、特に、モビリティの高い都市圏においては、その動態を精緻に把握することは容易ではない。

これまでに活用されてきた人口統計データには、それぞれ優れた特徴が備わる反面、限界や制約も存在する。たとえば、国勢調査に基づく人口統計は、これまで夜間人口分布を表すデータとして長く活用されてきたが、夜間人口に基づく議論には限界がある。また、パーソントリップ調査データ (PT データ) は、詳細な属性情報と都市内移動の発着地・発着時刻・移動手段・移動目的などの情報を備えており、動的な人口分布を把握するデータとして優れている。しかし、サンプル率が 2% 程度のアンケート調査に基づくデータであり、また、おおむね 10 年に 1 度の調査であるため即時性は乏しい。

近年、携帯電話の位置情報 (基地局情報) から、都市内滞留人口 (本稿では、混乱を生じない範囲で移動人口を含めて滞留人口と呼ぶこととする) の時空間分布を推定した人口統計データが活用されつつある。全国の任意の場所について毎時のデータが整備されており、サンプル率が高いことから、汎用性の高いデータとしてさまざまな分野での活用が期待されている。株式会社 NTT ドコモが提供する人口統計の一つであるモバイル空間統計[®] は、携帯電話ネットワークを利用して携帯電話の台数を集計した値に、NTT ドコモの携帯電話の普及率を加味することで推計した人口統計

おさらぎ としひろ

東京工業大学環境・社会理工学院

〒152-8550 東京都目黒区大岡山 2-12-1-M1-25

osaragi.t.aa@m.titech.ac.jp

データ（1時間ごとの平均人数を算出した人口分布統計）である。2020年3月現在における国内のサンプルサイズは約8,000万台であり、サンプル率は50%を超えている。また、その精度に関する検証も試みられている [2]。しかし、プライバシー保護の観点から詳細な属性情報は秘匿される（データから削除される）ことが多く、また、移動者と滞留人口を区別することもできない [3]。

本稿では、モバイル空間統計[®]の即時性・継続性（毎時獲得可能な点）、PTデータの詳細な情報（年齢・性別・職業・滞在目的・移動手段など）、国勢調査人口統計の完全性（悉皆調査に基づく精緻な国勢調査に基づく点）を活かしながら統合することで、時空間人口統計データを高度化する方法について論述する。

1.3 携帯電話人口統計データに関する既往研究

清家ら [2] は携帯電話人口統計の基礎的な特性把握を試みており、有村ら [4] は建物用途別の延床面積と棟数を用いて入込人口を推定している。また、矢部ら [5] はシミュレーションとパーティクルフィルタを同化させることで災害時の人流推定を行い、瀬戸ら [6] は熊本地震時における避難所の混雑度推計を行うなど、国内における応用分野は拡大している。海外では Calabrese et al. [7] が携帯電話のGPS情報を用いて、イタリア・ローマ市内のリアルタイム監視システムの構築を試みている。また、Dewille et al. [8] は国勢調査などの既存人口統計との比較をとおして、携帯電話人口統計の可能性について論じている。そのほかにも、携帯電話の位置情報を活用して人々の都市内活動をモニタリングしようとする研究が広く試みられている。

携帯電話人口統計データに備わる特性を直接活用した研究事例は多く、また、国勢調査データと衛星データとを関連させようとする試みはあるものの、複数のデータを統合することで、データの価値を高めようとする視点からの取り組みは十分ではない。

2. 人口統計データの統合方法の視点

即時性・継続性に優れたモバイル空間統計[®]を基礎データとして、これにPTデータの詳細な滞留目的や属性情報などの情報を統合する方法について述べる。また、モバイル空間統計[®]で秘匿された滞留人口を国勢調査人口統計の完全性に基づき補完する方法について述べる。これら3種類の人口統計データの統合方法の概念を図1に示してある。

