

DiverCityMeter: 大規模移動データによる生活パターン分析を通じた都市空間の多様性算出手法

庄子 和之^{1,a)} 寺島 青¹ 田村 直樹¹ 片山 晋¹ 浦野 健太¹ 米澤 拓郎¹ 河口 信夫^{1,2}

概要: 都市空間において、あるエリアの訪問者の多様性が高いとは、そこが様々な人にとって魅力的な場所であることを指す。つまり、エリアの訪問者多様性の把握は、地域活性化を促す施策の判断基準になり得るため、住民の生活の質の向上に不可欠である。従来、都市の多様性とは、性別や年代、職業といった住民の属性情報を活用し、測られてきた。しかし、同じ属性を持つ人々でも、生活パターンの観点から見ると、その特徴は多岐にわたる。つまり、生活パターン多様性の把握も同様に、都市計画やマーケティングの分野において、重要な役割を果たすはずである。本論文では、GPS 位置データに基づく大規模な移動データのみを用いて、生活パターンの観点から人々をモデル化し、種類の豊富さと均一さからエリアごとの生活パターン多様性を算出する。そして、生活パターン多様性を算出するフレームワーク DiverCityMeter を提案する。DiverCityMeter は、エリアモデリング・行動モデリング・人モデリングの3つのモジュールから構成される。エリアモデリングでは、都市における各エリアの「使われ方」が埋め込まれたベクトル表現を生成し、これにより人々の移動行動をエリアの意味的な系列として表現する。次に、行動モデリングを通じて、各日の移動行動を「いつ、どのエリアに滞在したか」という系列特徴が埋め込まれたベクトル表現に変換する。最後に、行動ベクトルのベクトル演算を通して、個々人を「生活パターン」特徴が埋め込まれたベクトル表現で変換し、クラスタリングを通して生活パターンごとの人口分布からエリアごとの多様性を算出する。つまり、DiverCityMeter は、人の属性データを使う手法とは異なり、移動行動由来の人の性質に基づき多様性を算出する新しい手法である。我々は、DiverCityMeter を数万単位のス마트フォンユーザから収集された GPS 位置データから成る実世界データセットを用いて評価した。また、COVID-19 の影響により、エリアごとの多様性がどのように変化したかについても分析した。

DiverCityMeter: Calculating Resident Diversity through Life-Pattern Analysis using Large Mobility Dataset

1. はじめに

どこにどんな人がどれだけ来るかを把握することは、そのエリアの人の多様性を測ることであり、都市における住民の生活の質の向上に関わる重要な指標だと考えられる。例えば、多様性が高い場合、公共施設やサービスへのアクセスが容易なエリアである可能性がある。つまり、多様性情報は、都市をどのようにデザインするかの判断基準となりうる。従来の研究では、都市住民の性別 [1] や職業 [2]、

収入 [3], [4] などの属性を基に、都市の多様性を理解しようとしてきた。しかし、属性が同じ人々でも、生活パターンの観点から見ると、その特徴は多岐にわたるはずである。例えば、会社員の中でも「定時出社 or フレックスタイム制」がおり、高所得者の中でも「インドア or アウトドア」が存在する。つまり、人の生活パターンもまた、日々の移動や活動の選択、消費行動などの複合的な要素によって形成されるため、エリアごとに異なる多様性を示す。そのため、人々の生活パターンの多様性もまた、都市にとって経済的・社会的に重要な要素と見なせるはずである。

そこで本論文では、都市空間におけるエリアの多様性を、人々の生活パターン特徴の観点から算出する手法 DiverCityMeter を提案する。DiverCityMeter は、大規模かつ匿名化された GPS 位置データのみを活用し、人々の生

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

^{a)} shoji@ucl.nuee.nagoya-u.ac.jp

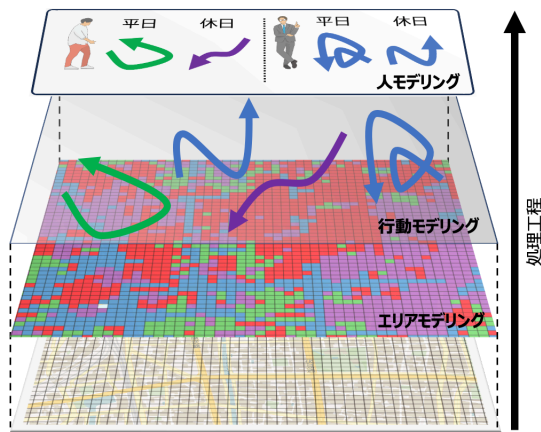


図 1: DiverCityMeter の全体像

活パターンの種類の豊富さと均一さから、エリアの多様性を算出する。DiverCityMeter の全体像を図 1 に示す。DiverCityMeter は、エリアモデリング・行動モデリング・人モデリングの 3 つのモジュールから構成される。エリアモデリングでは、滞在という人々のエリア内での活動を活用し、エリアを「使われ方」で特徴付け、低次元のベクトル表現 UAS (Usage of Area with Stay information) に変換する。我々は、このための表現学習モデル Area2Vec [5] を開発した。Area2Vec により、地理座標の系列で表現される移動は、エリアの意味が埋め込まれた UAS の系列に変換できる。そこで行動モデリングでは、どのような順序でどのような特性を持つエリアへ移動したかという「遷移パターン特徴」で UAS 系列を特徴付け、低次元のベクトル表現 PSU (Pattern of Sequence with UAS) に変換する。我々は、このための表現学習モデル Behav2Vec (Behavior to Vector) を開発した。人モデリングでは、どのようなスケジュールでどのような行動を取ったかという「生活パターン特徴」で個人々人をモデル化する。このモデル化は、行動モデリング結果のベクトルの連結や中心の算出といった単純なベクトル演算で達成される。最後に、人モデリング結果をクラスタリングして得られる人々の生活パターングループを使って、グループごとの滞在時間を集計することで、都市のエリアごとの多様性を算出する。つまり、DiverCityMeter とは、人の移動分析にはまず空間分析が必要という考えの下、GPS 位置データのみから個人々人の生活パターンをモデル化し、移動行動由来の人の性質でエリアの多様性を算出する新しい手法である。

我々は、DiverCityMeter を、数万単位のスマートフォンユーザから収集された GPS 位置データから構成される実世界データセットを用いて評価した。その結果、生活パターンに応じた人の分類が GPS 位置データのみを用いて可能なことと、都市のエリアごとの生活パターン多様性を算出できることを示した。また、COVID-19 前と COVID-19 中で、多くのエリアで生活パターン多様性が低くなったことを発見した。

