

滞在ビッグデータに基づくエリア分散表現の 加法構成性の分析と活用

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

概要：人々が都市や地域における各エリアをどのように利用するかといったエリア利用形態の把握は、商圏分析・出店戦略等の商業的利用をはじめ、都市計画、交通施策、感染予測など幅広い分野で重要な課題である。エリアの利用形態の定量的な分析のため、位置情報データを用い、エリアをモデル化する技術がさかんに研究されている。このようなエリアモデリング手法として、単語埋め込み手法である Word2Vec に倣い、それぞれのエリアを埋め込んだベクトル表現（分散表現）としてモデル化する手法が提案されている。ここで、Word2Vec による単語分散表現について、多義語の語義に対する分散表現の平均が多義語の分散表現となる性質が知られている。本性質は OR の加法構成性と呼ばれ、例えば「右」と「権利」の語義を持つ“right”の分散表現は、それぞれの語義の分散表現の平均に近似できる。本研究では、エリアごとの滞在傾向をモデル化する Area2Vec による分散表現について、エリアの包含関係と OR の加法構成性の関係について分析する。また、本性質を用いた、エリアの空間的な定義の変換手法や意味の合成手法を提案する。

キーワード：エリアモデリング、移動ビッグデータ、スマートシティ

Analysis and Application of Additive Compositionality of Area Distributed Representation Based on Stay Big Data

1. はじめに

人々が都市や地域における各エリアをどのように利用するかといったエリア利用形態の把握は、商圏分析・出店戦略等の商業的利用をはじめ、都市計画、交通施策、感染予測など幅広い分野で重要な課題である [1]。そこで、エリアの利用形態の定量的な分析のため、位置情報データを用い、エリアをモデル化する技術がさかんに研究されている。このようなエリアモデリング手法として、単語埋め込み手法である Word2Vec [2] に倣い、それぞれのエリアの傾向を埋め込んだベクトル表現（分散表現）を獲得する手法が挙げられる [3], [4], [5], [6]。本研究では、滞在傾向によってエリアごとの分散表現を獲得する Area2Vec [6] を用いる。Area2Vec の詳細については 2 章で説明する。

エリアモデリングにおける課題として、エリアの空間的な定義が固定される点が挙げられる。エリアモデリングでは、まずエリアを空間的に定義し、それぞれのエリアに対する傾向をモデル化していく。ここでエリアの空間的な定義方法は、図 1 に示すように、固定サイズのメッシュ、施設の空間的形状に合わせたポリゴンなど、使用するデータセットや分析対象によって様々である。例えば、都市全体を網羅的にモデル化したい場合は 250m 四方など固定サイズのメッシュを定義する必要がある、一方で公園などの一部のエリアの利用形態を分析したい場合は、公園の形状に従って公園のエリアのみを切り取る必要がある。また、エリアの空間的な定義は対象地域のデータ量にも依存する。例えば都市の中心部など、データが豊富なエリアでは、それぞれのエリアが豊富な特徴をもつためより細かくエリアを定義可能である。一方で、位置情報データが少ない山間部や郊外などの地域は、より広いエリアの特徴を集約する必要がある、より粗くエリアを定義する必要がある [7]。このように、個々の用途に従いモデル化するエリアの定義は

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)} tam@ucl.nuee.nagoya-u.ac.jp



図 1: エリア定義間の変換

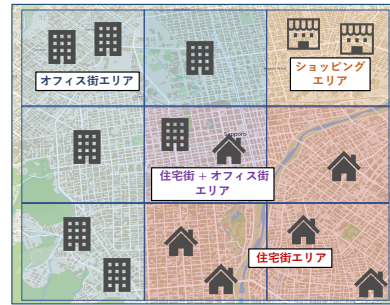


図 2: 多用途エリア

様々である。しかし、既存のエリアモデリング手法では、固定のエリア定義に限られ、モデル化後のエリア形状の相互変換ができない。

これに加えて、複数の用途が混在したエリアに対する分析手法は提案されていない。既存のエリアモデリング手法では、住宅街やオフィス街など、エリアをその地域の用途によって分類する。しかし、図 2 の中央のメッシュエリアのように、複数の用途を併せ持つエリアが生じる。このような多用途エリアでは、多様な需要が見込まれ、都市施策やマーケティングの上で非常に重要なエリアである。例えば、住宅とオフィスが混在するエリアでは、オフィスワーカーによるコンビニや飲食店等への需要に加えて、居住者によるスーパーマーケットや病院等に対する需要も生じる。また、これらの需要の時間帯による変化も予測される。このような多様な需要の把握によって、適切なサービスの提供やマーケティングが可能である。しかし、既存のエリアモデリング手法では、このような多用途エリアに対する分析が難しい。

一方、Word2Vec による単語分散表現について、多義語の語義に対して、語義の分散表現の平均が多義語の分散表現となる性質が知られている [8]。本性質は OR の加法構成性と呼ばれ、例えば「右」と「権利」の語義を持つ “right” の分散表現は、それぞれの語義の分散表現の平均に近似できる。本研究ではエリア分散表現における OR の加法構成性について分析し、本性質を用いたエリアの空間的な形状の変換手法や多用途エリアの分析手法を提案する。例えば、エリアの定義を固定サイズのメッシュから行政区分などのポリゴンに変化したい場合、エリア分散表現の演算によって、これを実現する。また、エリア分散表現の加減算による複数のエリアの意味の合成によって、多用途エリアの分析を行う。