モバイル空間統計[®]とPTデータには、両者を直接関連付けるための「キー」は備わっていない。そこで、

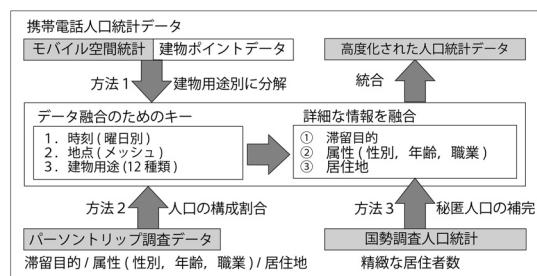


図1 人口統計データの統合方法の概要

本稿では「時刻」「地点」「建物用途」の組み合わせを「キー」として統合を図る（図1左）。これら3要素に着目する理由は、たとえば、平日の昼間（時刻）、丸の内（地点）の事務所系建物（建物用途）における滞留人口の属性や滞留目的について考えると、その多くは30～50歳代の就業者である。また、大岡山（地点）の学校施設（建物用途）であれば、20歳前後の学生と20～60歳代の教職員が多数を占める。一方、渋谷駅前（地点）に立地する飲食店（建物用途）であれば、滞留人口属性はさまざまであるが、曜日や時刻（時刻）に応じて特徴を有すると考えられる。

すなわち、「時刻」「地点」「建物用途」を限定すれば、そこに滞留している滞留人口の属性や滞留目的には、ある種固有の特性が備わっていると考えられる。本稿では、人口統計データを「時刻」「地点」「建物用途」ごとに分割し、その組み合わせに潜む滞留人口の特有（滞留目的や属性情報など）について考えることで、人口統計データの統合を図る方法について検討する。

3. 滞留人口の建物用途別への分解（方法1）

3.1 滞留人口分解モデルの定式化

モバイル空間統計[®]には滞留中の建物用途に関する情報が備わっていない。そこで、モバイル空間統計[®]から得られる各メッシュの滞留人口を建物用途別に分解する方法について考える（図1の方法1）。具体的には、PTデータとGISデータを組み合わせ、任意の時刻・エリアの滞留人口を推定した文献 [9] を応用する。

時刻 t 、メッシュ i における滞留人口 y_i^t [人] は、同一時刻、同一メッシュ内の建物用途 j の建物内外の滞留人口 y_{ij}^t [人] の合計として以下のように書ける。

$$y_i^t = \sum_j y_{ij}^t \quad (1)$$

滞留人口 y_{ij}^t [人] は、メッシュ i における建物用途 j の建物の合計延床面積 x_{ij} に比例すると仮定し、時

刻別単位面積当たりの滞留人口密度 D_{ij}^t [人/100m²] を用いて次式で記述する。

$$y_{ij}^t = D_{ij}^t x_{ij} \quad (2)$$

このとき、滞留人口密度 D_{ij}^t は、時刻 t 、建物用途 j に依存する「時刻変動成分」 α_j^t と、メッシュ i だけに依存する「場所性成分」 β_i の積で表現できると考える。

$$D_{ij}^t = \alpha_j^t \beta_i \quad (3)$$

未知パラメータ α_j^t 、 β_i は、重回帰分析を用いた収束計算により推定することができる。詳細は文献 [9] を参照されたい。

3.2 分析対象地域と使用データ

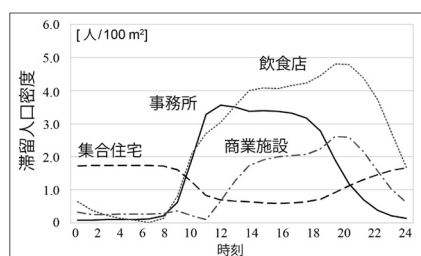
分析対象地域は東京都 23 区とし、目的変数である滞留人口 y_i^t [人] には、平日と休日のモバイル空間統計[®]を用いる（メッシュサイズは 500 m×500 m）。説明変数であるメッシュ i における建物用途 j の合計延床面積 x_{ij} は、建物ポイントデータ（株式会社 ZENRIN が提供する建物属性情報 [用途・延床面積] を有する POI (Point of Interest) データ）を用いて構成する。ただし、複数の用途が共存する複合施設などに関しては、用途別件数に基づき延床面積を用途別に案分した。また、住宅を含む複合施設に関しては、東京都における集合住宅の戸当たり平均延床面積（住宅・土地統計調査：総務省統計局）を参考として算出した。鉄道駅に関しては、国土数値情報鉄道データをもとに算出した。また、公園は建物ではないが、滞留人口の時空間分布を記述するうえで重要であると考え採用し、東京都 GIS データ（土地利用現況調査）から面積を算出した。

