

Development of nationwide synthetic people flow data based on agent model and statistical data

Takehiro KASHIYAMA, Yanbo PANG and Yoshihide SEKIMOTO

People movement data is used in various fields, such as urban planning, commercial development, and disaster management. However, due to privacy issues, the data that describes people's movement, such as GPS data and CDR data derived from the mobile phone, is usually hard to obtain and should be used cautiously. On the other hand, the people movement data derived from the PT survey, widely used in research, has a problem that the coverage area is limited. Therefore, in this study, we propose the generation method of nationwide pseudo people flow data based on public statistical data and existing PT survey data and evaluate generated pseudo people flow data with high accuracy.

Keywords: 人の移動データ、オープンデータ、エージェントモデル

1. はじめに

人の移動データへのニーズは高く、様々な分野で活用されている。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。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を除けば一般的に入手することは難しい状況にある。

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。

一方で、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。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となるため、だれもが容易に扱えるものではなく、また、研究の対象とされるエリアも極めて小さい。

そこで、筆者らは、PT 調査の範囲に限定されず、だれもが利用可能で、安定した精度を持ち、シームレスに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、「疑似人流データ」の開発を開始した。疑似人流データの開発では、一般的に入手可能なオープンデータとして公開される統計デ

*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@iis.u-tokyo.ac.jp

**正会員 東京大学空間情報科学研究センター (Center for Spatial Information Science, the University of Tokyo)

ータと既存の PT 調査データ、そして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疑似的に表現する人流データの実現を目指している。したがって、疑似人流では、リアルな人の移動を再現するものではないが、都市空間的な人の流動の傾向を把握することは可能であり、都市計画や感染症の広がりなどのシミュレーションへの活用を期待している。

先の研究報告[3]では、静岡県と富山県を対象に、県レベルの疑似人流データの作成し、OD量、500mメッシュ人口分布、道路断面交通量の観点で精度評価を行い、低くない精度が得られた。そこで、本研究では、これまでの知見をもとに、対象範囲を拡大し、最終目標である全国規模の疑似人流データの開発に着手した。本稿では、これまで作成した第一プロトタイプの生成方法とその精度評価について報告する。以降、2章において全国規模の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. 疑似人流の生成手法

2.1. 概要

本研究では、一般的に入手できるオープンなデータからの疑似人流の作成を目指しており、人流に関する非集計データを使用しない。しかし、オープンな集計データだけでは疑似人流の構築は困難である。そこで、人々の行動の開始時間や行動時間等の一部の項目については、非集計の PT 調査データより集計した値を活用することとした。

処理手順として、以下の7つのステップによって、個々の人々の1日の典型的な行動を表現する移動データを生成される。基本的には、各ステップにおいて、集計データとして得られる経験的分布に基づいてランダム選択することで、各行動の要素を決定していく。

1. 人の初期分布の生成
2. 役割の選択
3. 行動パターンの選択
4. 行動先の決定
5. 行動時間の決定

6. 交通手段の選択

7. 移動経路の選択

以降、2.2節では各ステップで使用するデータを示したのち、2.3節以降では各ステップの詳細を説明する。

2.2. 使用データ

疑似人流生成のために使用するデータを表1に示す。PT調査ベースのデータは東京大学 CSIS が整備する研究用空間データ基盤 Joras からデータを入手することが可能である。ゼンリンが提供する建物データと日本デジタル道路地図協会が提供するデジタル道路地図 (DRM) については、有償データではあるが、最低限必要な基盤空間データとして考えて使用した。なお、これらのデータも、研究目的であれば Joras よりダウンロード可能となっている。

2.2. 人の初期分布の生成

人の移動行動は、世帯構成によって影響を受ける。たとえば、世帯が保有する車両台数によって、世帯の構成員が選択できる交通手段にも制約がつく。そこで、本研究では、梶原らの研究[4]において、統計データから推計された世帯単位の人口分布データを用いた。本データは、ゼンリンの建物データに世帯データを割り当てたものであり、建物ごとに世帯数、家族構成、年齢、性別等の属性情報を把握できる。詳細は、梶原らの文献[4]を参照されたい。

