口頭発表

2025年11月1日(土) 9:20 ~ 10:40 | C会場(共通教育棟 D21番教室)

[C1] 空間解析理論

司会：西 颯人

9:20 ~ 9:40

[C1-01] ラプラシアン行列に基づく空間可変係数モデルの提案と街路単位の犯罪分析への応用

*許 芮溧[1]、井上 亮[1] (1. 東北大学)

9:40 ~ 10:00

[C1-02] 連続的な空間的異質性を考慮した空間相互作用モデルによる地域間関係分析

*安藤 夏樹[1]、井上 亮[1]、彭 湛[1] (1. 東北大学大学院 情報科学研究科)

10:00 ~ 10:20

[C1-03] LLMベースと従来型ジオコーディングの精度比較：日本の大・中・小規模地域における評価

*Lu Shiqi[1,2,3]、龐 岩博[2,1]、関本 義秀[2,1,3] (1. 関本研究室、2. 空間情報科学研究センター、3. 東京大学)

10:20 ~ 10:40

[C1-04] 大学病院の集中治療室における医療スタッフの動線分析

*三田村 和喜[1]、沖 拓弥[1]、仙頭 佳起[2]、里見 拓哉[1]、野坂 宜之[2]、野口 綾子[2]、海塩 渉[1]、若林 健二[2] (1. 東京科学大学、2. 東京科学大学病院)

©一般社団法人地理情報システム学会

ラプラシアン行列に基づく空間可変係数モデルの提案と街路単位の犯罪分析への応用

許 芮溧*・井上 亮**

Graph Laplacian-Based Spatially Varying Coefficient Model and Its Application to Street-Segment Crime Analysis

Ruili XU*, Ryo INOUE**

Abstract: Traditional spatial regression often relies on Euclidean proximity and analyses aggregated to grid or administrative units, which can overlook network-driven, link-level heterogeneity. Advances in sensing technology and open data collection have produced diverse, high-resolution spatial datasets, creating a demand for regression methods that accommodate heterogeneous, multi-scale structures. Moreover, many observations are inherently based on networks—e.g., crime concentrated along street segments or passenger flows on transit links—yet methods that explicitly model link or node level processes remain underdeveloped. Therefore, in this study, we propose a framework that leverages graph Laplacian eigenvectors as multiscale basis functions to model spatially varying coefficients (SVC) on road networks. By combining eigenvector spatial filtering with Laplacian spectra, our approach enables inference at the street segment scale, and better captures localized network heterogeneity. At last, we demonstrate feasibility and empirical utility using crime calls for service data in New York City.

Keywords: 空間的異質性（spatial heterogeneity），空間可変係数（spatially varying coefficients），ネットワーク（network），グラフラプラシアン（graph Laplacian）

1. はじめに

センサ技術の発達やオープンデータ施策の進展に伴い，空間データの多様化・大規模化は著しく，空間解像度の高い詳細な空間データが入手可能になり，それに応じた分布・精度・構造の異なる幅広い時空間データのための回帰法が求められている．

空間データの基本特性である空間的異質性を考慮した解析の理論や方法は，空間統計学を中心に発展を遂げてきた．空間的異質性は空間現象の生成要因・過程が場所によって異なるという性質であり，従来の分析手法として，空間全域で連続的な変化と特定の地域で不連続的な変化の二種類の仮定が提案される．前者は，地理的カーネルを用いて地理的局所的な分析に基づく地理的加重回帰法 (Fotheringham et al., 1998) や，位置によって生成要因の影響が徐々に変化する空間構造を捉える空間可変係数モデル (Spatially Varying Coefficient, SVC) が代表モデルとして挙げられる．これらの一例として，空間フィルタリング (Eigenvector Spatial Filtering, ESF) を応用したSVC モデル(ESF-SVC) (Griffith, 2008) が存在し，空間的な近接性を記述する行列の固有ベクトルの線形和をあてはめることで空間上での連続的な変化を推定しようというものである．一方，後者は最小分析空間単位とする地域ごとに各地域に異なる係数を設定したモデルを用い，スパースモデリング手法を活用して不連続に変化する空間構造を捉える分析が検討されてきた．

空間的異質性を統計学的に検証する際，モデルの選択を問わず，空間重み行列の設定が重要である．空間重み行列は空間オブジェクト間の隣接関係や距離をもとに，「近さ」を表現するための指標である．従来，地点間のユークリッド距離を元に地理的近接性を定義した分析が行われることが多い．しかし，高解像度の地理分析を行う場合，特に都市内の社会経済現象を対象とする際は，現象が道路網などのネットワーク構造の影響を受けて生起していることを踏まえる必要性が高い．また，分析対象の現象の観測は，道路区間単位や交差点単位での交通事故件数など，リンクやノード単位に行われていることもある．また，現象の影響要因も，沿道道路環境や交差点構造などリンクやノードの性質であることも多い．これらを考えた分析を行うためには，ネットワーク構造に基づく空間分析が不可欠である．

本研究では，ネットワーク構造を表す方法としてラプラシアン行列に注目し，空間的異質性分析への応用を検討する．ラプラシアン行列は，グラフの連結の強さの定量的評価として使われ，固有ベクトルの集合はグラフ分

* 学生会員 東北大学 大学院情報科学研究科（Tohoku University）
〒980-8579 宮城県仙台市青葉区荒巻字青葉 6-6-06 E-mail：xu.ruili.r5@dc.tohoku.ac.jp
** 正会員 東北大学 大学院情報科学研究科（Tohoku University）

割のために重要な情報を提供する(斎藤, 2015)．さらに，ラプラシアン固有ベクトルは大域から局所までの変動を表す基底になり得る(Belkin & Niyogi, 2003)．したがって，ラプラシアン固有ベクトルを説明変数の係数関数の基底として用いることで，異なる空間スケールにおける係数の変動を表現でき，ESF-SVC と同様に空間的可変係数モデルを構成することが可能である．

そこで，本研究は，このラプラシアン行列の固有ベクトルを用いた空間可変係数モデル (Laplacian-SVC) を提案する．道路網などのグラフの隣接関係や重みを考慮した SVC モデルによって，「線 (edge) 単位」「街路セグメント単位」など，より微視的・局所的な空間単位での分析が可能となる．

以降，第 2 章で提案した融合モデルの実行可能性とその推定方法を示し，第 3 章では実データを用いて提案モデルの有用性を示し考察する．最後に，第 4 章において本研究の結論を示す．

２．分析手法

２.１　ESF-SVC モデル

ESF-SVC は，空間的自己相関の代表的な検定統計量である Moran's I 統計量 (MC) に基づいた空間モデリングアプローチである (Griffith, 2008)．n 地点で得られた空間データ$\boldsymbol{z}$ ($n \times 1$ベクトル) の空間相関を表す尺度 MC は，$n \times n$の空間隣接行列$\boldsymbol{C}$，単位行列$\boldsymbol{I}$を用いると，

$$MC = \frac{\boldsymbol{z'MCMz}}{\boldsymbol{z'Mz}} \tag{1}$$

と定式化できる．なお，$\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{11'}/N$，すべての要素が 1 の$n \times 1$のベクトルを$\boldsymbol{1}$と表す．$\boldsymbol{MCM}$の各固有ベクトル$\boldsymbol{e}_l$は空間隣接行列$\boldsymbol{C}$の有する空間構造を表し，$\boldsymbol{e}_l$に対応する固有値$\lambda_l$は空間的自己相関の大きさを表す．ESF-SVC は，この固有ベクトルを説明変数のパラメータβ_kに導入して，β_kが有する連続的空間的異質性を表すモデルで

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{x}_k \circ \boldsymbol{\beta}_k^{ESF} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$
$$\boldsymbol{\beta}_k^{ESF} = \beta_k \boldsymbol{1} + \boldsymbol{E\gamma}_k \tag{2}$$

で表される．ここで，"∘"は要素ごとの積，$\boldsymbol{\beta}_k^{ESF}$は$k$番目の説明変数$\boldsymbol{x}_k$の空間可変パラメータベクトルである．定数$\beta_k$と各固有ベクトルにかかるパラメータベクトル$\boldsymbol{E\gamma}_k$は，それぞれ$\boldsymbol{\beta}_k^{ESF}$の平均値と空間的変動を示す．

２.２　提案手法

ネットワークをノードとそれぞれつなぐリンクで表現し，その隣接関係を隣接行列で記述する．ノード総数をnとすると，隣接行列$\boldsymbol{A}$は，$n \times n$の正方行列で，その成分A_{ij}は，ノードi, jの隣接関係を表し，リンク$\boldsymbol{E}$で結ばれているときに 1，結ばれていないときに 0 とする．

$$A_{ij} = \begin{cases} 1, (i,j) \in \boldsymbol{E} \\ 0, otherwise \end{cases} \tag{3}$$

各ノードに接続するリンクの総数を次数と呼び，次数を対角成分とする行列を次数行列$\boldsymbol{D}$と呼ぶ．その成分D_{ii}は$D_{ii} = \sum_{j=1}^{n} A_{ij} = \sum_{j=1}^{n} A_{ji}$である．

グラフ理論におけるグラフの行列表現であるラプラシアン行列は，$L = D - A$と定義される．この行列は定義から明らかに対称行列である．また，2 次形式は

$$\boldsymbol{x}^T \boldsymbol{Lx} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} \left(x_i - x_j\right)^2 \geq 0 \tag{4}$$

と書けるため，ラプラシアン行列は半正定値であり，固有値は必ず非負で，しかも固有ベクトルはお互いに直交することが知られているため，ラプラシアン行列は以下の形でスペクトル分解できる．

$$\boldsymbol{L} = \sum_{i=1}^{n} \lambda_i \left(\boldsymbol{a}_i \boldsymbol{a}_i^T\right) \tag{5}$$

ここで，λ_iは固有値でて，$\lambda_1 = 0 \leq \lambda_2 \leq \cdots \leq \lambda_n$を満たし，ベクトル$\boldsymbol{a}_i$はそれに対応する単位の固有ベクトルである．ラプラシアン行列のスペクトル解析より，最小固有値λ_1は必ずゼロとなり，最小の正の固有値はグラフの代数的連結度 (algebraic connectivity) と呼ばれ，グラフの連結の強さの定量的評価として使われている．さらに，$\boldsymbol{Lv}_i = \lambda_i \boldsymbol{v}_i$によって得られる固有値$\lambda_i$と固有ベクトル$\boldsymbol{v}_i$はグラフ上の空間モード（周波数）を表せる．すなわち，λ_iが小さいほど，その対応する$\boldsymbol{v}_i$はグラフ上で滑らか（大域に緩やかに変化）なモードを表し，逆に，λ_iが大きいほど，$\boldsymbol{v}_i$は局所的に急変するモードを表す(Belkin & Niyogi, 2003)．

更に，Yamada (2021) は，空間的自己相関測度指標の一つである Geary's C を用い，新たな視点からグラフラプラシアン固有ベクトルを用いた空間パターン表現を議論した．Geary's C は

$$C = \frac{(n-1)}{2\boldsymbol{W}} \frac{\sum_i \sum_j w_{ij} \left(x_i - x_j\right)^2}{\sum_i (x_i - \bar{x})^2}, \boldsymbol{W} = \sum_i \sum_j w_{ij} \tag{6}$$

