

全国擬似人流データの提供と評価

龐 岩博*・榎山 武浩**・関本 義秀*

The Provision of Nationwide Pseudo People Flow Data and its Evaluation

Yanbo PANG*, Takehiro KASHIYAMA** and Yoshihide SEKIMOTO*

With the rapid trend of developing Smart Cities and Digital Twins, a better understanding of how humans move and perform a daily routine in the city area is of vital importance. In the past few decades, the Person Trip Survey in metropolitan areas provides the most detailed whereabouts of people and their social activities in urban areas. However, because of the high data collection cost, these survey data only cover metropolitan areas with small sampling ratio. On the other hand, GPS data and CDR data derived from the mobile phone, is usually hard to obtain due to privacy issues. 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: 人の流れ(people flow), 擬似データ(pseudo data), データ基盤(data infrastructure)

1. はじめに

近年, デジタル化の急速な進展・高度化が進む中, さまざまなデータを迅速に組合せ, 社会課題を解決していく事が期待される. そのため, 現実世界をサイバー空間で再現するとともに, 個人・民間企業・国家の意思決定と政策立案へ基盤データの構築と整備は不可欠である. 空間情報学の分野は, 従来から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[1]. しかしながら, 大規模調査であるために多大な費用と労力を要する. また, データセットがカバーする範囲と人口比率が限定されるといった課題も存在している.

一方, 近年ではパーソントリップ調査データの他にも, 匿名化された携帯電話データの利用可能性が広がってきている. 例えば, モバイル空間統計は通信基地局通信履歴 (CDR : Call Detail Record) に含まれる位置情報から集計されたデータであり, 1 時間毎に日本全国のメッシュ別の滞在人数を 24 時間 365

日把握することができるデータである[2]. しかし, 携帯電話データは非常に高価であるため一般的に入手することは難しい状況にある. さらに, プライバシー保護の観点から, その実データから個人属性や移動属性が提供されることなく, それぞれの移動者の移動目的や地域変容による変化に関する分析は困難である.

そこで, 我々は実世界の人々の動きを俯瞰可能とし, 都市計画・交通計画等, 公益に資する活動のための基礎データの整備を目標として見据え, 擬似人流データの作成手法 [3]に基づいて, 安定した精度を持つ「全国擬似人流データ」を作成して提供する. 本データセットは下記の特徴を持つ.

- 擬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ータと既存の PT 調査データの集計結果, そして建物データ等の低廉に入手可能なデータのみを用いること.
- 上記の集計レベルの調査データを用いることで, 合成的人口を作成し, そして人々の典型的な日常の行動を擬似的に再現し, リアルな人の位置情報ではないため, その結果は研究目的で

* 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)
〒153-8505 東京都目黒区駒場 4-6-1 E-mail : pybdtc@iis.u-tokyo.ac.jp

** 正会員 大阪経済大学経済学研究科 (Osaka University of Economics)

公開可能。

- 位置情報だけでなく、全人口に対する一人一人の人口属性（世帯・年齢・性別・就業状態）、1日の生活活動、活動に伴う発生した移動トリップ（目的・時刻・起点・終点・交通手段）、また移動軌跡をすべて提供する事。

以降、2章で擬似人流データの生成手法、3章についてデータの提供状況と提供方法を述べる。そして、4章でデータセットの精度を評価する。最後に、5章では現状の課題と今後の方針について述べる。

2. 擬似人流データの生成手順

本研究では、オープンな調査データと低廉な価格で入手できる商業データから典型的な全国における平日の1日中の擬似人流を再現する。結果として、断片的な位置情報だけでなく、どのような人が、どのような目的で、いつ、どのような交通手段で、どこからどこへ移動するかといった情報を全般的に提供する。そして、今回提供する擬似人流データは下記の4種類のデータセットで構成される。

1. 擬似人口データ
2. 擬似活動データ
3. 擬似トリップデータ
4. 擬似軌跡データ

本章では、それぞれのデータの詳細と生成方法について紹介する。

2.1. 擬似人口の生成

個人の日常的活動（例えば、通勤・通学・食事・買物等）は、人口属性によって影響を受ける。例えば、個人の就業状態によって、昼間に行う生活活動に制限が付き、また、住所によって職場や生活圏が決められる。そこで、本研究の第一歩としては、国勢調査から提供する集計レベルの人口統計データに基づいて、個人の世帯構成・年齢・性別・就業状況・住所等の情報を推定し、擬似人口データを生成する。具体的には、梶原らの手法[5]により、統計データから世帯単位の人口分布を推定し、そして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・性別などの属性情報を個人に配分する。詳細は、梶原らの論文[4]