2. 関連研究

2.1 人の移動データ分析

従来、人の移動分析では、地理空間上での軌跡の類似度を算出してきた [6], [7] が、これらは地理的な制約を受けるため、物理的に離れた移動同士の比較が不可能という問題があった。そこで、意味情報が付与された移動データを活用した研究が盛んに行われている。Wan ら [8] は、POI (Point of Interest) チェックイン情報を使い、意味的な遷移パターンをマイニングする手法を提案した。Cao ら [9] は、POI をモデル化する手法 Habit2Vec の開発を通して、空間情報を排除した形で人々の POI 間の移動行動をモデル化するスキームを提案した。しかし、意味情報が付与されたデータの準備には、訪問 POI や訪問目的の推定が必要であり、時間とコストがかかるため、大規模なデータセットの作成が困難という課題がある。

長期間の移動データを使用し、生活パターンに応じた人々のクラスタリングを目的とする研究ももちろん存在する [10], [11], [12]。彼らは、個人々人の移動履歴から頻繁に訪れる場所や長時間滞在する地点を「重要な場所」と定め、重要な場所間の移動から個人々人の生活パターンをモデル化する手法を提案した。しかし、これらの手法は、滞在時間に応じたラベルが「重要な場所」に付与されるだけで、その場所がどんな意味を持つか (住宅、オフィス等) は把握できないという課題がある。

そこで本論文では、GPS 位置データのみを用い、滞在エリアがどんな意味 (住宅、オフィス、商業など) を持つかを明らかにしつつ、生活パターン特徴に基づき人々をモデル化する手法を開発する。

2.2 都市の多様性分析

都市空間における多様性について、その重要性をデータ駆動的に分析する研究が盛んに行われている。Yabe ら [4] は、COVID-19 中の行動変化が都市での出会いの所得多様性にどのような影響を与えたか、大規模な移動データを用いて分析した。その結果、都市での出会いの多様性は、パンデミック中に大幅に減少し、行動変化や新しい場所への探索意欲の低下がこの多様性の減少に寄与していることを明らかにした。Yuji ら [13] は、Jane Jacob の都市多様性の概要を定量化し、店舗タイプ (飲食店、ホテル等) の多様性が周辺の経済利益に貢献するかどうかを、クレジッドカードの購買履歴と組み合わせることで検証した。その結果、大都市において、多様性が高いほど店舗の売上高が高い傾向にあることを明らかにした。

3. DiverCityMeter

DiverCityMeter の処理工程を図 2 に示す。DiverCityMeter は、GPS 位置データのみを用い、生活パター

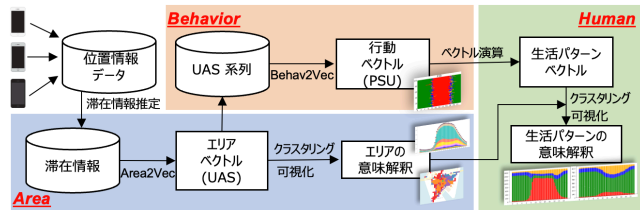


図 2: DiverCityMeter の処理工程

表 1: エリアモデリングに用いる滞在時間情報

曜日	平日, 休日 (祝日を含む)
到着時刻	0:00-1:59, 2:00-3:59, ..., 20:00-21:59, 22:00-23:59
滞在時間	-29, 30-59, 60-119, 120-239, 240-359, 360-719, 720-1079, 1080-1439, 1440- (unit:min)

ン特徴に基づいた人々のモデル化を通して、都市におけるエリアごとの生活パターン多様性を算出するためのフレームワークである。

3.1 エリアモデリング

本節では、GPS 位置データのみを用いて、エリアのベクトル表現を作成する手法 Area2Vec [5] について説明する。

Area2Vec のアーキテクチャを図 3 に示す。Area2Vec は、滞在という人々の活動から明らかになる「使われ方」が埋め込まれたエリアのベクトル表現 UAS (Usage of Area with Stay information) を作成する。学習フェーズでは、エリア情報とそこへの滞在時間情報のペアが学習データセットとして用意される。エリア情報が入力されたとき、そこで発生した滞在時間情報を予測し、その予測精度を上げるように学習が進む。エリア a が与えられたとき、滞在時間情報 s を予測する条件付き確率モデルは次のように表される： $P(s|a) = \frac{\exp(\mathbf{v}_s \cdot \mathbf{u}_a)}{\sum_{s' \in S} \exp(\mathbf{v}_{s'} \cdot \mathbf{u}_a)}$ 。 \mathbf{v}_s は滞在時間情報のベクトル（出力層側の重み）、 \mathbf{u}_a はエリア a のベクトル（中間層）、 S は全ての滞在時間情報の組み合わせを意味する。本論文では、表 1 に示す滞在時間情報を用いるため、曜日、到着時刻、滞在時間の全組み合わせは、 $S = 2 \times 10 \times 24 = 480$ となる。損失関数としては、負の対数尤度を取り、次のように定義される： $L = -\frac{1}{|D|} \sum_{(a,s) \in D} \log P(s|a)$ 。ただし、 D はトレーニングデータの集合を表す。Area2Vec の学習では、 L を最小化するように、モデル全体のパラメータが最適化アルゴリズム Adam によって調整される。最終的に、入力層と隠れ層の間の重み W の各行が各エリアの UAS として扱われる。

3.2 行動モデリング

エリアモデリングの結果、地理座標の系列で表現された人々の移動行動は、UAS 系列に変換可能となる。本節では、UAS 系列のベクトル表現 PSU (Pattern of Sequence with UAS) を作成する手法 Behav2Vec について説明する。

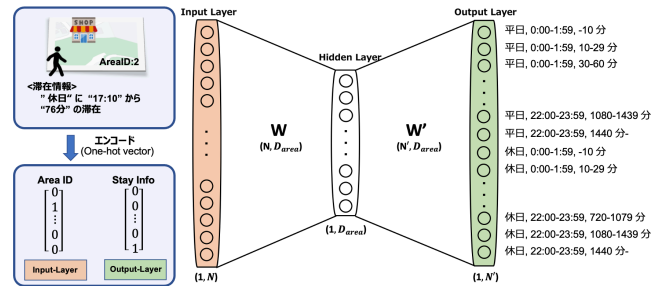


図 3: Area2Vec [5] のアーキテクチャ

Behav2Vec のアーキテクチャは、図 4 に示すように LSTM (Long Short Term Memory) の Autoencoder というシンプルな作りである。初めに、エンコーダ部に UAS 系列 $\{UAS_1^{100}, UAS_2^{54}, \dots, UAS_T^{22}\}$ (ただし、 UAS_t^i とは、時刻 t での滞在エリア ID が i であることを示す) が入力される。本研究では、1 日を 30 分区切りの系列としたため $T = 48$ となる。このとき、時刻 t における隠れ状態 h_t とセル状態 c_t は、次の式のように現在の入力 UAS_t と、前の h_{t-1} と c_{t-1} により更新される： $h_t, c_t = f_{LSTM}(h_{t-1}, c_{t-1}, UAS_t)$ 。そして、最後の UAS_T が処理された後に出力される h_T が入力系列全体の潜在表現として利用される。

