

# 都市を対象とした大規模移動履歴に基づく 疑似人流データ生成手法

田村 直樹<sup>1</sup> 浦野 健太<sup>1</sup> 青木 俊介<sup>3</sup> 米澤 拓郎<sup>1</sup> 河口 信夫<sup>2,1</sup>

受付日 xxxx年0月xx日, 採録日 xxxx年0月xx日

**概要:** 都市計画や研究に用いるために、交通シミュレーションや感染シミュレーションの需要が高まっている。一方、スマートフォンの普及により、高精度で大量のユーザの移動履歴収集が可能になり、このデータのシミュレーションへの活用が期待されている。しかし、いまだに都市の全人口分の移動軌跡を再現できる詳細な移動履歴の収集は困難である。同時にこのような個人の移動履歴は住所や職場などの個人情報を含むため、現実のデータのそのままの活用には制約がある。そこで、合成的な移動軌跡を生成する手法が盛んに研究されているが、「大量の教師データが必要」、「移動軌跡が取れる詳細な移動履歴が必要」といった課題があった。本手法では現実の移動履歴データを用いて、教師なし手法によりユーザの移動・滞在傾向をモデル化し、都市レベルの疑似的な人流データを生成する。本手法のポイントは、教師なし学習を用いたモデリングによって、移動履歴データに特別なラベル付けを必要としない点、また収集頻度の高いGPSデータのみを用いる点である。これに加えて、生成データに実際のユーザの移動軌跡は含まれず、全ての移動軌跡は確率的に生成されるため、現実のユーザの個人情報が含有されにくく、現実のデータに付随するプライバシー上の課題の解決が期待できる。また評価実験を行ったところ、本手法は従来手法よりも空間的に高い粒度の生成データで都市の人流を近似できることが確認できた。

**キーワード:** 人流モデリング, アーバンコンピューティング, シミュレーション

## Synthetic Human Flow Generation Based on Large-Scale Movement History for Urban Areas

NAOKI TAMURA<sup>1</sup> KENTA URANO<sup>1</sup> SHUNSUKE AOKI<sup>3</sup> TAKURO YONEZAWA<sup>1</sup>  
NOBUO KAWAGUCHI<sup>2,1</sup>

Received: xx xx, xxxx, Accepted: xx xx, xxxx

**Abstract:** There is a growing demand for traffic and infection simulations for use in urban planning and research. On the other hand, the widespread use of wearable devices has made it possible to collect a large amount of user location history with high accuracy, and it is expected that this data will be used for simulation. However, it is difficult to collect location histories for the entire population of a city, and detailed data that can reproduce trajectories is expensive. In addition, such personal location histories contain private information such as addresses and workplaces, which restricts the use of raw data. Therefore, methods to generate synthetic movement trajectories have been actively studied, but they require a large amount of teacher data and a detailed movement history from which movement trajectories can be obtained. Using real-world movement history data, this method models users' movement and stay tendencies with unsupervised learning, and generates synthetic human flow data at the city level. The key points of this method are that it uses unsupervised learning for modeling, does not require special labeling of the movement history data, and uses only frequently collected GPS data. Using this method, we generate synthetic human flow data without private information. Evaluation experiments confirmed that this method can approximate urban human flows with spatially more granular generated data than conventional methods.

**Keywords:** Human mobility modeling, Urban computing, Simulation

## 1. はじめに

都市計画のために交通シミュレーションや混雑予測が必要とされている [1][2]. 特に新型コロナウイルス流行の影響で、感染予測や影響分析に活用するための感染シミュレーションの需要が高まっている [3]. これらのシミュレーションには、人々が日々の生活の中でいつ、どのように移動・滞在を行なっているかを示す人流データが不可欠である. 一方、GPS (Global Positioning System) 機能を備えたスマートフォンやウェアラブル端末の普及により、大量のユーザの移動履歴を高精度で収集可能になった. そこで、人々が都市環境でどのように移動・滞在しているかという都市レベルの人流データの、シミュレーションへの活用が期待されている [4]. しかし、いまだに都市の全人口分の移動履歴の収集は不可能であり、移動軌跡を再現可能な詳細なデータは収集コストが高い. なぜならこのようなウェアラブル端末で収集する位置情報は、多くの場合アプリの使用に紐づいてデータを取得するため、ユーザ数が限られ、また端末の電力消費を抑えるために収集間隔は広がっている. 同時に、このような個人の移動履歴は住所や職場などの個人情報を含むため、現実のデータのそのままの活用には制約がある [5]. つまりシミュレーションに用いる人流データとして、現実のユーザの個人情報を含まず、かつ都市人口規模の移動特徴を再現できるデータが必要である. そこで、本研究では現実の移動履歴データに基づいて、都市レベルの人流を近似できる疑似的な人流データの生成を目的とする. さらに生成データは現実のユーザの移動履歴とは一致しないため個人情報を含まず、自由に加工、可視化できる.

疑似人流データ生成には一人一人の市民を表すエージェントが時間経過とともにどのように移動・滞在を行うかを定義する活動モデルが必要である. ユーザの活動モデリング手法は過去にも多く研究されてきたが、疑似人流データ生成に応用する上で、「大量のラベル付けされた正解データが必要」、「移動軌跡が取れる詳細な移動履歴が必要」という課題が存在した. 基本的に現実の GPS 移動履歴データは、滞在地点の座標と測位時刻を示すタイムスタンプの形で得られ、多くの場合測位頻度が低いため移動軌跡の再現は非常に難しい. 従来手法では、これらの滞在に対して、どのようなエリアへの滞在なのか、またどのような活動のための滞在なのかというラベル付けが必要であった [6][7][8]. しかし、人間の活動は高次元の情報であるため解釈が難しく、現実のデータへのこのようなラベル付けは困難である.

一方、現実の GPS データと教師無し学習を用いて、地域の使われ方、ユーザの活動傾向をモデリングする手法が研究されている [9][10]. これにより、従来よりも測位頻度が少なく、ラベルのないデータを元に、ユーザの活動をモデリングできる. これは日常的に収集されている GPS データに基づいたモデルであるため、社会政策の変更、パンデミックの発生、季節の変化など時間経過に伴う環境の変化を加味できる利点もある. [10] ではコロナ前後という環境の変化に伴うエリアの変化がモデリングできていることが示されている. しかし [9] におけるユーザモデリングは、ユーザの滞在傾向のみに注目したものであり、移動傾向を考慮できていない. そこで本稿では、疑似的な人流再現に不可欠な、ユーザがいつ、どのように移動するのかという移動傾向をモデリングし、滞在傾向モデリングと合わせることで、より高度なユーザモデリングを実現する. このような教師なし学習ベースのユーザモデリングを用いて、現実のユーザの移動・滞在傾向に基づく疑似人流データの生成を行う.

本手法の有用性を検証するために、生成元のデータや他のデータセットと比較して評価を行なった. 具体的には、生成データが現実の滞在人口分布を再現できているか、移動回数、移動距離などの移動特徴を再現できているか、といった点で評価した. 結果として、従来手法と比較して高精度で都市の滞在傾向を再現できているか、また空間的に高い粒度のデータ生成が実現できることが示された.

本論文の貢献は次の 3 つである.