本稿の構成は以下の通りである。2 章では Word2Vec や既存のエリアモデリング手法について紹介し、本研究で利用するエリアモデルである、Area2Vec について説明する。3 章ではエリア分散表現における OR の加法構成性、および活用手法について述べる。4 章では OR の加法構成性の分析を行い、本性質の活用手法について評価する。最後に 5 章で今後の展望について述べる。

2. 関連研究

モバイル端末の普及により、GPS、WPS (WiFi Positioning System) などによる、大量で時空間的に高粒度なモビリティデータが収集可能になっている [9]。このようなデータ利活用手法は多岐に渡り、都市計画 [1] やマーケティングなど様々な分野に有効である。例えば、人の行動予測 [10]、潜在顧客の推定 [11]、都市ダイナミクスの予測 [12]、POI(Point of Interest) レコメンド [13]、都市の異常検知 [14]、感染症のリスク推定 [15]、都市シミュレーション [16] などが挙げられる。

本研究では、滞在ビッグデータを利用し、人が都市や地域における各エリアをどのように利用するかといったエリア利用形態の把握を目的とした、エリアモデリングを行う。様々な機能を持つエリアのモデル化によって、効率的な分析が可能となり、多様なタスクに有益な情報を得ることができる。エリアモデリング手法は、使用するデータやエリアのもつ傾向をどのように抽象化するかといった点で様々なバリエーションがあるが、エリアを連続的なベクトルレベルでモデル化する手法として、Word2Vec[2] ベースの手法がさかんに研究されている。Word2Vec とは単語ごとの分散表現を作成するモデルである。自然言語処理の分野において、機械学習等のモデルへの入力として利用するために、単語をベクトルで表現する方法が研究されている。最もシンプルなのは、単語ごとの one-hot ベクトルとして表現する方法であるが、単語の総数に等しい高次元のベクトル表現となり、計算コストが高くなる。これに対して、Word2Vec による単語の分散表現は、より密で低次元のベクトル表現であり、単語間の共通性をベクトル間の類似度として反映している。

このような Word2Vec のアルゴリズムを活用し、エリアをモデリングする手法が研究されてきた。Yao ら [3] や Crivellari ら [4]、Liu ら [17] はエリア間の移動遷移から各エリアの関係を分散表現としてモデル化する手法を提案している。Zhai ら [5] は多様な POI を定義した上で、POI 同士の関係性を分散表現としてモデル化する Place2Vec を用いることでエリアをモデリングしている。庄子ら [6] は滞在時刻や滞在時間の長さなどの滞在情報を傾向として、エ

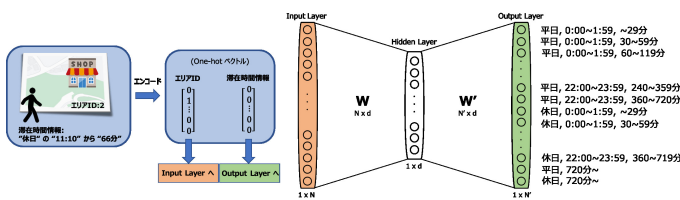


図3: Area2Vec([6] から引用)

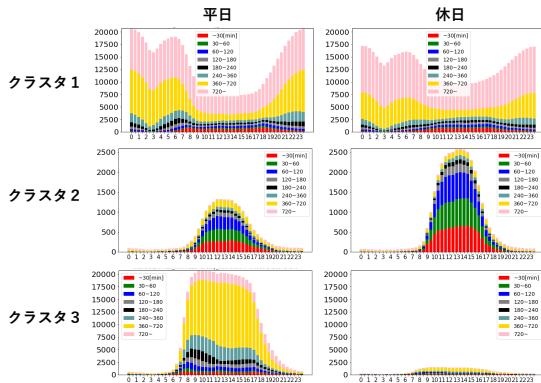


図4: Area2Vecによるエリアのクラスタリング

リア内での人の活動傾向を反映したエリア分散表現を生成している。これは、地域の使われ方はエリア内の滞在特微量に紐づくという考えに基づいており、これによって、滞在傾向としてエリア分散表現の解釈が容易になっている。本手法で生成したエリア分散表現をUAS(the Usage of Area with Stay information)と呼ぶ。Area2Vecのアーキテクチャを図3に示す。入力はいくつかのエリアに対応するone-hotベクトルであり、出力は分散化された各滞在特微量に対応するone-hotベクトルである。滞在特微量は、「平日/休日」、「滞在開始時刻」、「滞在時間の長さ」の3つを用いている。これにより、各エリアのUASは、滞在傾向が似ているもの同士は分散表現間の類似度が高く、滞在傾向が異なるもの同士は類似度が低くなる。従って、UASのクラスタリングによって図4のように、滞在傾向によるエリアの分類が可能になる。本研究では、エリアモデリング手法として、Area2Vecを用いる。