で表され，式(4)によりラプラシアンを用いて

$$C = \frac{(n-1)}{\boldsymbol{W}} \frac{\boldsymbol{x}^T \boldsymbol{Lx}}{\boldsymbol{x}^T \boldsymbol{Mx}} \tag{7}$$

と書き換えられる．すなわち，Geary's C とラプラシアン固有分解を結びつけると，ラプラシアン固有値は空間的自己相関のスケールを解釈可能となり，ESF へ拡張できる．そこで，ESF-SVC に基づき，グラフ構造を考慮できるラプラシアン行列に基づいた SVC モデルを提案する．

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{x}_k \circ (\beta_k \boldsymbol{1} + \boldsymbol{L\gamma}_k) + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}) \tag{8}$$

ここで，注意すべき点は，説明変数$\boldsymbol{x}_k$とラプラシアン固有ベクトル$\boldsymbol{L}$との交互項は次元が大きくなりがちで，そのままモデルに組み込むと過学習や計算負荷の増大を招きやすいことである．この問題に対処するため，Generalized Lasso に基づくl_1型の正則化を導入し，交互項の選択を同時に行う必要がある．

３．ニューヨーク市の犯罪通報データへの適用

本章は，ニューヨーク市内の犯罪多発地域を対象に，提案手法を用いて街路区間単位で犯罪通報件数の空間的異質性を分析し，提案手法の有効性を確認する．

犯罪は都市内で特定の場所に偏在する傾向があることが数々の研究で指摘されている．

例えば Weisburd (2015) は，米国の規模の異なる数都市で，街路セグメント単位で犯罪を集計し，都市規模によらず，全犯罪の 50%が 4%程度の街路区間に，25%が 1.5%程度の街路区間に偏在することを示した．

また，近隣物理環境や社会環境，さらには街路の構造的特性が犯罪分布に影響を及ぼす可能性が示唆されている．例えば Gill et al. (2017) は，小規模都市ほど犯罪集中度が高いことを観察した．小規模都市では郊外型の道路構造が多く，大半の道路が一戸建て住宅に面している一方，少数の道路に大型集合住宅が建っており，人口が集中し潜在的な犯罪ターゲットも多いことが影響している可能性があるとしている．また Park (2019) は，平均街路区間長が長く，かつ人口密度が低い地域ほど，犯罪が特定の場所に集中しやすいことを明らかにした．

しかし，都市内で地域や街路ごとに犯罪集中の程度が異なるメカニズムについての理解はまだ十分とは言えない．Weisburd et al. (2024) は，犯罪集中の原因解明が治安政策や犯罪予防の改善に不可欠であることを強調している．例えば，土地利用特性が犯罪集中に影響を与える場合には，土地利用の改変や用途規制を通じて犯罪集中度を緩和し，犯罪発生件数の多い道路の数を減らすことが期待される．そうした施策によって，限られた警察資源をより効果的に配分し，犯罪発生率の高い道路に対する重点的な介入を実現できる可能性があるとしている．従って，犯罪が集中する道路の土地利用特性や道路構造を把握することは，市当局が犯罪防止対策を改善するための有効な手段となり得る．

それらの既往研究を踏まえ，従来手法より柔軟に街路セグメント単位の空間的異質性を捉えることができる提案モデルを用いて，街路構造特性 (幅員や長さ) や近隣土地の土地利用特性が犯罪集中に与える影響を検証する．

３．１　データ

本研究では，ニューヨーク市の calls for service（通報データ）を対象とし，解析単位を街路区間 (street segment) とする．通報データは「通報された全ての事象（確認の有無を問わない)」を包含するため，逮捕記録や事件報告より包括的である．道路ネットワークは DCM (Digital City Map) の道路中心線データを用いる．散歩道の(path)データには幅員などの情報が欠落しているため，道路ネットワークの構築には通り(street)タイプのみを使用した．各街路上の通報件数（集計値）を被説明変数とし，「general crime，violent crime，property crime」をカテゴリ別に集計・分析する (Weisburd at el., 2024)．道路属性（街路区間長と街路幅員）や土地利用当該街路両側に接する土地利用タイプのカウント値（一戸建て・集合住宅・商業用地・空き地）を説明変数とする．また，通報記述に“ambulance”, “train”を含む救急や鉄道関連の記録は非犯罪的通報とみなし分析対象外とした．交差点で発生した通報は，交差点固有の発生機構を持つため街路区間上の事象と性質が異なると判断して除外した．

解析対象を「辺」に置くために，元の道路グラフ $G=(V,E)$ をライン・グラフ$L(G)$に変換した．変換後は元の各辺が$L(G)$の頂点に対応し，元の節点で接続する辺同士が$L(G)$上で隣接する関係となる．これにより「辺上データ」を頂点データとして扱い，ラプラシアン行列を構築し，得られた固有ベクトルを提案モデルに導入して分析する．なお，説明変数は，各パラメータに対する正則化の影響を同一にするため，説明変数の値を標準化した．

３．２　分析結果

本分析では，ニューヨーク市の中で犯罪通報数が最も多い管轄区であるブロンクス区の第 40 管轄区を対象地域として選定した(図 1). 2025 年 1 月から 2 月までの 2 か月間に，本地域では合計 74,499 件の犯罪通報が記録され，そのうち 2,499 件が財産犯であった（図 2）．対象地域の街路区間数は 945 である．ここでは窃盗・強盗など財産犯に関する分析の推定結果のみを示す．図 3 の上段は街路属性に関するパラメータ推定結果，下段は土地利用特性に関するパラメータ推定結果を示す．

街路幅員のパラメータは全域で正かつ一定に推定され，財産犯発生に対して一様な影響を与えており，空間的異

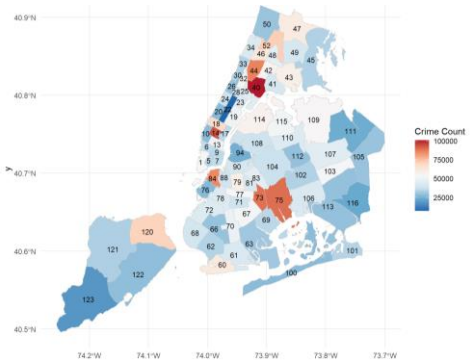


図 1　2025 年ニューヨーク市警察署管轄区ごとの犯罪通報件数



図 2　街路ごとの総通報件数(左)と財産犯(右)の分布

質性は検出されなかった．街路区間長の影響は場所によって正負の効果が現れ，空間的異質性を有することが確認された．特定の街路区間では区間長が延びたら，路上の窃盗や強盗などの財産犯発生が増加する傾向にある．一方，集合住宅件数の影響は正に推定され，集合住宅が増加すると財産犯が増加する傾向が確認された．集合住宅の集積が犯罪発生に影響を与える可能性があるという既往研究の結論を裏付けているが，集合住宅の規模や居住密度が分析に入っていないため，潜在的ターゲットの集中が犯罪発生を助長するのかは判断できない．一戸建て住宅件数のパラメータは負の定数に推定され，財産犯の増加には寄与しないことを示唆する．また，商店数が街路上の財産犯発生に与える影響は地域によって異なり，特に東部の一部街路においては，商店数の増加が財産犯発生の可能性を高める傾向が観察された．商業活動の集中に伴う人の集積や魅力的なターゲットの存在が影響している可能性がある．空き地は，北部および東部に多く存在するが，これらが増加すると一部の街路で財産犯発生が有意に増加する傾向が見られた．

４．おわりに

本研究は，ネットワーク構造の空間的異質性を捉え，より微視的・局所的な空間単位での分析が可能となる手法として，ラプラシアン行列の固有ベクトルを用いた空間可変係数モデルを提案した．

提案手法の有効性を確認するため，ニューヨーク市の犯罪多発地域内の犯罪通報データに適用した．分析結果から，既往研究の結論と一致して，街路区間長と街路両側の集合住宅数の増加が犯罪発生の増加に寄与することが確認できた．さらに，土地利用が街路ごとで犯罪発生への影響は，街路区間ごとで異なることが分かった．

今後は，犯罪のような係数データに適したポアソンモデルと提案モデルを統合し，発生頻度のモデリング精度を高める予定である．併せて，大規模ネットワークへの適用性と計算効率の改善を図るとともに，ユークリッド距離に基づく従来手法との比較を行い，ネットワーク解析の実用的有用性と限界を改めて検証する．

参考文献

斎藤直樹. (2015). グラフ・ネットワーク上での応用調和解析. 応用数理, 25(3), 102-111.

Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, *15*(6), 1373-1396.

Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 30(11), 1905-1927.

Gill, C., Wooditch, A., & Weisburd, D. (2017). Testing the “law of crime concentration at place” in a suburban setting: Implications for research and practice. *Journal of Quantitative Criminology*, 33(3), 519-545.

Griffith, D. A. (2008). Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A*, 40(11), 2751-2769.

Park, S. (2019). *Examining the “Law of Crime Concentrations” across Multiple Jurisdictions* (Doctoral dissertation, George Mason University).

Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology*, 53(2), 133-157.

Weisburd, D., Zastrow, T., Kuen, K., & Andresen, M. A. (2024). Crime concentrations at micro places: A review of the evidence. *Aggression and Violent Behavior*, 78, 101979.

Yamada, H. (2021). Geary’s C and spectral graph theory. *Mathematics*, 9(19), 2465.