を参照。

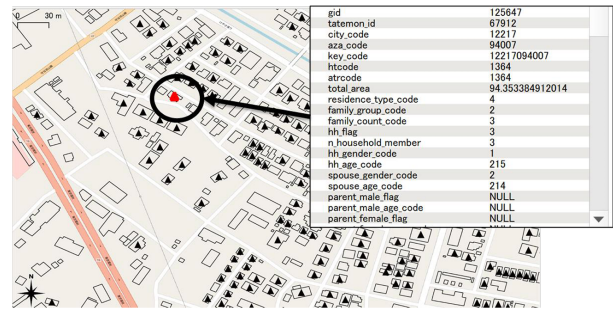


図1 擬似人口データの例

推定した世帯ごとの人口データに、個人の就業状況に対して、以下の10種類を設定する。

- ① 就業者
- ② 主夫・主婦と高齢者（65歳以上の方）
- ③ 専門学校・大学の学生（18歳以上の方）
- ④ 高校生（15歳以上また18歳未満の方）
- ⑤ 中学生（12歳以上また15歳未満の方）
- ⑥ 小学生（6歳以上また12歳未満の方）
- ⑦ 幼稚園児（3歳以上また6歳未満の方）
- ⑧ 幼児（3歳未満の方）

その内、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・年齢別に決定する。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、性別別に決定する。残りの⑥から⑧については、年齢属性より決定した。なお、生成した人口は実際の人間ではなく、本文中に「エージェント」と表す。

2.2. 擬似活動データの生成

地域交通計画策定や将来交通需要予測には、移動軌跡だけでなく、移動の目的、つまり各場所に行う活動についての情報を得る必要がある。一方、近年の非集計の交通行動分析モデルの研究結果により、従来のトリップベースモデルと比べると、アクティビティベースモデルは個人の活動の派生需要と時空間制約の考慮等、時空間の細かな表現ができるメリットがある。そして、擬似人流を再現する際、まず、個々の人間の1日の活動全体を推定する。結果として、擬似活動データは「誰が、いつ、何を」の情報を提供する。

本研究では、在宅・勤務・通学・買物・食事・通院・その他合計7種類の活動を定義し、各時間帯に

個々のエージェントは必ずこの7種類の中から1つの活動を選択する。

次に、各活動間の遷移確率をマルコフモデルによって推定する。

$$p_k(i, j) = P\{X_{k+1} = j | X_k = i\} \quad (1)$$

$p_k(i, j)$ は活動*i*から*j*へ遷移する確率。一方、個人の行動パターンは人口属性や地域性による異なるため、上記のマルコフモデルのパラメータについて、2.1の10種類の就業状態に対して、大(50万人以上)・中(30万人以上50万人未満)・小(30万人未満)の都市規模ごとにパーソントリップ調査から値を推定した。

2.2. 擬似トリップデータの生成

2.1で推定した活動内容により、活動は異なる場所に行くと新しいトリップが発生する。本研究では、各トリップに対して、「誰が、いつ、何の目的で、どこからどこへ」の情報を提供する。その内、移動者「誰」、移動目的「活動内容」、出発時間「活動時間」は前節で説明したが、行動先は、2.1で割り当てた就業状況と移動目的に応じて決定される。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。一方、対象者の従業値と自宅は同じ市区町村の場合、メッシュの就業者数と目的地までの距離を特徴量としてハフモデル[4]を用いて目的地への移動確率を求める。ハフモデルの式を以下に示す。

$$P_{ij} = \frac{\frac{S_j}{D_{ij}^\gamma}}{\sum_{i=1}^n \frac{S_j}{D_{ij}^\gamma}} \quad (2)$$

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、 D_{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。

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。

【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。

【自由行動】

最後に、主夫(婦)・高齢者のすべての行動と就業者及び学生の自由行動の目的地選択に対して、ゾーン(パーソントリップ調査ゾーン)レベルロジットモデルを用いて各メッシュの選択確率を計算して行う。各ゾーンの効用は目的に応じて、該当商業分類の従業員数と事務所数によって算出する。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。例えば、通院目的のトリップに対して、各ゾーンの医療機関事務所数と医療従業員数を使ってゾーンの効用を算出する。また、効用関数のパラメータは、大(50万人以上)・中(30万人以上50万人未満)・小(30万人未満)の都市規模ごとにパーソントリップ調査から値を推定した。移動先のゾーンが決定した後は、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。