表1 使用データ

ステップ	データ
1	梶原らが開発した世帯推計データ[5]
2	H27 国勢調査, R2 学校基本調査
3	H28 社会生活基本調査 2011 年中京都市圏 PT 調査
4	H27 国勢調査, H28 経済センサス ゼンリン建物データ (Zmap TOWN II) 国土数値情報: 学校・小学校区・中学校区
5	2011 年中京都市圏 PT 調査 2016 年東駿河湾 PT 調査
6	2015 年全国都市交通特性調査
7	DRM 道路ネットワーク 金杉らが開発した鉄道ネットワーク[5]

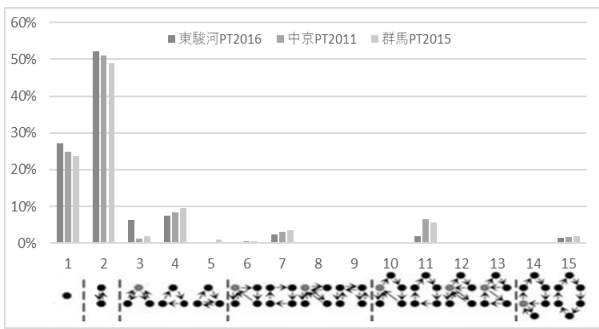


図1 モチーフ分析結果

2.3. 役割の選択

人々の行動を定義する際、無限にバリエーションを定義することは難しい。そこで、本研究では、それぞれの人に、以下の10種類からいずれかの役割を割り当てる。

- ① 就業者
- ② 主婦・主婦と高齢者（65歳以上）
- ③ 大学+専攻学生（18歳以上）
- ④ 高校生（15歳以上，18歳未満）
- ⑤ 中学生（12歳以上，15歳未満）
- ⑥ 小学生（6歳以上，12歳未満）
- ⑦ 幼稚園児（3歳以上，6歳未満）
- ⑧ 幼児（3歳未満）

①，②，③から④の割合については、国勢調査の就業状態等基本集計の結果より地域・性別・年齢別に決定した。④と⑤の高等教育の学生数について、学校基本調査の結果より都道府県，性別別に決定した。残りの⑥から⑧については、年齢属性より決定した。以降、役割別に、詳細な行動要素を決定する。

2.4. 行動パターンの選択

PT調査からモチーフ分析を行った結果を図1に示す。モチーフ分析は、訪問場所を示すノードと場所間の移動を示す有向エッジのネットワーク構造によって、人の行動パターンを表現するための分析である。この分析結果から終日自宅滞在の人（パターン1）と一つだけの行動先を持つ人（パターン2）だけで70%以上の人をカバーすることがわかる。本研究では、第一プロトタイプ疑似人流として、上記のパターン1とパターン2だけを再現を対象とした。この場合、自宅外での行動先が一つ以下となるため、アクティビティモデル等で課題となるスケジ

ューリング問題を考慮する必要がない。

パターンの割り当てについては、2.3節の役割を考慮して設定する。①就業者については、社会生活基本調査の就業状態別行動者率をもとに推計した平日の就業者率に基づいて、パターンを割り当てた。②主婦・主夫と高齢者については、PT調査より主婦・主夫，高齢者別に算出した外出率をもとに、パターンを割り当てた。なお、外出率は、異なるPT調査間で大きな差がないため、本研究では、2011年の中京PT調査から集計した値を使った。なお、③から⑦の学生には全員にパターン②を、⑧の幼児にはパターン1を割り当てている。

2.5. 行動先の決定

行動先は、2.2節で割り当てた役割が持つ目的に応じて決定される。①就業者の行動目的は通勤であり、勤務先を決定する。②主婦+高齢者の行動目的は自由行動として、スーパー，飲食店や公園等の多様な行動先を設定する。③以降の学生の行動目的は通学であり、学校が行動先となる。具体的な行動先の決定方法は、次の通りである。

【①就業者】

国勢調査の従業地・通学地集計より取得した市区単位の通勤OD量に従って、確率的に勤務先の市区を決定する。次に、経済センサスより取得した市区内のメッシュ就業者数に応じて確率的に勤務先メッシュを決定する。最後に、ゼンリンの建物データを用いて、メッシュ内の建物の床面積をもとに、確率的に最終的な勤務先建物を決定する。