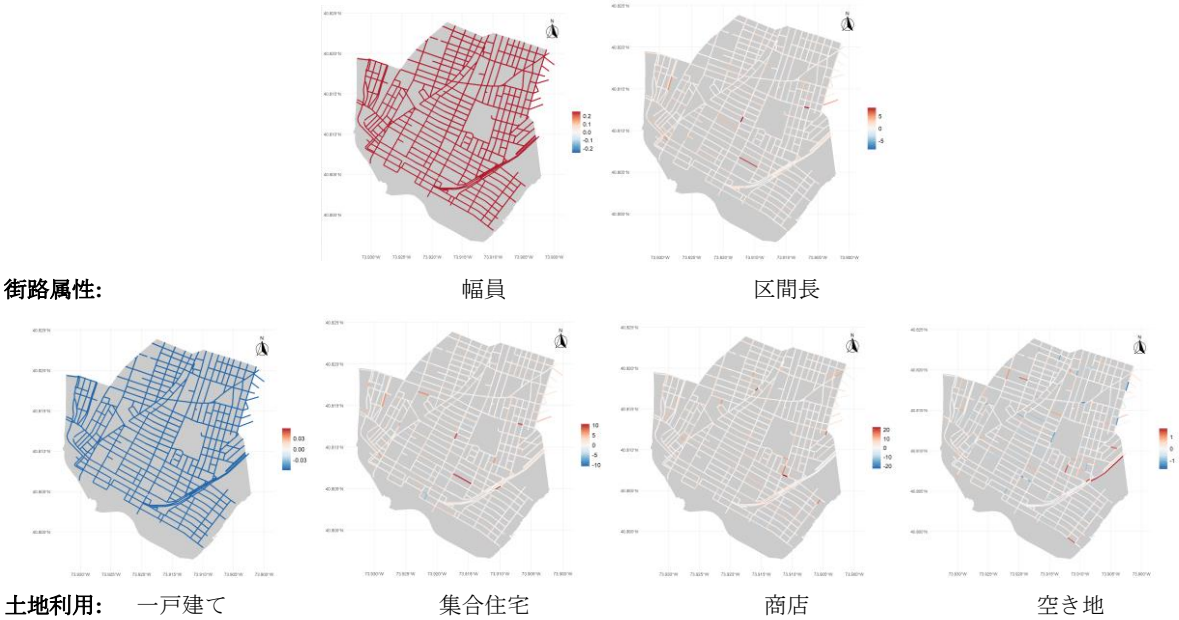


図３　財産に関する犯罪に対するパラメータの推定結果

連続的な空間的異質性を考慮した空間相互作用モデルによる地域間関係分析

安藤 夏樹*・井上 亮・彭 湛****

Spatial interaction model with continuous spatial heterogeneity to analyze interregional relationships

Natsuki Ando*, Ryo Inoue**, Zhan Peng**

Abstract: Continuous spatial heterogeneity is that generation process of spatial phenomena changes gradually in space. While various approaches have been developed to estimate spatially varying coefficients (SVC), their application to spatial interaction models has been limited. This study constructs a gravity model with spatially varying coefficients using the eigenvector spatial filtering-based SVC (ESF-SVC) method and evaluates its effectiveness for analyzing interregional relationships. An application to commuting and schooling trips from the Tokyo Metropolitan Person Trip Survey demonstrates that the model can capture heterogeneity in trip generation and attraction that depends on origins and destinations. The results confirm the potential of incorporating SVCs into interaction models, while also highlighting an identification issue: distance decay cannot be estimated independently from the coefficients representing trip generation and attraction.

Keywords: 空間的異質性（Spatial heterogeneity），ESF-SVC，空間相互作用モデル（Spatial interaction model）

１．はじめに

近年，情報化社会の進展により，様々な空間現象を記録した空間データが大量に蓄積されている．その一種である，人や物の地域間の移動を記録したデータも入手可能性が向上している．これらの移動量データに基づいて，地域間関係の強弱を表す空間相互作用を分析する研究は，社会構造の理解や都市計画において重要である．

空間相互作用分析の最も基本的なモデルは重力モデルである．t_{ij}を地域 ij 間の移動量，P_iを発地 i の発生量の多寡を表す指標，Q_jを着地 j の集中量の多寡を表す指標，d_{ij}を地域 ij 間の距離指標，k, α, β, γ を係数とすると，

$$t_{ij} = kP_i^{\alpha} Q_j^{\beta} d_{ij}^{-\gamma} \quad (1)$$

で表される．

重力モデルでは通常，全地域で共通の係数を推定する．しかし，もし郊外ベッドタウンの地域からの通勤移動の発生量が，同一人口規模の他の地域の通勤移動の発生量よりも多いなら，発生量指標を人口としたモデルでは，ベッドタウンの α は他の地域よりも大きいと考えるのが適切であろう．このように地点・地域によって被説明変数と説明変数の関係が異なる様子を表現するモデルを空間可変係数 (Spatially Varying Coefficient: SVC) モデルといい，そこではモデルで表されている空間現象の生成過程・要因が，地点・地域によって異なる「空間的異質性」を仮定している．

移動量には，以下の 2 種類の空間的異質性があると考えられる．1 つは距離が移動量に与える影響が発着地の組によって異なるという性質である．これは，歴史的・文化的な結び付きなどによって移動量の距離減衰効果が発着地の組毎に異なる場合を表す．もう 1 つは発地や着地の属性が移動量に与える影響が異なるという性質である．これは同規模の人口を持つ地域であっても，発生量や集中量の大小がある場合を表す．

本稿では，地点や地域の場所によって係数が徐々に変化する連続的な空間的異質性を仮定した分析に焦点を絞る．連続的な空間的異質性分析の手法は，Geographically Weighted Regression (GWR) (Brunsdon et al., 1996) や Eigenvector Spatial Filtering based Spatially Varying Coefficients model (ESF-SVC) (Griffith, 2008) が代表的である．これらの SVC モデルは 1 地点・地域に紐づけられた空間データを対象とするが，2 地点・地域間の関係性を表すデータである移動量の分析へ拡張する試みもなされてきた．重力モデルに GWR を組み合わせた研究は多く存在する(例えば, Nakaya, 2001; Kordi et al., 2016) が，ESF-SVC を組み合わせた研究は未だ見られない．

GWR は，全ての係数で同じ空間スケールの空間的異質性構造しか推定できず，データ件数が多いと推定に計算を要する課題を有する．GWR の拡張として係数毎に異なるスケールの空間的異質性構造を推定できる Multiscale GWR (MGWR) (Fotheringham et al., 2017)も提案されてい

* 学生会員 東北大学 大学院情報科学研究科（Tohoku University）
〒980-8579 宮城県仙台市青葉区荒巻字青葉 6-6-06 E-mail:ando.natsuki.s2@dc.tohoku.ac.jp
** 正会員 東北大学 大学院情報科学研究科（Tohoku University）

るが，更に推定に要する計算量が増加し，実行可能性に制約がある．一方，ESF-SVC は，係数毎に異なるスケールの空間的異質性の構造を推定できる利点があり，データ件数が大きくても推定計算の実行可能性は MGWR より大きい．空間相互作用モデルのデータ件数が，地域数の 2 乗であることを考えると，ESF-SVC に基づく重力モデルの検討は空間相互作用の空間的異質性分析の可能性を広げる上で有用だと考える．

そこで，本研究は移動量が有する 2 種類の連続的な空間的異質性を表現する方法として，ESF-SVC を基盤に，重力モデルに SVC を導入するモデルを提案する．地域 i に固有の発地属性係数α_i，着地属性係数β_iを，また地域 ij 間の距離減衰係数γ_{ij}を発地 i の影響と着地 j の影響の和で表現し，発着地の組毎に異なる係数として設定する．東京都市圏パーソントリップ調査の移動データを用いて提案モデルを推定し，提案モデルの実行可能性を評価するとともに，推定結果に基づく地域特徴や地域間関係の把握に対する有用性を検証する．

２．連続的な空間的異質性の分析手法

２.１　GWR

地点 i における GWR のモデルは，y_iを被説明変数，$\beta_{0,i}$を切片，$x_{k,i}$を k 番目の説明変数の値，$\beta_{k,i}$を$x_{k,i}$に対する係数とすると，下式で表される．

$$y_i = \beta_{0,i} + \sum_{k=1}^{K} x_{k,i}\,\beta_{k,i} + \epsilon_i \qquad (2)$$

推定は，地点iに近い観測点に対しより大きな重みを設定した重み行列 $\mathbf{W}_i$ を用いた，重みつき最小二乗法で行う．$\mathbf{y}$を被説明変数ベクトル，$\mathbf{X}$を説明変数行列とすると，推定式は

$$\beta_{ik} = (\mathbf{X}^{\mathrm{T}}\mathbf{W}_i\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}_i\mathbf{y} \qquad (3)$$

と表される．$\mathbf{W}_i$ は非対角要素が 0，対角要素$w_{i,jj}$が地域 ij 間の距離d_{ij}を距離減衰関数に代入した値で与えられる行列で，距離減衰関数はガウス型の関数

$$w_{i,jj} = \exp\left(-0.5\left(\frac{d_{ij}}{b}\right)^2\right) \qquad (4)$$

が用いられることが多い．なお，バンド幅 b は近傍範囲を定めるハイパーパラメータで，この値は交叉検定や修正赤池情報量規準を用いて決定する．バンド幅決定では，その候補に対して推定を繰り返す必要があり，地点毎の推定と併せて多くの計算を要する．

GWR の拡張手法である MGWR は，全ての係数で共通のバンド幅を用いる GWR に対し，係数毎にバンド幅を決定し，空間的異質性のスケールが異なることを表現できる．しかし，係数毎にバンド幅を変えた推定を繰り返す必要があるため，GWR 以上に計算を要する．

さて，GWR や MGWR を重力モデルに導入した研究として，Spatially Weighted Interaction Models (SWIM) (Kordi et al., 2016) や Multiscale flow-focused GWR (MFGWR) (Zhang et al., 2019) が挙げられる．SWIM には重み$w_{i,jj}$の定義に用いる距離を「発地間距離」「着地間距離」「発着地対の距離」とした 3 種類のモデルが提案されており，GWR で推定する．MFGWR は「発着地対の距離」を定義して重みを与え，MGWR で推定する．これらは，GWR や MGWR と同様に推定に要する計算量に課題がある．

２.２　ESF-SVC

ESF-SVC は Eigenvector Spatial Filtering (ESF) (Griffith, 1996) を基にしたモデルである．n×n の空間重み行列$\mathbf{C}$と Centering Matrix $\mathbf{M} = (\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ で構成される行列 $\mathbf{MCM}$ の固有ベクトルを用いて SVC を表す．n は地域数，$\mathbf{I}$は n×n の単位行列，$\mathbf{1}$は全要素が 1 の n×1 ベクトルである．$\mathbf{MCM}$ の固有ベクトルは空間スケールの異なる空間相関構造を表し，値が大きい固有値に対応する固有ベクトルは大域的な構造，値が小さい固有値に対応する固有ベクトルは局所的な構造を表現する．

ESF-SVC のモデルは$\mathbf{x}_k$を k 番目の説明変数ベクトル，$\circ$を要素積，$\beta^{(k)}$を$\mathbf{x}_k$に対する非可変係数，$\mathbf{E}^{(k)}$を $\mathbf{MCM}$ の固有ベクトルから$\mathbf{x}_k$に対して選択されたL_k個の固有ベクトルで構成される n×L_kの行列，$\boldsymbol{\gamma}^{(k)}$を固有ベクトルに対する$L_k$×1 の係数ベクトルとすると，下式で表される．

$$\mathbf{y} = \sum_{k=0}^{K} \mathbf{x}_k \circ \left(\beta^{(k)}\mathbf{1} + \mathbf{E}^{(k)}\boldsymbol{\gamma}^{(k)}\right) + \boldsymbol{\epsilon} \qquad (5)$$

SVC は全域共通の$\beta^{(k)}\mathbf{1}$と空間パターンを表す$\mathbf{E}^{(k)}\boldsymbol{\gamma}^{(k)}$の和で表現される．説明変数毎に固有ベクトルの行列$\mathbf{E}^{(k)}$が異なるため，各係数は異なるスケールの空間構造を表現する．説明変数に対する固有ベクトルの選択方法として，Lasso などが提案されている (Seya et al., 2014).