- (1) 教師なし学習によるモデリングを用いて、現実の GPS 移動履歴データセットに基づく疑似人流データ生成手法を開発したこと.
- (2) 従来の滞在傾向のみに注目したユーザモデリングに対して、移動傾向を考慮した手法を開発したこと.
- (3) 生成データを滞在・移動の二つの観点から評価し、本手法の有用性を証明したこと.

本論文の構成は次に示す通りである. まず 2 章で既存のユーザモデリング手法やシミュレーション用の人流データ生成に関する研究における課題を整理する. 3 章では提案する疑似人流データ生成手法について解説し、4 章では現実のデータセットを用いて疑似データ生成実験、生成データの評価を行う. 最後に 5 章でまとめと今後の展望について述べる.

## 2. 関連研究

合成的な人流データの生成には、人々が日々どのように移動・滞在をしているかのモデル化が必要である. このようなモデル化の試みとして、パーソントリップ調査をベースとしたものが古くから知られている [11][12]. しかしパーソントリップ調査によるデータは収集コストが高く、収集頻度が低いため、典型的な活動パターンしかモデ

<sup>1</sup> 名古屋大学大学院 工学研究科  
Graduate School of Engineering, Nagoya University

<sup>2</sup> 名古屋大学 未来社会創造機構  
Institutes of Innovation for Future Society, Nagoya University

<sup>3</sup> 国立情報学研究所  
National Institute of Informatics

ル化できず、また環境の変化に伴う人々の活動の変化をモデリングできない。これに対し、近年のモバイル端末の普及により、個人の CDR (Call Detail Record) や GPS 移動履歴が大量に得られ、これを用いた活動モデリングが盛んに行われている。Yin ら [13] は CDR による移動履歴を用いて、隠れマルコフモデルによる活動モデリングを行っている。Ouyang ら [7] は GPS の軌跡データを用いて、人間の移動性をモデリングし、移動軌跡を合成的に生成している。しかし、CDR は GPS と比較して空間的粒度が粗いため疑似人流生成には不向きであり、[7] で用いる測位頻度の高い GPS データは日常的な収集が難しくユーザ数が限られる欠点がある。

人間の活動傾向をモデリングする手法として、深層学習を用いた手法の研究も盛んである。Song ら [8] は、LSTM マルチタスク学習による人間の移動パターンや交通パターンの学習、都市レベルのシミュレーションシステムを実装している。Borysov ら [14] は深層学習ベースの手法を用いて、より多くの、より多様なユーザモデルを生成しており、またその要素の組み合わせによってサンプルにないモデルの生成を目指している。また、多様で複雑な要因下での、ユーザの日々の活動スケジュールをモデリングする研究も挙げられる [15][16]。他にも強化学習ベースの活動モデリングによって、より自然な移動軌跡を生成する研究も存在する [6][17]。これらの手法の課題としては、モデルを学習するために、詳細で大量のラベル付けされたデータや適切な報酬関数の設定が必要ということである。しかし、一般的に個人の移動は様々な要因に基づく高次元な情報であるため、移動軌跡のラベル付けはコストが高く、広範囲の大規模なデータセットの生成が困難となる。また恣意的なラベル付けによって、データに本来存在する典型的でない少数のユーザやエリアを考慮できなくなる恐れがある。

現実の人流データにおける個人のプライバシーを侵害する危険性は度々指摘されており [19]、疑似的に都市の人流を近似できるデータセットが求められている [4][20]。都市レベルの疑似人流データを実際に生成している研究として、GPS、CDR に加えて、国勢調査による人口分布や交通量などのデータセットを総合的に用いたものが挙げられる [22]。この手法における生成データの空間的な粒度は 250-500m メッシュ程度であるが、混雑予測や感染シミュレーションのためにはより詳細な粒度の人流データが必要である。また国勢調査のデータなど、収集頻度の低いデータを用いるため、日常的な環境変化を考慮できないという課題がある。本研究ではより詳細な粒度の人流生成を目指しており、ラベルのない GPS 移動履歴から個人情報を含みしない疑似人流を生成することを目指す。GPS 移動履歴データは収集頻度が高いため、日常的な環境変化も考慮した疑似データを生成できる。

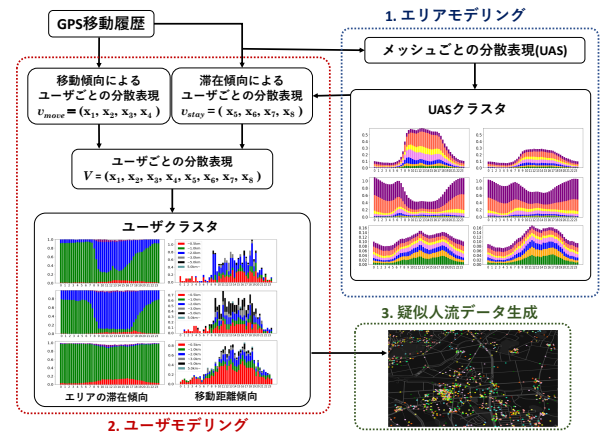


図 1: 提案手法の流れ

Fig. 1 Overview of the proposed method

### 3. 提案手法

#### 3.1 手法の概要

本研究では元となる GPS 移動履歴データのみを用いて、現実の人流を近似できるような疑似的な人流データを生成する。疑似人流データ生成の流れを図 1 に示す。入力となる移動履歴データは、端末を持つユーザの緯度、経度、タイムスタンプとして得られる。疑似人流データも同じく疑似ユーザの緯度、経度、タイムスタンプとして生成する。まず最初に、既存手法 [10] を用いて対象のエリアをモデリングする。次に現実のユーザを、いつ、どんなエリアに移動・滞在を行うかという特徴を用いてモデリングする。最後にユーザをモデルとして、エージェントによる移動履歴を出力し、疑似人流データを生成する。以降ではそれぞれの手順について順に詳しく説明する。

#### 3.2 Area2Vec によるエリアモデリング

まず、Area2Vec[9][10] を用いたエリアモデリングを行う。具体的には、対象地域をメッシュ状に分割し、Area2Vec を用いて各メッシュに対する分散表現を獲得後、クラスタリングによってメッシュを分類する。Area2Vec のアーキテクチャを図 2 に示す ([10] 図 2 から引用)。図中の  $N$ ,  $N'$ ,  $d$  はそれぞれモデリング対象のメッシュ数、出力層のノード数、分散表現の次元数である。最終的に入力層から隠れ層への重み  $W_{N \times d}$  をメッシュごとの分散表現として獲得する。この分散表現は、メッシュごとの出力層ノード特徴量の傾向を反映した、 $d$  次元のベクトルとなる。つまり、いつ、どんな滞在が行われたかというメッシュごとの滞在傾向を反映した分散表現となる。本論文では [10] に従って、以降この分散表現を UAS (the Usage of Area with Stay information) と表記する。獲得した UAS のクラスタリングによって、メッシュそれぞれの滞在傾向による分類が可能となる。図 3 に実際に 3 クラスタに分類した場合の各

