

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¹ 浦野 健太¹ 青木 俊介³ 米澤 拓郎¹ 河口 信夫^{2,1}

概要：近年，都市計画のための交通シミュレーションや混雑予測，特に新型コロナウイルス流行の影響で，感染シミュレーションの需要が高まっている．一方，ウェアラブル端末の普及により，高精度で大量のユーザの移動履歴収集が可能になったため，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．しかしながら，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．また，このような個人の移動履歴は住所や職場などのプライバシー情報を含むため，現実のデータのそのままの活用には制約がある．そこで本論文では現実の移動履歴データからユーザの移動，滞在傾向を抽出，モデル化し，個人情報を含まない疑似的な人流データを生成する手法について提案する．

A Method for Generating Virtual People Flow Data Based on Large-Scale Movement History in Urban Areas

NAOKI TAMURA¹ KENTA URANO¹ SHUNSUKE AOKI³ TAKURO YONEZAWA¹
NOBUO KAWAGUCHI^{2,1}

1. はじめに

近年，都市計画のための交通シミュレーションや混雑予測，特に新型コロナウイルス流行の影響で，感染シミュレーションの需要が高まっている．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．例えば，大規模な移動履歴データを用いた，パンデミックの際の政策効果の分析などが行われている [1]．しかしながら，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．また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，本研究では現実の移動履歴データを元に，都市レベルの人流を再現できる疑似的な人流データを生成する．この疑似人流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．

疑似ユーザの人流生成には疑似ユーザ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，主に「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」といった課題があった．一方，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 [2]．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ル化出来る．また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．本論文では，この教師無し学習ベースの活動モデリングを用いた，疑似人流データ生

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

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

³ 国立情報学研究所
National Institute of Informatics

成手法を提案する。

2. 関連研究

前章で示したように、疑似的な人流を生成するためには、それぞれの疑似ユーザがどのように移動、滞在するかといった活動モデルが必要である。人々が日々どのように移動、滞在しているかをモデリングする試みとして、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 [3][4]。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なっている。例えば Song ら [5] は、主に GPS データを用いて、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。Yin ら [6] 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。Borysov ら [8] 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 [9][10]。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 [11][12]。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。しかしながら、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はコストが高い。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は、GPS、CDR に加えて、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 [13]。しかしこのようなデータは一般的に高コストであり、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。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、50m メッシュの粒度で疑似人流を生成することを目指す。

3. 提案手法

本研究では元となる移動履歴データから、現実の人流を再現できるような疑似的な人流データ生成を目指す。入力となる移動履歴データは、端末を持つリアルユーザの緯度、経度、タイムスタンプとして得られる。疑似人流デー