地点 i における ESF-SVC のモデルは，$\mathbf{row}_i\left(\mathbf{E}^{(k)}\right)$を$\mathbf{E}^{(k)}$の i 行目ベクトルとすると

$$y_i = \sum_{k=0}^{K} x_{k,i}\left(\beta^{(k)} + \mathbf{row}_i\left(\mathbf{E}^{(k)}\right)\boldsymbol{\gamma}^{(k)}\right) + \epsilon_i \qquad (6)$$

と表せ，また式(5)は式(7)と書き直すことができる．

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{k=0}^{K}\sum_{l_k=1}^{L_k}\left(\mathbf{x}_k \circ \mathbf{e}_{l_k}^{(k)}\right)\boldsymbol{\gamma}_{l_k}^{(k)} + \boldsymbol{\varepsilon} \qquad (7)$$

ここで，$\boldsymbol{\beta}$は非可変係数ベクトル，$\mathbf{e}_{l_k}^{(k)}$は$\mathbf{E}^{(k)}$の l_k 列目ベクトル，$\boldsymbol{\gamma}_{l_k}^{(k)}$は$\mathbf{e}_{l_k}^{(k)}$に対する係数ベクトルである．

ESF-SVC は GWR や MGWR と比較して，利点が 2 点ある．1 点目は計算量の少なさである．GWR は，地点毎の推定とバンド幅決定に多くの計算を要するのに対し，ESF-SVC は一回の最小二乗推定で空間スケールも含めて推定ができる．2 点目は，係数毎に異なる空間スケールの空間的異質性の構造を推定できることである．MGWR も可能だが，係数毎に異なるバンド幅を探索する推定過程

に，GWR 以上の計算を要する．一方，ESF-SVC は説明変数毎に与えられた固有ベクトルの係数の推定を通して，係数毎に異なる空間構造を表現できる．

このように，ESF-SVC は計算量の少なさと異なる空間構造表現の容易さという点で GWR より優れている．そこで本研究では ESF-SVC を基盤に SVC を導入した重力モデルを提案する．提案手法では，重力モデルの 3 種類の係数 $\alpha \cdot \beta \cdot \gamma$ を SVC とし，発地属性・着地属性・距離が移動量に与える影響の空間的異質性を区別して表現することを目指す．

3. 提案手法

本研究は 3 種類のモデルを提案する．モデル 1 は発地属性係数 α, 着地属性係数 β, 距離減衰係数 γ 全てが SVC, モデル 2 は α，β が SVC，モデル 3 は γ が SVC である．モデル 1 は以下の式で表される．

$$t_{ij} = kP_i^{\alpha_i} Q_j^{\beta_j} d_{ij}^{-\gamma_{ij}}$$

$$\alpha_i = \alpha_{const} + \mathbf{row}_i(\mathbf{E})\boldsymbol{\eta}$$
$$\beta_j = \beta_{const} + \mathbf{row}_j(\mathbf{E})\boldsymbol{\mu}$$
$$\gamma_{ij} = \gamma_{const} + \mathbf{row}_i(\mathbf{E})\boldsymbol{\nu} + \mathbf{row}_j(\mathbf{E})\boldsymbol{\xi} \quad (8)$$

ここで，α_{const}，β_{const}，γ_{const}は非可変係数，$\boldsymbol{\eta}$，$\boldsymbol{\mu}$，$\boldsymbol{\nu}$，$\boldsymbol{\xi}$は固有ベクトルに対する係数ベクトルである．γ_{ij}は非可変係数，発地 i の項$\mathbf{row}_i(\mathbf{E})\boldsymbol{\nu}$，着地 j の項$\mathbf{row}_j(\mathbf{E})\boldsymbol{\xi}$の和で表現する．式(8)を対数化した式(9)の線形回帰モデルを，通常最小二乗法で推定する．

$$\log t_{ij} = \log P_i \cdot \alpha_{const} + \log P_i \cdot \mathbf{row}_i(\mathbf{E})\boldsymbol{\eta} + \log Q_j \cdot \beta_{const} + \log Q_j \cdot \mathbf{row}_j(\mathbf{E})\boldsymbol{\mu} - (\log d_{ij})\gamma_{const} - \log d_{ij} \cdot \mathbf{row}_i(\mathbf{E})\boldsymbol{\nu} - \log d_{ij} \cdot \mathbf{row}_j(\mathbf{E})\boldsymbol{\xi} \quad (9)$$

このモデルの説明変数は，$\log P_i$ ・ $\log P_i \cdot \mathbf{row}_i(\mathbf{E})$ ・ $\log Q_j$ ・ $\log Q_j \cdot \mathbf{row}_j(\mathbf{E})$ ・ $\log d_{ij}$ ・ $\log d_{ij} \cdot \mathbf{row}_i(\mathbf{E})$ ・ $\log d_{ij} \cdot \mathbf{row}_j(\mathbf{E})$である．

4. 通勤・通学目的の移動量データへの適用を通した提案手法の検証

提案手法の適用可能性を，平成 30 年実施・第 6 回東京都市圏パーソントリップ調査の，計画基本ゾーン間の通勤・通学目的移動量を用いて検証する．対象範囲は同調査の計画基本ゾーン全 615 地域で，東京都・神奈川県・埼玉県・千葉県の全域と茨城県南部を含む．移動量 t_{ij}は「目的種類：自宅–通学」と「目的種類：自宅–通勤」の移動量を集計した．発生量指標P_iは，同調査の基礎集計項目「居住地ゾーン別性別年齢階層別人口」から，性別・年齢階層を合計した居住地ゾーン人口を用いる．集中量指標Q_jは，令和 3 年度経済センサス–活動調査の全事業所数を用いた．距離d_{ij}は計画基本ゾーンの重心間距離により算出した．本研究では，同ゾーンを発着する内々移動と，居住地人口が 0 の 4 ゾーンは分析対象外とする．

4.1 説明変数特性と提案モデルの比較

適用の結果，説明変数 $\log P_i \cdot \mathbf{row}_i(\mathbf{E})$ と $\log d_{ij} \cdot \mathbf{row}_i(\mathbf{E})$間と，$\log Q_j \cdot \mathbf{row}_j(\mathbf{E})$と$\log d_{ij} \cdot \mathbf{row}_j(\mathbf{E})$間に強い正の相関が確認された．通常，ESF-SVC モデルでは説明変数に相関は発生しないため，提案モデルの説明変数の特徴が原因と考えられる．

ここで，固有ベクトルの要素積算出前の変数$\log P_i$と$\log d_{ij}$の関係に注目する（図 1）．$\log P_i$は発地毎に値が異なるのに対し，$\log d_{ij}$は発着地ペア毎に値が異なる．このため 1 つの$\log P_i$に対し複数の$\log d_{ij}$が対応し，更に$\log d_{ij}$の平均は発地に依らずほぼ同水準の値となる．$\log P_i$と$\log d_{ij}$の相関係数は 0.01 と無相関で，$\log Q_j$と$\log d_{ij}$の相関も同様に無相関である．しかし，$\mathbf{row}_i(\mathbf{E})$との要素積を取った $\log P_i \cdot \mathbf{row}_i(\mathbf{E})$ と$\log d_{ij} \cdot \mathbf{row}_j(\mathbf{E})$の相関係数は 0.97 となり，強い正の相関が生じる（図 2）．平均が同水準の$\log d_{ij}$に同じ固有ベクトルの要素をかけることが説明変数の相関の原因である．α, β, γ 全てが SVC のモデル 1 は多重共線性のため，係数を正しく推定できない．

また，モデル 2 の説明変数の$\log P_i \cdot \mathbf{row}_i(\mathbf{E})$とモデル 3 の説明変数の$\log d_{ij} \cdot \mathbf{row}_i(\mathbf{E})$間，モデル 2 の$\log Q_j \cdot \mathbf{row}_j(\mathbf{E})$とモデル 3 の$\log d_{ij} \cdot \mathbf{row}_j(\mathbf{E})$間にも強い正の相関が存在する．結果としてモデル 2・3 によって推定される係数の空間構造はほぼ同一であると言え，提案モデルでは発地・着地属性が移動量に与える影響の空間的異質性と，距離が移動量に与える影響の空間的異質性を区別できないことが明らかとなった．

赤池情報量規準や自由度調整済み決定係数による比較の結果，僅かながらモデル 2 の方が適合度や説明力に優れていた．そこで，次節の検証ではモデル 2 を用いる．

4.2 地域分析への適用可能性

発地属性係数α_i，着地属性係数β_jそれぞれの空間分布を図示し，その解釈可能性を考察する．α_i，β_jにはそれぞれ 2 種類の解釈が考えられる．1 種類目は発地・着地属性が移動量に与える影響の空間的異質性を表すとの解釈である，$\alpha_i \cdot \beta_j$が大きいと，人口や事業所数の割に発生量や集中量が多いと解釈される．2 種類目は距離が移動量に与える影響の空間的異質性を表すとの解釈である．$\alpha_i \cdot \beta_j$が大きいと，遠距離移動の割合が高いと解釈される．

図 3 にα_iの空間分布を示す．値が小さい 23 区などの都心とその周縁地域は，人口の割に発生量が少ない，または発生量内の遠距離移動割合が低いと解釈される．値が大きい郊外は，人口の割に発生量が多い，または遠距離移動割合が高いと解釈される．通勤移動は郊外から都心への移動が多く，また，発生量内の遠距離移動割合は郊外が高く，都心が低い．α_iの分布は移動の特徴と合致し，解釈可能性が高い．

次に，図 4 にβ_jの空間分布を示す．値が小さい都心の周縁地域は，集中量内の遠距離移動割合が低い，または

事業所数の割に集中量が少ないと解釈される．値が大きい都心と郊外は，遠距離移動割合が高い，または事業所数の割に集中量が多いと解釈される．通勤移動では，集中量内の遠距離移動割合は都心で高く，周縁地域や郊外は低いであろう．この実態を踏まえて考えると，都心やその周縁では解釈可能性が高いが，郊外は解釈が難しい．郊外はゾーン面積が大きくゾーン間距離が長いため，遠距離移動割合が大きく評価され，β_jが大きく推定された可能性が考えられる．以上より，β_jの推定結果は，着地属性と距離の相反する効果が影響した可能性が考えられ，推定結果の解釈が難しくなった可能性が示唆される．

5. おわりに

本研究では，連続的な空間的異質性の分析手法である ESF-SVC を基に，空間可変係数を導入した重力モデルを提案した．東京都市圏パーソントリップ調査の通勤・通学移動データを用いて提案モデルを推定した結果，移動データの生成過程が有する空間的異質性を表現できる可能性を確認した．ただし，提案モデルでは，距離が移動量に与える影響の空間的異質性と，発地属性や着地属性が移動量に与える影響の空間的異質性を分離して推定できないことが明らかとなった．

今後，発地・着地間の相互作用を表せるよう距離減衰係数の空間的異質性の新たなモデル化を検討する．また，本研究では重力モデルに GWR を導入した既往手法と比較しておらず，それらとの比較分析が必要である．

参考文献

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28(4), 281–298.

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, 107(6), 1247–1265.

Griffith, D. A. (1996). Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. *Canadian Geographer/Le Géographe canadien*, 40(4), 351–367.

Griffith, D. A. (2008). Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A*, 40(11), 2751–2769.

Kordi, M., & Fotheringham, A. S. (2016). Spatially weighted interaction models (SWIM). *Annals of the American Association of Geographers*, 106(5), 990–1012.