以上の準備の下で、平日・休日別に滞留人口分解モデルを推定した。おおむねモデルの記述精度は良好であり、特に、12:00、18:00 の推定精度は決定係数が 0.95 以上であり優れていることを確認した。

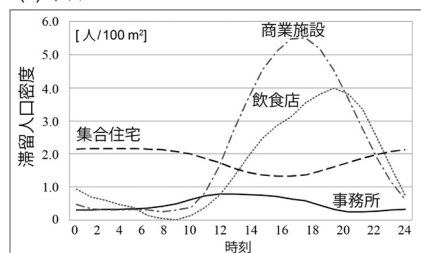
3.3 時刻変動成分・場所性成分パラメータの解釈

時刻変動成分パラメータ α_j^t の推定値の一部を図 2 に示してある。床面積当たりの滞留人口密度の時刻変動は、建物用途に応じて、また、平日・休日に応じて、大きく異なることがわかる。たとえば、平日の事務所系建物では 8:00 過ぎに密度が急上昇し、18:00 過ぎに緩やかに下降する。集合住宅はこれと相反する変化を示している。また、飲食系建物では 20:00 頃にピークとなる。一方、休日には事務所系建物の滞留人口密度は平日の 1/4~1/5 程度であり、また、商業系建物のピークは 17:00 頃であり平日の 2.5 倍程度である。

場所性成分パラメータ β_i の推定値を図 3 に示して



(a) 平日



(b) 休日

図 2 建物用途別滞留人口密度 α_j^t の時刻変動

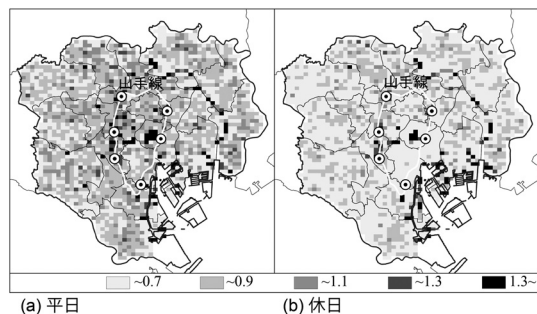


図 3 場所性成分 β_i の空間分布

ある。平日の β_i の値は、東京駅周辺地域などの商業業務集積地で高く、滞留人口密度の時刻変動幅はほかの地区と比較して相対的に大きいことを示している。また、平日と休日を比較すると、休日の β_i の値は平日よりも変動幅が小さく一様に分布している。すなわち、式 (3) が示すように、滞留人口密度に及ぼす場所性の影響の程度は、休日よりも平日の方が大きいことがわかる。

3.4 滞留人口の建物用途別への分解

上記の滞留人口分解モデルを用いて、モバイル空間統計[®]の滞留人口を建物用途別の滞留人口に分解する（モデルによる建物用途別滞留人口の推定値の大きさに応じてモバイル空間統計[®]の滞留人口を案分する）。平日と休日における 14:00 の建物用途別滞留人口の空間分布の一部を図 4 に示してある。平日の (i) 滞留人口（モバイル空間統計[®]）が、(ii) 都心部の事務所系建物、(iii) 周縁部の独立住宅、(iv) 商業系建物に、それぞ