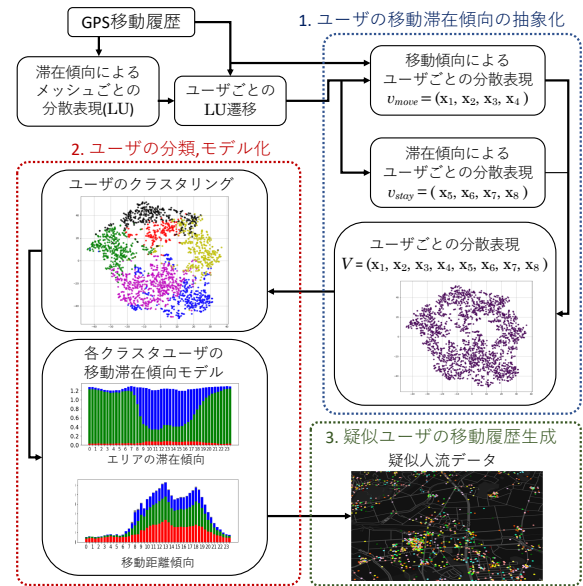


図 1 手法の流れ

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。疑似的な人流データの生成には、疑似ユーザが時間経過に伴い、それぞれどのように移動、滞在するかという活動モデルが必要である。この活動モデルについて、本研究では主として教師なし学習を用いてモデリングする。手法の概要を図 1 に示す。まず [2] の手法を用いて、滞在傾向によるメッシュごとの分散表現を生成し、ユーザの移動を LU(Location Usage) 遷移の形で表現する。これによって、ユーザの移動情報が単なる地理的な移動ではなく、POI(Point of Interest) を含む移動情報として表現できる。次に、リアルユーザの移動履歴から移動、滞在情報を抽出し、ユーザごとに分散表現の形で抽象化する。そして、ユーザごとの分散表現のクラスタリングによってユーザを移動、滞在傾向で分類する。分類された各クラスターのユーザはそれぞれ移動、滞在傾向の似ているユーザ群となる。例えば主婦のようなクラスターのユーザは一日の多くの時間を自宅で滞在し、サラリーマンのようなクラスターのユーザは朝と夕方にそれぞれ出勤、退勤を行う傾向を持つ。最後に、このように分類された各クラスターのユーザの移動、滞在傾向に基づいて疑似ユーザの活動をモデリングし、疑似人流データを生成する。これらの手法について以降で詳しく説明していく。

3.1 リアルユーザの移動、滞在傾向の抽象化

ここではリアルユーザの移動、滞在傾向による分類のために、それぞれのユーザの移動、滞在傾向を分散表現 V として抽象化する。この V の計算手順について説明していく。まず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、各スロットにおける滞在メッシュを割り当てる。次に [2] の手法を用いてメッシュごとの時間的な使わ

れ方の傾向を表した分散表現である LU を計算する。これをクラスタリングすることで、[2] に示されているように、そのメッシュの使われ方による分類が実現できる。本論文では以降この LU によるメッシュのクラスタを LU クラスタと呼ぶ。この LU クラスタは、例えば朝夕の長時間の滞在が多いクラスタは住宅街クラスタ、昼間の長時間の滞在が多いクラスタはオフィス街クラスタといったものである。そして、各ユーザの滞在傾向、移動傾向を反映した分散表現、 v_{stay} , v_{move} をそれぞれ独立に Agent2vec で計算する。Agent2Vec とは Word2Vec を応用し、ユーザの移動傾向、滞在傾向、それぞれについて学習することでユーザごとの分散表現を生成するものである。図 2 は、Agent2Vec のアーキテクチャを図示したものである。入力層は各ユーザごとの one-hot ベクトル、出力層は各移動、または滞在の特徴ごとに当てはまる次元にフラグがたった one-hot ベクトルである。これを学習することで得られた中間層の重みによって、各ユーザごとの滞在傾向が反映された分散表現が計算できる。このユーザごとの分散表現の次元数は隠れ層の次元数 N と等しく、今回は v_{stay} , v_{move} 共に次元数は $N = 4$ とした。滞在傾向としては、各滞在中、平日か休日か、滞在時間帯、滞在 LU クラスタの 3 つの情報を用いて Agent2Vec で学習する。これによって各ユーザがいつ、どんな属性のメッシュに滞在したかという滞在傾向に基づく分散表現 v_{stay} が計算できる。移動傾向としては各移動について、平日か休日か、滞在時間帯、移動距離、主要メッシュへの移動かという 4 つの情報を用いて Agent2Vec で学習する。この主要メッシュは、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、ユーザごとに得られた期間のデータ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定義する。これにより、いつ、どんな距離の、またそのユーザにとってよく訪れるメッシュへの移動かといった移動傾向による分散表現 v_{move} が計算できる。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 を作成する。

3.2 リアルユーザの分類

ここでは、ユーザごとの分散表現 V をクラスタリングすることでユーザを移動、滞在の傾向によって分類する。クラスタリングには Kmeans++ を用いる。図 3～図 5 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 3 は平日、休日それぞれの各時間帯に滞在した LU クラスタの割合を示しており、図 4、図 5 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 4、図 5 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