2.4. 擬似軌跡データの生成

擬似軌跡データの生成については、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。なお、作成した軌跡データは、数秒ごとに移動者の位置を記録し、GPS データと同じように扱うことは可能となる。図 2 に示すように、疑似軌跡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ができる。

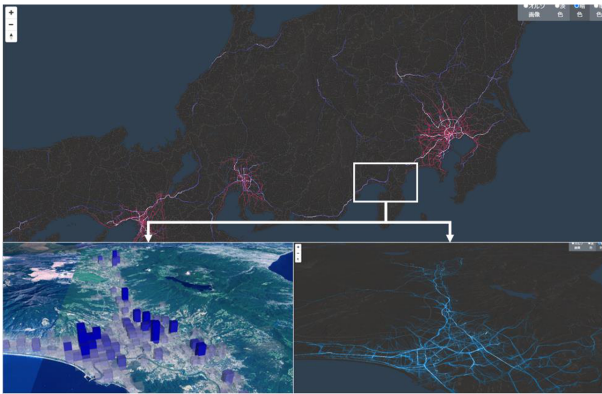


図 2 作成した疑似人流の可視化の例.

(上) リンクベース交通量

(下) 左：メッシュ滞在人口 右：軌跡データ

2.5. 使用データ一覧

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、研究目的であれば東京大学空間情報科学研究センター共同研究利用システム JoRAS(<https://joras.csis.u-tokyo.ac.jp/>)よりダウンロード可能となっている。

表 1 使用データ

	データ
2.1	<ul style="list-style-type: none"> H27 国勢調査 ゼンリン建物データ (Zmap TOWN II)
2.2	<ul style="list-style-type: none"> H28 社会生活基本調査 2011 年中京都市圏 PT 調査 2015 年全国都市交通特性調査
2.3	<ul style="list-style-type: none"> 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・小中学校区 2011 年中京都市圏 PT 調査 2016 年東駿河湾都市圏 PT 調査
2.4	<ul style="list-style-type: none"> DRM 道路ネットワークデータ 鉄道ネットワークデータ

3. 疑似人流データの提供状況

本研究では、日本全国 47 都道府県 1724 市区町村、約 1 億 3 千万の人口に対して、疑似人口・疑似活動内容・疑似トリップ・疑似軌跡 4 種類のデータをそれぞれに作成して提供している。表 2~4 に各データの内容についてまとめている。各データセットには、移動者の個人 ID を紐づけている。

表 2 疑似人口データ

項目	説明
世帯 ID	世帯を識別するユニーク ID
世帯種類	単独・夫婦のみ・夫婦子供等 16 種類
市区町村番号	住所の市区町村
住所	住所の経度・緯度
個人 ID	個人を識別するユニーク ID
年齢	移動者の年齢
性別	1.男性 2.女性
就業状態	8 種類 (2.2 章を参照)

表 3 疑似活動データ

項目	説明
個人 ID	個人を識別するユニーク ID
年齢	移動者の年齢
性別	1.男性 2.女性

就業状態	8種類 (2.2章を参照)
活動開始時間	活動開始時刻
活動持続時間	秒単位
活動内容	在宅・勤務・通学・買い物・外食・通院・その他7種類
活動場所	活動場所の経度・緯度
市区町村番号	活動場所の市区町村

表4 擬似トリップデータ

項目	説明
個人ID	個人を識別するユニークID
出発時間	出発時刻
出発位置	出発地の経度・緯度
目的地位置	目的地の経度・緯度
交通手段	徒歩・自転車・車・電車4種類
移動目的	在宅・勤務・通学・買い物・外食・通院・その他7種類
就業状態	8種類 (2.2章を参照)

表5 擬似軌跡データ

項目	説明
個人ID	個人を識別するユニークID
Unixtime	からの秒数
timestamp	yyyy-MM-dd HH:mm:ss で表す時刻
経度	該当時刻の経度
緯度	該当時刻の緯度
交通手段	徒歩・自転車・車・電車4種類
リンクID	経度緯度に応じる道路・鉄道ネットワーク上のリンクID

