強化学習を用いた都市圏レベルの人の流れの再現手法の構築

Pang Yanbo*・樫山武浩・関本義秀**

Reconstruction of people flow in metropolitan area via reinforcement learning approach

Yanbo Pang, Takehiro Kashiyama*, Yoshihide Sekimoto*

Abstract: Recently, the depopulation and aging of society have significantly changed urban structure and policy especially in transportation domain. Such changes challenge urban planners in various of issues such as traffic congestion, longer commuting, accidents, loss of public space and disaster management. Therefore, how to accurately capture the demand and leverage existing infrastructures effectively is essential for urban planning. However, most existing studies only reconstruct people flow in limited area. In this study, we propose an agent-based simulation method based on reinforcement learning approach to reconstruct people flow on Tokyo metropolitan area. We leverage People Flow data as real data to validate our results.

Keywords: （3～5 語） 人の流れ（people flow）， 強化学習（reinforcement learning）， モビリティー（human mobility）

１．はじめに

近年，人口の急激な減少と高齢化を背景として，都市政策や都市全体の構造が変化し，個人レベルから都市圏レベルまでの人の流れをボトムアップに把握することが重要である．特に，近年ではビッグデータが大きな注目を集める中で，大規模な人口の流動を捉えることができるデータも,その解析として，社会課題への適用が可能となっている．

しかし，こうした人の流れのビッグデータ，例えば携帯電話の基地局通信履歴やアプリケーション GPS，車載ナビなど様々な端末による収集したデータが存在するものの，個人情報を保護するため，データの匿名・集計処理に伴う個人レベルの移動情報が消えてしまうことがなるし，集計ベースのデータであっても高価であるし，特に都市圏レベルやその以上の大規模なエリアの人流を把握したい場合，サービス提供者以外の主体がデータを入手することは簡易ではない．

一方，この問題を解決する 1 つのアプローチとしてのエージェントシミュレーションが存在している．人間の日々の行動ルールを仮定してエージェントモデルを構築し，コンピュータ上で交通シミュレータを用いて擬似の人流データを生成する事例が行われてきている．さらに，エージェントモデルに現実の行動を反映させるために，実際の個人レベル行動データを解析して，その結果をエージェントモデルに実装する研究がなされている（Pang ら，2010）．しかし，人々の移動を再現する際に，モデルの時空間的粒度が細かいほど再現性が高いと考えられるが，粒度が不変の前提で，対象エリアの拡大と伴い，モデルの自由度や状態空間を膨大となり，シミュレーションの計算速度やモデルの予測パフォーマンスの低下は回避することはできない。そして，数多くの既存手法は，限られたエリア（都市・ゾーンレベル）を対象としており，都市を跨ぐ往復を取り除き，都市圏レベルの人流を全面的に反映することはできない．特に，総務省統計局 3 大都市への流入人口（通勤・通学者）によると，日々東京都特別区部を従業地・通学地として他市区町村から流入する人口は 333 万人（特別区部を従業地・通学地とする者（735 万人）．

そこで，本研究では，都市圏レベルの人の流れに着目し，エージェントシミュレーションによる擬似人流データを作成することを目指す．この時，強化

* 正会員 東京大学 生産技術研究所（Institute of Industrial Science, the University of Tokyo）
〒**151-5566** 東京都目黒区駒場
E-mail：**pybdtc@iis.u-tokyo.ac.jp**

学習を用いるエージェントベースモデルを取り込んで，首都圏などの広範囲における人の流れの再現手法を構築する．また，エージェントモデルに現実の行動を反映させるために，それぞれのエージェントの属性や行動パターンに対して，一致する実際の人の位置情報を用いてパラメータを学習させ，エージェントモデルに実装する。構築したモデルに基づいて、首都圏レベルの人の流れの再現に適用する．

２．提案手法

2.1. 都市を跨ぐ移動エージェントの提案

個々のエージェントは以下の基本属性を持つ：

- 初期状態：エージェント行動開始の初期位置.
- 状態空間：各時刻におけるエージェントの位置（メッシュに区切りした目的地集合）を表す.
- 行動空間：エージェントが環境に対して行う働きかけの種類を表し、一つのトリップ（起点・終点・交通手段から構成）がこれに相当する.
- 報酬関数：ある状態 ss が与えられるとその良さを返す関数です。逆強化学習では、エキスパートの行動系列をいくつかサンプルし、そのサンプルから報酬関数 R(s,a)を求めます。