既存のエリアモデリング手法では、エリアの定義が統一されておらず、相互比較が難しい。エリアの定義方法は、特定の地点の抽出[4], [5], [17], 固定サイズのメッシュ[6], [7], [18]や、施設の空間的形状に合わせたポリゴン[3]など、使用するデータセットや分析対象によって様々である。また、エリアの定義は対象地域のデータ量にも依存する。これに対して、Haikalら[7]はいくつかのサイズの階層的なメッシュの定義により、次の滞在地点の予測やエリアの分類タスクでの精度向上を実現した。しかし、あくまで予め定義されたサイズのメッシュを活用しており、より高度な分析のためには、異なるエリア定義によるモデル間の相互比較が必要である。

これに加えて、複数の用途が混在したエリアに対する分析手法は提案されていない。既存のエリアモデリング手法では、住宅街やオフィス街など、エリアをその地域の用途によって分類する。しかし、多くのエリアは複数の用途が混在し、細分化されたエリアでも、それぞれのユーザーにとってエリアの用途は異なる。例えば、ある飲食店は来店者にとっては飲食の場として使われるが、店主にとっては職場として使われる。Linら[19]は、このような人によるエリアの用途の違いを、用途ごとに別々の表現として埋め込むことを提案している。しかし、単一用途エリアと多用途エリアの関係や、意味の合成には言及しておらず、このような多用途エリアに対する分析は難しい。

そこで、本研究ではエリア分散表現におけるORの加法構成性について分析し、本性質を用いたエリアモデリング手法における課題を解決する。具体的には、エリア定義の相互変換や、分散表現の算術演算による多用途エリアの分析手法を提案する。

3. 提案手法

本章では、Area2Vecを用いたエリア分散表現であるUASについて、ORの加法構成性とその活用手法について説明する。

3.1 エリア表現におけるORの加法構成性

まず、自然言語処理の分野における多義語と語義の分散表現間の関係を紹介する。多義語とは複数の語義を持つが、具体的な使用においては特定の語義のみが現れる単語である。例えば、「right」は「右」や「権利」の単語を語義として持つ多義語であり、2つの語義が重複して用いられることはない。Naitoら[8]によると、多義語 w が k 個の語義 w_1, w_2, \dots, w_k を持つ時、多義語の分散表現は語義の分散表現の重み付き平均に近似できる。すなわち、多義語の分散表現を v_w 、語義の単語分散表現を $v_{w_1}, v_{w_2}, \dots, v_{w_k}$ とすると、以下の関係が成り立つ。

$$v_w \approx \sum_{i=1}^k \frac{p(w_i)}{p(w)} v_{w_i} \quad (1)$$

$p(w), p(w_i)$ はそれぞれ多義語と語義に対応する単語のコーパス内での出現頻度である。つまり、多義語の分散表現はそれぞれ語義の単語出現頻度による重み付き平均に近似できる。

次に、階層的なエリアメッシュを定義した上で、階層的なエリア間のUAS同士の関係を説明する。図5のように、空間的粒度の細かい子メッシュ m_1, m_2, \dots, m_k と、 k 個の子メッシュを包含する空間的粒度の粗い親メッシュ M を定義する。親メッシュの面積は子メッシュの面積の和に等しく、メッシュ m の面積を S_m とすると、

$$S_M = \sum_{i=1}^k S_{m_i} \quad (2)$$

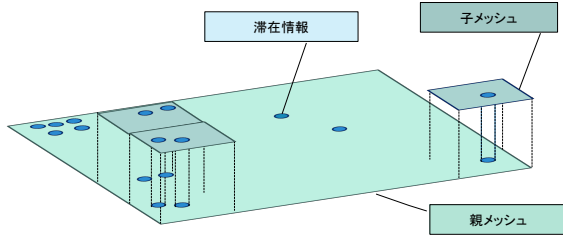


図5: 階層的なエリアメッシュ

で表せるとする。この時、親メッシュにおける滞在情報集合を I_M とすると、以下の式が成り立つ。

$$I_M = I_{m_1} \cup I_{m_2} \cup \dots \cup I_{m_k} \quad (3)$$

$$v_M \approx \sum_{i=1}^k \frac{p(m_i)}{p(M)} v_{m_i} \quad (4)$$

$p(M)$, $p(m_i)$ はそれぞれ親メッシュと子メッシュに対応する滞在情報量である。つまり、親メッシュの分散表現はそれぞれの子メッシュの滞在数による重み付き平均に近似できる。これをエリア分散表現における OR の加法構成性とする。本性質については4章で詳しく分析する。

3.2 エリア定義の変換

式(4)が成り立つと仮定した上で、多様な形状、大きさのエリアに合わせた UAS の獲得、相互変換手法について説明する。エリアは様々な形状で定義されるが、本研究では図1に示すポイント、メッシュ、ポリゴンの3つの形状を扱う。