デコーダの目的は、入力系列の潜在表現を受け取り、元の入力系列を再構築することである。初めに隠れ状態とセル状態は $h_1^{dec} = h_T$ と $c_1^{dec} = 0$ で初期化される。時刻 t における隠れ状態 h_t^{dec} とセル状態 c_t^{dec} は、次の式のように前の状態である h_{t-1}^{dec} と c_{t-1}^{dec} 、それからゼロベクトルにより更新される： $h_t^{dec}, c_t^{dec} = f_{LSTM}(h_{t-1}^{dec}, c_{t-1}^{dec}, 0)$ 。そして、 h_t^{dec} が全結合層に渡され \hat{UAS}_t が生成される。ゼロベクトルをデコーダへの入力とする理由は、LSTM Autoencoder の無条件デコーダ [14] を参考にしたためである。つまりこれは、デコーダが前の時刻の情報を入力として使用しないことを意味する。このテクニックは、特定の詳細や変化に対して敏感になることを抑える効果があり、つまり、エンコーダの出力が全体的なパターンや一般的な特徴を捉えた表現になることが期待できる。Behav2Vec の損失関数は、元の入力データと再構築されたデータとの間の誤差を最小にするように定義され、次のように平均二乗誤差を用いる： $L = \frac{1}{T} \sum_{t=1}^T \|UAS_t - \hat{UAS}_t\|^2$ 。Behav2Vec の学習では、損失関数 L を最小化するように、エンコーダとデコーダのパラメータが、最適化アルゴリズム Adam によって調整される。学習終了後に使用されるのはエンコーダ部のみであり、モデル化対象の行動 (UAS 系列) を入力したとき、エンコーダの出力である潜在表現が、その行動の PSU となる。

3.3 人モデリング

本節では、どのように個々人をモデル化するかを説明する。本論文において、個々人は、生活パターンに基づいて

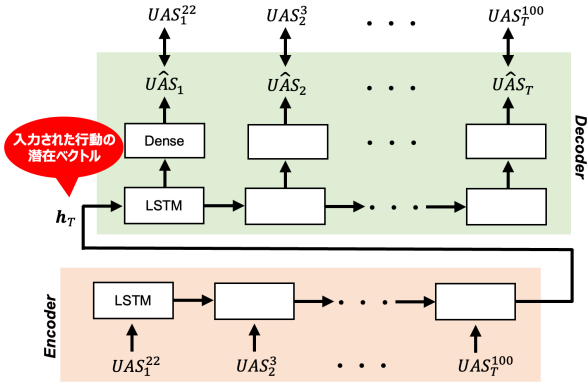


図 4: Behav2Vec のアーキテクチャ

モデル化される。そして、生活パターンとは、どのようなスケジュールでどのような行動を取ったか、と定義する。

生活パターンに応じた人の表現は、行動モデリング結果を活用することで、次の数式で表されるシンプルなベクトル演算で作成される。

$$\begin{aligned} \mathbf{v}_u &= \text{concat}(PSU_{u, \text{Weekday}}, PSU_{u, \text{Weekend}}) \quad (1) \\ &= \text{concat}\left(\frac{1}{|M|} \sum_{m \in M} PSU_{u, m}, \frac{1}{|N|} \sum_{n \in N} PSU_{u, n}\right) \end{aligned}$$

$PSU_{u, x}$ は、ある人 u が持つ行動 x に対応する行動ベクトルを表す。 M と N はそれぞれ、平日と休日に人 u が発生させた行動の集合を表す。つまり、平日の行動ベクトルの平均と、休日の行動ベクトルの平均を連結したものが、その人の表現ベクトルとなる。ただしこれは一例であり、行動ベクトルを活用することで、柔軟に人ベクトルを構築できる。例えば「月曜日から日曜日の各曜日に対応する行動ベクトル」を連結 $\mathbf{v}_u = \text{concat}(PSU_{u, \text{Monday}}, PSU_{u, \text{Tuesday}}, \dots, \mathbf{h}_{u, \text{Sunday}})$ すれば、一週間単位の生活パターンで人のモデル化が可能である。また「平日は行動ベクトルの平均」と「休日は土・日・祝日の行動ベクトル」を連結 $\mathbf{v}_u = \text{concat}(PSU_{u, \text{Weekday}}, PSU_{u, \text{Saturday}}, PSU_{u, \text{Sunday}}, PSU_{u, \text{Holiday}})$ することで、平日の大まかな生活パターンを考慮しつつも、土・日・祝日の生活パターンの違いを考慮した、人のモデル化が可能である。

3.4 多様性の算出

本節では、生活パターン特徴に応じたエリアごとの多様性算出手法について説明する。

まず初めに、人モデリング結果をクラスタリングし、人々を生活パターン特徴に応じて C 個のグループに分割する。本論文では、 $C = 10$ に設定した。次に、エリア $a \in A$ (A は都市をメッシュ化した際の全グリッドの集合) における滞在者のデータを収集し、以下の数式を適用することで、エリアごとの多様性を算出する。

$$D_a = 1 - \frac{5}{9} \sum_c \left| \tau_{a, c} - \frac{1}{10} \right| \quad (2)$$