一方，都市圏レベルの移動をシミュレートする際には，非常に膨大な状態空間と行動空間が必要となり，行動空間での行動選択および強化学習に計算コストがかかる．近年，深層学習の導入することで，高次元および連続な行動空間での行動政策を解決できても，大規模な人流データを作成することが不可能となる．一方，エージェントに余計な選択肢を与えると生成する軌跡の精度が落ちる．従って，シミュレーションのパフォーマンスを向上させるために，個々のエージェントに対して適切な行動空間を設ける必要がある．本研究は，エージェントの行動パターンによって分類を行い，それぞれにエージェントモデルおよび行動空間を設定する．具体的に，エージェントの初期位置と昼間の主要行動場所による，終日滞在者・県内行動者・県外移動者でそれぞれに区分する．各種類のエージェントに対して，該当するエリアだけの状態空間を設定する．詳細は第 3 章実験設定に説明する．

また，エージェントモデルに現実の行動を反映させるために，実際の人の位置情報を用いてパラメータを学習させ，エージェントモデルに実装する必要がある．本研究では，実際の人の行動履歴から，報酬関数パラメータを求める，そのパラメータをエージェントモデルに実装することで，エージェントを実際の人間と同じような行動ルールを自律的に生成させる．具体的には，Ziebart et al. (2008)の最大エントロピー逆強化学習(Maximum Entropy Inverse Reinforcement Learning)アルゴリズムを用いて，位置情報から抽出した移動軌跡によって報酬関数のパラメータを学習させる.

2.2. 強化学習に基づく意思決定ルール

人々の交通行動意思決定は不確実性を持ち，特に都市における複雑な環境のもとで数えきれないくらいの軌跡が生成できる．このような現象を再現するには，一つ一つに行動ルールを設計してすべての行動パターンをモデル化することは不可能である．一方，不確実性を伴う意思決定のモデリングにおける数学的枠組みとして、強化学習など動的計画法が適用される．そこで，本研究では，環境をマルコフ決定過程(Markov Decision Process, MDP)として定式化され，状態空間 S：メッシュに区切りした目的地集合，行動空間 A：潜在トリップ（起点・終点・交通手段から構成）の集合，報酬 R：その行動の即時的な良さを表す. 遷移関数$P_{sa}(\cdot)$は状態 s において行動 a を取った時の次の状態への状態遷移確率を表す. 割引因子は現在の報酬と未来の報酬との間における重要度の差異を表す．エージェントの最適な行動政策を得るため，強化学習を行う．その方策の評価指標として，状態価値関数

$$V^{\pi}(s) = R^{\pi}(s) + \gamma \sum_{a \in A} \sum_{s' \in S} \pi(s,a) Pr(s,a,s')\, V^{\pi}(s')$$

を表す．最も良い方策は次式で更新する:

$$\pi(s) \leftarrow \underset{a \in A_s}{argmax} \sum_{s' \in S} T(s,a,s') \big(R(s,a,s') + \gamma V^{\pi}(s') \big)$$

これらの操作πがすべての状態に対し, 変化しなくなるまで繰り返すことで最適な方策を得る.

３．実験

3.1. 実験設定

本研究は東京都市圏（東京都・神奈川県・埼玉

県・千葉県・茨城県）を対象とし，人の流れを再現する．しかし，このような広いエリアの実際の人の流れをすべて観測することは不可能ため，CSIS で提供している「人の流れデータ」https://pflow.csis.u-tokyo.ac.jp/)約 50 万人分のデータを観測とする．一方，2 章で説明したシミュレーション対象の種類を表 1 に示す．対象の 1 都 4 県に対して，移動者が一日中「県内通勤通学」,「他県通勤通学」を分類する上で，東京への流入・流出移動者をさらに細分化する（東京以外の県の間の移動が少ないため，今回省略する）．なお，個々エージェントの行動開始の初期値位置は観測データから生成する．エージェントは 1 km メッシュ単位で移動し，30 ごとに一回交通行動の意思決定を行う．