まず、ポイントの UAS を他の形式のものに変換する場合など、包含関係に重複がない場合を考える。この時、全てのポイントは重複なくエリアに対応付けることができ、1つまたは複数のポイントに対応するエリアの UAS は、式(4)を用いて、対応したポイントの UAS の滞在重み付き平均として得られる。例えばメッシュやポリゴンで定義されたエリア A の範囲内にポイントとして定義されたエリア a_1, a_2, \dots, a_k が存在する場合、 A の UAS, v_A は、

$$v_A \approx \sum_{i=1}^k \frac{p(a_i)}{p(A)} v_{a_i} \quad (5)$$

$$p(A) \approx \sum_{i=1}^k p(a_i)$$

として得られる。このようにエリアの包含関係に重複がない場合、式(5)によって、エリアの相互変換ができる。

次に、メッシュの UAS をポリゴンの UAS へ変換する場合など、包含関係に重複がある場合を考える。図6の例では、赤色のポリゴンが複数のメッシュにまたがっている。ここで、エリア内の滞在密度が一樣であるとすると、ポリゴンとメッシュの共通部分における滞在数はその面積に比例する。従って、面積 S_A のポリゴン A が複数のメッシュ m_1, m_2, \dots, m_k にまたがっている場合、それぞれのメッシュ

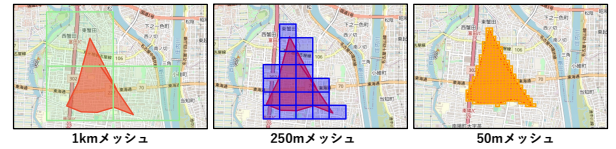


図6: あるポリゴンと重なるメッシュの比較

の面積 $S_{m_1}, S_{m_2}, \dots, S_{m_k}$ と、それぞれのメッシュとの共通部分の面積 $S_{A \cap m_1}, S_{A \cap m_2}, \dots, S_{A \cap m_k}$ を用いて、

$$v_A \approx \sum_{i=1}^k \frac{S_{A \cap m_i} p(m_i)}{S_{m_i} p(A)} v_{m_i} \quad (6)$$

$$p(A) \approx \sum_{i=1}^k \frac{S_{A \cap m_i} p(m_i)}{S_{m_i}}$$

と表せる。このように、式(4)やその拡張を用いて、エリアの定義に合わせて UAS を変換する。

3.3 エリアの意味の合成

最後に、式(4)が成り立つと仮定した上で、UAS の加減算によるエリアの意味の合成手法について説明する。各エリアは、夜間の長時間滞在が見られる「住宅街傾向」や昼間の長時間滞在が見られる「オフィス街傾向」など様々な意味を持つ。これらの滞在傾向は、エリアが細分化されるにつれ、エリアごとに単一の用途に相当する傾向をもつ。例えば、各エリアに単一の住宅や店舗のみ含まれる場合、それらの住宅や店舗の使われ方のみを反映した滞在傾向となり、これが自然言語における単一の意味の語義にあたる。多くのエリアでは、その範囲が広くなるにつれて、複数の単一用途のエリアを含む多用途エリアとなる。このような多用途エリアにおける滞在情報集合を I_M 、その範囲内の各単一用途エリアに相当する滞在情報集合を $I_{m_1}, I_{m_2}, \dots, I_{m_k}$ とすると、

$$I_M = I_{m_1} \cup I_{m_2} \cup \dots \cup I_{m_k} \quad (7)$$

となる。従って、式(4)と同様に、多用途エリアの UAS は、各用途エリアの UAS の滞在重み付き平均として得られる。すなわち、多用途エリア B の UAS を v_B 、範囲内の単一用途エリア b_1, b_2, \dots, b_k の UAS を $v_{b_1}, v_{b_2}, \dots, v_{b_k}$ とすると、

$$v_B \approx \sum_{i=1}^k \frac{p(b_i)}{p(B)} v_{b_i} \quad (8)$$

$$p(B) \approx \sum_{i=1}^k p(b_i)$$

が成り立つ。このように範囲内の単一用途エリアの滞在重み付き平均によって、多用途エリアの UAS を計算できる。

4. 評価実験

本章では、実験によって3章における OR の加法構成性について分析し、エリア定義の変換手法、エリアの意味の合成手法について評価する。

表 1: データセット

期間	2019	2020	2021
ユーザ数	970328	572180	245259
滞在数	23454763	17787184	7732395

表 2: 滞在特徴量の離散化

曜日	平日, 休日
滞在開始時刻	0:00~0:59, 1:00~1:59, ... 23:00~23:59
滞在時間	5 min~29 min, 30 min~59 min, 60 min~119 min, 120 min~179 min, 180 min~239 min, 240 min~359 min, 360 min~719 min, 720 min~

4.1 実験設定

我々が使用するデータは事前に同意を得たユーザのスマートフォンにインストールされたアプリから収集された GPS データセットである。本研究では滞在特徴を用いてエリアをモデリングするため、[20] の手法を用い、滞在情報 $s = (lat, lon, t_a, t_l)$ として抽出する。 lat, lon, t_a, t_l はそれぞれ、緯度、経度、到着時刻（滞在開始時刻）、出発時刻（滞在終了時刻）であり、ある地点への 1 回の滞재를指す。実験では、2019/04/01~2019/04/30, 2020/04/01~2020/04/30, 2021/04/01~2021/04/30 の期間における、名古屋市の滞在データを用いる。データの詳細を表 1 に示す。

本研究における Area2Vec のパラメータについて説明する。まず、モデル化するエリアは 50m, 250m, 1km 四方のメッシュとし、2019 年 4 月, 2020 年 4 月, 2021 年 4 月と異なる期間のメッシュを別のメッシュとして定義した。従って、それぞれのメッシュは 2019 年 4 月, 2020 年 4 月, 2021 年 4 月と、3 つの期間に対応する 3 つの分散表現が学習される。次に、出力ノードとして利用する離散化滞在特徴量を表 2 に示す。本研究では、[6] を参考に、曜日を 2, 滞在開始時刻を 24, 滞在時間を 8 のカテゴリに離散化している。従って、出力ノードの数は $2 \times 24 \times 8 = 384$ となる。分散表現の次元数（隠れ層の次元数）は 4 とした。

