

県レベルの疑似人流データの作成と精度評価

桒山 武浩*・龐 岩博*・関本 義秀*

Development of prefecture-level people flow data

Takehiro KASHIYAMA, Yanbo PANG and Yoshihide SEKIMOTO

Understanding people's movement is important in urban planning, commercial development, and disaster management. However, there is no high-quality person flow data covering rural areas and is available to everyone. Mobile carriers monopolize mobile phone data, such as GPS and base station logs. In addition, due to privacy issues, it is impossible to obtain disaggregated people flow data easily. Therefore, we aim to develop nationwide synthetic people flow data based on public statistical data such as the national census and economic census. In this paper, we describe the generation method of prefectural level person flow data and evaluation results.

Keywords: 人流データ, オープンデータ, 行動モデル

1. はじめに

人の移動を理解することは、都市計画、商業開発や災害対策において重要である。近年では、スマートフォンに搭載された GPS や基地局ログなどの携帯電話から取得した人の移動データを分析する研究が注目されている。たとえば、コロナ状況下の現在では、ソーシャルディスタンスやウィルス感染の広がりなどの分析のために広く活用されている。

一方、携帯電話から取得されるデータは、プライバシー保護の観点から、その実データが携帯キャリアから外部に提供されることはなく、入手できたとしても、その取扱いは非常に難しいといった課題がある。最近では、一部、匿名化された携帯電話データであれば購入できる状況になってきたが、それでも非常に高価なため限られたエリア、かつ限られたサンプルレートのデータ以外では入手することは一般的に難しい状況にある。

筆者らは、このような状況を改善したいという思いから、これまで、人の流れデータ[1]や Open PFLOW[2]といった人流データセットの開発してきた。前者は、非公開のパーソントリップ調査（以降、PT 調査と呼ぶ）から得られたトリップデータに対して、

時空間内挿入を行うことで生成したデータであり、現在 36 都市圏、延べ約 700 万人分の人流データを研究者に提供している。後者は、集計された形であるがオープンデータ化された PT 調査と統計データを組み合わせることで生成した人流データで、オープンデータとして公開している。しかし、いずれの人流データセットも、PT 調査をベースとしているため、データセットがカバーする範囲が限定されるといった課題がある。

研究レベルでは、携帯電話データから機械学習を用いて行動モデルを構築し、それをもとに人流データを構築する取り組みも行われている[3]。しかし、行動モデルは、シミュレーション対象の環境に合わせて、多くの入力データを用意する必要があり、パラメータ調整等のテクニカルな作業も必要となるため、だれもが容易に扱えるものではない。また、我々の人流データセットと共通の課題として、都市圏などの限られたエリア内の人の移動を対象としている点が挙げられる。具体的には、対象エリアの通過、内外や外内交通を考慮しないため、どうしても生成された人流と実際の人流とに決定的なズレが発生してしまう。

* 正会員 東京大学生産技術研究所 (Institute of Industrial Science, the University of Tokyo)
〒153-850 東京都目黒区駒場 4-6-1 生産技術研究 Ce509
E-mail : [ksym, pybdt, sekimoto]@iis.u-tokyo.ac.jp

また、さまざまな場面で、断面交通量などの人の移動に関する状況が現地計測などで把握できるようになってきた現在、これらの計測データを活用することで人流データを高精度化するような仕組みを構築することが重要だと考える。その場合、人流データとしては、特定の商用データに依存するのではなく、オープンで汎用的なものをベースとすることが望ましい。

そこで、本研究は、PT 調査範囲に限定されず、だれもが利用可能で、安定した精度を持ち、全国規模で整備された人流データの実現することを目標として見据え、まずは、県レベルの疑似人流データの開発を行った。この過程では、オープンデータとして公開される統計データと、建物データ等の低廉に入手可能な地理空間情報のみを用いることを前提条件として、典型的な日常の行動を表現する人流データの構築を目指した。以降、2章において疑似人流データの生成手法について説明し、3章において精度評価の結果を示す。そして、最後に、4章で現状の課題と今後の方針について述べる。

2. 提案手法

2.1. 使用データ

本研究では、国勢調査などの統計データと、低廉に入手可能な地理空間情報を使って、疑似人流データを生成することを目的とする。本研究で使用するデータ一覧を表1に示す。人流データの生成のために、3つのオープンな統計データと2つの商用データを使用する。