Nakaya, T. (2001) Local spatial interaction modelling based on the geographically weighted regression approach. *Geo Journal*, 53, 347–358.

Seya, H., Murakami, D., Tsutsumi, M., & Yamagata, Y. (2015). Application of LASSO to the eigenvector selection problem in eigenvector-based spatial filtering. *Geographical analysis*, 47(3), 284–299.

Zhang, L., Cheng, J., Jin, C., & Zhou, H. (2019). A multiscale flow-focused geographically weighted regression modelling approach and its application for transport flows on expressways. *Applied Sciences*, 9(21), 4673.

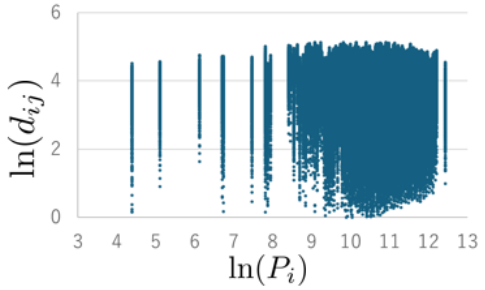


図 1　$\log P_i$と$\log d_{ij}$の値の散布図

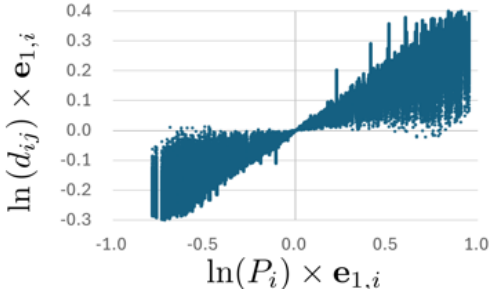


図 2　第 1 固有ベクトルと要素積をとった値の散布図

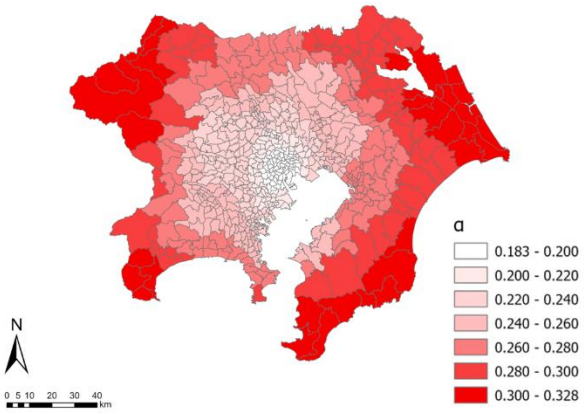


図 3　発地属性係数α_iの空間分布

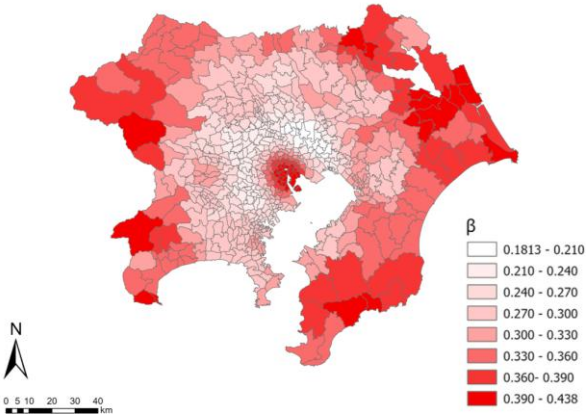


図 4　着地属性係数β_jの空間分布

LLM ベースと従来型ジオコーディングの精度比較：日本の大・中・小規模地域における評価

卢　诗琪*・龐　岩博・関本　義秀**

Comparing the Accuracy of LLM-based and Geocoding with Traditional Methods: Evaluation across Large, Medium, Small Regions in Japan

Shiqi Lu*, Yanbo Pang**, Yoshihide SEKIMOTO

Abstract: In recent years, large language models (LLMs) have shown promise for geocoding, i.e., deriving geographic coordinates from textual place descriptions. While traditional methods, such as the Google Maps Geocoding API, achieve high accuracy for structured addresses, LLMs can better interpret unstructured data, though their geocoding accuracy remains uncertain. In this study, we systematically compare LLM-based geocoding with traditional methods using Zenrin Telepoint data as reference coordinates. Across representative large, medium, and small regions in Japan, LLMs show competitive performance, achieving mean errors of 0.94 km, 2.03 km, and 4.85 km, respectively, with an overall average processing time of 1.65 seconds. These results demonstrate that LLMs can offer comparable or improved geocoding accuracy relative to baseline methods, while maintaining efficiency.

Keywords: 大規模言語モデル（Large Language Models, LLMs），ジオコーディング（Geocoding），位置情報解析（Geographic information analysis），精度比較（Accuracy comparison），日本の地域（Japanese regions）

１．はじめに

近年，LLM（大規模言語モデル）の進展により，地理空間情報の自動生成・理解を前提とする応用（デジタルツイン，都市 OS，政策立案支援など）が急速に広がっている．しかし，その信頼性評価，とりわけ「位置が正しいか」を問う基礎能力の検証は十分とは言えない．高精度な空間参照が担保されなければ，意思決定への適用はリスクとなる．本研究は，LLM が地理空間情報をどこまで理解し，実座標を正しく復元できるのかを検証する最初の一歩として，ジオコーディングに焦点を当て，従来型手法と比較しながら精度・頑健性・処理時間を体系的に評価する．

１.１　研究の目的

本研究の目的は, LLM を用いたジオコーディング手法と従来型手法を，日本の大・中・小規模地域における多様な住所表現に適用し，その位置推定精度，頑健性，および処理時間を体系的に評価・比較することである．本研究により，各手法の精度特性と限界を明らかにするとともに, LLM の地理空間情報理解能力を実座標復元の観点から検証し，従来型手法との補完的関係や実務応用上の可能性を示すことを目的とする．

１.２　先行研究

地名解決（トポニム解決）とジオコーディングに関する先行研究では，主に従来型手法と大規模言語モデル（LLM）を活用した手法の発展が確認される．

従来の地名解決手法では，ルールベースや統計的手法が広く用いられてきた．例えば，ルールベースのアプローチでは住所体系の階層構造を利用して文字列マッチングを行う方法があり (Zandbergen, 2009)，統計的手法では隠れマルコフモデル（HMM）や条件付き確率場（CRF）が地名のシーケンスタギングに応用されている．近年では, 双方向 LSTM-CRF モデルが高い性能を発揮し, 住所要素の抽出精度を向上させたことが報告されている）(Zhengcong Yin, 2023).

一方, Transformer ベースのモデル (BERT, roBERTa など）や LLM（GPT-3 など）の応用が注目されている．GENRE や BLINK といったエンティティリンキングモデルは, Wikipedia などの大規模知識ベースを活用して地名の曖昧性解消を試みているが，非構造化テキストや誤入力に対する頑健性に課題が残る．また，GPT-3 を含む LLM は自然言語理解能力に優れるものの, 内部に地理データベースを持たないため，高精度な座標出力が難しいことが明らかにされている (Xuke Hu, 2024).

近年の研究では, LLM と従来型手法の融合が提案されている．Hu（2024）は，軽量なオープンソース LLM（Mistral，Llama2 など）を微調整し，地名の明確な参照先（市区町村，国など）を推定した後，GeoNames や ArcGIS などのジオコーダーで座標を取得するハイブリッド手法を開発．これにより，従来の投票システムや GENRE よりも高精度（Accuracy@161km で 0.91）かつ計算効率に優れる結果を示した．

*　学生会員　東京大学大学院新領域創成科学研究科（The University of Tokyo）
〒153-0041　東京都目黒区駒場 4-5-29　E-mail：6056775605@edu.k.u-tokyo.ac.jp
**　正会員　東京大学空間情報科学研究センター（The University of Tokyo）

２．検証対象と研究方法

本研究では，まず川崎市をテスト地域として選定し，100 件の実住所サンプルを抽出して，現行の GPT モデルのジオコーディング能力を検証した．ジオコーディング過程における課題を解決し精度を向上させるため，適切なプロンプトを設計し，処理時間も記録した．その後，最も性能の良いプロンプトを用いて，日本国内の大規模・中規模・小規模の各 5 地域（合計 15 地域）における GPT モデルのジオコーディング能力を評価した．

２.１　使用データとベースライン設定

本研究では，日本国内の大・中・小の規模別に分類された各 5 地域，合計 15 地域を対象とする（表 1 参照）．各地域からランダムに 100 件の実住所サンプルを抽出し，区・町丁目・番地レベルまでの詳細情報を含むデータを収集した．

表 1　対象地域表

地域規模	都道府県・市町村名
大規模	東京都，大阪府，横浜市，札幌市（北海道），福岡市（福岡県）
中規模	金沢市（石川県），松山市（愛媛県），熊本市（熊本県），高松市（香川県），盛岡市（岩手県）
小規模	白川村（岐阜県），与那国町（沖縄県），大間町（青森県），小豆島町（香川県），鶴居村（北海道）

本研究における正解座標は，ゼンリンテレポイントデータを用いた．これは全国の住所を号・番地レベルまで高精度に位置付け可能な座標データである．

さらに，従来型ジオコーディング手法との比較のため，Google Maps API による住所変換結果をベースラインとして設定し，LLM 出力との比較を行った．

２.２　使用モデル

本研究では，Azure OpenAI の GPT-4，GPT-4o，および GPT-O1 / GPT-O1 Mini を対象に，地理座標推定能力を検証した．初期検証の結果，GPT-4o は精度面で最も優れていたが，異なる住所に対して同一座標を出力する重複問題が確認された．一方，GPT-O1 および GPT-O1 Mini では，この重複問題は観察されなかった．

そこで，GPT-4o に対してプロンプト設計の改善を行い，重複問題の解消と精度維持を確認した．最終的に，本研究の正式テスト段階では，改善済みプロンプトを適用した GPT-4o を使用した．

２.３　プロンプト設計

本研究では，LLM に対するジオコーディングの指示文（プロンプト）を設計し，精度向上および出力の安定化を図った．初期のプロンプトでは GPT-4o において 59％の重複座標が出力されたが，最終的なプロンプトでは 0 対に低減され，重複問題の解決が確認された．最終的に改善したプロンプトの特徴は以下の通りである．

1)　住所の階層的解析指示

都道府県，市区町村，町丁目，番地・号といった住所の階層を逐次解析するよう明示し，各階層を対応させることで住所の理解を促進した．

2)　精度の明示的制約

出力される緯度・経度の精度を少なくとも小数点以下 6 桁以上と指定し，高精度な座標生成を保証した．

3)　出力形式の厳格化

座標および付随情報の出力形式を厳密に指定し，解析・可視化処理への適用を容易にした．

4)　反復性制約の導入

過去に出力された座標を再利用しないよう制約を設けることで，不確実性や重複座標の問題を低減した．

5)　Few-shot 学習の活用

簡単，中程度，複雑な住所のサンプルを提示する few-shot 形式により，モデルが多様な住所表記に対応できるよう誘導した．

これらの改良により，最終プロンプトを用いた GPT-4o において，精度向上と重複座標の低減が確認された．

基本プロンプト：

> I have a place in Japan: "{place_name}".
> Please provide the approximate latitude and longitude in decimal degrees.
> Format your answer as: "Latitude: xx.xxxx, Longitude: yy.yyyy".
> If you don't know, make your best guess.