4.2 評価指標

2 つのエリア分散表現群の間の類似度として、AMND (Average Mutual Norm Distance), AMCS (Average Mutual Cosine Similarity) を用いる。これは、[18] の評価手法を参考にしており、前者は同一のエリアに対する分散表現間のユークリッド距離の平均、後者はコサイン類似度の平均である。すなわち、あるエリア a に対する、異なる分散表現をそれぞれ $\mathbf{v}_a, \mathbf{v}'_a$ とすると、

$$AMND = \frac{1}{N} \sum_{a=1}^N \|\mathbf{v}_a - \mathbf{v}'_a\| \quad (9)$$

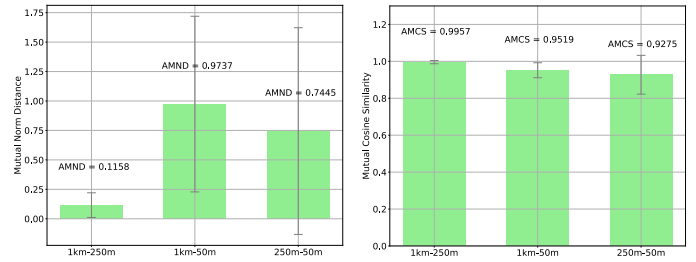


図 7: メッシュ階層間の OR の加法構成性

$$AMCS = \frac{1}{N} \sum_{a=1}^N \frac{\mathbf{v}_a \cdot \mathbf{v}'_a}{\|\mathbf{v}_a\| \|\mathbf{v}'_a\|} \quad (10)$$

となる (N はメッシュの総数)。従って、AMND は低いほど、AMCS は高いほど、それぞれのベクトル群の間の類似度が高いといえる。

4.3 エリア分散表現における OR の加法構成性

3.1 節では、単語分散表現における多義語と語義の関係から、エリアの階層性と UAS の OR の加法構成性の間に成り立つ式 (4) を考察した。ここでは、1km 四方, 250m 四方, 50m 四方の 3 種類のメッシュを定義し、それぞれの間で式 (4) が成り立つか確認する。具体的には、親メッシュの UAS と子メッシュの重み付き平均ベクトルの間の Norm-Distance と Cosine-Similarity を各親メッシュごとに計算する。例えば、1km-250m メッシュ間であれば、それぞれの 1km メッシュに対して、実際に学習した UAS とその範囲内の 250m メッシュの重み付き平均ベクトルとの距離を計算する。図 7 にその結果を示す。

特に、1km-250m 間では類似度が高く、50m メッシュを含む階層間で類似度がやや低くなっている。これは、50m メッシュは 250m メッシュや 1km メッシュと比較してメッシュあたりの滞在数が少なく、分散表現の学習に対する情報量が足りていないためと考えられる。これを検証するため、各親メッシュについて、対応する子メッシュの平均滞在数と、階層間の UAS 類似度を図 8 に示す。図から子メッシュの平均滞在数と類似度の間には正の相関があり、より滞在数の多い子メッシュと親メッシュの間の類似度が高くなっている。従って、滞在数が少ないメッシュにおける UAS は安定せず、学習によって確率的に変化し、滞在数が少ない子メッシュ UAS の平均によるベクトルと親メッシュ UAS の類似度が低くなっていると考えられる。従って、十分な滞在数をもつエリア同士であれば、式 (4) のように、親メッシュの UAS は子メッシュの重み付き平均に近似できると言える。

4.4 エリア定義の変換手法の評価

次に、メッシュとポリゴンの変換式 (6) を評価する。ここでは、図 9 のようにランダムなポリゴンエリアを定義し、

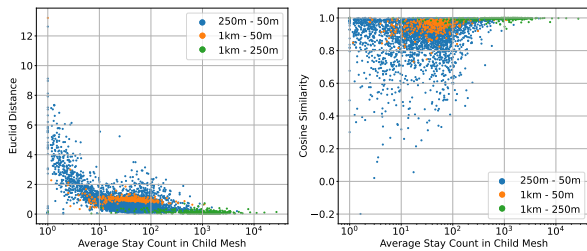


図 8: 子メッシュの平均滞在数と類似度の関係

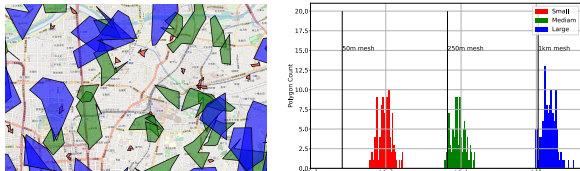


図 9: ランダムポリゴンの定義

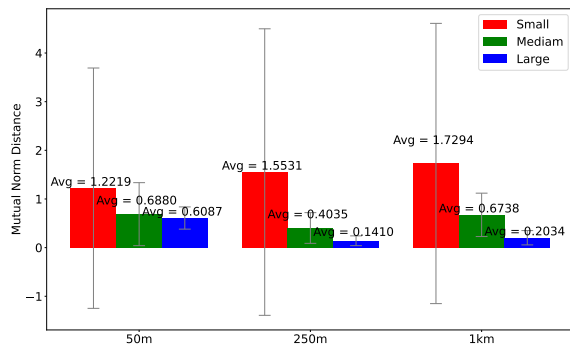


図 10: メッシュ-ポリゴン変換 (AMND)