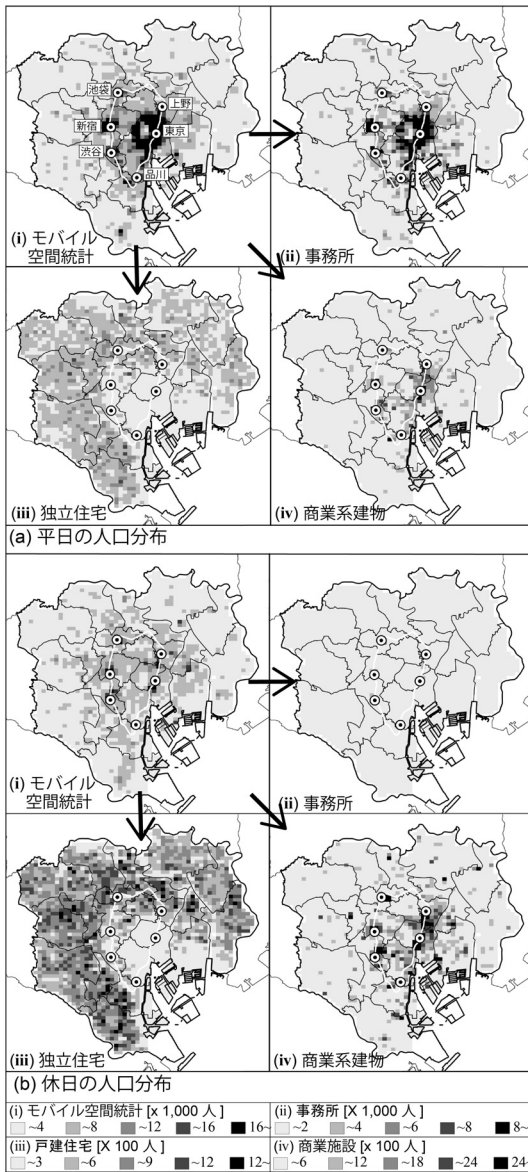


図4 滞留人口の建物用途別への分解例

れ分解されていることがわかる。休日を見ると、(iii) 周縁部の独立住宅の滞留人口は多く、在宅者が平日より多いことが現れている。また、(ii) 都心部の事務所系建物で滞留人口は少なく、(iv) 都心部および大規模商業施設が立地するメッシュの商業系建物に滞留人口が集中している。

4. 人口統計データの統合

4.1 PT データの前処理

PT データは、町丁目を 10 数個合算した程度（夜間人口が 15,000 人程度）の小ゾーンと呼ばれるエリア

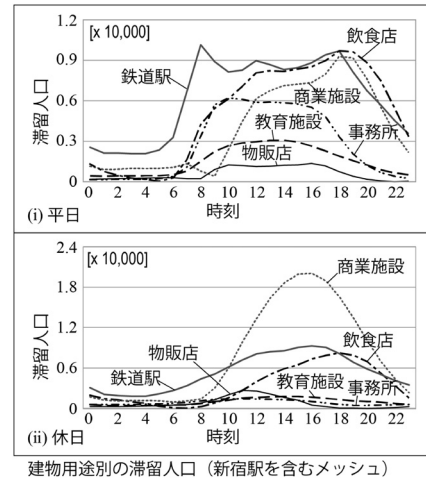


図5 特定メッシュの建物用途別滞留人口

で集計されている。そこで、時刻 t 、小ゾーン s 、建物用途 j 、属性情報 k の滞留人口 M_{sjk}^t を、用途別建物面積に応じてメッシュ i に案分することで滞留人口 y_{ijk}^t を算出した。ただし、移動者については、小ゾーンの代表点を頂点とする TIN (Triangulated Irregular Network) ネットワーク上の最短経路を一定速度で移動していると仮定し、各時刻における通過中の小ゾーンを求めた [10]。次に、小ゾーン内の移動者をメッシュに案分し、建物周辺の移動者とみなして、上記の方法で推定した滞留人口 y_{ijk}^t の大きさに応じて滞留人口 y_{ijk}^t に合算した。また、休日の PT データについては、文献 [11] を用いて推定し、同様の前処理を施した。

4.2 建物用途別・目的別滞留人口（方法 2 ①）

滞留建物の用途が同一でも滞留目的は異なる可能性がある。そこで、PT データから時刻 t 、メッシュ i 、建物用途 j 、滞留目的 l の滞留人口 n_{ijl}^t を求め、時刻 t 、メッシュ i 、建物用途 j を「キー」とする目的別滞留人口割合 m_{ijl}^t を以下のように求めた。

$$m_{ijl}^t = \frac{n_{ijl}^t}{\sum_l n_{ijl}^t} \quad (4)$$

算出した m_{ijl}^t を建物用途別滞留人口 y_{ij}^t に乗じれば、建物用途別・目的別滞留人口 Y_{ijl}^t [人] を求めることができる。

$$Y_{ijl}^t = m_{ijl}^t y_{ij}^t \quad (5)$$

新宿駅を含むメッシュの計算例を図 5 に示してある。平日には多様な建物用途に多くの人々が滞留しており、その時刻変動は大きい。しかし、休日には滞留人口の時刻変動は穏やかであり、商業系建物における変動の