これは、Moro らが提案した場所ごとの所得分離を算出す

る手法 [3] をベースにしたものである。以降はそれぞれの変数と数値が何を意味するかを説明する。 D_a は、場所 a における生活パターン多様性指標を表す。 D_a が 1 の場合、各生活パターングループが完全に均等な分布を示し、0 の場合、1 つのグループだけが存在する完全に偏った状態を示す。 $\tau_{a, c}$ は、生活パターングループ $c \in C$ に属する人々が場所 a で過ごした滞在時間の割合を表す。つまり、 $\tau_{a, c}$ はエリア a における生活パターングループ c に属する人々の滞在総時間を、そのエリアへの全滞在者の滞在総時間で割ったものである。例えば、エリア a での滞在総時間の内、生活パターングループ c に属する人々が 30% の時間を占める場合、 $\tau_{a, c} = 0.3$ となる。 $\frac{1}{10}$ は、 $C = 10$ であるため、理想的な均等分布の割合を意味する。すなわち、各生活パターングループの人々が全体の時間の 10% ずつを占める場合を示す。この値は、すべての生活パターングループが均等に分布していると仮定した場合の理想的な状態である。 $|\tau_{a, c} - \frac{1}{10}|$ は、生活パターングループ c の人々のエリア a での滞在時間の割合と、理想的な均等分布の割合との差の絶対値を取ることを意味する。これにより、実際の分布と理想的な分布との偏りの大きさを測定する。 \sum_c は、すべての生活パターングループについて上記の偏りの大きさを合計することを意味する。各生活パターングループの偏りを合計することで、そのエリア全体の偏りの大きさを示す。 $\frac{5}{9}$ は、スケーリング係数である。この係数により、 D_a が 0 から 1 の範囲に収まるように調整される。この数式は、エリア a における各生活パターングループの滞在時間の割合をもとに、その場所がどれだけ異なる生活パターンの人々によって均等に利用されているかの指標を提供する。

4. 実験

4.1 エリアモデリング

本節では、Area2Vec が都市におけるエリアをその使われ方が埋め込まれた適切な表現 UAS に変換できることを示す。

我々は、GPS 位置データに滞在推定アルゴリズム [15] を適用し、Area2Vec の学習に用いる滞在情報のデータセットを作成した。本論文では、名古屋市を 50 m 四方のメッシュに変換した際の各グリッドをエリアと定め、滞在が 10 回以上観測されたエリアをモデル化対象と定めた。その結果、2019 年と 2020 年のそれぞれで、181,650 エリアと 170,540 エリアがモデル化対象となった。本論文では、[16] で述べられている式に基づき、USA のベクトルサイズを 5 次元に設定した。

図 5a は、2019 年の名古屋市の全 UAS を 4 つのグループにクラスタリングし、各グループの特徴を可視化したものである。クラスタリングには、k-means++ [17] を使用した。この積み上げ棒グラフは、各グループに含まれるエリア内での滞在時間長別の人数を表している。横軸は時間

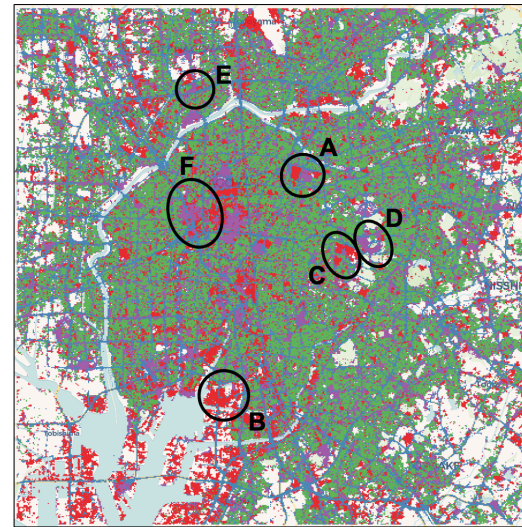
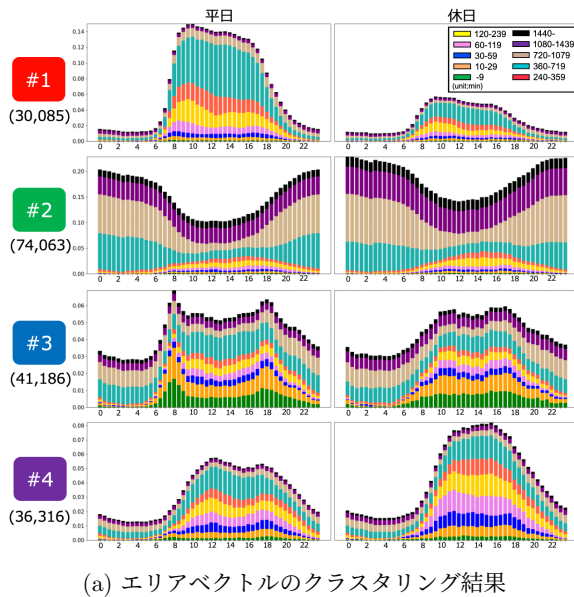


図 5: エリアモデリング結果の可視化

を表して、ピンは 30 分である。縦軸は人数を表している。各層の色が滞在時間長を表している。左カラムが平日、右カラムが休日についてである。グループ番号の下括弧の中には、そのグループに含まれるエリア数を記した。縦軸に関しては、各グループに割り当てられたエリアの数に違いがあるため、エリアの数で割ることで正規化している。また、平日と休日は日数が違うため、それぞれの収集期間内の日数で割ることで正規化している。つまり、このグラフは 1 エリアあたり、かつ 1 日あたりの滞在時間長別の人数を表している。この処理により、グループ間でのグラフの比較と、平日と休日間でのグラフの比較が可能になる。注意として、長時間の滞在者は複数のピンでカウントされている。例えば、午前 10 から午後 12 の滞在进行した人は、ピンが 30 分であるため、20 個目のピンから 24 個目のピンの全てに同一人物がカウントされている。このグラフを見ることで、各エリアがどんな特性を持つか解釈可能になる。

我々は、エリアを 4 つのグループに分類した。グループ #1～#4 の解釈とその理由は次の通りである。

#1 : 「オフィスエリア」

- 8 時くらいからの長時間の滞在。
- 夜中と休日の人の少なさ。

#2 : 「住宅エリア」

- 平日・休日に夜から朝方にかけての長時間の滞在。
- 日中の人の減少。減少幅は休日より平日の方が大きい。これは仕事に行く人の影響だと考えられる。

#3 : 道路や駅といった移動に関係する「交通エリア」

- 短時間の滞在割合が多い。
- 8 時と 18 時あたりの通勤時間帯に発生する短時間の滞在。長時間の滞在も発生しているが、これはグ

ループ間の境界が曖昧で「住宅エリア」の滞在もここにカウントされたためだと考えられる。

#4 : 飲食店・ショッピング・エンターテインメントエリアが含まれる「商業エリア」

- 休日の方が人が多い。
- 日中の短～中時間の滞在割合が多い。

続いて、このクラスタリング結果をマップに反映したものを図 5b に示す。エリアの色は、図 5a のクラスタリング結果と対応している。この地図から、どのあたりが中心地で、どこに住宅街が広がっているかなど、都市全体の様子を把握できる。A には、企業の工場とイオンモールが道路を隔てて隣接しており、それらがそれぞれ #1 と #4 に割り当てられている。C には、名古屋大学がある。こちらも #1 に割り当てられており、使われ方としては大学キャンパスもオフィスと同じ部類に入ることが分かる。D と E にはそれぞれ、大規模な動植物園と mozo ワンダーシティがある。これらは、#4 に割り当てられており、我々の直感にマッチした結果となっている。最後に、F には名古屋市で最も大きな名古屋駅があり、#1、#3、#4 が混在する結果となった。この周辺は駅だけでなく、オフィスビルやショッピング施設が存在するため、納得がいく結果である。