【②主婦+高齢者】

複数のPT調査を分析した結果、自由行動の約80%以内が自宅から5km圏内であった。そこで、本研究では、まず、自宅から5km圏内のメッシュからハフモデルを用いることで、移動先のメッシュを確率的に求める。ハフモデルの式を以下に示す。

$$P_{ij} = \frac{S_j}{D_{ij}^\gamma} \frac{1}{\sum_{j=1}^n \frac{S_j}{D_{ij}^\gamma}} \quad (1)$$

式内の P_{ij} はiからjへの移動確率であり、 S_j はjの魅力度、 T_{ij} はiからjへの距離、 γ は距離抵抗のパラメータを示す。本研究では、距離抵抗のパラメータとし

て、一般的に使用される2の値を設定した。

メッシュの魅力度の算出のためには、2011年の東京PT調査から得られた自由行動トリップの集中量を目的変数として、経済センサスから得られる「宿泊業、飲食サービス業」「生活関連サービス業、娯楽」「教育、学習支援業」「医療、福祉」の事業所数を説明変数として重回帰分析することで、事業所種別が与える吸引力への影響度を係数として求める。この係数をつかって、全国分のメッシュの魅力度を算出している。移動先のメッシュが決定した後は、就業者の場合と同様に、メッシュ内の建物の床面積をもとに、確率的に最終的な移動先の建物を決定する。

【③と④15歳以上の学生】

従業者と同様に、国勢調査の従業地・通学地集計より取得した市区単位の通勤OD量に従って、確率的に通学先の市区を決定する。そして、通学先の市区内の学校のいずれかをランダムに選択する。

【⑤と⑥小中学】

国土数値情報の校区データをもとに、通学先の学校を決定する。ただし、本データには掲載されていない地区が多くあるが、本研究ではそれらを個別に収集する作業は行っていない。

【⑦幼稚園児】

幼稚園児をランダムな順番で、定員をオーバーしないように、最寄りの施設を選択するようにした。なお、今回はすべての施設の定員を300とした。

2.6. 行動時間の決定

行動時間に関する統計データとして、社会生活基本調査があるが、人々の行動の再現に必要な行動開始時間と行動時間を直接的に把握できる結果を取得することはできない。そこで、本研究では、PT調査より役割、性別ごとの集計した分布データを用いた。就業者の行動時間に関する分布データを図2に示す。

2.7. 交通手段の選択

全国都市交通特性調査では、都市種別（三大都市圏、地方都市圏）、性別、目的（通勤、通学、それ以外）、年齢と距離別の交通分担率のデータが公開されている。交通手段は、集計データと2.5節で決定した目的地までの距離に基づいて、確率的に決定した。上記調査では交通手段が7種類に分類されているが、

本研究では鉄道、自動車（バス、自動車-運転-有無、自動2輪）、自転車と徒歩の4種類に集約した。

2.8. 移動経路の選択

移動経路については、交通シミュレータを用い、混雑度や信号制御などを考慮して計算することが最適であるが、既存の交通シミュレータで全国分の人流を対象とすることは難しい。そこで、第一プロトタイプ疑似人流の開発では、ダイクストラ法による最短経路探索により移動経路を算出した。交通手段が徒歩、自転車、自動車の場合にはDRMの道路ネットワークデータを、鉄道については金杉ら[6]が開発した鉄道ネットワークデータを使用した。道路ネットワークのコストとしては、道路リンク毎に設定された制限速度とリンク長から算出した移動時間を、鉄道ネットワークについては鉄道のリンク長(路線の長さ)を設定して計算した。

3. 精度評価

3.1. 疑似人流データの生成

本研究では、2章で示したステップ1から7の生成手順に従って、全国分の疑似人流データを生成した。ステップ1は本研究内で処理していないが、ステップ2から6については、Core i7-8565U(Core:4)のCPUを搭載したPC上でシングルスレッド処理を行い、いずれも1時間以内で処理が終了した。ステップ7の経路探索処理では、単純な最短経路探索を用いたにもかかわらず、全人口分を処理するためには、膨大な計算リソースが必要となった。本研究では、AWSのEC2(c5a.24xlarge:96vCore)を用いて処理を行った。その結果、約70時間で処理を完了することができた。今後、交通シミュレーションの導入を考慮した処理方法について検討が重要となる。