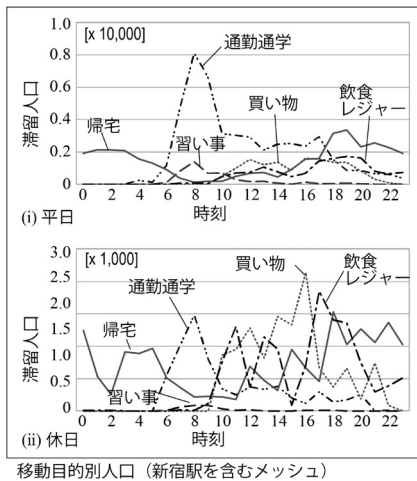


図6 特定メッシュの鉄道駅における移動目的別滞留人口

みが卓越している。

具体的な建物用途として鉄道駅に着目し、滞留人口を目的別に分類した例を図6に示してある。平日の8:00には通勤・通学目的の滞留人口が卓越しており、18:00以降には帰宅目的の滞留人口が多くなる。また、休日には食事、娯楽、買い物など多様な目的を有する滞留人口の時刻変動が見て取れる。

4.3 建物用途別・性別年齢別滞留人口（方法2②）

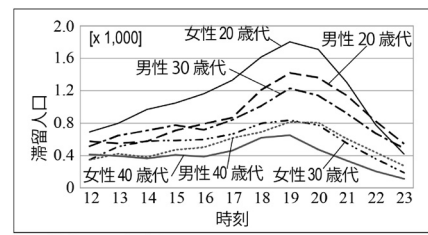
PTデータから時刻 t 、メッシュ i 、建物用途 j 、性別年齢階層 u の滞留人口 s_{iju}^t を求め、その滞留人口割合 r_{iju}^t を算出する。

$$r_{iju}^t = \frac{s_{iju}^t}{\sum_u s_{iju}^t} \quad (6)$$

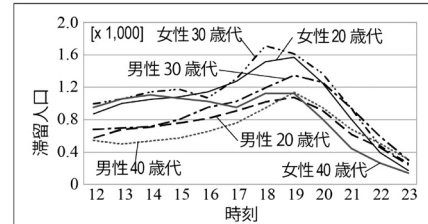
建物用途別滞留人口 y_{ij}^t に r_{iju}^t を乗じれば、建物用途別・性別年齢階層別滞留人口 Y_{iju}^t [人]を求めることができる。ただし、モバイル空間統計[®]でも時刻 t 、メッシュ i 、性別年齢階層 u の滞留人口 y_{iu}^t が得られている。このデータと整合性を保って推定するためには、次式を用いて推定すればよい。

$$Y_{iju}^t = \frac{y_{iu}^t}{\sum_j r_{iju}^t y_{ij}^t} r_{iju}^t y_{ij}^t \quad (7)$$

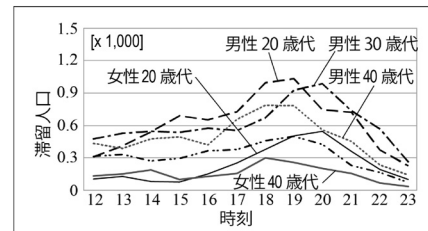
特定のメッシュに注目し、平日の商業系建物における性別年齢階層別滞留人口の時刻変動の例を図7に示してある。渋谷109を含むメッシュの商業系建物では20～30代の男女が多く、銀座駅を含むメッシュでは20～30代の女性が多い。また、秋葉原駅を含むメッシュの商業系建物では、20～40代の男性が多く、特に夕刻に多くなるなど、場所の特性に応じた滞留人口の