2.2. 初期人口分布の生成

人の流動を再現するためには、人々の行動開始の初期値位置を決定する必要がある。人の行動をシミュレーションする際、個々のエージェントが年齢や性別といった細かな属性情報を保持していれば、属性に応じて、緻密な人の行動制御が可能となる。そこで、本研究では、梶原らの研究[4]によって提案された人口分布の推定手法を用いて、人口の初期分布データを生成した。なお、この初期分布データには、性別、年齢といった一般的な属性に加え、住宅床面積、世帯種別、家族構成、自宅位置（経度、緯度）といった属性データが含まれている。

表1. 使用データ一覧

No	データ名	種類
1	統計局-国勢調査	オープン
2	統計局-経済センサス	オープン
3	統計局-社会生活基本調査	オープン
4	DRM(道路ネットワーク)	商用
5	ゼンリン住宅地図	商用

表2. 交通モード

距離	種類
距離<500m	徒歩
距離<1000m	自転車
距離>=1000m	ビジネスマン・その他：自動車
距離>=1000m	学生：自転車

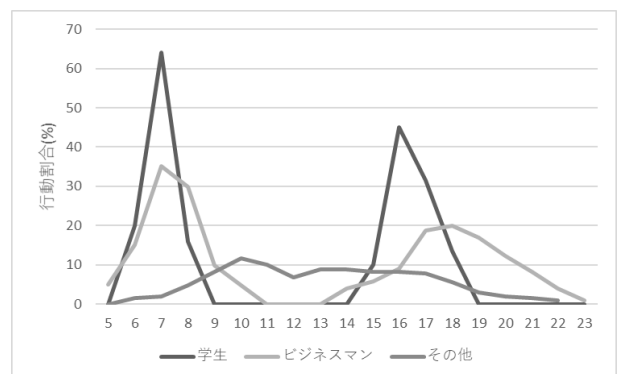


図1 時間帯別の行動割合

2.3. 行動モデル

本稿では、学生、ビジネスマン、その他（基本的には、主夫・主婦の行動を表現することを目的とする）の3つのエージェントを定義する。なお、それぞれのモデルは、【自宅 - 目的地 - 自宅】の行動パターンを持つことにした。

それぞれの行動パターンは、年齢に応じて、エージェントに割り当てる。4歳以上かつ18歳以下の場合には学生、65歳以下の60%にはビジネスマン、それ以外の80%にはその他の行動パターンを割り当てた。それ以外のエージェントは、行動パターンを持たず、一日中自宅に滞在するものとした。

2.4. 交通モード

本研究では、鉄道での移動は考慮せず、徒歩（時速5km）、自転車（時速10km）、自動車（時速30km）の交通モードでの移動に限定した。具体的には、表

2に示すように、現在地から目的地までの距離に応じて、交通モードを設定した。

2.4. 行動開始時刻

学生、ビジネスマンの通勤・通学時間は、社会生活基本調査のうち、時間帯別の行動種別「通勤・通学」の行動者率の割合を参考に、出勤・通学、帰宅時間をそれぞれのエージェントに割り当てた。その他の行動時間は、行動種別「移動（通勤・通学を除く）」の行動者率の割合を参考に、目的地への出発時間を割り当てた。なお、目的地での行動時間は、0時間から2時間でランダムに決定した滞在時間を設定することとした。時間帯別、行動モデル別の行動割合を図1に示す。図1では、たとえば、学生の場合、午前の山が通学時間、午後の山が下校時間のための行動割合を意味している。

2.5 目的地の選択

学生の場合は、目的地としては最寄りの学校とした。もちろん、年齢別に小学生であれば小学校、中学生であれば中学校、高校生であれば高校に通学するように目的地を設定した。ビジネスマン、その他については、離散選択モデルの一つである多項ロジットモデル(Multinomial Logit : MNL)を用いて算出した選択確率に応じて目的地を決定する。なお、目的地選択の過程では、対象エリアを3 kmメッシュごとに分割し、分割したエリアごとに目的地の選択確率を算出する。MNLの選択確率式を式(1)に、その効用関数を式(2)に示す。