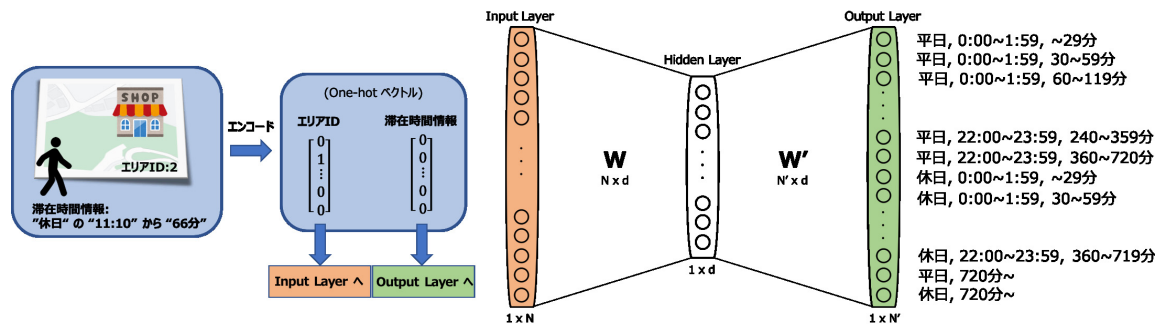


図 2: Area2Vec のアーキテクチャ

Fig. 2 Architecture of Area2Vec

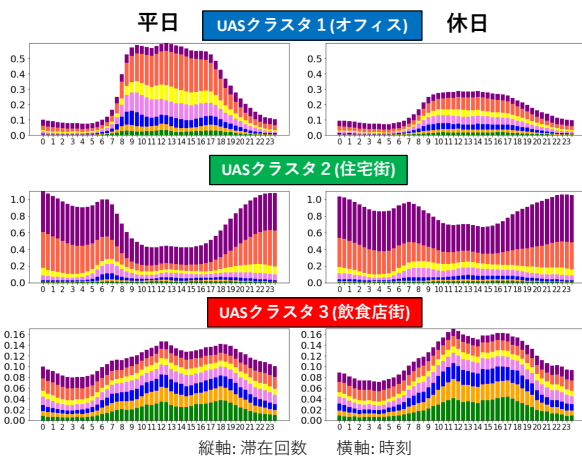


図 3: UAS クラスタごとの滞在傾向

Fig. 3 Stay tendency by UAS cluster

クラスタの滞在傾向を示す。グラフの色はそれぞれ滞在時間の長さを示し、緑が最も短く、紫が最も長い滞在時間の滞在である。また、本論文では以降 UAS のクラスタリングによる各クラスタを UAS クラスタと表記する。図 3 の UAS クラスタ 1 は、平日昼間の長期滞在が目立ち、休日はその割合が減少していることからオフィス街のような滞在傾向のエリアと見て取れる。一方 UAS クラスタ 2 は、夜間の長期滞在が目立ち、休日は昼間の滞在も増加していることから住宅街のようなクラスタである。このように、メッシュごとの滞在傾向を反映した UAS を獲得し、クラスタリングを行うことでエリアモデリングが可能である。

### 3.3 ユーザモデリング

次に、疑似人流データ中のエージェントのモデルとするために、現実のユーザをモデリングする。具体的には移動・滞在傾向をユーザごとの分散表現  $V$  として抽象化し、クラスタリングによってユーザ进行分类する。そしてクラスタごとの確率的な移動・滞在特徴を疑似人流データ中のエージェントのモデルとする。

ユーザごとの分散表現  $V$  の計算手順を図 4 に示す。まず最初に各ユーザの移動履歴を 30 分ごとのタイムスロッ

トに分割し、各スロットごとの滞在メッシュを抽出する。そして、[9] の User2Vec アーキテクチャに従って、ユーザごとの分散表現  $v_{stay}$  を獲得する。これは、図 4 の左側にあたる。入力層は各ユーザごとのノード、出力層は量子化された滞在特徴ごとのノードである。滞在特徴として、平日か休日か、滞在時間帯、滞在 UAS クラスタの 3 つの情報を用いた。これにより  $v_{stay}$  はユーザが、いつ、どんなエリアに滞在したかという滞在傾向に従った分散表現となる。しかし、疑似人流データ中のエージェントのモデルとしては、現実のユーザの移動傾向も考慮する必要がある。なぜなら、シミュレーションに利用する上で、エージェントがどんな移動を行うかは重要な情報になり得るからである。そこで、本手法では新たに現実のユーザの移動特徴を抽出し、分散表現  $v_{move}$  を獲得する。移動特徴として、平日か休日か、移動時間帯、移動距離、主要メッシュへの移動かという 4 つの情報を用いた。ここで主要メッシュとは、それぞれのユーザごとに多くの日で滞在が確認できた複数のメッシュであり、そのユーザにとっての自宅、職場などのメッシュである。本手法では、ユーザごとにデータが得られた日数の半分以上の日において滞在が観測されたメッシュをそのユーザにとっての主要メッシュとした。これにより  $v_{move}$  はユーザごとのいつ、どんな移動を行なったかという移動傾向に従った分散表現となる。最後に  $v_{stay}$ ,  $v_{move}$  の結合によってユーザごとの移動・滞在傾向による分散表現  $V$  を計算する。従って  $v_{stay}$ ,  $v_{move}$  とともに次元数を  $N$  とした時、 $V$  は  $2N$  次元の分散表現となる。

次に、ユーザごとの分散表現  $V$  を k-means を用いてクラスタリングする。図 5～図 7 には実際に 40 クラスタにクラスタリングしたものからいくつかのクラスタを例として示す。図 5 は平日、休日それぞれの各時間帯に滞在した UAS クラスタの割合を示しており、図 6、図 7 は、各時間帯の移動回数とその移動距離の割合、主要メッシュへの移動の割合を示している。図 6、図 7 の縦軸は、各時間帯 (30 分ごと) における移動の割合であり、各ユーザクラスタの人数、平日、休日の日数について正規化している。図 5 より、クラスタ 1, 2 は夜の時間帯に住宅街クラスタ、昼の時