以上より、Area2Vec を使用することで、「使われ方」が似たエリア同士がベクトル空間上で近くに配置されるように UAS が学習されていることが示された。

4.2 行動モデリング

本節では、Behav2Vec が UAS 系列を移動パターン特徴が埋め込まれた表現 PSU に変換できることを示す。

本論文では、Behav2Vec 学習のためのデータセットとして、0 時から 24 時間を 30 分区分切りにした長さが 48

の UAS 系列を用いる。我々は、滞在情報が無い時間スロットは欠損として扱い、その時間帯は前後の滞在情報で補間した。それに伴って、本論文では 70%以上の時間 (24 時間 \times 0.7 = 17 時間) がどこに滞在したか推定できる UAS 系列のみを使用した。その結果、モデル化対象となる UAS 系列は、2019 年と 2020 年でそれぞれ、5,156,745 個と 4,573,142 個抽出された。ただし、学習データセットについては、多様な系列パターンを網羅するために「0 時から」だけではなく「10 時から」「22 時から」など様々な開始時間からの 24 時間の UAS 系列を作成した。また、UAS 系列は、曜日を区別していないため、平日と休日の UAS 系列が混在していることに注意されたい。これにより、様々な入力 UAS 系列に対応したモデルになると期待できる。その結果、学習データセットとしては、2019 年と 2020 年で合わせて 22,989,216 個の UAS 系列が作成された。PSU のベクトルサイズは 50 次元に設定した。

図 6 は、2019 年の PSU をクラスタリングし、グループごとの特徴を可視化したものである。このグラフは、1 行に 1 UAS 系列として積み上げたものである。ただし、各グループの中心からユークリッド距離が近い上位 100 個の UAS 系列だけを使って、グラフを作成している。横軸は時間を表しており、ピンは 30 分である。縦軸は系列数を表している (どのグループも最大値は 100)。各色は、図 5a の UAS クラスタリング結果に対応している。200 個のグループに分類したが、本論文ではその内の特徴的な 6 グループを示す。クラスタリング手法は、エリアモデリングの時と同じ理由で、k-means++を使用する。

グループ #1~#6 の解釈は次の通りである。#1 と #2 に割り当てられた UAS 系列はそれぞれ、住宅エリアと交通エリアに終日留まる行動であると読み取れる。#3 は、基本的には住宅エリアで過ごす但、途中で 1~2 時間商業エリアで過ごす行動であると読み取れる。つまり、主婦の買い物を伴う日常、または、働いている人の外出を伴う休日で発生するものと予想できる。#4 は、8 時から 18 時にオフィスエリアで過ごし、朝方と夜中は住宅エリアで過ごしていることが読み取れる。また、オフィスに行く前に交通エリアに滞在することも読み取れる。つまり、駅を使ってオフィスに行く会社員から生成されたものと予想できる。#5 は、#4 と同様に会社員から生成された行動だと分かるが、住宅エリアに移動する前に商業エリアに滞在することが読み取れる。つまり、会社員の中でも仕事帰りに外食や買い物をしてから帰宅する人々から生成されたものと予想できる。#6 は、10 時から 21 時に商業エリアで過ごし、朝方と夜中は住宅エリアで過ごしていることが読み取れる。会社員に比べると働く時間帯が遅いことから、つまり、店舗や飲食店などで働く人々から生成されたものと予想できる。

以上より、Behav2Vec を使用することで、似た特徴を

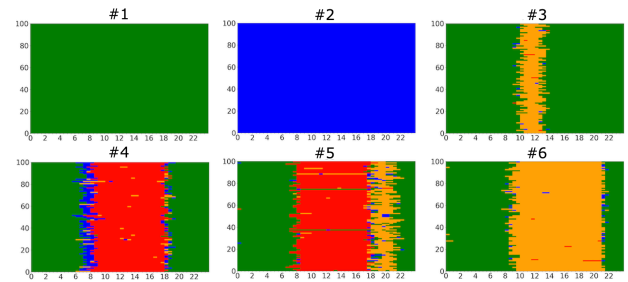


図 6: 移動行動ベクトルをクラスタリングした際の特徴的なクラスタの可視化

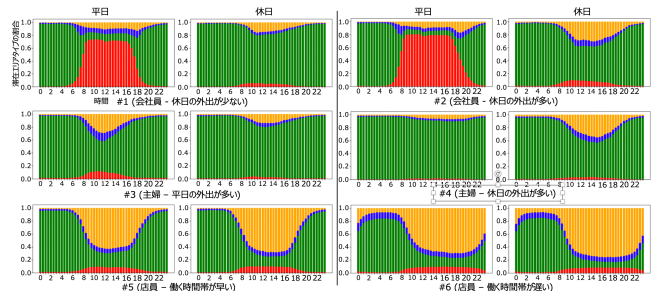


図 7: 人ベクトルをクラスタリングした際の特徴的なクラスタの可視化

持つ行動同士がベクトル空間上で近くに配置されるように PSU が学習されていることが示された。

4.3 人モデリング

本節では、Behav2Vec から生成された行動ベクトル PSU のベクトル演算で、生活パターン特徴が埋め込まれた人ベクトルを作成可能なことを示す。

本論文では、平日と休日という変数のみを使用し、個人を生活パターンでモデル化する。つまり、PSU は 50 次元であるため、式 (1) に基づき、個人は 100 次元のベクトルで表現される。本論文では、行動モデリングで設けた有効行動 (24 時間中 70%以上どこに滞在したか推定できる UAS 系列) を、平日と休日でそれぞれ 10 日と 5 日以上持つ人をモデル化対象とした。その結果、2019 年では 72,970 人、2020 年では 62,637 人が抽出された。