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_k e^{V_{nk}}} \quad (1)$$

$$V_{ni} = \alpha S_i + \beta d_{ni} \quad (2)$$

P_{ni} は目的地*i*を選択する確率、 V_{ni} は目的地*i*への効用を示す関数である。 S_i はビジネスマンの場合は、目的地における全業種の従業員数、その他の場合は卸売の従業員数とした。 d_{ni} は自宅から目的地までの距離である。目的地のメッシュごとの従業員数は、統計局の経済センサスの調査結果から取得した。また、効用関数のパラメータ α, β については、パーティクルフィルタを用いて、最適なパラメータ値を算出した。この際、各パーティクルの尤度計算には、式(3)の式を用いた。

$$L = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(\frac{-(RMSE)^2}{2\sigma^2}\right) \quad (3)$$

L は各パーティクルの尤度、 σ は誤差の標準偏差、RMSEは2乗平均平方根誤差を示す。まず、選択確率を用いて目的地メッシュごとに、到来するエージェント数を集計する。そして、経済センサスから取得した従業員数を観測データとして、RMSE値を算出する。なお、その他のモデルのパラメータ計算時は、その総数が従業員数に一致するようにスケールを調整している。最後に、式(3)を用いて、各パーティクルの尤度を計算した。

なお、上記の選択モデルにより、3 kmメッシュが目的地として選択される。メッシュ内での詳細な目的地については、メッシュ内の事業所からランダムに選択した。学校、事業所等の建物情報については、ゼンリン住宅地図を使用している。

2.6 交通シミュレーションの導入

前項までの手法を用いて、個々のエージェントに1日分のトリップを割り当てたのち、交通シミュレータの一つであるMatsim(Multi-Agent Transport Simulation)を用いて、混雑度を考慮して自動車の移動ルートを決めた。なお、徒歩と自転車については、道路ネットワーク上を最短経路で移動するものとした。

3. 精度評価

3.1 疑似人流データ生成の評価方法

本研究では、静岡県と富山県を対象に、疑似人流データの試作を行った。その中では、試作データの精度評価として、まず、静岡県裾野市の住民のみを対象として、市レベルでの疑似人流データを作成(以降、疑似人流Aと呼ぶ)し、その精度評価を行った。次に、県全域の住民を対象として、市間の人の移動を考慮して作成した県レベルの疑似人流データ(以降、疑似人流Bと呼ぶ)を用いることで、人流データAに比べ、裾野市の人流の再現率がどの程度向上するかを確認した。最後に、疑似人流Bを対象に、県レベルにおいて、観測データとの比較を行った。なお、県レベルの評価では、研究の進捗の都合上、富山県だけを対象にその精度を評価していることに留意していただきたい。