(a) 渋谷109を含むメッシュ



(b) 銀座駅を含むメッシュ



(c) 秋葉原駅を含むメッシュ

図7 特定メッシュの商業施設における属性別滞留人口

性別・年齢階層の違いが読み取れる。

4.4 建物用途別・居住地別滞留人口（方法2③）

PTデータから時刻 t 、メッシュ i 、建物用途 j 、居住地 h の滞留人口 v_{ijh}^t を求め、建物用途別居住地別滞留人口割合 w_{ijh}^t を算出する。

$$w_{ijh}^t = \frac{v_{ijh}^t}{\sum_h v_{ijh}^t} \quad (8)$$

建物用途別滞留人口 y_{ij}^t に w_{ijh}^t を乗じることで、建物用途別居住地別滞留人口 Y_{ijh}^t [人]を求めることができる。モバイル空間統計[®]でも時刻 t 、メッシュ i 、居住地 h の滞留人口 y_{ih}^t が得られていることから、これと整合するように、次式を用いて推定する。

$$Y_{ijh}^t = \frac{y_{ih}^t}{\sum_j w_{ijh}^t y_{ij}^t} w_{ijh}^t y_{ij}^t \quad (9)$$

ただし、モバイル空間統計[®]における時刻 t 、メッシュ i 、居住地 h の滞留人口 y_{ih}^t については、居住地 h の分類数が多いことからデータは秘匿されやすく、原データを直接活用することには問題がある。そこで、次節では、滞留人口 y_{ih}^t を補完する方法について検討する。

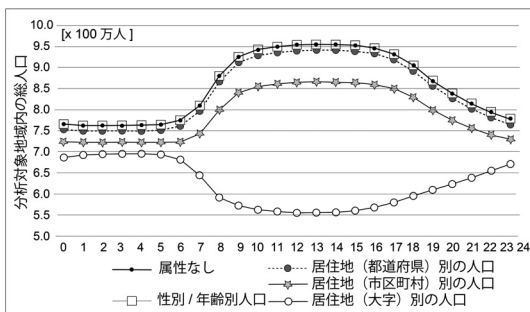


図8 居住地付きデータから求めた総人口

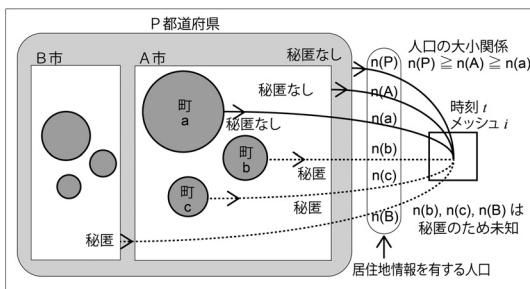


図9 居住地付きデータにおける秘匿処理

4.5 秘匿人口の補完 (方法3)

モバイル空間統計[®]の時刻 t 、メッシュ i 、居住地 h の滞留人口 y_{ih}^t は、居住地 h の空間範囲の違いにより、都道府県別、市区町村別、大字別の3種類が利用可能である。ただし、滞留人口 y_{ih}^t は、ある下限値 N を下回るとプライバシー保護の観点からデータが秘匿される（滞留人口としてカウントされない）。分析対象地域全域の各時刻 t における総人口を求めて図8に示してある。一つのメッシュ内に同一都道府県に住まう滞留人口は比較的多くなるので秘匿されにくい。このことは、居住地情報を含まないデータ（属性なし）との差がわずかであることから理解できる。一方、大字別については、一つのメッシュ内に同一大字に住まう滞留人口は限られるため秘匿されてしまう（滞留人口としてカウントされない）。そのため、大字別の滞留人口の総計は実際の滞留人口よりも少なくなる。この現象は、移動が活発な昼間に発生しやすく、多くの人が居住地メッシュにカウントされる夜間には、この差は小さくなる（図8）。

秘匿情報を補完する方法の概念図を図9に示してある。時刻 t 、メッシュ i 、居住地 h がP県A市a町である滞留人口は多く、秘匿されていないとする。一方、居住地がb町、c町の滞留人口は下限値 N を下回り秘匿されているとする。このとき、居住地がA市である

滞留人口の数 $n(A)$ には a, b, c 町の滞留人口（それぞれ $n(a), n(b), n(c)$ ）が含まれるので下限値 N を上回り秘匿されず、 b, c 町の滞留人口は合計で $[n(A) - n(a)]$ であると考えることができる。秘匿された $n(b), n(c)$ の値を正確に知ることはできないが、「秘匿人口は夜間人口（国勢調査人口統計の値）に比例する」と仮定して、 b, c 町の夜間人口に応じて、 $[n(A) - n(a)]$ を $n(b)$ と $n(c)$ に案分すれば、秘匿された滞留人口を補完することができる。