改善済みプロンプト：

> You are tasked with geocoding a Japanese address to latitude and longitude coordinates.
> Instructions:
> Read and analyze the entire Japanese address carefully, including all hierarchical levels.
> In Japan, addresses generally follow this order:
> 「都道府県 → 市（市区町村）→ 区 → 町 → 丁目 → 番地 → 号 → 建物名・部屋番号」.
> Note:
> - For 政令指定都市 (designated major cities such as Yokohama, Kawasaki, Osaka), "区 (ku)" refers to administrative wards under the city.
> - For 東京都特別区部 (Tokyo 23 Special Wards, e.g., 新宿区, 港区), each "区" functions as a municipality itself, equivalent in status to a city.
> Use all details provided (from prefecture down to the smallest available unit, such as block, building number, or apartment room).
> Provide coordinates that best match the complete address provided.

Do not truncate or ignore any part of the given address; use all provided details to achieve maximum accuracy.

Here, maximum accuracy means:

Latitude and longitude must be reported with at least 6 decimal places (≈ within 0.1 meter precision).

All hierarchical components of the Japanese address must be analyzed.

The output coordinates should correspond to the smallest available unit of the address, not a higher-level centroid (e.g., not just the ward or city center).

Avoid duplicating coordinates from previous queries; treat each address as unique.

If uncertain, provide your best possible guess, clearly based on the complete address.

Output Format:

Exactly format your response as follows:

Latitude: xx.xxxxxx, Longitude: yy.yyyyyy

Do NOT use any Markdown formatting such as ** or *; output plain text only.

Examples:

Example 1:

Address: 東京都　目黒区　青葉台　１丁目　１６－９

Latitude: 35.647322, Longitude: 139.696684

Example 2:

Address: 静岡県　静岡市葵区　相生町　１４－２８－１Ｆ

Latitude: 34.977019, Longitude: 138.392405

Example 3:

Address: 岐阜県　大野郡白川村　大字荻町　２４９９

Latitude: 36.255081, Longitude: 136.9023

Example 4:

Address: 京都府　京都市中京区　木屋町四条上

Latitude: 35.003803, Longitude: 135.770556

Now, please geocode the following address with careful attention to the full details provided:

Address: "{place_name}"

Provide the coordinates below:

2.4　評価指標

本研究では，従来型手法（Google Maps API）をベースラインとし，LLM 出力との相対比較によって性能を評価する．評価指標は以下の 3 点である．

1)　座標精度（距離誤差）

モデルが出力した緯度・経度およびベースライン（Google Map API による座標出力）を，参照座標（ゼンリンテレポイントデータ）と比較し，各モデルおよび地域ごとの誤差分布，中央値，標準偏差を算出・評価した．

2)　処理時間

各モデルが住所入力から座標出力までに要する平均処理時間（秒）を計測し，モデルの計算効率を評価した．

3)　座標の一意性（重複座標の有無）

出力された座標が異なる住所に対して同一座標となる事象の発生を確認し，モデルの出力の一意性・頑健性を評価した．

3. 実験結果

3.1　予備実験

本研究では，最終的に GPT-4o を採用する前に，複数のモデルを対象として精度，処理時間，およびデータ重複率の 3 つの指標で性能比較を行った．各モデルの結果は以下の通りである（表 2～表 4, 単位は km および秒）．

表 2　各モデルの数値指標

モデル	平均誤差	標準偏差	最小誤差
GPT-4	1.76	0.742	0.002
GPT-4o	1.041 ～ 1.289	0.683 ～ 0.830	0.000 ～ 0.001
GPT-O1	1.091	0.721	0.003
GPT-O1 Mini	4.417	3.595	0.002

表 3　各モデルの平均処理時間

モデル	平均時間（秒）
GPT-4	2.48
GPT-4o	1.83
GPT-O1	15.51
GPT-O1 Mini	4.73

表 4　各モデルの座標の一意性

モデル	繰返数
GPT-4	78%
GPT-4o	59%→0
GPT-O1	0
GPT-O1 Mini	10%

表 2～表 4 より，GPT-4o は平均誤差が最小で精度に優れると同時に，処理時間も短く，プロンプト改善によりデータ重複率を 0 に抑えることが可能であった．一方，GPT-O1 は精度は比較的高いものの，処理時間が非常に長く, GPT-O1 Mini は誤差が大きく不安定であった. GPT-4 は精度は中程度であるものの，重複率が高い点が課題であった．

以上より，本研究の本実験では，改善済みプロンプトを適用した GPT-4o を採用した．

3.2　本実験

3.2.1　平均精度の比較

図 1 に，日本大・中・小代表地域における Google map API と GPT によるジオコーディング精度の平均比較を示す．横軸は各地域，縦軸は平均誤差距離（km）である．

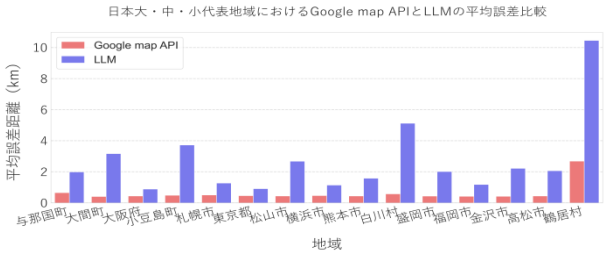


図 1　代表地域における平均誤差

図 1 より，Google map API は全地域で平均誤差が 1km 程度かそれ以下で，精度が安定して高い．GPT は東京都，大阪府など大都市では比較的小さい誤差を示すが，白川村や鶴居村など郊外・農村地域では平均誤差が 5km を超え，鶴居村では 10km を超えるなど，地域の規模や都市化の程度に応じて精度に顕著な差異が見られる．

3.2.2　誤差分布の詳細分析

図 2 は，同地域における Google map API と GPT の誤差分布比較を示す．横軸は各地域，縦軸は誤差距離（km）である．

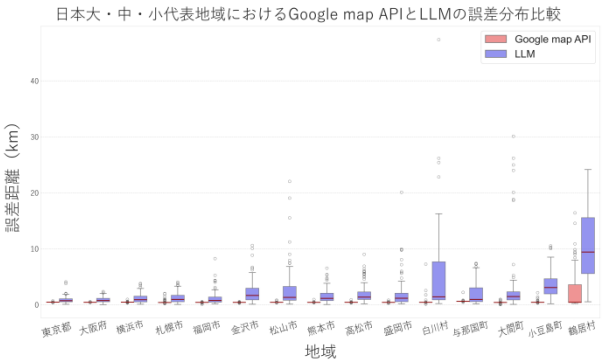


図 2　代表地域における誤差分布比較（箱線図）

図 2 より，Google map API は全地域で誤差が小さく，分布も比較的狭い．GPT については，東京都，大阪府，横浜市など大都市では誤差が小さく，分布も狭いが，白川村や鶴居村などの一部郊外・農村地域では誤差の分布が非常に広がり，大きな誤差も見られる．これにより，GPT の誤差分布は地域の規模や都市化の程度に応じて顕著な差異が存在することが示された．

3.2.3　処理時間および座標の一意性の比較

各地域における GPT-4o のジオコーディング処理時間およびデータ重複率を測定した．全地域の平均処理時間は 1.65 秒 で，概ね 2 秒以内に収まった．出力の一意性は多くの地域で確保されているが，与那国町では重複が顕著（18 件）であった．

4．考察

本研究では，GPT-4o をはじめとする LLM が日本国内の多様な地域において住所から座標を推定できることを示した．大都市圏では平均誤差が 1 km 前後と比較的高精度であった一方，白川村や鶴居村など一部の小規模自治体では誤差が大きく，また与那国町においては繰返し出力が多く観察されるなど，LLM の地理的常識・地名解釈能力には限界も確認された．これらの結果は，LLM の適用可能性が地域特性や訓練データの偏りに左右されることを示唆している．他方，適切なプロンプト設計を行うことで重複出力を回避でき，出力の安定性を改善できることが確認された点は重要である．また，従来のジオコーディング API が入力形式に依存しやすいのに対し，LLM は文脈理解に基づく柔軟な推定が可能であるため，両者を補完的に活用することで精度と頑健性の向上が期待できる．さらに，本研究で測定した処理時間はいずれの地域でも 2 秒以内に収まり，実用性の観点からも有望であることが示された．

5．結論と今後の課題

本研究は，LLM によるジオコーディングの可能性と限界を実証的に明らかにし，特に GPT-4o が従来手法を補完する形で活用し得ることを示した．加えて，プロンプト設計の工夫により精度と一意性が改善されることを確認し，LLM の柔軟性が地理情報処理において有効であることを示した点に意義がある．

今後の課題としては，第一に LLM のジオコーディング精度をさらに高める手法の探究，第二に GPT-4o 以外の最新 LLM を含むモデル間比較による最適モデルの検証，第三に LLM と既存ジオコーディング API を統合したハイブリッド手法の開発が挙げられる．また，少数ショット学習や追加情報（郵便番号・都道府県名など）の付与による精度向上の可能性についても検討が必要である．これらの課題に取り組むことで，LLM に基づく地理情報処理は実用的かつ信頼性の高い次世代技術へと発展することが期待される．

謝辞

研究室の皆様にご支援をいただき，感謝申し上げます．

参考文献

Zandbergen, P. A. (2009). Geocoding quality and implications for spatial analysis. *Geography Compass*, **3**(2), 647-680. DOI: 10.1111/j.1749-8198.2008.00205.x.

Yin, Z. C., Li, D., Goldberg, D. W. (2023). Is ChatGPT a game changer for geocoding - a benchmark for geocoding address parsing techniques. *In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Searching and Mining Large Collections of Geospatial Data*, 1-8. DOI: 2310.14360.

Hu, X., Kersten, J., Klan, F., Farzana, S. M. (2024). Toponym resolution leveraging lightweight and open-source large language models and geo-knowledge. *International Journal of Geographical Information Science*, https://doi.org/10.1080/13658816.2024.2405182.

大学病院の集中治療室における医療スタッフの動線分析

三田村和喜*・沖拓弥・仙頭佳起***・里見拓哉****・野坂宜之***・野口綾子***・海塩渉****・若林健二*****

Traffic Flow Analysis of Medical Staffs in the Intensive Care Unit of a University Hospital

Kazuki MITAMURA*, Takuya OKI**, Yoshiki SENTO***, Takuya SATOMI****, Nobuyuki NOSAKA***, Ayako NOGUCHI***, Wataru UMISHIO****, Kenji WAKABAYASHI***

Abstract: This study focuses on addressing congestion and traffic line issues in hospital intensive care units (ICUs) by analyzing the movement trajectories of medical staff. Previously, we collected 24 hours of 2D point cloud data using 18 small laser scanners at T University Hospital. We then developed a method from this data to extract individual movement trajectories for each scanner using a fine-tuned YOLO v8 object detection model in conjunction with a Kalman filter. However, because the scanning areas of the individual scanners overlapped, the same person was detected by multiple scanners simultaneously, resulting in duplicate trajectory data. To address this issue, this study aims to use the trajectory data from all scanners to reconstruct continuous movement paths across the ICU space. Analyzing these unified trajectories, we analyze the actual conditions of traffic congestion and flow conflicts. The results will offer insights into improving operational efficiency in critical care settings.