表 1．人の流れの再現対象

	観測者数	訓練データ数
東京都内	193,532	2,000
神奈川県内	124,489	2,000
埼玉県内	83,394	2,000
千葉県内	98,005	2,000
茨城県内	13,725	2,000
東京から神奈川	10,517	2,000
東京から埼玉	967	500
東京から千葉	592	-
東京から茨城	325	-
神奈川から東京	6,980	1,000
埼玉から東京	5,571	1,000
千葉から東京	6,209	1,000
茨城から東京	259	-

3.2．報酬関数の特徴量選定

2.1 章紹介した報酬関数は，行動をフィーチャ$f_a \in R$が特徴づけ，重みベクトルθを用いて、これらフィーチャをパラメータ化した線形関数に従う．軌跡のフィーチャf_ζ は，軌跡ζに含まれる動作のフィーチャの合計である．フィーチャに適用される軌跡の報酬の線形和は次のように定義する：

$$reward(s,a) = \theta^T f_{(s,a)}$$

$$reward(f_\zeta) = \theta^T f_\zeta = \sum_{s_i \in \zeta} \theta^T f_{sj}.$$

というのは，エージェントは選択した動作がもたらす報酬は，個々の状態・動作ペアは目的地該当メッシュ内の情報（夜間人口，従業員数，事務者数，学校数，集客施設数）と移動コストからなる特徴量ベクトルと報酬関数のパラメータによる決まる．

3.3．結果

3.3.1 行動軌跡から推定した報酬関数

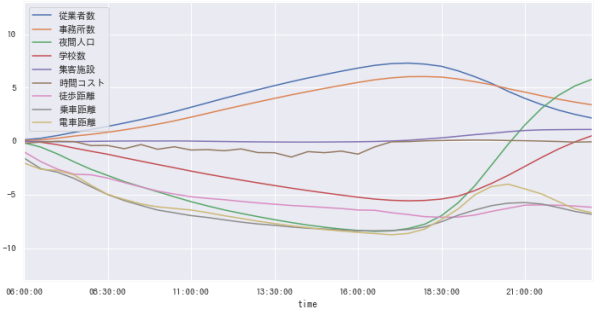

図 1　報酬関数推定結果

表 1 のように報酬関数を推定するために，個々の行動パターンに対して「人の流れデータ」から 2000 人分の 1 日分を抽出し，収束するまで約 1，000 回の繰り返し計算を行った．東京に在住し，神奈川へ通勤通学する対象者の報酬関数パラメータ（各特徴量の重み）について図−1 に示す．各パラメータは時間ごとに変化しつつ，例えば，目的地の従業者数と事務所数は 6 時から 16 時まで向上し，商業施設が多いエリアは移動者に対して吸引力が高く，一方，昼間に夜間人口が多いエリアの吸引力が低下していく傾向が見える．それに対して，移動にかかるコスト（時間・距離）は基本的にマイナスだが，朝・夜の通勤時間帯では移動への対抗が弱くなっている．また，「集客施設」や「乗車距離」に関するパラメータは終日ゼロに近くとどまり，訓練データのなか「車で移動する」と「集客施設に訪問する」サンプルが少ない原因である．

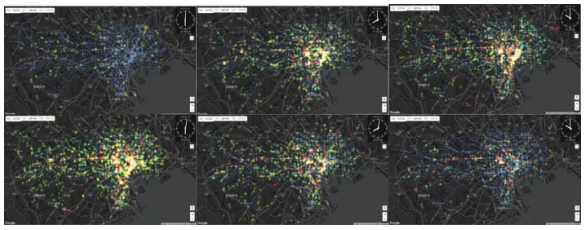

図 2　擬似人流データの可視化結果（東京都例）

このような報酬関数を用いて作成した擬似人流デ

ータの可視化例を図２に示す．提案手法はメッシュ単位に行動するため，具体的位置はメッシュ内の道路ノードにランダムに配分する．各点はエージェントとし，当時点の交通手段により色分けしている（青：滞在；黄：徒歩；緑：車；赤：電車）．