それぞれのポリゴンの UAS と、式 (6) で計算した UAS の間の類似度を計算する。ランダムなポリゴンは、面積について 3 つの異なる分布から生成しており、その分布を図 9 の右図に示す。また、式 (6) によるポリゴン UAS の計算には、図 6 のように、1km, 250m, 50m 四方のメッシュのいずれかのみをそれぞれ利用し、比較した。これらによって、ポリゴンの大きさと式 (6) の精度の関係、またどの大きさのメッシュを用いて変換するべきかを評価する。その結果を図 10, 11 に示す。

まず、面積が Large, Medium のポリゴンについては、1km メッシュ、250m メッシュの変換による UAS との類似度が 50m メッシュのものと比較して高くなっている。これは図 8 から分かるように、メッシュあたりの滞在数が多いほど、式 (4) の関係が成り立つためと考えられる。一方、Small のポリゴンについては、50m メッシュによる UAS との類似度が高くなっている。これは、図 6 から分かるように、250m メッシュや 1km メッシュと、Small ポリゴンとの面積のずれが原因であると考えられる。また、250m

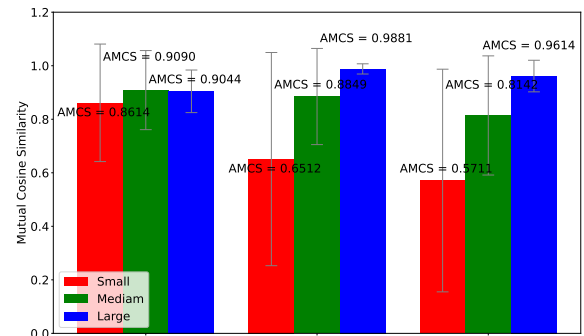


図 11: メッシュ-ポリゴン変換 (AMCS)

メッシュや 1km メッシュにおける UAS の平均化も原因として考えられる。例えば、より大きい面積のエリアは範囲内に多様なエリアを含む可能性が高いため、より平均化された UAS をもち、逆により小さい面積のエリアは単一の用途による滞在傾向をもつ可能性が高い。従って、Small のポリゴンにおける UAS はより偏った滞在傾向による UAS の可能性が高く、250m メッシュや 1km メッシュなどの平均化された UAS では表現できない。これが Small で 50m メッシュによる UAS との類似度が高い原因と考えられる。これらの結果から、メッシュ-ポリゴン変換など、異なる形状への重み付き平均による変換は、メッシュあたりの滞在数や、変換先のエリアの面積を考慮した上で行う必要があるが、十分な滞在数のメッシュ UAS と、適切な大きさ同士であれば、式 (6) は成り立ち、メッシュ-ポリゴン間の変換が可能である。

4.5 エリアの意味の合成手法の評価

最後に、UAS の加減算によって、多用途エリアの分析を行う。まず単一用途エリアの UAS を獲得するため、50m メッシュによるエリアを 128 クラスタにクラスタリングする。これは、エリアをより細分化し、単一用途の滞在傾向のクラスタを得るためである。これらのクラスタから 4 つのクラスタを抽出したものを図 12 に示す。横軸は時刻、縦軸は 1 日ごとの各クラスタにおける滞在数である。例えば、住宅街クラスタでは夜間の長時間の滞在が多くなっており、オフィス街クラスタでは、平日昼間の長時間滞在が多くなっている。次に、単一用途エリア UAS を、これらのクラスタに属する UAS 群の滞在重み付き平均ベクトルとして獲得する。これらの単一用途エリア UAS の加減算で多用途エリアの UAS を獲得する。ここでは、「住宅街」+「オフィス街」、「駅」+「飲食店街」、「駅」+「飲食店街」-「オフィス街」の多用途エリア UAS を獲得した。これらの UAS から Norm 距離の小さいエリアを抽出し、多用途エリアに類似したエリアを検索する。このようなエリアの滞在傾向をそれぞれ図 13 に示す。また、これらのエリアの各