図 7 は、2019 年の人々をクラスタリングし、グループごとの特徴を可視化したものである。横軸は時間を表しており、ピンは 30 分である。縦軸は滞在エリアタイプの割合を表している。エリアタイプとは、4.1 節で示したオフィス・住宅・交通・商業エリアを指し、各色は図 5a の UAS クラスタリング結果に対応している。200 個のグループに分類したが、本論文ではその内の特徴的な 6 グループを示す。クラスタリング手法は、エリアモデリングの時と同じ理由で、k-means++を使用する。

グループ #1~#6 の解釈は次の通りである。#1 の人々は、平日の日中は大半をオフィスエリアで過ごし、夕方から朝方にかけては住宅エリアで過ごす傾向にあることが読

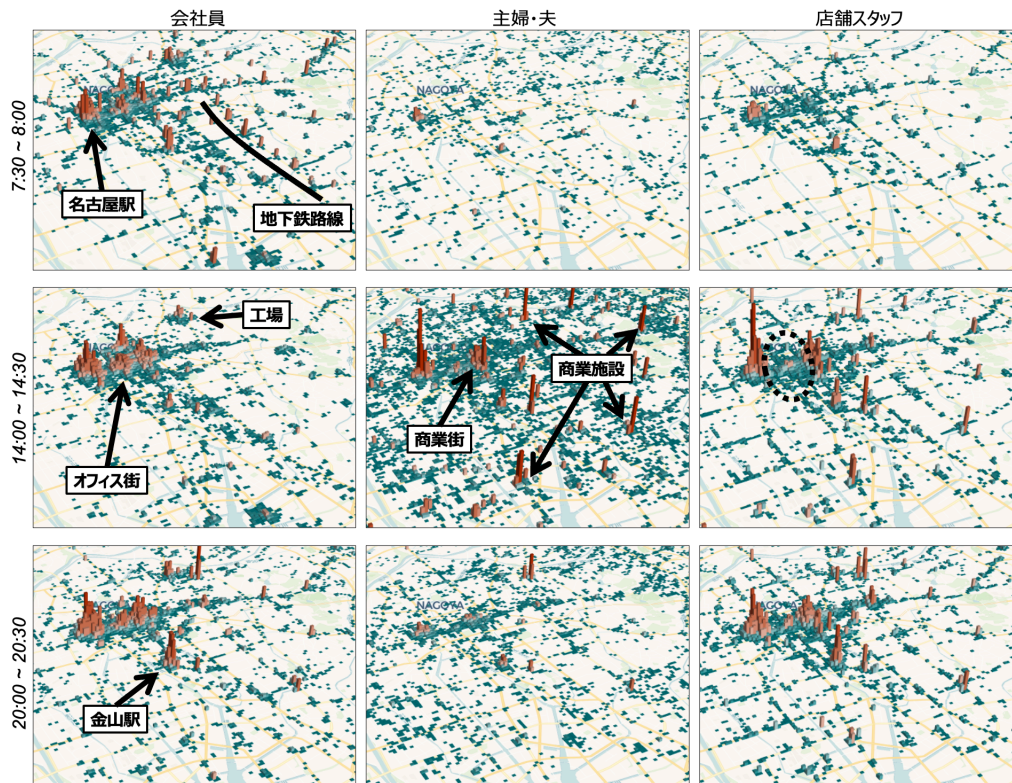


図 8: 都市における推定属性別人口密度の可視化

み取れる。また、休日はほとんどを住宅エリアで過ごす、日中に商業エリアに移動することも読み取れる。つまり、このグループの人々は、会社員であると解釈できる。しかしここで、クラスタ#2も同様に会社員だと解釈できるが、クラスタ#1の会社員とは少し違い、休日に外出する確率が高いことが読み取れる。つまり、会社員の中でも、休日の過ごし方の違いでさらに分類可能なことが分かる。

#3と#4の人々は、1日の大半を住宅エリアで過ごす、日中に商業エリアに行くことも多いことが読み取れ、主婦であると解釈できる。しかし、外出するタイミングについて、#3の人々は平日に多く、#4の人々は休日に多い。つまり、主婦の中でも外出のタイミングの違いでさらに分類可能なことが示された。

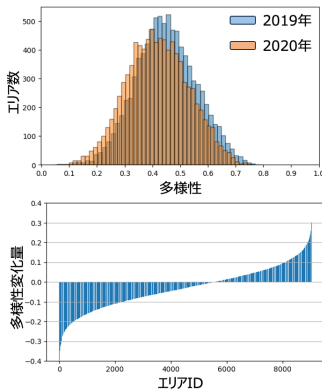
#5は、会社員とは異なり、日中を過ごす場所が商業エリアであることが読み取れる。つまり、このグループの人々は、店舗や飲食店の店員である可能性が高いことが分かる。#6は、#5と似ているが、商業エリアで過ごす時間帯が少し遅いことが分かる。これは、居酒屋やバーといった夜中まで営業している店舗で働いている店員である可能性が高いと解釈できる。つまり、店員の中でもさらに働く店の種別が分類可能なことが示された。

次に、彼らの実空間上での移動を見ることで、上記属性推定結果の妥当性を示す。図8は、2019年の7:30~8:00、14:00~14:30、20:00~20:30における、会社員、主婦、店員の名古屋市内での人口分布である。グリッドサイズは100

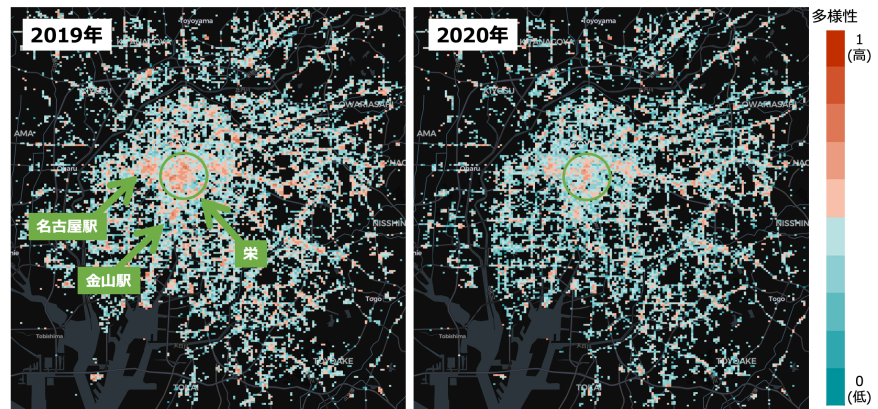
m四方である。バーが高いほど人口が多いことを示す。ただし、ユニークユーザー数が5人未満のグリッドは表示しない設定としており、また3ヶ月間で同じグリッドに同じ人はカウントされないことに注意されたい。