現在、研究目的の使用に限り、上記の擬似人流データを東京大学 CSIS が整備する研究用空間データ基盤 JoRAS(<https://joras.csis.u-tokyo.ac.jp/>)からデータを無料に入手することが可能である。各データセットは市区町村ごとに csv 形式で整備され、そのサイズについて表6で整理する。

表6 提供データサイズ

	データ量(GB)
人口データ	7.02
活動データ	29.78
トリップデータ	30.63
軌跡データ*	1360

* JoRAS システムの制限により、一部の地域（北海道，東京，大阪等）はデータが都道府県レベルで整備され、約 2GB /ファイルで分割して提供している。詳細は人の流れプロジェクト擬似人流データのページをご参照ください。

(<https://pflow.csis.u-tokyo.ac.jp/data-service/pseudo-pflow/>)

4. 擬似人流データの精度評価

4.1. 評価方法

本研究では、人口分布・トリップ数・トリップカバー率3つの集計量から作成した擬似人流データを評価する。その中、人口分布に対して富山県と静岡県（図3の緑色エリア）のドコモ社モバイル空間統計データを用いて評価する。トリップ数に対して近畿都市圏（図3の青色エリア）PTデータを用いて評価する。トリップカバー率は東駿河湾都市圏（図3のオレンジ色）PTデータを用いて評価する。

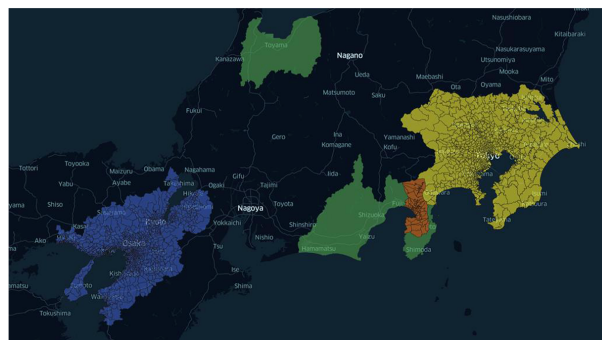


図3

4.2. 人口分布に関する評価

図4(a-c)には、ドコモモバイル空間統計（静岡県と富山県）を用いて、12時における人口分布を市区町村・1kmメッシュ・500mメッシュ3つのレベルで評価した。比較結果として、作成した擬似人流データは市区町村レベルにおいて携帯電話データとほぼ同じ人口分布であるが（相関 $R^2 = 0.98$ ），空間解像度を上げつつ相関が低くなる傾向がある。現段階の擬似人流データは、1kmメッシュレベルでは実世界の観測データと同じように扱うことが可能と考えられる。