以上の大字を例にした補完方法は、居住地 h が市区町村の場合でも同様に適用できる。図8でみたように、市区町村別滞留人口は大字別滞留人口よりも秘匿されにくい。すなわち、上記の補完方法は、まず、秘匿の少ない都道府県から始め、市区町村、大字の順に補完を繰り返して行えばよい。

5. まとめ

本稿では、「時刻」「地点」「建物用途」の組み合わせを「キー」として、モバイル空間統計[®]、パーソントリップ調査データ、国勢調査人口統計を統合することで時空間人口統計データの付加価値を高める方法について論じた。

まず、携帯電話の位置情報から推定されるモバイル空間統計[®]と、建物用途と延床面積などの情報を備えた建物ポイントデータを用いて、時刻別、地点別、建物用途別の都市内滞留人口の時空間分布を推定する滞留人口分解モデルを構築した。東京都23区を対象地域としてモデル検証を行い、推定モデルの良好な記述精度を確認した。この滞留人口分解モデルを用いれば、モバイル空間統計[®]から得られる時刻別地点別の滞留人口を建物用途別に分解することが可能である。

次に、「時刻」「地点」「建物用途」の組み合わせを「キー」として、PTデータから得られる滞留目的・属性情報・居住地情報を時刻別地点別建物用途別滞留人口に組み込む方法を示した。東京都23区の平日・休日のデータを用いて、目的別滞留人口、性別年齢別滞留人口の推定事例を示した。滞留人口の時空間分布を、滞留建物用途別、滞留目的別、性別年齢階層別、居住地別に把握することが可能となり、人口統計データの付加価値を大きく高めることが可能となった。

謝辞 本研究はJSPS 科研費17H00843の助成による研究成果の一部について解説したものである。

参考文献

- [1] 大佛俊泰, “詳細な滞留者属性情報の組み込みによる時空間人口統計データの高度化,” 日本建築学会計画系論文集, **85**, pp. 2375–2383, 2020.
- [2] 清家剛, 三牧浩也, 原裕介, 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,” 都市計画論文集, **46**, pp. 451–456, 2011.
- [3] 大佛俊泰, 早坂遼, “携帯電話人口統計を用いた都市内移動者の時空間分布推定,” 日本建築学会計画系論文集, **84**, pp. 1853–1862, 2019.
- [4] 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 D3, **72**, pp. L515–L522, 2016.
- [5] 矢部貴大, 関本義秀, 樫山武浩, 金杉洋, 須藤明人, “パーティクルフィルタを用いた災害時におけるリアルタイムな人流推定手法,” 交通工学論文集, **2**, pp. A_19–A_27, 2016.
- [6] 瀬戸寿一, 樫山武浩, 関本義秀, “平成 28 年 4 月熊本地震における携帯電話位置情報を用いた避難所の混雑度推計,” 地理情報システム学会講演論文集 (CD-ROM), C-3-2, 2016.
- [7] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata and C. Ratti, “Real-time urban monitoring using cell phones: A case study in Rome,” *IEEE Transactions on Intelligent Transportation System*, **12**, pp. 141–151, 2011.
- [8] P. Deville, C. Linard, S. Martin, M. Gilbert, R. F. Stevens, E. A. Gaughan, D. V. Blondel and J. A. Tatem, “Dynamic population mapping using mobile phone data,” *PNAS*, **111**, pp. 15888–15893, 2014.
- [9] 大佛俊泰, “数値地図を用いた都市内滞留人口の時空間分布推定,” 日本建築学会計画系論文集, **73**, pp. 2673–2678, 2008.
- [10] 大佛俊泰, 村上彩夏, “居住者の時空間共存性に基づく距離測度,” 日本建築学会計画系論文集, **80**, pp. 2001–2010, 2015.
- [11] 大佛俊泰, 島田廉, “平日と休日における都市内滞留人口の時空間分布推定と地震被害想定への応用,” 日本建築学会計画系論文集, **74**, pp. 145–152, 2009.