次に，鉄道の利用者人数と乗用車の利用人数についてシミュレーションの結果比較を行った．図 2 は住宅地―通勤通学地別の鉄道乗車人数の結果である．ここで，乗り換えは考慮していないものの，時間は 30 分に離散化したので，30 分以上のトリップは区切りした. 全体的に見ると, 県を跨ぐトリップより，各県内また東京都内の移動が多く予測される傾向が見える．提案手法は報酬関数を用いてエージェントの行動政策を計算し，1 日中の高い累積報酬をもたらす行動を選択し，自律的行動ルールを生成する．そして，滞在に対して短い距離で頻繁に高い吸引力の場所を訪問するパターンになる．一方，県を跨ぐ長距離トリップは移動コストの制限で正しい OD 量を予測した．

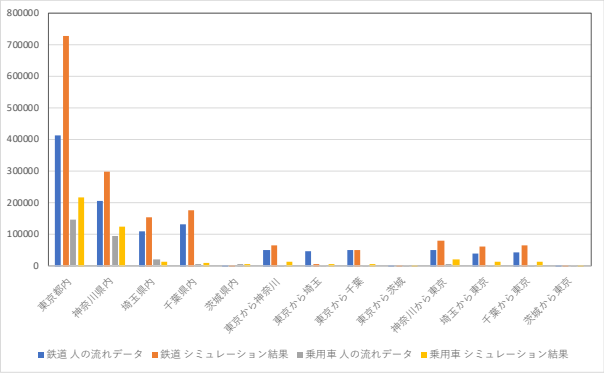

図 3 県レベル OD 別の鉄道・道路利用者数

また, 図 4 には,「人の流れデータ」を観測値とし，時間ごとのシミュレーション結果と推定人口の分布を散布図で示す（横：真値，縦：推定値）. 図の中の数字は真値と推定値との相関係数である. エージェント分布の誤差は通勤通学時間帯の開始から大きくなり，昼間に 0. 6 ぐらいの相関にとどまり，帰宅行動の後に回復した．エージェントは自律的に行動開始・終了の時間を正しく制御できるが，行動場所・およそ行動回数のプランニングは十分高い精度と至ってない．これは，エージェントは目的地を選択する際，平均的 2，000 個ぐらいの潜在目的地選択肢があり，設計した強化学習アルゴリズムは

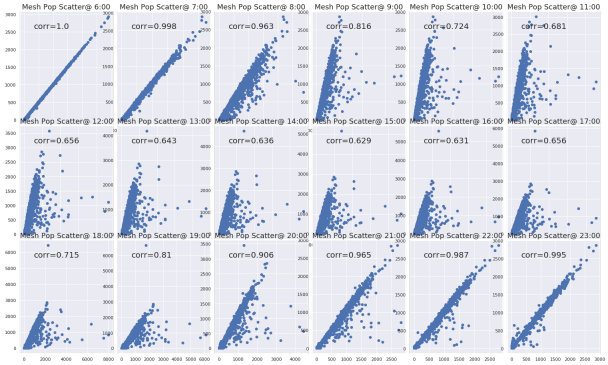

図 4 メッシュごと人口分布の比較結果

4．まとめ

本研究では，強化学習に基づくエージェントシミュレーション手法を用いた都市圏レベルの人の流れを再現した．その成果として，擬似人流データを生成し．また，メッシュ人口と OD 交通量で生成した擬似人流を実際の人の流れデータと比較した．

今後の課題として，まずエージェントモデルのパフォーマンスを改善することが挙げられる. さらに，提案手法の実用性を向上することが必要である．本研究では，強化学習の計算量を削除ため，個々のエージェントを初期化する際に目的地を県内・県外に指定した．しかし，実際応用する際，このような詳細な初期情報が利用できるシナリオほとんどない．エージェントはより多く属性（年齢・職業・収入など）情報を考慮し，自律的に広い範囲で現実的な行動できると，実用化に向けて大きく役立つと期待できる．

参考文献

Pang, Y., Kashiyama, T., Yabe, T., Tsubouchi, K., & Sekimoto, Y. (2020). Development of people mass movement simulation framework based on reinforcement learning. Transportation Research Part C: Emerging Technologies, 117, 102706.

Ziebart, Brian D., et al. (2008). Maximum entropy inverse reinforcement learning. AAAI. Vol. 8.

中村敏和， 関本義秀， 薄井智貴， & 柴崎亮介. (2013). パーティクルフィルターを用いた 都市圏レベルの人の流れの推定手法の構築. 土木学会論文集 D3（土木計画学），69(3), 227-236.