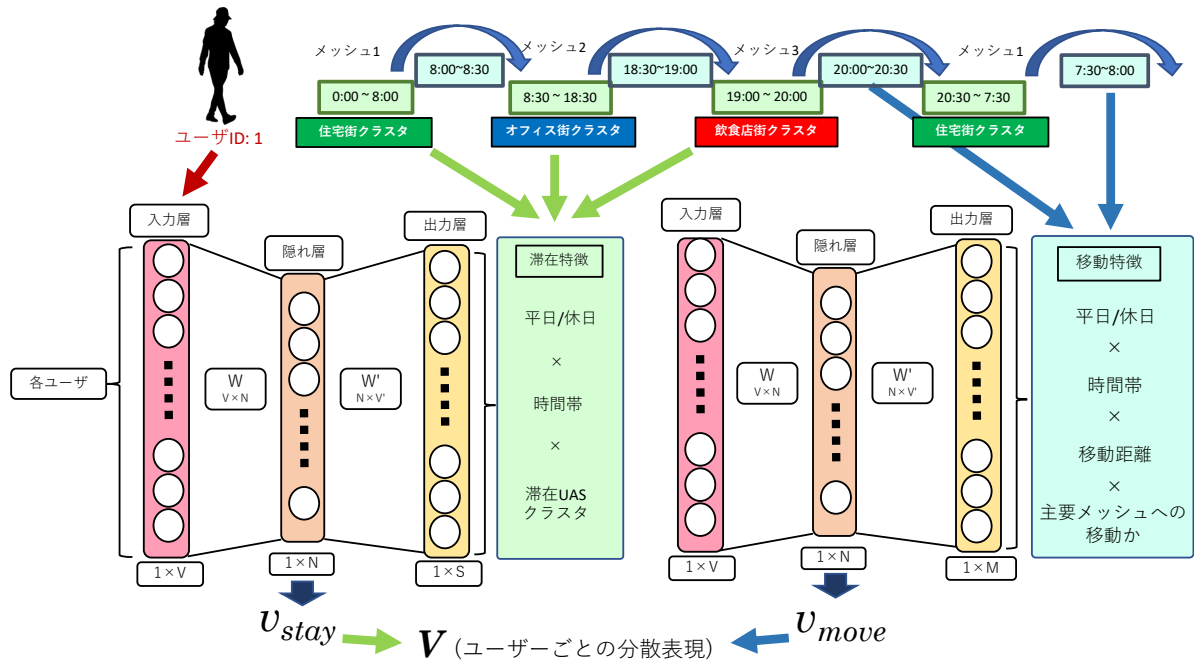


図 4: ユーザごとの分散表現の獲得

Fig. 4 Acquisition of distributed representation for each user

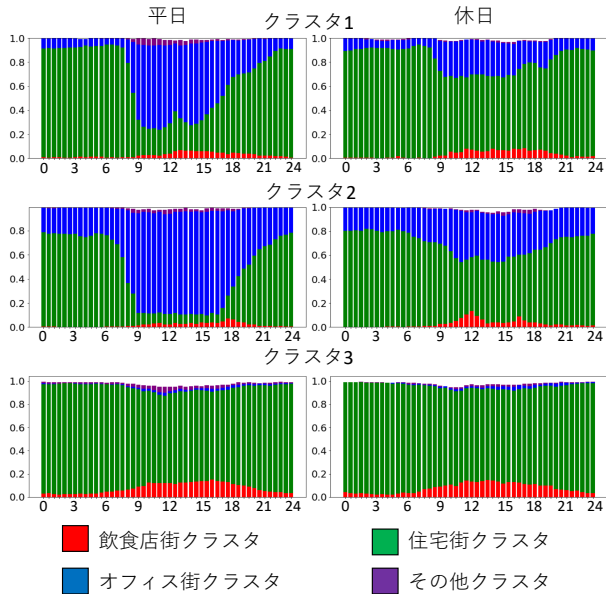


図 5: 滞在 UAS クラスター

Fig. 5 Stayed UAS cluster

間帯にオフィス街クラスターに滞在する割合が高く、一般的なサラリーマンに近い滞在傾向である。一方クラスター3は昼間に飲食店街クラスターに滞在することがある以外は住宅街クラスターに滞在しており、主婦のような滞在傾向を示している。図6を見ると、クラスター1は1km~2kmの移動が多く、また昼時の12時~14時に移動が多くみられる。それに対してクラスター2は朝の7時~9時と夕方17時~20時に移動のピークがあり、3km~5kmの移動が目立つ。クラスター3はクラスター1, 2と比較すると移動が少なく、移動距離も0.5km以下の割合が高い。また図7から、クラ

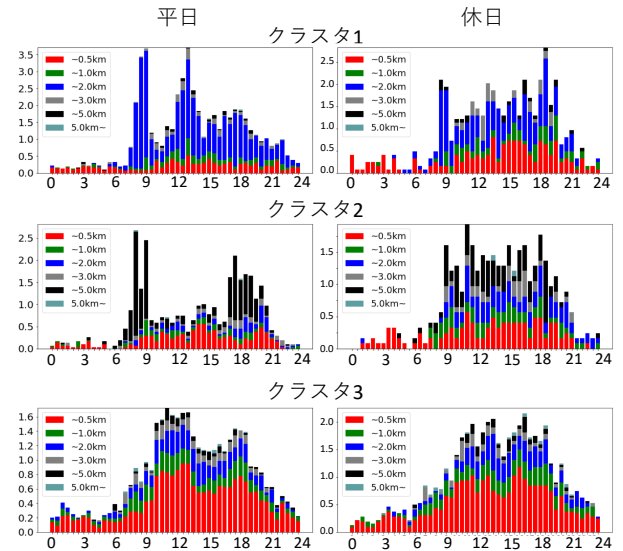


図 6: 移動距離の傾向

Fig. 6 Travel distance tendency

スタ1, 2の朝, 夕方の移動は主要メッシュへの移動割合が高く、職場への出勤、自宅への帰宅が見て取れる。クラスター3はクラスター1, 2と比較して昼の時間帯に主要メッシュ以外への移動が多い。これらのクラスターは一例であるが、このようにVをクラスタリングすることでユーザの滞在傾向だけでなく、ユーザの移動傾向も考慮した分類ができる。各クラスターは移動・滞在傾向が似ているユーザ群となり、それぞれのクラスターの時間帯ごとの移動・滞在傾向をエージェントのモデルとして利用する。

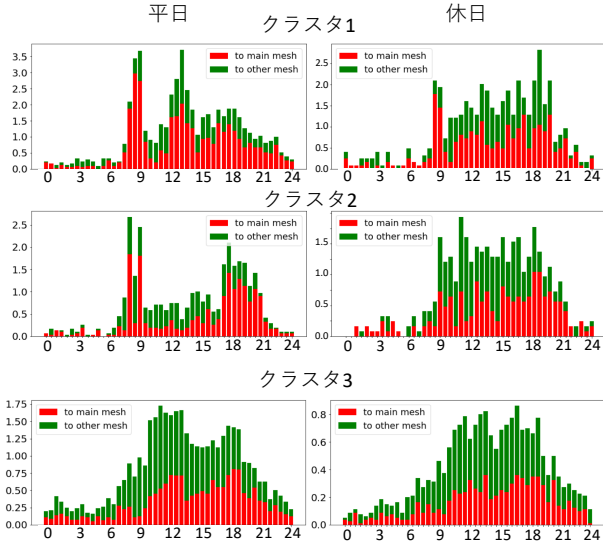


図 7: 主要メッシュへの移動傾向

Fig. 7 Travel to main mesh tendency

### 3.4 疑似人流データの生成

最後に、分類したそれぞれのユーザー群からエージェントの活動をモデル化し、各エージェントの移動履歴を出力する方法について説明する。まず最初に、分類したどのユーザークラスタをモデルとして利用するか、エージェントごとに決定する。あるエージェントがユーザークラスタ  $u$  をモデルとして利用する確率  $P(u)$  は、ユーザークラスタ  $u$  に属するユーザーの人数  $N_u$  を用いて、

$$P(u) = \frac{N_u}{\sum_{k=0}^U N_k} \quad (1)$$

とする ( $U$  はユーザークラスタ数)。すなわち、属するユーザーの多い典型的なクラスタはエージェントのモデルとして選ばれやすい。

次に、ユーザーモデルに従ってエージェントの移動履歴を生成する。具体的には現滞在中のメッシュに対して、30 分ごとの各タイムスロット中に「1. 移動するか留まるか」、「2. 移動する場合はどのメッシュに移動するか」を確率的に選択していく。まず現滞在メッシュから移動するかについては、リアルユーザーの平均移動回数  $\lambda$  に基づくポアソン分布に従って決定する。すなわちあるタイムスロット  $t$  で移動しない確率  $P_{stay}$  は

$$P_{stay} = \frac{\lambda_t^0 \exp(-\lambda_t)}{0!} = \exp(-\lambda_t) \quad (2)$$