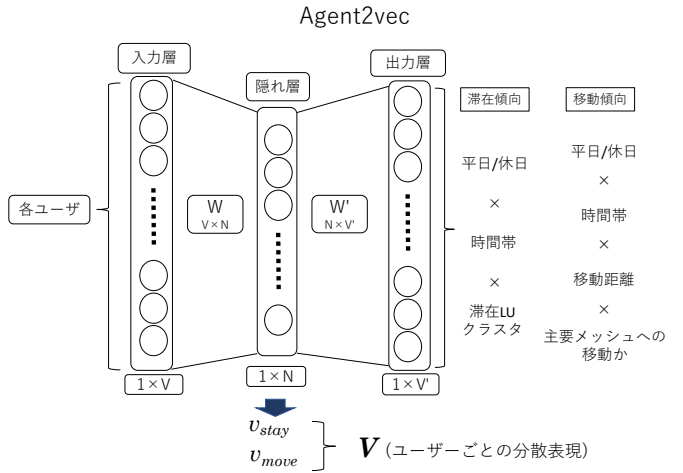


図 2 Agent2Vec による分散表現作成

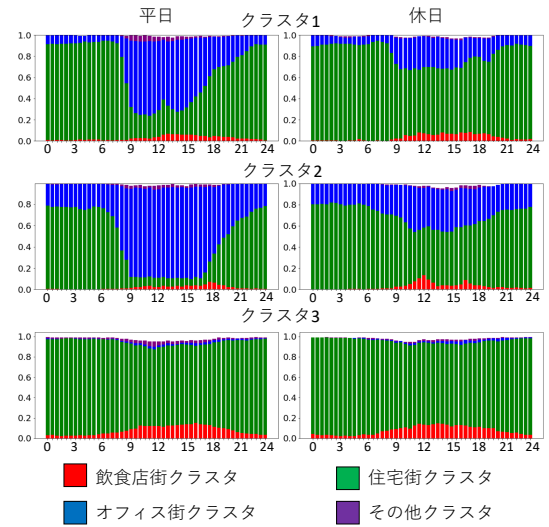


図 3 滞在 LU クラスタ

化している。図 3 より、クラスタ 1, 2 は夜の時間帯に住宅街クラスタ、昼の時間帯にオフィス街クラスタ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 3 は昼間に飲食店街クラスタに滞在することがある以外は住宅街クラスタに滞在しており、主婦のような滞在傾向を示している。図 4 を見ると、クラスタ 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 2 は朝の 7 時～9 時と夕方 17 時～20 時に移動量のピークがあり、3km～5km の移動が目立つ。クラスタ 3 はクラスタ 1, 2 と比較すると移動量が少なく、移動距離も 0.5km 以下の割合が高い。また図 5 から、クラスタ 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 3 はクラスタ 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出来る。

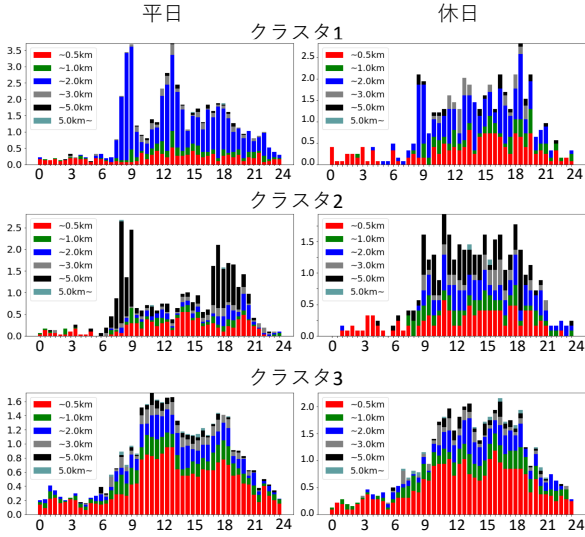


図 4 移動距離の傾向

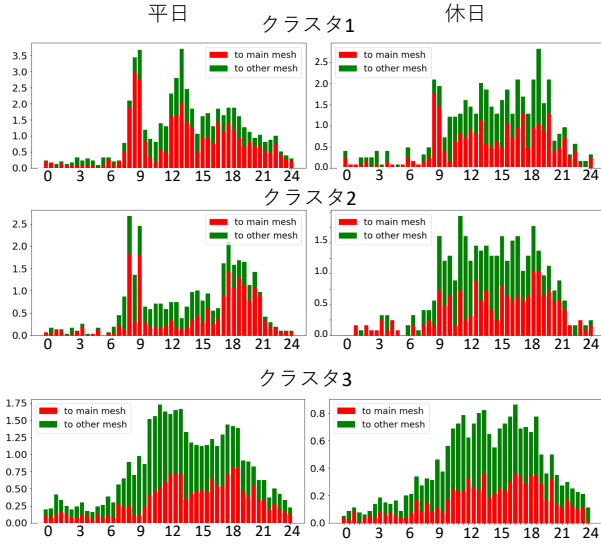


図 5 主要メッシュへの移動傾向

3.3 疑似ユーザの生成

ここでは、分類したそれぞれのユーザ群から疑似ユーザの活動をモデリングし、どのように各疑似ユーザの移動履歴を生成するか説明する。まず最初に、疑似ユーザが分類したどのユーザクラスタをモデルとして利用するか決定する。ある疑似ユーザがユーザクラスタ u をモデルとして利用する確率 $P(u)$ は、ユーザクラスタ u に属するユーザの人数 N_u を用いて、