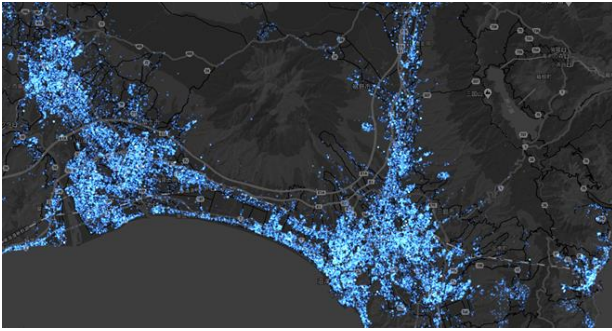


図2 人流データの可視化結果（東駿河湾都市圏）

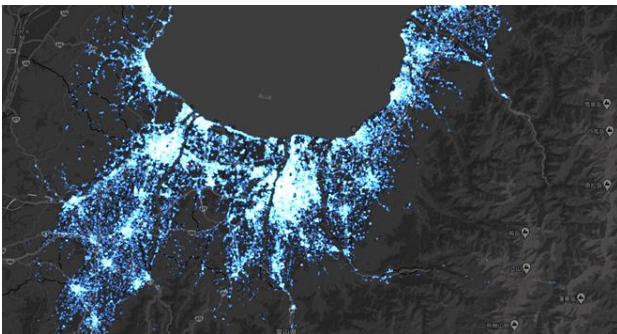


図3 人流データの可視化結果（富山県全域）

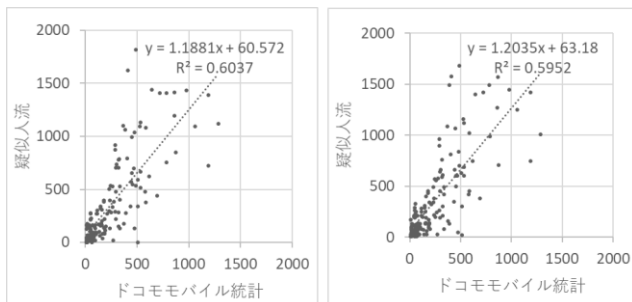


図4 メッシュ人口の比較結果

（裾野市，疑似人流 A，左 6 時，右 12 時）

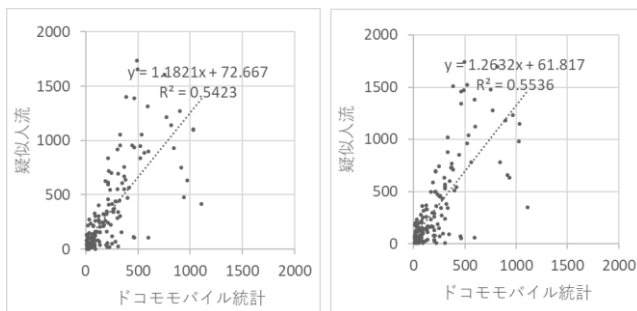


図5 メッシュ人口の比較結果

（裾野市，疑似人流 B，左 6 時，右 12 時）

疑似人流データの可視化例を図2と図3に示す。可視化結果において、個々の青い点が各エージェントを示し、エージェントの密集地は白い点として表現されている。

3.2 裾野市でのメッシュ人口評価

裾野市において、1日の行動がほぼ開始されていない6時と、多くの人が移動する時間帯である12時の2つの時間帯で、500mメッシュの人口分布について、観測データとの比較実験を行った。観測データには、ドコモ社が提供するモバイル空間統計を使用した。なお、観測日については、コロナの影響がない、2019年10月15日平日のデータを使用した。比較結果は、図4と図5のとおりである。

比較結果より、疑似人流 A、疑似人流 B とともに観測データに対して、0.5～0.6 程度の相間があることがわかった。また、両疑似人流において、6時と12時において、相間の変化量が小さいことがわかった。このことから、疑似人流データが、人の行動によるメッシュ人口の増減を再現できているといえる。

一方、疑似人流 A、疑似人流 B を比較すると、疑似人流 B の精度が低いことが分かった。この現象は、6時と12時ともに共通する結果である。6時の段階では、ほとんどのエージェントが行動を開始していないため、精度に差は発生しないはずである。考えられる理由は、拡大係数の設定がある。今回、疑似人流 A では全数、疑似人流 B では、県全域の人口データを作成するため、コンピュータリソースの問題により、10分の1のエージェントのみを用い、拡大係数を設定した。これにより、精度が低下したと考えられるが、その差は小さく、疑似人流 A と疑似人流 B の差はほぼないといえる。

3.3 裾野市における断面交通量評価

裾野市において、自動車の交通量を再現できているか評価するため、観測データと時間帯別の断面交通量の比較を行った。観測データには、日本道路交通情報センター（JARTIC）が提供する断面交通量データを使用した。このデータは、全国の都道府県警察が設置する車両感知器等の機器を使用して収集されたデータである。本研究では、裾野市内の1地点において、国道246号線の上下線合計値で比較を行った。断面交通量の集計場所を図6に示す。また、断面交通量の比較結果を図7に示す。

比較結果から、疑似人流 A から集計した断面交通量は、観測データの10分の1程度のボリュームし

かないことが分かった。これは、疑似人流 A が裾野市だけを対象とし、都市間の移動を考慮しないためだと考えられる。平成 27 年度の国勢調査によると、裾野市から他市区町村へ通勤する人数は約 1 万人、他市区町村から通勤してくる人数は約 1 万 6 千人となっている。また、特に、断面交通量の計測地点は、裾野市の北部の御殿場市との境界近くにあり、多くの通勤者が利用する道路である。このことから、疑似人流 A では、断面交通量が過小となったと考えられる。