同時に、移動する確率は

$$P_{move} = 1 - \exp(-\lambda_t) \quad (3)$$

となる。この平均移動回数  $\lambda$  は各タイムスロット、移動前 UAS クラスタごとに平均をとる。これは移動の生起は時間帯だけでなく、移動前の UAS クラスタにも相関があるという考えに基づく。例えば、サラリーマンクラスタのエージェントが朝の 8 時に住宅街クラスタにいる場合とオ

フィス街クラスタにいる場合の移動確率は異なると考えられる。

ここから、移動する場合どのメッシュに移動するかの決定について説明する。これには主に、(1) 時間帯ごとの滞在密度、(2) UAS 類似度、(3) 移動距離の 3 つを考慮しており、順に説明する。

- (1) 滞在密度について、時間帯ごとのリアルユーザーの滞在密度を前もって計算しておく。あるタイムスロットにおける、滞在密度  $D$  は、メッシュ  $m$  内のユーザー数  $C(m)$  を用いて、

$$D(m) = \frac{C(m)}{\sum_{k=0}^M C(k)} \quad (4)$$

で計算する ( $M$  はメッシュの総数)。

これはエリア全体に対する、各メッシュにおける滞在割合を示し、無次元量となる。移動確率としてこれを考慮することで、滞在が多いメッシュへの移動確率を高く設定できる。

- (2) UAS 類似度については、移動先 UAS クラスタと類似した UAS のメッシュへの移動確率を高く設定する。移動先 UAS クラスタは属するユーザークラスタの時間帯ごと、現滞在 UAS クラスタごとの次滞在 UAS クラスタの割合 (図 5) に従って決定する。そして、移動先 UAS クラスタとの類似度に従って移動先メッシュを決定する。ここでは計算量の削減のために、UAS クラスタごとに UAS の平均をとり代表ベクトルとして利用する。例えば、ある UAS クラスタ  $c$  の代表ベクトル  $r_c$  は、

$$r_c = \frac{\sum UAS(cluster = c)}{M_c} \quad (5)$$

で表せる。(ただし  $M_c$  は UAS クラスタ  $c$  に属するメッシュの総数。) この各 UAS クラスタの代表ベクトル  $r_c$  とのコサイン類似度を元にそれぞれのメッシュへの移動確率を計算する。すなわち、移動先メッシュとしてあるメッシュ  $m$  を選択する確率  $P_{UAS}(m)$  はコサイン類似度を用いて、

$$P_{UAS}(m) = \frac{\cos(UAS_m, r_c)}{\sum \cos(UAS, r_c)} \quad (6)$$

と計算できる。これにより、移動先 UAS クラスタと類似度が高い UAS を持つメッシュへの移動確率を高く設定できる。すなわち、個々のエージェントはエリアの使われ方を反映した滞在遷移を行う。例えばサラリーマンクラスタによるエージェントは、平日の昼間はオフィスエリアに滞在し、夜間は住宅街エリアに滞在する確率が高くなる。

- (3) 移動距離については、モデルごとのリアルユーザーの移動距離の傾向 (平均距離と距離の分散) を元に、移動距離が適切なものへの移動確率を高く設定する。例え



図 8: 疑似人流データの可視化

Fig. 8 Visualization of synthetic data

ば、あるタイムスロット  $t$  における平均移動距離が  $\mu$ 、移動距離の分散が  $\sigma^2$  であったとする。この時、移動先メッシュとして移動距離が  $d_m$  であるメッシュ  $m$  を選択する確率  $P_{distance}(m)$  は、

$$P_{distance}(m) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(d_m - \mu)^2}{2\sigma^2}\right) \quad (7)$$

で表せる。

このように滞在密度、UAS 類似度、移動距離のそれぞれを考慮した各メッシュへの移動確率を計算し、その積によってメッシュへの移動を生起する。従って、エージェントのあるタイムスロットにおける、特定のメッシュ  $m$  への移動確率は

$$P_t(m) = D(m) * P_{UAS}(m) * P_{distance}(m) \quad (8)$$

と計算できる。

つまり、エージェントは各移動ごとに、滞在密度、UAS 類似度、移動元メッシュとの距離から移動先のメッシュを決定する。エージェントごとに各ユーザモデルから確率的に滞在遷移を生成することで、任意のエージェント数分の疑似人流を生成する。実際に生成したデータの可視化を図 8 に示す。図中のプロットが個々のエージェントを示しており、各エージェントの所属ユーザクラスごとに色付けしている。個々のエージェントは時間経過に伴ってモデル化された移動・滞在を行い、同時に生成元データで滞在密度の高いエリアにはエージェントが集まりやすくなる。

## 4. 評価実験

### 4.1 実験の目的

本章では提案手法の有効性を評価するため、実際に疑似人流データを生成し、生成データが現実の人流を近似できているか確認する。以下では実際に生成したデータを滞在密度、移動回数、移動距離といった都市人流の特徴を再現

できているか、という観点で評価する。評価手法については、[22] における手法を参考にした。都市部の人流データに関する正解データは存在しないため、基本的に元となったデータと、滞在密度については商用の人口統計データセットとも比較した。

### 4.2 データセット

実験に使用するデータセットについて説明する。本実験では、ブログウォッチャー社提供の GPS 滞在履歴データセットを用いた。これは、事前に同意を得たユーザのスマートフォンアプリから収集された滞在履歴である。ユーザごとの滞在履歴  $S$  は次に示す滞在情報  $I$  の連続として得られる： $I = (x, y, t, d)$ ,  $S = I_0, I_1, I_2, \dots, I_n$  ( $n$  は滞在地点の個数でありユーザごとに異なる)。 $x, y, t, d$  はそれぞれ、緯度、経度、滞在開始時刻、滞在時間の長さである。本データのサンプリングレートは数分～数日であり、GPS データであるため空間的な誤差も生じる。したがって本実験においては 30 分ごとにどのメッシュに滞在したかのメッシュ滞在履歴： $I' = (m, t', d)$ ,  $S' = I'_0, I'_1, I'_2, \dots, I'_n$  として扱う。 $m$  は 50m メッシュコード、 $t'$  は 30 分ごとのタイムスロットである。対象地域は愛知県日進市、期間は 2020/03/01～2020/03/31 のデータとし、対象地域で 1000 件以上位置情報が得られた 2155 人のユーザのデータを用いた。タイムスロット数は一日当たり 48、期間全体で  $48 \times 31 = 1488$  となる。したがって 1000 件のサンプルが得られたユーザは、サンプリングレートがほぼ均一と仮定した場合、 $1000/1488 = 67.2[\%]$  程度の滞在履歴を持つと推測できる。これをモデリングに十分なデータ量とし、これらのユーザのデータを抽出して実験データとした。UAS クラスタ数は 4、ユーザクラスタ数 40 とした。UAS クラスタ数を大きくすると、ユーザの滞在傾向による分散表現  $v_{stay}$  獲得において出力層ノード数  $S$  が大きくなる。これによって  $v_{stay}$  の学習が不安定になる可能性があるため、住宅街やオフィス街などエリアクラスタとして最低限必要と考えられる数に設定している。ユーザクラスタ数については、クラスタ数を大きくするほど多様なエージェントを定義できる。しかし、各クラスタに所属するユーザの数が減少するため、各時間帯における確率的な遷移のモデル化が不十分となる。ここでは、各クラスタに所属する平均ユーザ数が 50 人程度であればモデル化に十分と考え、ユーザクラスタ数を 40 としている。エージェント数は、30000 として疑似データを生成した。生成元データはスマートフォンアプリから収集しているため、主に 15～64 歳程度のユーザのサンプルであると考えられる。したがって本実験では、生成データの統計的な評価に十分な値として、日進市の 15～64 歳の人口 59128 の半数程度である 30000 エージェント分のデータを生成した。また、評価用としてモバイル空間統計 (NTT ドコモ) のデータセットを用いた。これは携帯基地局との通