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

で計算できる (U はユーザクラスタ数)。すなわち、属するユーザの多い典型的なクラスタは疑似ユーザのモデルとして選ばれやすい。

次に、選択したユーザモデルに従って疑似ユーザの移動履歴を生成していく。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、リアルユーザの平均移動回数 λ 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。この平均移動回数は各タイムスロット、移動前 LU 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 LU 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疑似ユーザが朝の 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、時間帯ごとの滞在密度、LU 類似度、移動距離の 3 つを考慮しており、順に説明していく。まず滞在密度について、時間帯ごとのリアルユーザ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 t における、滞在密度 D_t は、メッシュ m 内のユーザ数 $C(m)$ を用いて、

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

で計算する (M はメッシュ数)。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。次に LU 類似度について、ユーザの POI に沿った移動先を生起するために、移動先 LU クラスタと類似度が高い LU を持つメッシュへの移動確率を高く設定する。移動先 LU クラスタは属するユーザクラスタの時間帯ごとの滞在 LU クラスタの割合 (3) に従って決定する。ここでは計算量の削減のために、LU クラスタごとに代表として LU の平均を取り、各 LU クラスタの平均 LU との類似度を元に移動確率を計算している。最後に移動距離については、リアルユーザの移動距離の傾向 (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。具体的には、平均距離と分散を元に正規分布に従った移動確率を各メッシュに割り当てている。

このように滞在密度、LU 類似度、移動距離のそれぞれについて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、疑似ユーザは各移動ごとに、滞在密度、POI、移動元メッシュとの距離から移動先のメッシュを決定していく。各ユーザモデルから疑似ユーザの移動を確率的に生成していくことで、任意のユーザ数分の擬似人流を生成できる。

4. 実験と評価

提案手法の評価のために、プログウォッチャー社提供の GPS 移動履歴データセットを用いて実際に疑似人流データを生成した。対象地域は愛知県日進市、期間は 2020 年 3 月のデータとし、対象地域で十分に位置情報が得られた 2155 人のユーザのデータを用いた。疑似ユーザ数は 30000, LU クラスタ数 4, ユーザクラスタ数 40 とした。以下では実際に生成したデータを滞在密度, 移動量, 移動距離といった観点から評価していく。評価手法については, [13] 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モバイル空間統計の各時間帯における人口分布データセットとも比較した。

4.1 滞在密度

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 6 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), プログウォッチャー (500m メッシュ), プログウォッチャー (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。双方のデータにおける滞在密度の正の相関が高ければ, 生成データは現実の人口分布を再現できている可能性が高いと考えられる。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。プログウォッチャー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 7 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。プログウォッチャーデータとは強い相関があり, 特に 50m メッシュの粒度でも 0.75~0.8 と強い相関があることから, 高粒度な人口分布を再現できている。しかしながら, モバイル空間統計との相関は全時間帯で低い。これは, 生成元のプログウォッチャー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原因として挙げられる。また, 図 8 はモバイル空間統計データとの滞在密度差を地理的に描画したものであり, 滞在密度のズレが大きいメッシュを濃く色付けしてある。図のプロットは対象エリアの主要駅である赤池駅と日進駅であり, 滞在密度差が大きいメッシュは駅周辺及び対象エリアの周縁部に位置していることが分かる。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