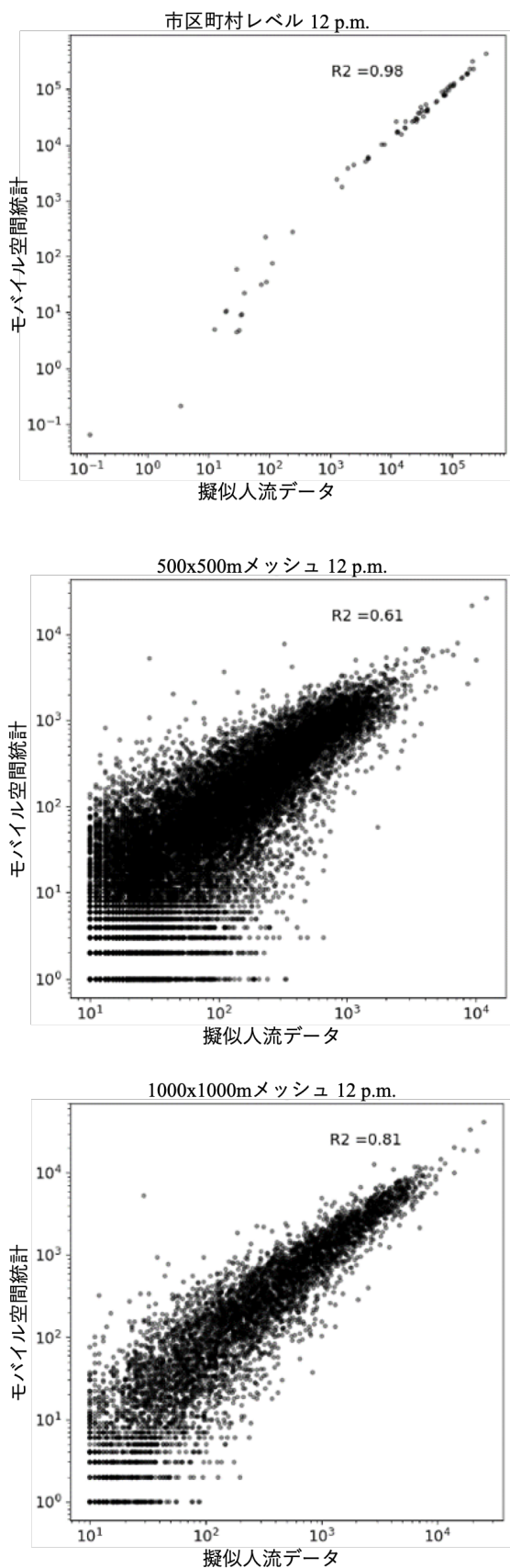


図4 解像度別人口分布の評価結果

((a)市区町村別 (b)1kmメッシュ (c)500mメッシュ)

4.3. トリップ数に関する評価

通勤・通学と私用(買物・通院・食事・その他)トリップのOD量について、大都市圏と地方都市圏別に精度を検証した。2008年東京都市圏を大都市 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。

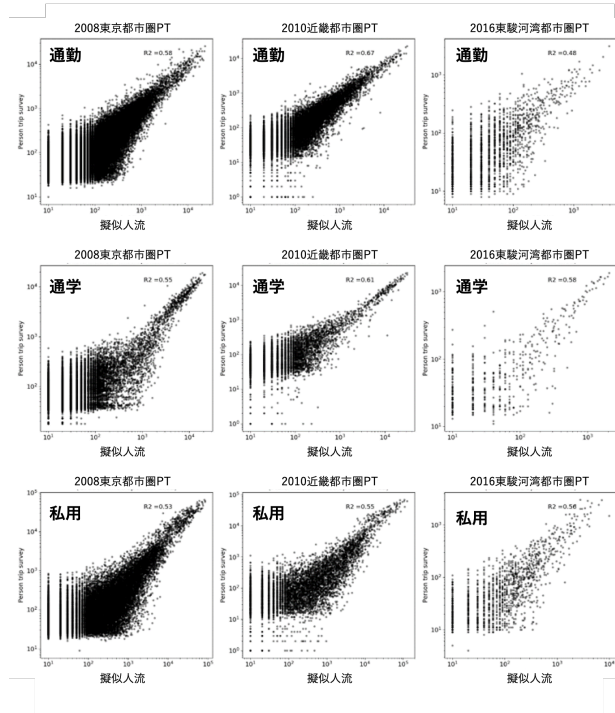


図5 目的別トリップ数の評価結果

4.4. トリップカバー率に関する評価

本研究では、2008年東京都市圏PTデータ・2010年近畿都市圏PTデータ・2016年東駿河湾都市圏PTデータを用いて疑似人流データの生成トリップカバー率を評価した。表7の結果より、東京・近畿都市圏ではほぼ全数を再現できている。一方、東駿河湾都市圏の結果はPTデータより10-17%不足していることがわかる。本データセットの行動パターンは大都市圏の調査データから抽出したため、地方の結果でも大都市へ偏り傾向という課題が存在している。

表7 都市圏別目的別トリップカバー率

移動目的	東京都市圏	近畿都市圏	東駿河湾
通勤	95%	105%	83%
通学	104%	113%	102%
娯楽	96%	106%	90%
帰宅	93%	97%	84%

5. 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擬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、擬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。

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、再現性の高い擬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. 布施孝志, 原田遼, 詳細な交通行動推定のための アクティビティシミュレーションと 観測データの統合に関する研究, 土木学会論文集D3 (土木計画学), Vol.75, No.5 (土木計画学研究・論文集第36巻), I_575-I_583, 2019.
3. Takehiro Kashiya, Yanbo Pang, Yoshihide Sekimoto, Takahiro Yabe, Pseudo-PFLOW: Development of nationwide synthetic open dataset for people movement based on limited travel survey and open statistical data, arXiv preprint, [arXiv:2205.00657](https://arxiv.org/abs/2205.00657), 2022.
4. David L. Huff, A probabilistic analysis of shopping center trade areas, Land Economics, Vol. 39, No. 1, pp. 81-90, 1963, doi:10.2307/3144521
5. 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.
6. 金杉洋, 関本義秀, 樫山武浩, 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.