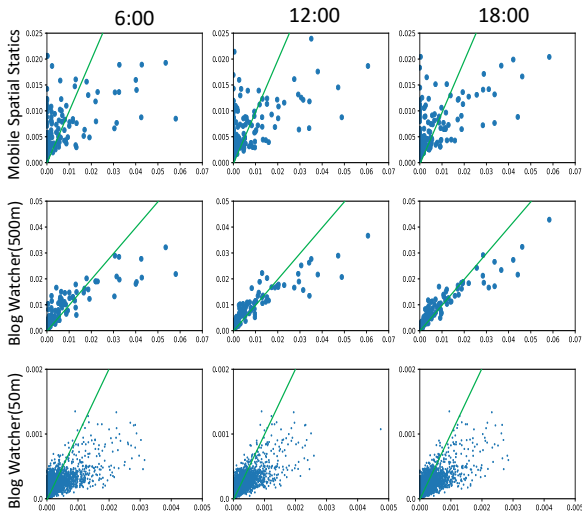


図 9: 滞在密度の比較

Fig. 9 Comparison of stay density

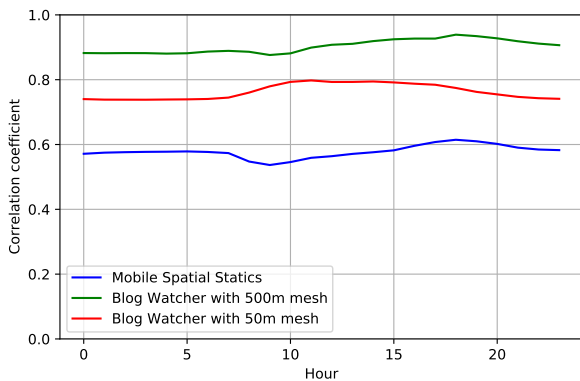


図 10: 時間帯ごとの相関の比較

Fig. 10 Comparison of correlation

信による時間帯ごとの滞在人口データであり、1時間ごとに500mメッシュごとのエリア内の人口データとして得られる。

### 4.3 結果

#### 4.3.1 滞在密度

滞在密度とは式(4)で計算される、各メッシュにおける滞在割合である。生成データの滞在密度を評価し、各時間帯における人口分布が再現できているかを確認した。図9は6:00, 12:00, 18:00の時刻で生成データの滞在密度を各データと比較したものである。上から、モバイル空間統計(500mメッシュ)、生成元データ(500mメッシュ)、生成元データ(50mメッシュ)の滞在密度と、生成データの滞在密度をメッシュごとにプロットしている。モバイル空間統計の人口分布データは500mメッシュの粒度のものしか存在しないため、このメッシュ粒度に合わせて比較している。直線は縦軸と横軸の値が等しい、45度の直線である。双方のデータにおける滞在密度の正の相関が強ければ、生

成データは現実の人口分布を再現できていると考えられる。モバイル空間統計とのプロットについては、モバイル空間統計データにおいて滞在密度が高いが、生成データにおいて滞在密度が低くなっているメッシュが存在している。また疑似データで滞在密度が高く、比較データで低いメッシュが目立つことから、疑似データにおいては一部のメッシュに滞在が偏っていることが分かる。生成元データとのプロットについては50m, 500mメッシュ両方で正の相関がみられる。図10は各時間帯ごとの相関係数をそれぞれのデータセットについて計算したものである。生成元データとは強い相関があり、500mメッシュの粒度では多くの時間帯で0.9前後と強い相関がある。これは[22]の手法における0.8前後の値と比較して高い値となっている。また50mメッシュの粒度でも0.75~0.8程度の値であり、空間的に高い粒度で人口分布を再現できている。しかし、モバイル空間統計との相関は0.5~0.6と低い値となっていることがわかる。

#### 4.3.2 移動回数と移動距離

図11は疑似データと生成元データの各時間帯のユーザーの移動回数の比較である。横軸は1時間ごとの時刻、縦軸は折れ線グラフが一日、一人当たりの平均移動回数、エラーバーが標準偏差を示している。ここで、移動とはタイムスロットごとのメッシュ単位での移動を指す。例えば0時における移動回数は、0:00~0:30と0:30~1:00のタイムスロットでのメッシュ間遷移のカウントである。各時間帯で0~2の値を取るため、平均に対して標準偏差の値が大きくなっている。

元データと比較して生成した疑似データは、ピークの時間帯などは再現出来ているが、夜中の時間帯における移動量が生成元データより少なくなっている。また昼間の移動については、平均、標準偏差ともに値が大きくなっていることが分かる。図12は疑似データと生成元データの移動距離の比較である。これより疑似データの移動は元データよりも全体的に長距離の移動が多い傾向にあり、特に0.5km以下の移動が少なくなっていることがわかる。

### 4.4 考察

実験によって得られた結果について考察する、まず滞在密度について、生成元のデータとの相関は高いが、比較用のデータとの相関は低いという結果となった。この原因として、生成元データがモバイル空間統計と比較してサンプル数が少なく、滞在履歴が偏ってしまうことが挙げられる。図13はモバイル空間統計データと生成元データの滞在密度の比較を地理的に描画したものである。左からモバイル空間統計、生成元データ、モバイル空間統計と生成元データの滞在密度差を描画している。滞在密度差は、モバイル空間統計の値から生成元データの値を引いたものである。これより、生成データは特にエリアの周縁部の滞在密度が