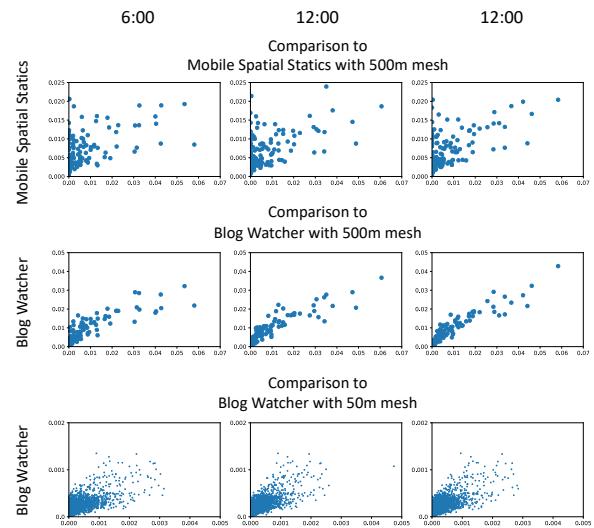


図 6 滞在密度の比較

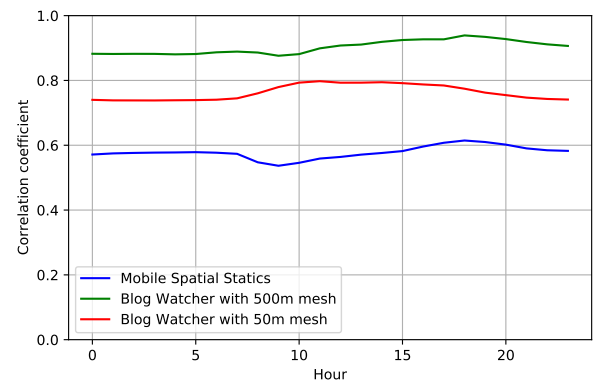


図 7 時間帯ごとの相関の比較

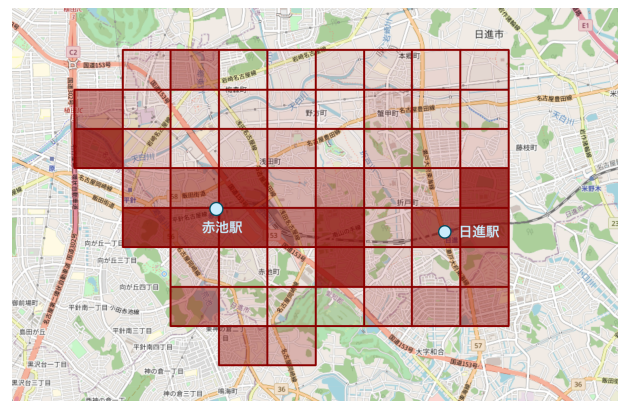


図 8 モバイル空間統計との RMSE が高いメッシュの分布

ザは省かれている。これがエリアの駅周辺や周縁部において, 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, 図の赤いメッシュはユーザのエリアの出入りが多い地域であると思われる。