疑似人流 B では、7 時から 9 時、16 時以降の時間帯において、観測データと近い値となっていることがわかる。一方、それ以外の時間帯では、観測データと比べ、交通量が半数以下となっている。本研究では、ビジネスマンについて、通勤行動のみを表現する行動パターンを設定した。しかし、実社会では、出社後に営業活動や他社への訪問等の目的で、多くの人が移動する。疑似人流 B では、これらのビジネスストリップを含んでいないため、交通量が過小となったと考えられる。事実、PT 調査結果を分析すると、昼間については、総トリップのうち、半数近くをビジネスストリップが占めることがわかる。しかし、ビジネスストリップの点を除けば、疑似人流 A と比較して、疑似人流 B は、交通量断面量という評価指標において、高い再現性があることがわかった。

3.4 富山県全域でのメッシュ人口評価

富山県全域を対象に、疑似人流 B と観測データを用いて、時間帯別の 500m メッシュ人口の比較を行った。観測データについては、裾野市での評価と同様のモバイル空間統計データを使用した。比較結果を図 8 に示す。



図 6 断面交通量の観測場所

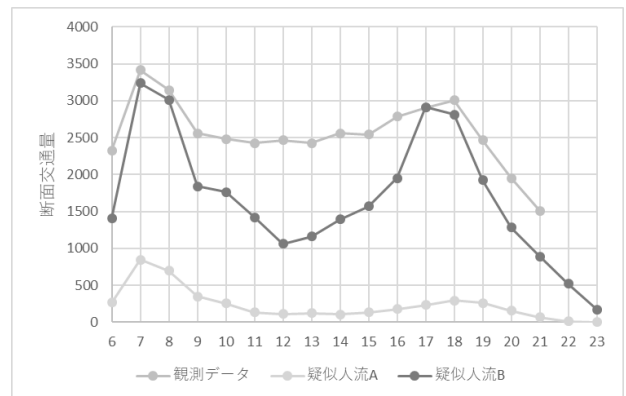


図 7 断面交通量の比較結果

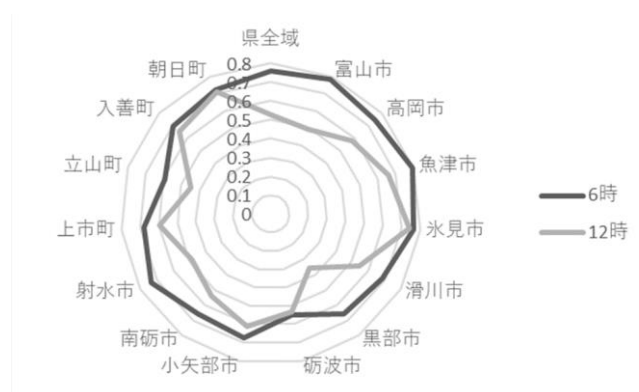


図 8 県全域でのメッシュ人口の比較結果

比較結果より、6 時の時点では、ほぼ全ての市で、0.7 以上の相間があるが、12 時になると観測データとの相間が低下していることがわかる。また、相間の低下率は、市によって異なっている。特に、富山市と黒部市では約 40% 近く低下している。一方、砺波市、小矢部市や入善町では低下率が 10% 以内に収まっていることがわかった。

県全体で評価すると、相間の低下率が約 30% と小さくない値となっている。このことから、疑似人流データが正しく人の行動を再現できていない可能性がある。この点については、地域別に低下量が異なる原因を詳しく調査することで、疑似人流データの精度向上につなげる予定である。

3.5 富山県全域での日 OD 量評価

疑似人流データが、メッシュ人口だけではなく、多種の観測データとの相間を確保できれば、高いレベルで品質が保証されることになる。そこで、本研究では、1999 年の富山・高岡都市圏の PT 調査から生成された人の流れデータを使用して、疑似人流データとの日 OD 量の比較を行った。