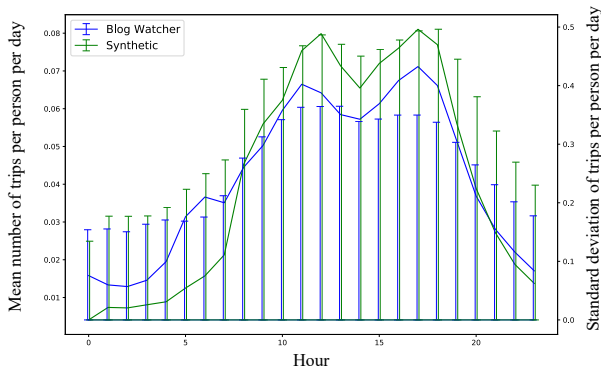


図 11: 移動量の比較

Fig. 11 Comparison of amount of travel

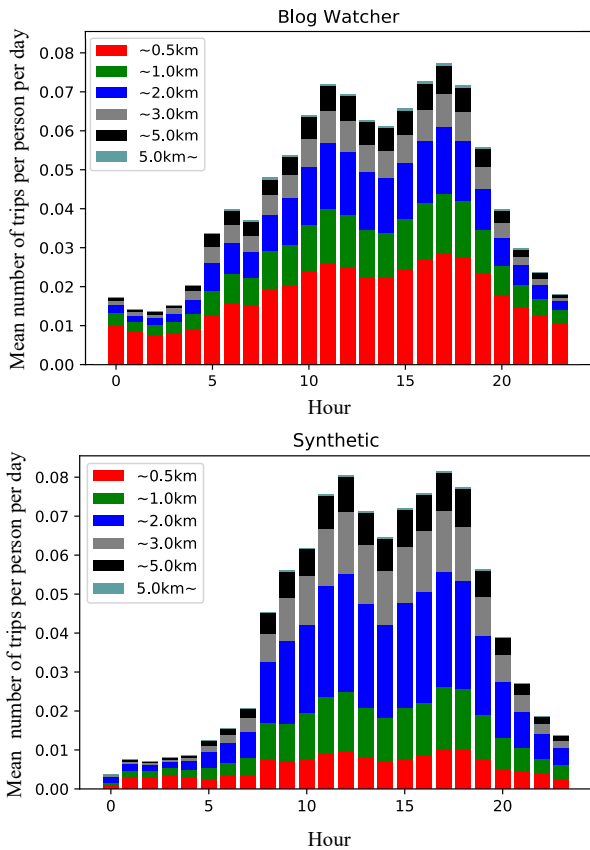


図 12: 移動距離の比較

Fig. 12 Comparison of travel distance

低く、それ以外の滞在密度が高くなっている。疑似データの生成元のデータセットは、エリア内で期間内の多くの時間帯に位置を測位できているユーザに絞っており、一日の多くの時間エリア外で滞在するユーザは省かれている。これがエリアの周縁部において、疑似データの滞在密度とモバイル空間統計の滞在密度の値が解離している要因と考えられる。生成元データによる都市人流の近似という観点では、従来より空間的な粒度の高い 50m メッシュレベルで、

従来手法と同程度の滞在再現ができている。

次に移動について、疑似データは生成元データと比較して、夜間の移動が少なく、昼間の移動が多くなっている。また、疑似データにおける 0.5km 以下の移動回数が元データと比較して少ない結果となった。これはエージェントの活動モデリングにおいて、地理的な特徴を十分にモデリングされていないことが原因と考えられる。例えば日々の活動における移動距離は、ユーザの自宅の位置、自宅と職場の位置関係、行きつけのスーパーへの距離などの地理的な位置に依存する。従ってこのような現実のユーザの地理的な特徴のモデリングが必要と考えられ、今後の課題である。

## 5. まとめと今後の展望

本論文では、大規模 GPS 移動履歴データを活用して、都市環境における疑似的な人流データを生成する手法を提案した。特に教師なし学習によるエリアモデリング、ユーザモデリングによってデータの生成を実現したことで、ラベルのない GPS 移動履歴データのみを用いて、疑似人流データを生成することが出来た。生成データは 30 分ごとの 50m メッシュレベルで都市人流を近似した滞在遷移として出力される。本データは例として、各時間帯におけるメッシュレベルでの感染確率計算や、メッシュ間の交通需要としてシミュレーションに利用できる。この疑似人流データは、現実の滞在密度を従来より空間的に詳細な粒度で都市の人流を近似している。また、個々のエージェントは UAS 類似度によって滞在履歴を生起するため、[22] のように主に滞在密度のみから生起する場合と比較してエリアの意味を考慮できる可能性が高い。従って個々のエージェントはより人間らしい移動を行うと考えられる。

今後の展望としては、今回の評価において精度の低かった移動について高精度な人流データを生成するために、ユーザの活動モデリングに地理的な要素を加えるアプローチを考えている。さらに、各エージェントの移動経路や移動速度、移動手段なども都市人流の近似に不可欠な要素と考えられるため、これらを再現できるデータを生成するために、手法を改良していきたいと考えている。

また、感染シミュレーションや交通シミュレーションへの利用可能性を測るために、生成データのさらなる評価が必要であると考えている。例えば、[3] では、各タイムスロットで、100m 以内の滞在ユーザを接触リスクのあるユーザ群としている。[1] では、移動手段や経路ごとの交通需要をシミュレーションしている。従って本手法における生成データについても、近傍エージェント数の評価、交通手段や経路ごとの交通量の評価等が必要である。また本論文では、モバイル空間統計以外のデータセットと比較出来なかったが、他の多様なデータセットと比較することで、疑似データの評価を進めていく予定である。特に、個々のエージェントがどの程度人間らしい動きをしているかと