4.2 移動量

図 9 は疑似データと生成元データの各時間帯のユーザの移動量の比較であり, 横軸は 1 時間ごとの時刻, 縦軸は移

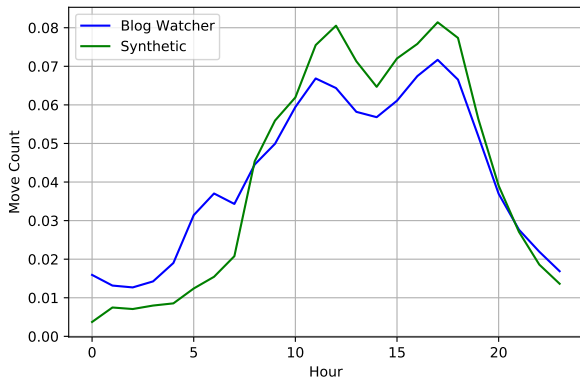


図 9 移動量の比較

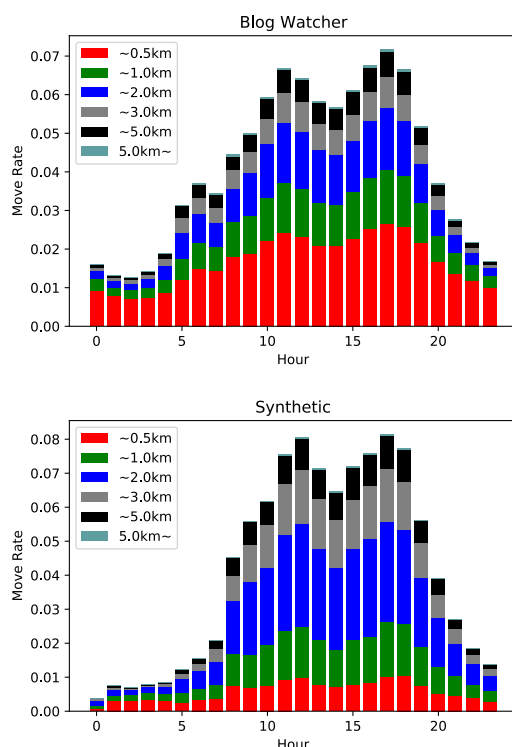


図 10 移動距離の比較

動量の割合を示している。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。図 10 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。これは移動先メッシュの選択時に、滞在密度、LU 類似度を重視し過ぎているためと考えられる。また、移動距離はユーザの自宅の位置や、自宅と職場の位置関係などの地理的な位置に依るところがあるため、疑似ユーザの活動モデリングにおいてこれらが十分にモデリングされていないことが原因と思われる。

5. まとめ

本論文では、大規模 GPS 移動履歴データを活用して、都市環境における擬似的な人流データを生成する手法を提案した。特に、Agent2Vec による分散表現作成、Kmeans++ によるクラスタリングを用いた、教師なし学習によるユーザの活動モデリングを実現した。これにより、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。この疑似人流データは、現実の滞在密度を従来より詳細な粒度で再現しており、また LU 遷移によって疑似ユーザの移動をモデリングしている。従って、現実のユーザの POI に沿った移動を生起し、よりリアルな人流を再現出来る。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距離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。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。これに加えて、各疑似ユーザの移動経路や移動速度、移動手段なども都市人流の再現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。

6. 謝辞

本研究の一部は、AMED 委託研究 (20he0622042h0001)、JST CREST (JPMJCR1882)、NICT 委託研究「アフターコロナ時代の超多様性社会を実現するダイバーシティ・ドリブン・トランスフォーメーション技術の研究開発」の支援を受けています。

参考文献

- [1] T. Yabe, K. Tsubouchi, N. Fujiwara, T. Wada, Y. Sekimoto, S. Ukkusuri. Non-compulsory measures sufficiently reduced human mobility in Tokyo during the COVID-19 epidemic. Scientific Reports volume 10, Article number: 18053, 2020.
- [2] 庄子 和之, 廣井 慧, 米澤 拓郎, 酒田 理人, 河口 信夫. 滞在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. 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.
- [3] C.R. Bhat, S.K. Frank. 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.
- [4] J.L. Bowman, E.B. Moshe. Activity-based disaggregate travel demand model system with activity schedules. Transport. Res. Part A: Policy Pract., 35 (1) (2001), pp. 1-28, 2001.
- [5] X. Song, K. Hiroshi, S. Ryosuke. DeepTransport: Prediction and simulation of human mobility and transportation mode at a citywide level. IJCAI (2016), p. 16, 2016.
- [6] M. Yin, M. Sheehan, S. Feygin, J.F. Paiement, A. Pozdnoukhov. A generative model of urban activities from

- cellular data. *IEEE Trans. Intell. Transp. Syst.*, 19 (6) pp. 1682-1696, 2017.
- [7] K. Ouyang, R. Shokri, D.S. Rosenblum, W. Yang. A Non-Parametric Generative Model for Human Trajectories. *IJCAI(2018)*, pp. 3812-3817, 2018.
 - [8] S.S.Borysov, J. Rich, F.C.Pereira. 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.
 - [9] J. Drchal, M. Čertický, M. Jakob. Data-driven activity scheduler for agent-based mobility models. *Transportation research part C: emerging technologies*, vol. 98, pp.370-390, 2019.
 - [10] P.D. Vecchio, G. Secundo, Y. Maruccia, G. Passiante. 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.
 - [11] Y. Pang, K. Tsubouchi, T. Yabe, Y. Sekimoto. Development of people mass movement simulation framework based on reinforcement learning. *Transportation research part C: emerging technologies*, vol. 117, 2020.
 - [12] Y. Pang, K. Tsubouchi, T. Yabe, Y. Sekimoto. 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.
 - [13] T. Kashiya, Y. Pang, Y. Sekimoto. Open PFLOW: Creation and evaluation of an open dataset for typical people mass movement in urban areas. *Elsevier*, Vol. 85, pp.249-267, 2017.