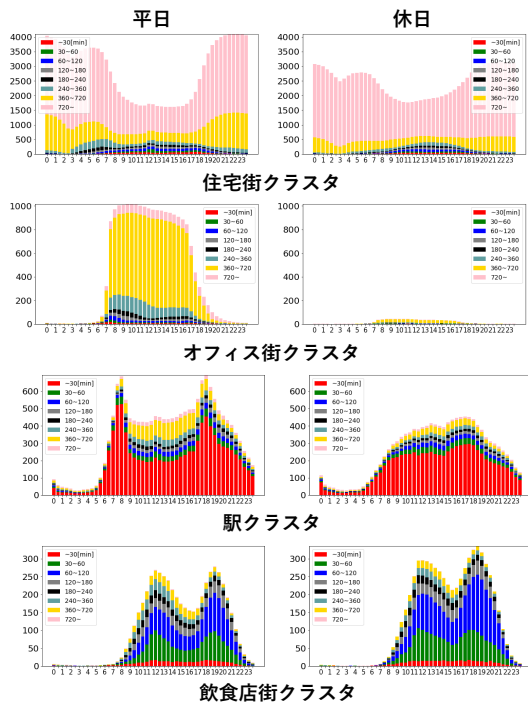


図 12: 単一用途エリアクラスターの滞在傾向

年ごとの地理分布をそれぞれ図 14 に示す。

まず、「住宅街」+「オフィス街」の滞在傾向については、平日昼間の滞在と、平日休日の日中から夜間にかけての長時間滞在が多く見られる。これは、オフィス街による滞在傾向と住宅街の滞在傾向の足しあわせのようになっている。また、地理分布の変化を見ると、2019 年から 2020 年にかけて増加しており、2021 年では再び減少している。これは 2020/4/10 の愛知県における緊急事態宣言の発令による、人々の外出の自粛の結果を反映していると考えられる。一方、2021 年 4 月は緊急事態宣言下になく、このような傾向のエリアは減少している。「駅」+「飲食店街」の滞在傾向については、平日の 8 時、12 時、19 時頃に短時間滞在の増加が見られ、駅と飲食店街の滞在傾向の合成的な傾向となっている。しかし、平日昼間の長時間滞在も多くみられ、ここではより駅と飲食店街のみを含むエリアを抽出するため、「オフィス街」UAS を減算した。その結果、平日昼間の長時間滞在がやや減り、より飲食店街の傾向が強くなっていることがわかる。地理分布に関しては、2019 年の名古屋駅、栄、金山周辺等、駅周辺の飲食店街に分布している。また、「オフィス街」UAS の減算によって、名古屋駅東側のオフィス街や金山駅北側の一部が除かれていることがわかる。また、2020 年や 2021 年ではこれらのエリアの分布は減少している。これは新型コロナウイルス流行の影響で、これらの駅前の飲食店街への滞在が減少したためと考えられる。このように、UAS の加減算によって、多用途エリアの分析への活用が期待できる。

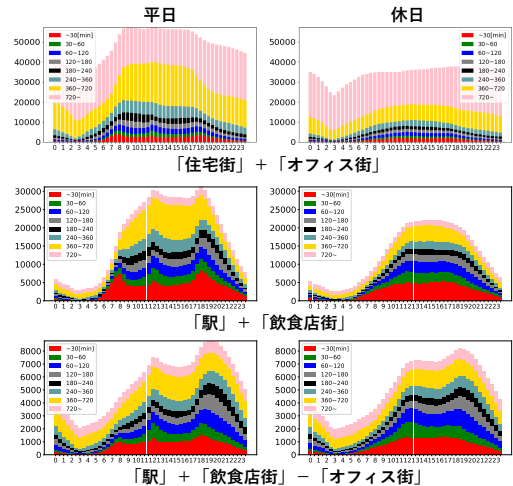


図 13: 各多用途エリアの滞在傾向

5. 今後の展望

本研究では、エリア分散表現における、OR の加法構成性を分析し、これらを用いたエリア定義の変換や、多用途エリアの分析手法を提案した。今後の展望としては、エリア分散表現の変化の分析が挙げられる。今回の実験で、エリアの滞在傾向の変化について多用途エリアの分布から分析したが、これらの変化について、エリア分散表現の時系列的な変化を用いて分析できると考える。例えば、飲食店への滞在の減少や、自宅滞在の増加などは、それらの変化前後のエリア分散表現間の差に反映される可能性が高い。このような変化についての定量的な分析によって、より実用的な活用が期待できる。

謝辞 本研究の一部は NICT 委託研究 (222C01, 22609), JST CREST(JPMJCR21F2, JPMJCR22M4), 科学研究費補助金 (22H03696) に支援いただいています。また、データ提供にご協力いただきましたブログウォッチャー社に感謝いたします。

参考文献

- [1] Xiaohui Mou, Fei Cai, Xin Zhang, Jie Chen, and Rongrong Zhu. 2020. Urban Function Identification Based on POI and Taxi Trajectory Data. In Proceedings of the 3rd International Conference on Big Data Research (ICBDR '19). Association for Computing Machinery, New York, NY, USA, 152–156. <https://doi.org/10.1145/3372454.3372468>
- [2] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'13). Curran Associates Inc., Red Hook, NY, USA, 3111–3119.
- [3] Zijun Yao, Yanjie Fu, Bin Liu, Wangsu Hu, and Hui Xiong. Representing urban functions through zone embedding with human mobility patterns. In Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI'18). AAAI Press, pp.3919–3925, 2018.