まずは7:30~8:00を見てみる。会社員については、人口の多いエリアが一定間隔で並んでいる箇所が観察された。そこには、地下鉄の駅が存在することが確認された。また、オフィス街も人口が多いことが観察された。それに対して、主婦や店員の場合、人口の多いエリアはほとんど存在しなかった。この時間帯は通勤時間であり、会社員が駅に集中するのは我々の直感に合致する。一方で、主婦や店員にとっては外での活動にはやや早い時間帯であるため、彼らが多く集まる場所はまだまだ少ないのも自然だと言える。

次に14:00~14:30を見てみる。会社員は、朝とは異なり、主にオフィス街での人口集中が観察された。一方、主婦は、人口が多いエリアが点在しており、それらのほとんどが商業施設(イオンモールやアピタなど)であることが確認できた。店員については、主婦と同じ場所で人口が集中していることが観察された。この時間帯は、会社員はオフィスで仕事をし、主婦は買い物に出かけたり家で作業をし、店員は店舗で働くのが一般的である。つまり、我々が観測した属性別の人口分布傾向は、我々の直感に合致している。また、店員については、図中の点線で囲まれたオフィスエリアに人が集まっていないことが確認できた。これは、店員と会社員の働く場所の違いを明確に表しており、



(a) 多様性分布と共通エリア間での多様性の変化量



(b) 地図上での多様性の可視化

図 9: 都市の「休日」の生活パターン多様性を可視化

生活パターンが適切に差別化されていると言える。

最後に 20:00～20:30 を見てみる。会社員は、オフィス街にまだ人が多いのが観察された。また、朝と同様に、地下鉄沿いに人が多いことも観察された。主婦は、朝と同様に、人が多いエリアをほとんど持たないことが見て取れた。店員については、朝とは違い、商業街や大きな駅に人が多いのが観察できた。会社員と主婦にとってこの時間帯は、一般的に帰路につく時間であるため、人口が多いエリアは少ないと我々は想像していた。主婦はその考えにマッチした結果になったが、会社員は実際には大きな駅を中心に多くの人が街に残るといった結果になった。これはおそらく、残業をしている、または仕事終わりの飲み会などが影響していると考えられる。店員は一般的に、レストランやバーなどはまだ営業を行っているはずであり、商業エリアや駅に人が集まっていると想像でき、実際にそのような状態になっており、我々の直感に合った結果となった。

以上より、PSU のベクトル演算を通して、人を生活パターンに基づいてモデル化できることが示された。

4.4 多様性算出

本節では、DiverCityMeter がエリアごとの生活パターン多様性を分析するためのフレームワークとして有効なことを示す。

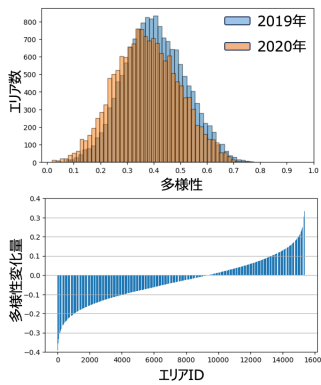
我々は、人モデリング結果をクラスタリングし、人々を生活パターンに応じた 10 グループに分けた。そして、式 (2) を適用することで、エリアごとの生活パターン多様性を算出した。このとき、COVID-19 の前 (2019 年) と途中 (2020 年) を比較することで、都市の各エリアにおける生活パターン多様性の変化を分析した。

本論文では、125m 四方のグリッドごとに多様性を算出した。ただし、平日と休日の 10 時～16 時を分析対象とし、滞在時間が 10 分以上 12 時間未満の滞在のみを使用した。また、ユニーク滞在者が 10 人以上のエリアを多様性

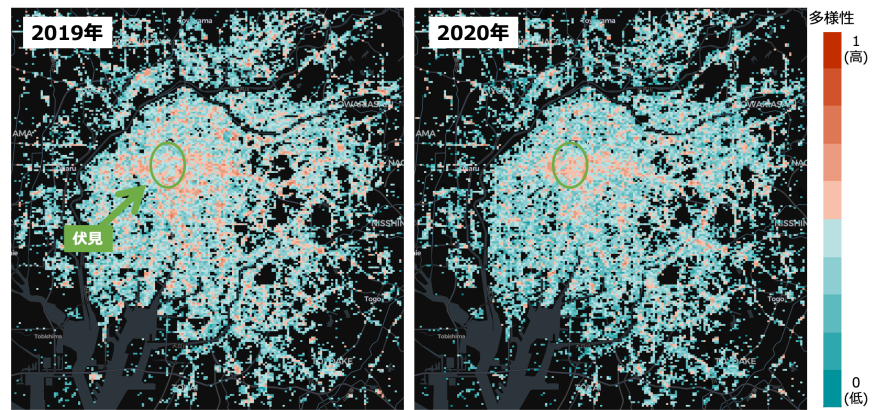
算出の対象とした。2019 年と 2020 年では、滞在レコード数に大きな差が存在した (2019 年: 18,160,489, 2020 年: 11,286,923) ため、ランダムサンプリングで 2019 年の滞在レコード数を 2020 年のものと一致させた。その結果、2019 年と 2020 年のそれぞれの総滞在時間は 33,316,809 時間と 33,354,166 時間となった。エリアの多様性に関しては、1 ヶ月ごとの多様性を算出し、その平均をそのエリアの多様性として扱った。

図 9 は、休日 10 時～16 時のエリアごとの多様性を 2019 年と 2020 年で比較したものである。上のグラフは、2019 年と 2020 年のそれぞれの多様性の分布を表す。2020 年は 2019 年と比べると、都市全体として多様性が低くなったことが読み取れる。下のグラフは、エリアごとの 2019 年から 2020 年への多様性の変化量を示す。つまり、2020 年の多様性から 2019 年の多様性をエリアごとに引いたものである。このグラフからも、多くのエリアで多様性が低くなったことが読み取れる。図 9b は、エリアごとの多様性を地図上に描画したものである。2019 年と比べると 2020 年では、栄 (商業街や繁華街) 周辺の多くのエリアで多様性が低下した様子が観察できた。また、それ以外の名古屋駅や金山といった主要ハブ周辺も多様性が低下した。これは、COVID-19 の影響により、人々が不要不急の外出を控えたためだと考えられる。また、人々が滞在時間を短くするように努めた結果の可能性もある。つまり、仕事として長時間滞在する人と、短時間滞在を意識した訪問者との間の隔たりが大きくなったため、エリアとしては多様性が下がったと考えられる。