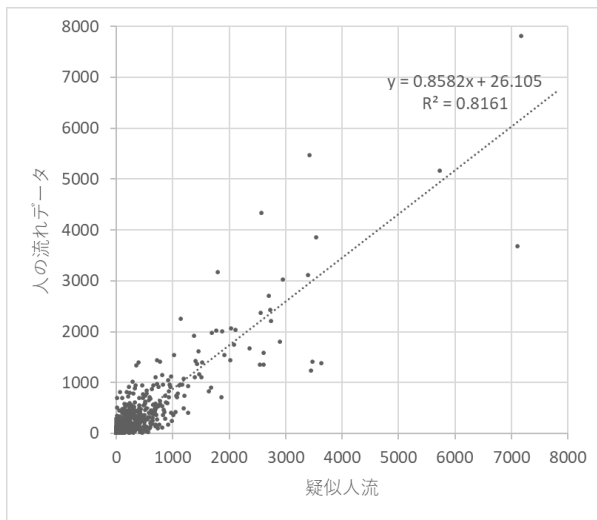


図9 通勤者の日 OD 量の比較結果

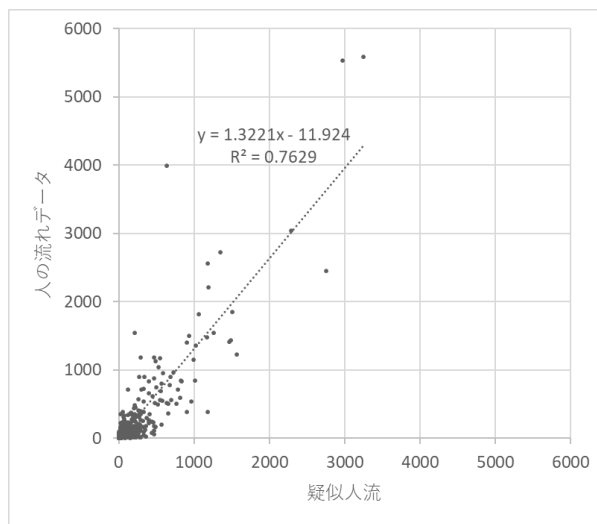


図10 その他の日 OD 量の比較結果

本比較では、人の流れデータをトリップ分割したのち、移動目的に応じて、通勤 OD、その他 OD 量を集計した。具体的には、移動目的が「通勤」のものを通勤者 OD に、「買い物、娯楽・食事、その他私用」のものをその他 OD とした。この集計された OD と疑似人流データから集計した OD 量を比較した結果を図9、図10に示す。なお、O と D は、対象エリアを5 kmメッシュで分割したエリアとした。

図9より、通勤者 OD 量と PT 調査から求めた OD 量には、高い相間があることを確認した。また、OD の総量も近い値を示していることがわかる。また、図10より、その他の OD 量についても、PT 調査の OD 量と高い相間を示すことがわかる。一方、OD の総量については、疑似人流が過小となっていること

がわかる。これは前述したように、行動モデルの表現する行動パターンが限定されていることに起因すると考えられ、今後、改善すべき点である。

4. おわりに

本稿では、統計データと限定された商用データだけから、平常時の行動を表現する疑似人流データを生成する手法を提案した。また、メッシュ人口、断面交通量と日 OD という指標で、疑似人流データと観測データを比較することで、その精度を評価した。その結果、疑似人流データは、観測データに対して、高い相間を持ち、一定のレベルで人の移動を再現できていることを確認したが、いくつかの点で課題も発見された。

今後は、行動モデルの表現力を拡張するとともに、より多種の統計データを組み合わせることで、疑似人流データの精度改善を図っていく。また、県レベルから都市圏、さらには全国規模にカバー範囲を広げていく予定である。

参考文献

1. Yoshihide Sekimoto, Ryosuke Shibasaki, Hiroshi Kanasugi, Tomotaka Usui, Yasunobu Shimazaki, PFLOW: Reconstruction of people flow recycling large-scale social survey data, IEEE Pervasive Computing, Vol.10, No.4, pp.27-35, 2011.
2. Takehiro Kashiya, Yanbo Pang, and Yoshihide Sekimoto, Open PFLOW: Creation and evaluation of an open dataset for typical people mass movement in urban areas, Transportation Res. Part C, 2017.
3. Yanbo Pang, Takehiro Kashiya, Takahiro Yabe, Kota Tsubouchi, Yoshihide Sekimoto, Development of people mass movement simulation framework based on reinforcement learning. Transport. Res. Part C, 2020.
4. 梶原健人, 金杉洋, 小川芳樹, 瀬戸寿一, 関本義秀, 世帯単位の動向に着目したエージェントベースの人口分布予測モデルに関する研究, 第28回地理情報システム学会講演論文集, 2019.