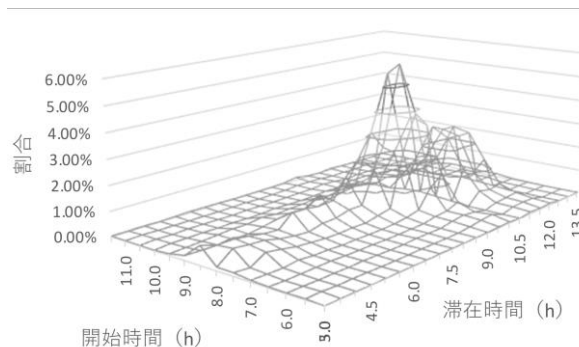


図2 行動時間に関する分布

表2 目的別トリップ数の比較

種別	2011年 中京都市圏 PT調査	疑似人流	カバー率
通勤	3,576,318	3,850,204	107%
通学	1,523,428	1,518,586	99%
自由	5,808,812	1,237,312	21%
業務	1,852,934	0	0%
帰宅	9,227,568	7,260,536	78%
合計	21,989,060	13,866,638	63%

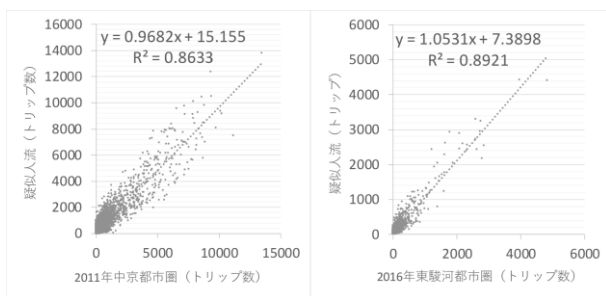


図3 通勤・通学トリップのOD比較

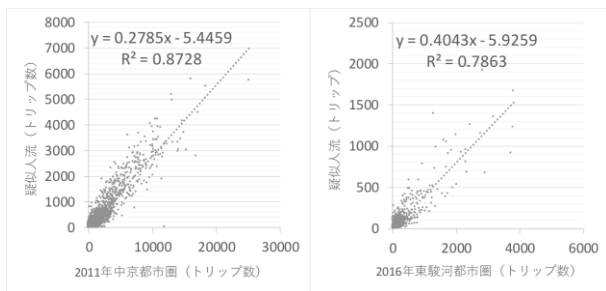


図4 自由行動トリップのOD比較

3.2. トリップカバー率

PT調査データに対する疑似人流データのカバー率を評価した。一人当たりのトリップ数は、都市圏によって違うが、そこまで大きな違いはない。本研究では、比較対象として、2011年の中京都市圏PT調査を用いた。評価結果を表2に示す。

表2の結果より、通勤と通学トリップはほぼ全数を再現できている。一方で、自由行動トリップについては、21%しか再現できていない。なお、業務トリップについては、今回対象としなかったため、0%となっている。合計では、66%のカバー率となっており、不足分が多いことが分かる。

外出率を分析すると、中京都市圏PTは81.2%であるのに対して、疑似人流は73.6%となっており、トリップのカバー率に比べて、不足率は低くない。

本研究では、行動パターンとして、図1のパターン1と2だけを考慮し、それ以外のパターンにはパターン2を割り当てた。これがトリップ数の不足の原因であることは明らかであり、行動パターンのカバー範囲の拡大を早急に進める予定である。特に、自由行動トリップについては、PT調査の分析結果によると、37.5%が就業者により、11.8%が学生により、50.6%がその他の主婦や高齢者により行われていることが分かる。したがって、自由行動のカバー率を高めるには、すべての役割において、行動パターンを拡張する必要がある。

3.3. OD量に関する評価

通勤・通学と自由行動のOD量について、大都市圏と地方都市圏の精度を検証するために、2011年の中京都市圏PT調査と2016年の東駿河湾PT調査と疑似人流の比較を行った。

通学・通勤ODについては、市区レベルで比較すると、相関値がほぼ1.0の値となる。これは、疑似人流生成に使用した国勢調査のデータとPT調査データの値が一致していることを示す。PTゾーンレベルで比較すると、図3に示す結果となった。