図 10 は、平日 10 時～16 時のエリアごとの多様性を 2019 年と 2020 年で比較したものである。図 10a を見ると、休日同様、2020 年は 2019 年と比べると、都市全体として多様性が低くなったことが読み取れる。図 10b を見ると、2019 年と比べると 2020 年では、伏見 (オフィス街) の多様性が上昇した様子が見て取れる。定常時は定時出勤が主だっ



(a) 多様性分布と共通エリア間での多様性の変化量



(b) 地図上での多様性の可視化

図 10: 都市の「平日」の生活パターン多様性を可視化

た会社員の生活パターンに、リモートワークやフレックスタイムという新しい生活様式が出現した。そしてそれに伴い、定時出勤の会社員の割合が減少した。これらの現象が重なったことで、結果として生活パターン分布が均一になり、オフィス街の多様性が上がったと考えられる。

以上より、DiverCityMeter というフレームワークを使用することで、データ駆動的に都市の生活パターン多様性が分析可能なことが示された。

5. まとめと今後の展望

本研究では、GPS 位置データのみを使い、生活パターンに応じて人をモデル化することで、都市の各エリアの生活パターン多様性を算出するための手法 DiverCityMeter を提案した。そして、2019 年と 2020 年の名古屋市を比較し、COVID-19 の影響による生活パターン多様性の変化を確認した。今後の展望としては、まず初めに、目的とデータの特性に沿った多様性指標の適切な選択が挙げられる。本論文では、Moro らが提案した指標 [3] を使用したが、彼らが対象にしているのは所得情報であり、生活パターン情報とはデータの性質が異なる可能性がある。多様性指標には様々なものが存在するため、この選択は分析結果に影響を及ぼす可能性があり慎重に行う必要がある。次に、多様性変化の要因の追求が挙げられる。本論文では、多様性の値のみを観察したが、生活パターングループごとの影響度を分析することや、どのような POI が影響を受けやすいかを調査することで、より詳細な洞察が得られるはずである。

謝辞 本研究の一部は、NICT 委託研究 (22609)、内閣府 SIP3 (JPJ012495)、JST CREST (JPMJCR22M4) により支援して頂いております。また、データ提供にご協力頂きました株式会社ブログウォッチャーに感謝いたします。

参考文献

- [1] Herring, C.: Does Diversity Pay?: Race, Gender, and the Business Case for Diversity, *American Sociological Review*, Vol. 74, No. 2, pp. 208–224 (online), DOI: 10.1177/000312240907400203 (2009).
- [2] Bettencourt, L., Samaniego, H. and Youn, H.: Professional diversity and the productivity of cities, *Scientific Reports*, Vol. 4, p. 5393 (online), DOI: 10.1038/srep05393 (2014).
- [3] Moro, E., Calacci, D., Dong, X. and Pentland, A.: Mobility patterns are associated with experienced income segregation in large US cities, *Nature Communications*, Vol. 12 (online), DOI: 10.1038/s41467-021-24899-8 (2021).
- [4] Yabe, T., Bueno, B., Dong, X., Pentland, A. and Moro, E.: Behavioral changes during the COVID-19 pandemic decreased income diversity of urban encounters, *Nature Communications*, Vol. 14 (online), DOI: 10.1038/s41467-023-37913-y (2023).
- [5] 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).
- [6] Li, X., Zhao, K., Cong, G., Jensen, C. S. and Wei, W.: Deep Representation Learning for Trajectory Similarity Computation, *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pp. 617–628 (online), DOI: 10.1109/ICDE.2018.00062 (2018).
- [7] Olive, X., Basora, L., Viry, B. and Alligier, R.: Deep Trajectory Clustering with Autoencoders, *ICRAT 2020, 9th International Conference for Research in Air Transportation*, ICRAT 2020 (2020).
- [8] Wan, Y., Zhou, C. and Pei, T.: Semantic-Geographic Trajectory Pattern Mining Based on a New Similarity Measurement, *ISPRS International Journal of Geo-Information*, Vol. 6, No. 7 (online), DOI: 10.3390/ijgi6070212 (2017).
- [9] Cao, H., Xu, F., Sankaranarayanan, J., Li, Y. and Samet, H.: Habit2vec: Trajectory Semantic Embedding for Living Pattern Recognition in Population, *IEEE Transactions on Mobile Computing*, Vol. 19, No. 5, pp. 1096–1108 (online), DOI: 10.1109/TMC.2019.2902403 (2020).
- [10] Xu, F., Xia, T., Cao, H., Li, Y., Sun, F. and Meng, F.: Detecting Popular Temporal Modes in Population-scale

- Unlabelled Trajectory Data, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., Vol. 2, No. 1 (online), DOI: 10.1145/3191778 (2018).
- [11] Ben-Gal, I., Weinstock, S., Singer, G. and Bambos, N.: Clustering Users by Their Mobility Behavioral Patterns, ACM Trans. Knowl. Discov. Data, Vol. 13, No. 4 (online), DOI: 10.1145/3322126 (2019).
- [12] Li, W., Zhang, H., Chen, J., Li, P., Yao, Y., Shi, X., Shibasaki, M., Kobayashi, H. H., Song, X. and Shibasaki, R.: Metagraph-Based Life Pattern Clustering With Big Human Mobility Data, IEEE Transactions on Big Data, Vol. 9, No. 1, pp. 227–240 (online), DOI: 10.1109/TB-DATA.2022.3155752 (2023).
- [13] Yoshimura, Y., Kumakoshi, Y., Milardo, S., Santi, P., Arias, J. M., Koizumi, H. and Ratti, C.: Re-visiting Jane Jacobs: Quantifying urban diversity, Environment and Planning B: Urban Analytics and City Science, Vol. 49, No. 4, pp. 1228–1244 (online), DOI: 10.1177/23998083211050935 (2022).
- [14] Srivastava, N., Mansimov, E. and Salakhutdinov, R.: Unsupervised Learning of Video Representations Using LSTMs, Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15, JMLR.org, p. 843–852 (2015).
- [15] Iwata, S., Nitta, T., Takayama, T., Yanagisawa, M. and Togawa, N.: A stayed location estimation method for sparse GPS positioning information, 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE), pp. 1–5 (online), DOI: 10.1109/GCCE.2017.8229230 (2017).
- [16] GoogleDevelopers: Blog of our latest news, updates, and stories for developers, <https://developers.googleblog.com/2017/11/introducing-tensorflow-feature-columns.html> (2017).
- [17] Arthur, D. and Vassilvitskii, S.: K-Means++: The Advantages of Careful Seeding, Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07, Society for Industrial and Applied Mathematics, p. 1027–1035 (2007).