Keywords: 動線分析（traffic flow analysis），レーザスキャナ（laser scanner），コンピュータビジョン（computer vision），大学病院（university hospital），集中治療室（intensive care unit）

１．はじめに

病院の集中治療室（ICU）は，大手術後や重篤な状況など濃密な医療とケアを要する患者を収容する場である．多領域にわたる専門性を有する様々な医療スタッフ（医師・歯科医師・看護師・薬剤師・管理栄養士・理学療法士・臨床工学技士など）や機器類が頻度高く出入りするほか，患者のベッドやストレッチャーの往来も多く，人やモノの動線は病院内でも特に複雑である．その結果，至る所で動線の交錯や滞留が発生し，スタッフの業務効率性に影響を及ぼしているとされるが，改善につながる定量的なエビデンスは乏しい．

昨今では，人工知能（AI）をベースとしたコンピュータビジョン技術が普及し，動画から自動抽出した移動軌跡データを用いた研究も盛んに行われ，実用化もなされている．しかし，病院では患者やスタッフのプライバシー保護等の理由により，研究目的で動画を撮影・活用することは容易でない．

これをふまえ筆者らは，T 大学病院の ICU（図 1）を対象に，18 台の小型 2 次元レーザスキャナ（以下，スキャナ）を用いた点群計測を行い（図 2：沖ほか，2024a），計測で得られた点群のうち 24 時間（2023/12/19 午前 3 時～2023/12/20 午前 3 時）分のデータから，スキャナごとにスタッフの移動軌跡を高解像度に抽出する手法を構築した（沖ほか，2024b；三田村ほか，2025）．しかし，複数のスキャナの計測範囲が重複する領域での統合処理や，人

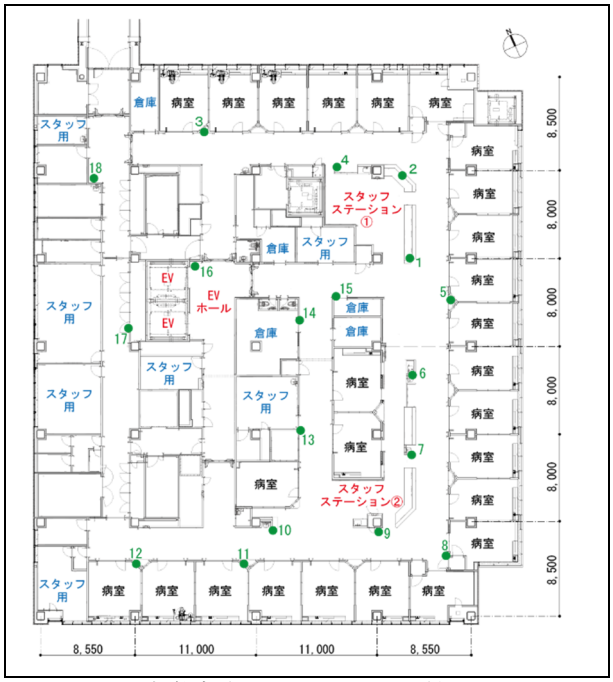


図 1　T 大学病院 ICU の平面図と計測ユニット（●で示す No.1～No.18）の設置位置

*　学生会員　東京科学大学 環境・社会理工学院 建築学系（Institute of Science Tokyo）
〒152-8550　東京都目黒区大岡山 2-12-1-M4-2　E-mail：mitamura.k.d283@m.isct.ac.jp
**　正会員　東京科学大学 環境・社会理工学院 建築学系（Institute of Science Tokyo）
***　非会員　東京科学大学病院（Institute of Science Tokyo Hospital）
****非会員　東京科学大学 環境・社会理工学院 建築学系（Institute of Science Tokyo）

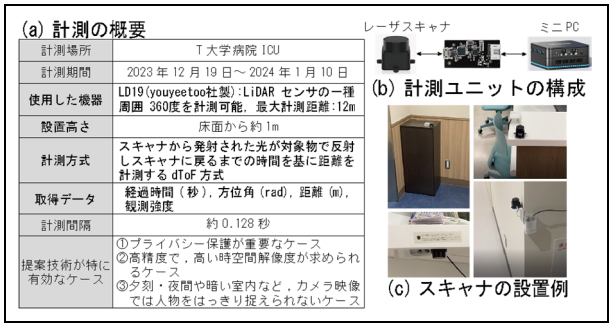


図 2　T 大学病院における計測の概要

物以外のノイズ（扉の開閉など）の処理に課題があった．

そこで本稿では，複数スキャナによる計測データをより正確に統合し，ノイズ処理を行うことで，ICU 全体でスタッフの移動軌跡を生成する．さらに，生成した軌跡データを用いて，ICU におけるスタッフの歩行量を分析するとともに，動線交錯の実態を明らかにする．

2．軌跡データの生成手法

2.1　点群データに基づく人物と軌跡の抽出

計 18 台のスキャナから得られた点群データから，図 3(a)に示す方法でスタッフを抽出・追跡し，スキャナごとの軌跡データを作成した．抽出されたスタッフを表す領域はバウンディングボックス（以下，ボックス）で囲まれ，追跡中は同一 ID が付与される．軌跡の可視化例を見ると，追従やすれ違い等の様子が確認できる（図 3(b)）．

2.2　ノイズの除去

前節の手法により抽出された軌跡データには，短時間に動きの見られた物品や扉などが人物として誤検出されている場合が含まれていた（図 4(a)）．そこで，ノイズの発生箇所のうち，扉など位置が固定の箇所を矩形領域（図 4(b)の赤枠）として指定し，この領域内に一定時間以上とどまり続ける軌跡をノイズとして除外した．

2.3　スキャナ間の位置・時刻合わせ

抽出した軌跡データは，各スキャナの位置を原点とするローカル座標で記録されている．そこで，各スキャナから得られた点群の可視化結果をもとに，Illustrator 上で ICU 図面（図 1）と目視で位置合わせを行うことで，18 台全ての軌跡データをグローバル座標に変換した．

また，スキャナの点群計測間隔は本来約 0.128 秒であるが，スキャナごとに計測間隔に若干のずれが見られた．そこで，図 5 の方法で移動軌跡を線形補間し，0.1 秒間隔の共通の時間フレームに変換した．

さらに，この共通の時間フレームにおいて，スキャナ間で観測時刻にずれが生じていた．これは，各スキャナには独立したミニ PC が接続されているが（図 2），それらは LAN から切り離されており，今回の計測環境ではミニ PC 間の時刻同期が困難であったことが原因と考えられる．そこで，特定の時間帯における時刻のずれを目視

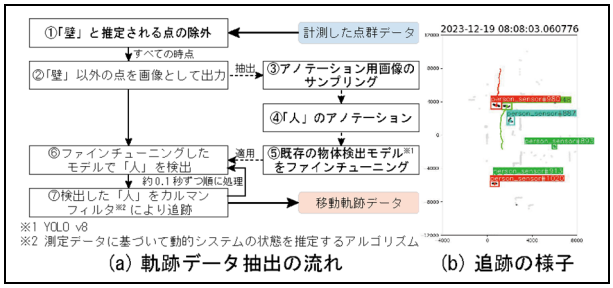


図 3　点群データに基づく軌跡データの抽出

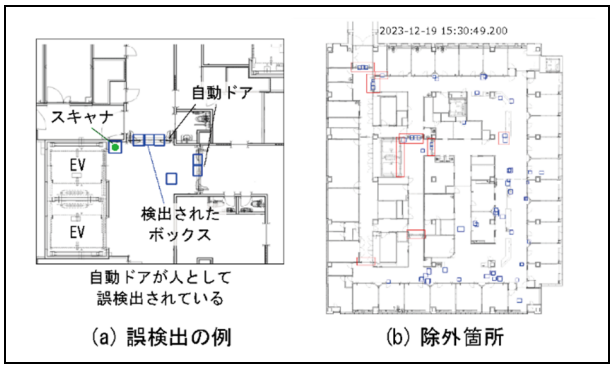


図 4　誤検出の例とノイズ除外箇所

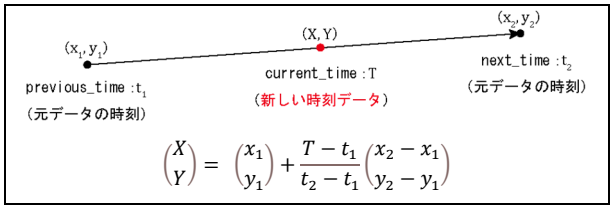


図 5　時間フレームの変換

表 1　スキャナ間の計測時刻のずれ

スキャナ	時刻の遅れ（秒）（スキャナ6を基準）	スキャナ	時刻の遅れ（秒）（スキャナ6を基準）
1	23.5	10	24.9
2	23.5	11	32.4
3	1.1	12	29.6
4	16.5	13	2.9
5	5.8	14	24.7
6	0.0	15	24.7
7	26.5	16	20.2
8	22.2	17	– ※
9	23.1	18	18.1

※スキャナ 17 に関しては，他のスキャナとの検知範囲の重複が確認できなかったため，ここでは計測時刻のずれを評価していない．

により把握し（表 1），ずれが時間帯によらず一定であることを確認した上で，得られた差に基づいて各スキャナの時刻を補正した．

2.4　軌跡の統合手法

スキャナの最大計測距離は 12 m であり，本 ICU における 18 台のスキャナの計測範囲には，互いに重複する領域が存在する．そのため，同一人物が複数のスキャナから検出される場合があるほか，人物がスキャナ間を移動