目的地決定の手法では、PTゾーンレベルでのOD精度を無視した単純な手法を使用した。両都市圏にて相関値が0.85以上の高い値を達成している。

次に、自由行動ODの比較結果を図4に示す。自由行動ODも通勤・通学ODと同様に、相関値が0.8以上の高い値を達成していることがわかる。しかし、東駿河湾PTの結果は、中京都市圏PTより低く、分散が大きいことがわかる。2.5節で説明したが自由行動の目的地選択には、ハブモデルを使ったが、魅力度の計算には中京都市圏PTより算出したパラメータを使用した。しかし、パラメータの地域差によって、東駿河湾都市圏では精度が低下したと考えられる。今後の課題としては、地域差を加味して移動先の魅力度を算出するモデルを構築することで、すべての都市圏で安定した精度を達成することを目指す。

3.4. 人口分布に関する評価

ドコモモバイル空間統計(静岡県と富山県)と平成22年近畿圏PT調査結果を用いて、500mメッシュ人口分布を比較した結果を図5に示す。図中の一

番下の初期分布は、疑似人流の初期分布との比較結果を示しており、行動モデルの有無による違いを比較するために示したものである。

比較結果として、ほとんどの人が自宅に滞在している6時では0.9程度の相関があるが、人の移動が始まるとともに、相関が低下して、12時と14時ごろには0.8程度になっている。0.8の相関係数は低い値ではないが、行動モデルが完全には人の移動を再現できないことを示している。ただ、初期分布では、0.7程度まで相関が低下していることを考えると、行動モデルによる効果も十分に確認できる。今後は、時間経過に伴う相関の低下を発生させず、高い相関の維持できるように行動モデルの改良を行っていく予定である。

3.4. 疑似人流の可視化結果

疑似人流データを道路や鉄道の種別を基準として、ズームレベルに応じた可視・不可視調整を行うことで、図7に示すように、全国分の人流のスムーズな可視化を実現した。詳細は、笠原らの文献[6]を参照していただきたい。

4. おわりに

本稿では、既存のPT調査データとオープンな統計データをもとに、平常時の典型的な行動を表現する全国分の疑似人流データの生成手法を提案した。また、トリップカバー率、OD量とメッシュ人口という指標で、疑似人流データの精度評価を行った。その結果、国勢調査で得られたODデータから作成した通勤・通学者の移動行動については、高い精度で再現できていることは分かったが、自由行動については、行動パターンを単純なものだけに限定したため、再現精度が低いことが分かった。

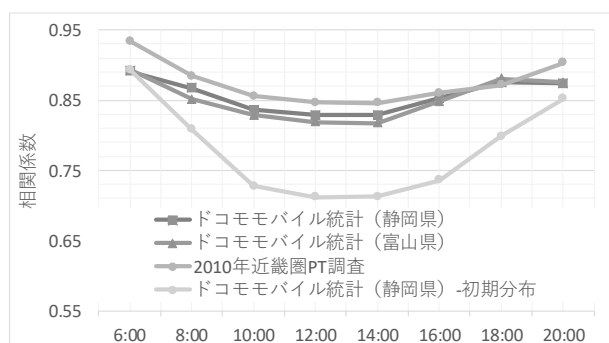


図6 500mメッシュ人口分布の比較結果



図7 疑似人流の可視化結果

今後は、行動パターンの多様化を図ることで、自由行動のカバー率を向上する。また、自由行動を買い物、通院や外食等に分類し、目的地の選択精度の向上を図る。また、交通シミュレータを導入することで、交通環境を加味した移動経路の決定することで、再現性の高い疑似人流データの目指す。

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, 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, Vol. 85, 2017.
3. 榎山武浩, 龐岩博, 関本義秀, 県レベルの疑似人流データの作成と精度評価, 29回地理情報システム学会講演論文種, 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.
5. 金杉洋, 関本義秀, 榎山武浩, 人々の流動再現へ向けたオープンな鉄道インフラデータの構築, 第22回地理情報システム学会講演論文集, 2013.
6. 笠原有貴, 関本義秀, 榎山武浩, ベクトルタイル技術を用いた全国規模の疑似人流データの効率的な可視化, 30回地理情報システム学会講演論文種, 2021.