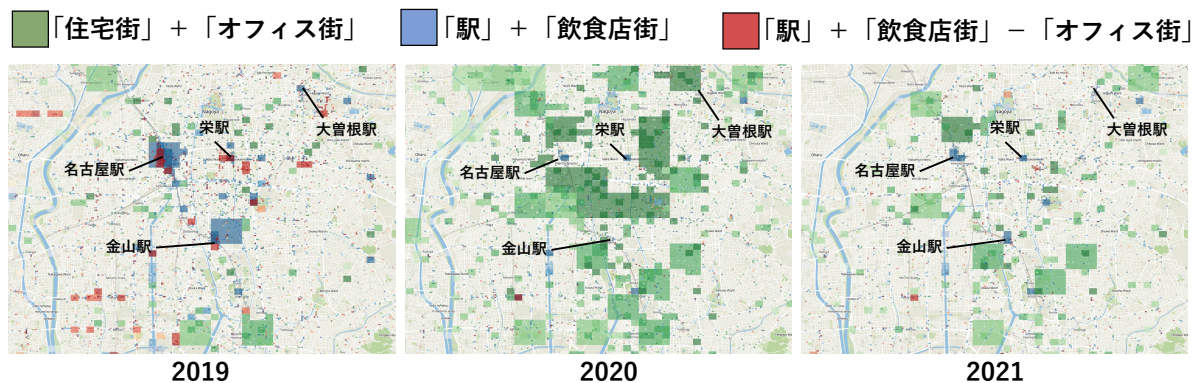


図 14: 各多用途エリアの地理分布

- [4] Crivellari A, Beinat E. From Motion Activity to Geo-Embeddings: Generating and Exploring Vector Representations of Locations, Traces and Visitors through Large-Scale Mobility Data. ISPRS International Journal of Geo-Information, Vol.8, no.3: pp.134, 2019.
- [5] Bo Yan, Krzysztof Janowicz, Gengchen Mai, and Song Gao. From ITDL to Place2Vec: Reasoning About Place Type Similarity and Relatedness by Learning Embeddings From Augmented Spatial Contexts. In Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '17). Association for Computing Machinery, New York, NY, USA, Article 35, pp.1–10, 2017.
- [6] 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.
- [7] Christophe Haikal, Pegah Alizadeh, Christophe Rodrigues, and Bi Chongke. 2022. Place embedding across cities in location-based social networks. In Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing (SAC '22). Association for Computing Machinery, New York, NY, USA, 539–546. <https://doi.org/10.1145/3477314.3506992>
- [8] Kiperwasser, E. and Goldberg, Y. (2017) Revisiting Additive Compositionality: AND, OR and NOT Operations with Word Embeddings. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2.
- [9] Liu H, Darabi H, Banerjee P, Liu J. Survey of wireless indoor positioning techniques and systems. IEEE Tran. Syst. Man Cybern Part C (Appl Rev) 37(6):pp.1067–1080, 2007.
- [10] Y. Zhou and Y. Huang, "DeepMove: Learning Place Representations through Large Scale Movement Data," in 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018 pp. 2403–2412.
- [11] Feng, S., Cong, G., An, B., Chee, Y. M. POI2Vec: Geographical Latent Representation for Predicting Future Visitors. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1), 2017.
- [12] Masamichi Shimosaka, Keisuke Maeda, Takeshi Tsukiji, and Kota Tsubouchi. Forecasting urban dynamics with mobility logs by bilinear poisson regression. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '15, p. 535–546. Association for Computing Machinery, 2015.
- [13] Huayu Li, Yong Ge, Richang Hong, and Hengshu Zhu. Point-of-interest recommendations: Learning potential check-ins from friends. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, p. 975–984. Association for Computing Machinery, 2016.
- [14] Huichu Zhang, Yu Zheng, and Yong Yu. Detecting urban anomalies using multiple spatio-temporal data sources. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., Vol. 2, No. 1, 2018.
- [15] Tessema S, Wesolowski A, Chen A, Murphy M, Wilhelm J, Mupiri AR, Ruktanonchai NW, Alegana VA, Tatem AJ, Tambo M, Didier B, Cohen JM, Bennett A, Sturrock HJ, Gosling R, Hsiang MS, Smith DL, Mumbengegwi DR, Smith JL, Greenhouse B. Using parasite genetic and human mobility data to infer local and cross-border malaria connectivity in Southern Africa. Elife. 2019.
- [16] 田村 直樹, 浦野 健太, 青木 俊介, 米澤 拓郎, 河川 信夫: 都市を対象とした大規模移動履歴に基づく擬似人流データ生成手法. 情報処理学会論文誌, Vol.64, No.1, pp.123–133, 2023.
- [17] Xin Liu, Yong Liu, and Xiaoli Li. 2016. Exploring the context of locations for personalized location recommendations. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16). AAAI Press, 1188–1194.
- [18] Takahiro Yabe, Kota Tsubouchi, Toru Shimizu, Yoshihide Sekimoto, and Satish V. Ukkusuri. 2019. City2City: Translating Place Representations across Cities. In Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '19). Association for Computing Machinery, New York, NY, USA, 412–415. <https://doi.org/10.1145/3347146.3359063>
- [19] Lin, Y., Wan, H., Guo, S., and Lin, Y. (2021). Pre-training Context and Time Aware Location Embeddings from Spatial-Temporal Trajectories for User Next Location Prediction. Proceedings of the AAAI Conference on Artificial Intelligence, 35(5), 4241–4248. <https://doi.org/10.1609/aaai.v35i5.16548>
- [20] S. Iwata, T. Nitta, T. Takayama, M. Yanagisawa and N. Togawa, "A stayed location estimation method for sparse GPS positioning information," 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE), Nagoya, Japan, 2017, pp. 1–5, doi: 10.1109/GCCE.2017.8229230.