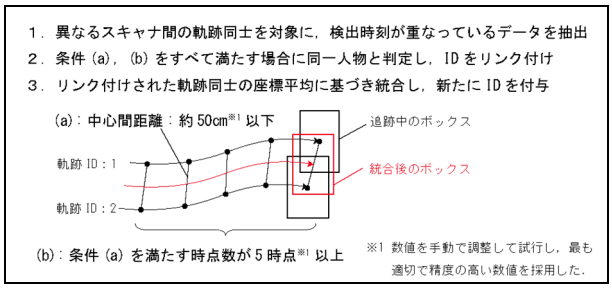


図 6　重複する軌跡の統合方法

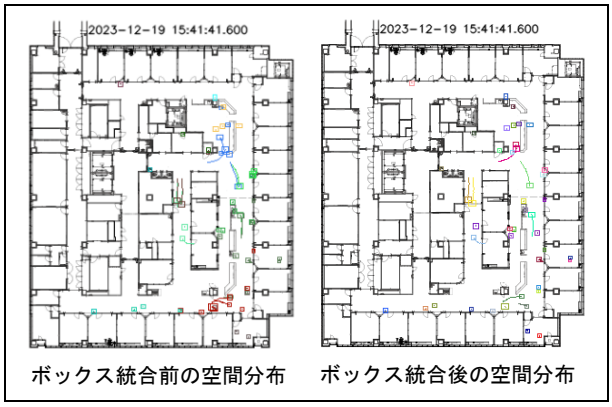


図 7　統合前後の軌跡の比較

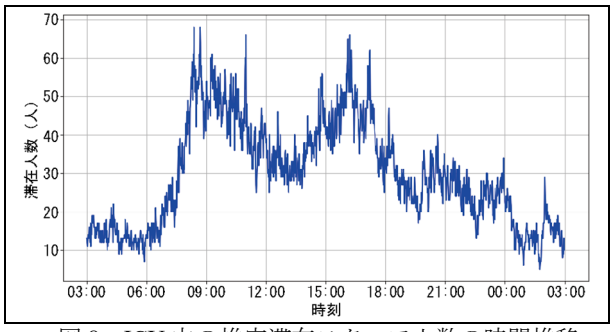


図 8　ICU 内の推定滞在スタッフ人数の時間推移（1 分単位）

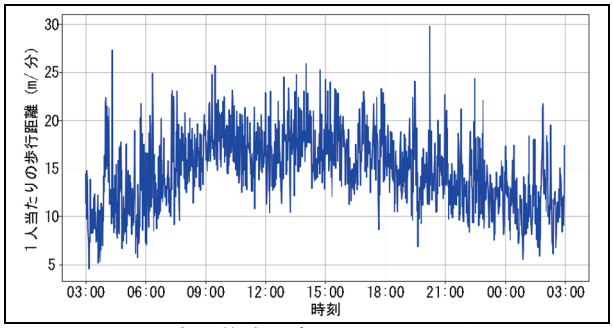


図 9　ICU 内の推定滞在スタッフ 1 人当たりの平均歩行距離の時間推移（1 分単位）

する際に，追跡 ID がスキャナごとに分断されてしまう問題が生じる．そこで，図 6 に示す方法で，個別に抽出された軌跡の統合処理を行った．

統合処理前後の比較例を図 7 に示す．同一人物と推定されるボックスの重複がある程度除外され，統合された軌跡として扱われていることが確認できる．

3．スタッフの動線分析

3.1　歩行量の評価

対象期間内（12/19 午前 3 時～12/20 午前 3 時）に ICU 内に存在するボックス数を 1 分ごとにカウントすることで，ICU 内に滞在するスタッフ人数の時間推移を推定した．病室内やスタッフルームに在室している場合には，スキャナに検出されないため，人数は過小に推定されている可能性がある．それでも，回診や，日勤・夜勤の引継ぎのために，ICU 内のスタッフ（主に看護師）の人数が多くなる朝方と夕方にピークが見られるなど，時間推移の傾向を良好に捉えられている（図 8）．

ここで，前章で生成した統合済みの軌跡をもとに，1 分ごとに ICU 全体の合計歩行距離を算出し，それを 1 分ごとの推定滞在者数（図 8）で除すことで，1 人当たりの平均歩行距離の時間推移を得た（図 9）．これを 24 時間で平均すると，スタッフ 1 人当たりの歩行量は以下のように推定できる．

- 1 人当たり・1 分当たりの移動距離：14.7 m/分
- 1 人当たりの総移動距離（8 時間勤務を想定）：7,039 m
- 1 人当たりの総歩数：11,635 歩（日本人の平均歩幅を 60.5 cm として計算：一般社団法人日本健康ツール開発機構の資料に基づく）

看護師の 1 日の歩行量は，多い人で，8 時間換算で 7,000～9,000 歩とされる（楠田ほか，2016）．また，T 大学病院で別途実施した，10 日間ずつ ICU/HCU リーダーの歩数をカウントしたデータによれば，ICU で平均 7,230 歩，HCU で平均 7,818 歩，一番多いスタッフで 11,625 歩であった．本節の分析では，特定の場所での目的を持った滞留時に生じる微動を除外しておらず，実際には歩いていない場合でもデータ上は軌跡が生成されることで，移動距離が加算される分，歩数がやや多めに推定されていると考えられる．それでも本節の結果は，本研究の手法で抽出した軌跡データが，ICU スタッフの日々の歩行量の把握にも有効であることを示している．従来から用いられている身体活動強度（METs）とは別の観点から労働負荷を測ることで，スタッフの働き方改革を推進する上での指標の一つとなり得る．将来的に個人の移動軌跡を追跡可能となれば，個人単位の時間ごと・日ごとの歩行量が詳細に推定可能になるものと期待される．

3.2　動線の交錯評価

個々の軌跡同士の距離が閾値（ここでは 1 m）以下となった地点を求め，その時点における各軌跡の移動方向ベクトルのなす角度を算出した（図 10(a)）．接近時における相対的な方向の分布傾向をローズダイアグラムで可視化すると（図 10(b)），後方から接近して追い抜く，あるいは，追従するケースが多いことが読み取れる．この傾向は，ICU 内では複数人がまとまって移動することが多い

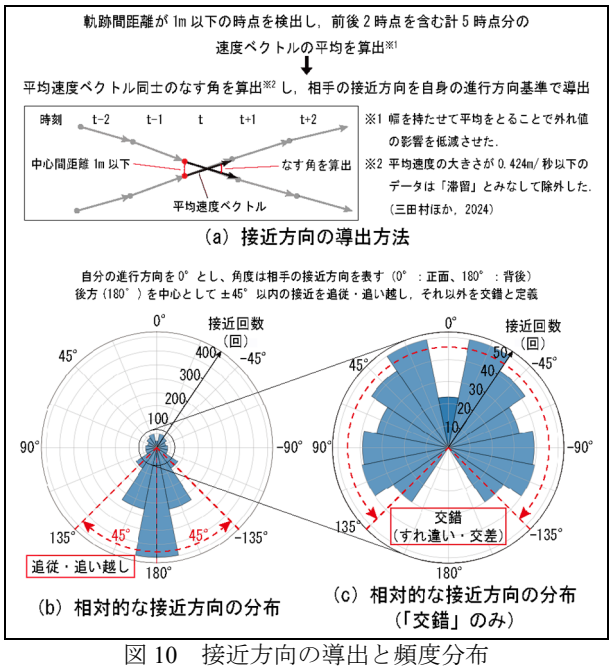


図 10　接近方向の導出と頻度分布

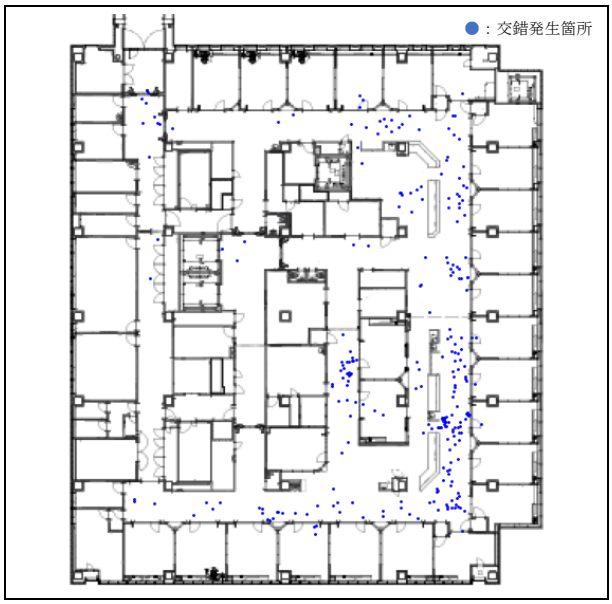


図 11　24 時間における交錯発生箇所の空間分布

ためと推察される．

特に進行方向が大きく異なる場合（閾値：45° ）を「交錯」と定義し，交錯時のみローズダイアグラムを作成した（図 10(c)）．いずれの方向でも一定程度の交錯が存在することが見てとれる．なお，0° 付近のデータが少ない理由として，1m の距離閾値に達する前に，相手との接触を避ける回避行動をとっているためと考えられる．

さらに，24 時間における交錯発生地点をプロットしたところ（図 11），図面上部に位置する病室周辺と比べ，図面下部の病室周辺において交錯が多く発生していることが確認された．また，スタッフステーションの出入り口付近や図面中央の廊下においても交錯が頻繁に発生している様子が見てとれる．

4．まとめと今後の課題・展望

T 大学病院 ICU において，18 台の小型 2 次元レーザスキャナで計測した 2 次元点群データのうち，24 時間分を対象として，スキャナごとにスタッフの検出・追跡を行い，それらを統合することで，ICU 全体における移動軌跡の抽出を行った．そして，得られた軌跡に基づき，ICU 内の歩行量や動線交錯の実態を分析した．

今後は，人物追跡精度の向上が課題である．①追跡中のスタッフが何らかのものに隠れて一時的に検出されなかった場合に，ID が切り替わってしまう場合や，②複数のスタッフがまとまって移動する時に，スキャナの性質上，時点ごとに観測される点群の形状が変化し，同一人物でも連続した ID が付与されない場合が散見された．解決策として，YOLO v8 のファインチューニングに用いた学習データの増加や，人物以外のモノなどを別のクラスタラベルとして学習させることなどが考えられる．

謝辞

本研究の実施にあたり，2023 年度東京工業大学・東京医科歯科大学研究マッチングファンド，および，JSPS 科研費 25K01391 の助成を受けた．また，ICU での計測に快くご協力いただいた T 大学病院のスタッフの皆様に，この場を借りて謝意を表する．

参考文献

一般社団法人日本健康ツール開発機構，きれい健康ネット，https://kirei-kenko.or.jp/walking/stride-average（参照：2025.08.28）

沖拓弥・仙頭佳起・野坂宜之・野口綾子・海塩渉・若林健二（2024a）大学病院の集中治療部におけるレーザスキャナを用いた行動分析の試み．2024 年度人工知能学会全国大会（第 38 回）．

沖拓弥・仙頭佳起・山中理沙・野坂宜之・野口綾子・海塩渉・若林健二（2024b）小型 2 次元レーザスキャナと画像認識技術を用いた時空間解像度の高い歩行軌跡の検出手法：大学病院の集中治療部におけるケーススタディ．第 33 回地理情報システム学会学術研究発表大会，地理情報システム学会講演論文集，33．

楠田将一朗・池田智紀・緒方悠元・小杉和希・榮智徳・森山凌・吉田耕輔（2016）看護師の働き方と健康，1-6，https://www.shiga-med.ac.jp/~hqpreve/kyouiku/socmed_fw/pdf/2016/2016_4.pdf（参照：2025.08.28）

三田村和喜・沖拓弥・仙頭佳起・野坂宜之・野口綾子・海塩渉・若林健二（2025）大学病院の集中治療室における医療スタッフの移動軌跡の抽出手法．2025 年度人工知能学会全国大会（第 39 回）．

Jocher, G., Chaurasia, A., & Qiu, J.（2023）Ultralytics YOLO (Version 8.0.0) [Computer software] . https://github.com/ultralytics/ultralytics.

Tzutalin（2015）LabelImg. https://github.com/tzutalin/labelImg.