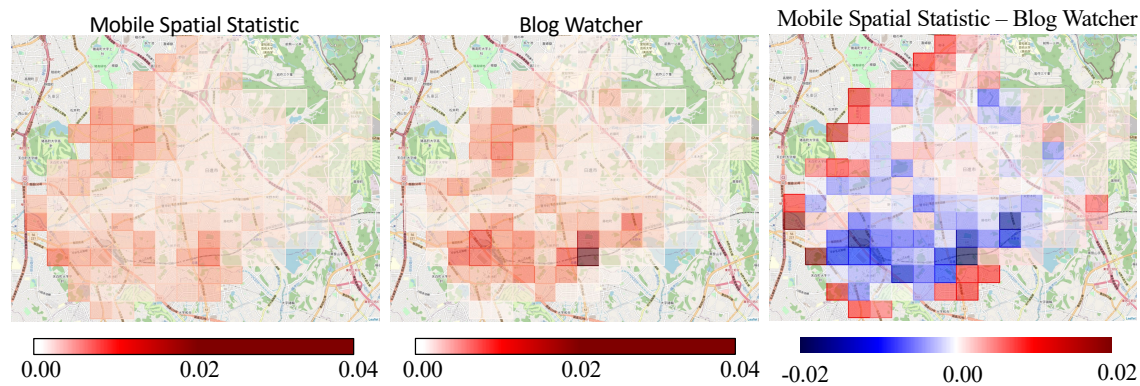


図 13: 滞在密度の地理的分布

Fig. 13 Geographic distribution of stay density

いった評価も必要と考える。これに加えて、疑似データが本当に個人情報を含んでいるか、といった点についてもさらなる評価と議論が必要である。

## 参考文献

- [1] Delhoum Y., Belaroussi R., Dupin F., Zargayouna M.: Activity-based demand modeling for a future urban district Sustainability, 12 (14), p. 5821, 2020.
- [2] Erath, A. Fourie, P.J. Eggermond, M.A. Ordonez, M.S.A. Chakirov, A. Axhausen, K.W.: Large-scale agent-based transport demand model for Singapore. Arb. Verk.-Raumplan., 790 2012.
- [3] Yabe, T. Tsubouchi, K. Fujiwara, N. Wada, T. Sekimoto, Y. Ukkusuri, S.: Non-compulsory measures sufficiently reduced human mobility in Tokyo during the COVID-19 epidemic. Scientific Reports volume 10, Article number: 18053, 2020.
- [4] Zhou, M. Li, J. Basu, R. Ferreira, J.: Creating spatially-detailed heterogeneous synthetic populations for agent-based microsimulation. Computers, Environment and Urban Systems 91, 2022.
- [5] Anda, C. Medina, S.A.O. Axhausen, K.W.: Synthesizing digital twin travellers: Individual travel demand from aggregated mobile phone data Transportation Research Part C: Emerging Technologies, 128, pp. 103-118, 2021.
- [6] Pang, Y. Tsubouchi, K. Yabe, T. Sekimoto, Y.: Development of people mass movement simulation framework based on reinforcement learning. Transportation research part C: emerging technologies, vol. 117, 2020.
- [7] Ouyang, K. Shokri, R. Rosenblum, D.S. Yang, W.: A Non-Parametric Generative Model for Human Trajectories. IJCAI(2018), pp. 3812-3817, 2018.
- [8] Song, X. Hiroshi, K. Ryosuke, S.: DeepTransport: Prediction and simulation of human mobility and transportation mode at a citywide level. IJCAI (2016), p. 16, 2016.
- [9] 庄子和之, 廣井慧, 米澤拓郎, 酒田理人, 河口信夫: 滞に関する時系列情報をういたエリア毎の分散表現の検討. マルチメディア, 分散, 協調とモバイル (DICOMO2020) シンポジウム. pp.1008-1013, 2020.
- [10] 庄子和之, 青木俊介, 米澤拓郎, 河口信夫: 大規模ユーザの滞在情報に基づくエリアの特徴付けと COVID-19 による影響分析. 情報処理学会論文誌, vol.62, No.10, pp.1644-1657, 2021.
- [11] Bhat, C.R. Frank, S.K.: Activity-based modeling of travel demand. Handbook of transportation Science, International Series in Operations Research & Management Science, vol 23, Springer, Boston, MA (1999), pp. 35-61, 2003.
- [12] Bowman, J.L. Moshe, E.B.: Activity-based disaggregate travel demand model system with activity schedules. Transport. Res. Part A: Policy Pract., 35 (1) (2001), pp. 1-28, 2001.
- [13] Yin, M. Sheehan, M. Feygin, S. Paiement, J.F. Pozdnoukhov, A.: A generative model of urban activities from cellular data. IEEE Trans. Intell. Transp. Syst., 19 (6) pp. 1682-1696, 2017.
- [14] Borysov, S.S. Rich, J. Pereira, F.C.: How to generate micro-agents? A deep generative modeling approach to population synthesis. Transportation research part C: emerging technologies, vol. 106, pp.73-97, 2019.
- [15] Drchal, J. Čertický, M. Jakob, M.: Data-driven activity scheduler for agent-based mobility models. Transportation research part C: emerging technologies, vol. 98, pp.370-390, 2019.
- [16] Vecchio, P.D. Secundo, G. Maruccia, Y. Passiante, G.: A system dynamic approach for the smart mobility of people: Implications in the age of big data. Technological Forecasting and Social Change, vol. 149, 2019.
- [17] Pang, Y. Tsubouchi, K. Yabe, T. Sekimoto, Y.: Replicating urban dynamics by generating human-like agents from smartphone GPS data. In Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 440-443, 2018.
- [18] Jeong, S. Kang, Y. Lee, J. Sohn, K.: Variational embedding of a hidden Markov model to generate human activity sequences. Transportation Research Part C: Emerging Technologies, vol. 131, 2021.
- [19] Bindschaedler, V. Shokri, R.: Synthesizing Plausible Privacy-Preserving Location Traces. IEEE Symposium on Security and Privacy (SP), pp. 546-563, 2016.
- [20] Borysov, S.S. Rich, J. Pereira, F.C.: How to generate micro-agents? A deep generative modeling approach to population synthesis Transportation Research Part C: Emerging Technologies, 106, pp. 73-97, 2019.
- [21] Chen, C. Liao, C. Xie, X. Wang, Y. Zhao, J.: Trip2Vec: a deep embedding approach for clustering and profiling taxi trip purposes Personal and Ubiquitous Computing, 23 (1), pp. 53-66, 2019.
- [22] Kashiyama, T. Pang, Y. Sekimoto, Y.: Open PFLOW: Creation and evaluation of an open dataset for typical

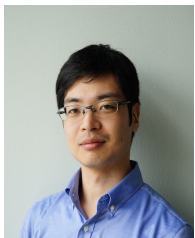
people mass movement in urban areas. Elsevier, Vol. 85, pp.249-267, 2017.

- [23] Ectors, W. Reumers, S. Lee, W.D. Choi, K. Kochan, B. Janssens, D. Bellemans, T. Wets, G.: Developing an optimised activity type annotation method based on classification accuracy and entropy indices Transportmetrica A: Transport Science, 13 (8), pp. 742-766, 2017.
- [24] Pappalardo, L. Simini, F.: Data-driven generation of spatio-temporal routines in human mobility Data Mining and Knowledge Discovery, 32 (3), pp. 787-829, 2018.
- [25] Hörll, S. Balac, M.: Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data Transp. Res. Part C: Emerging Technologies Elsevier Ltd, 130, 2021.
- [26] Diallo, A., Doniec, A., Lozenguez, G., Mandiau, R.: Agent-based simulation from anonymized data: An application to Lille metropolis.: The 12th International Conference on Ambient Systems, Networks and Technologies, Warsaw, Poland. pp.164-171, 2021.



**田村 直樹**

2021 年名古屋大学工学部電気電子情報工学科卒業。同年 4 月より同大学大学院工学研究科情報・通信工学専攻。



**浦野 健太** (正会員)

2016 年名古屋大学工学部電気電子・情報工学科卒業。2018 年に同大学大学院修士課程，2021 年に博士課程を修了。博士(工学)。屋内位置推定，生体信号のエンターテインメント利用に関する研究に従事。



**青木 俊介** (正会員)

2020 年カーネギーメロン大学計算機工学科 (Electrical and Computer Engineering) PhD 取得。船井情報科学振興財団 奨学生。2020 年 11 月より名古屋大学未来社会創造機構 特任助教。

2021 年 4 月より国立情報学研究所 助教。自動運転システム，サイバーフィジカルシステムに関する研究に従事。



**米澤 拓郎** (正会員)

2010 年慶應義塾大学大学院政策・メディア研究科後期課程博士号取得後，同大学院特任助教，特任講師，特任准教授を経て，2019 年より名古屋大学大学院工学研究科准教授。主に，ユビキタスコンピューティングシステム，ヒューマンコンピュータインタラクション，センサネットワーク等の研究に従事。ACM，各会員。



**河口 信夫** (正会員)

1990 年名古屋大学工学部電気電子工学科卒業。1995 年同大学大学院工学研究科情報工学専攻博士課程満了。同年同大学工学部助手，同大学講師，准教授を経て，2009 年より同大学大学院研究科教授。NPO 位置推定サービス研究機構 Lisra 代表理事。モバイルコミュニケーション，ユビキタスコンピューティング，行動センシングの研究に従事。博士(工学)。ACM，IEEE，人工知能学会，日本ソフトウェア科学会，電子情報通信学会，日本音響学会各会員。本会シニア会員。