大規模人流データに基づくユーザ固有の移動文脈を考慮した都市空間モデリング

寺島 青[1,a)]　坪内 孝太[2]　山口 修司[2]　田村 直樹[1]　庄子 和之[3]　河口 信夫[1,3]　米澤 拓郎[1]

概要：都市空間における各場所の意味や役割は，訪問者の持つ背景や生活スタイルによって異なる．都市サービスの高度化に向けては，各場所の意味の多様性を考慮したベクトル表現を学習する都市空間モデリングが求められる．近年，モバイル端末の普及により大規模な人流データの収集が可能となり，各ユーザの移動履歴を活用した都市空間モデリング手法が数多く提案されている．しかし，一般的な人流データは座標や時間といった数値情報から構成されており，移動の意味に関する情報は明示的に含まれていない．そのため，既存手法の多くは，統計情報や時間的特徴のモデリングにとどまり，ユーザごとの移動文脈に由来する場所の意味の多様性を十分に捉えられていない．本研究では，ユーザ固有の移動文脈に基づき，各場所の意味の違いを捉える都市空間モデリング手法を提案する．提案手法では，移動の時間的な周期性を系列モデルにより表現するとともに，移動を訪問頻度や時間情報に基づくテキストで表現し，言語空間に埋め込むことで，数値ベースの移動データだけでは表現が難しい場所の意味や役割に関する特徴を補う．最終的にこれらの特徴を統合し，各場所の埋め込み表現を構築する．実世界の人流データを用い，移動場所予測，移動軌跡検索，移動軌跡補間の 3 つの下流タスクを通じて既存手法と性能を比較した．実験の結果，提案手法により構築された埋め込み表現は全ての下流タスクにおいて既存手法を上回る性能を示し，最も競合的な既存手法に対し Top-1 Accuracy でそれぞれ，4.3%，3.6%，3.9%の精度向上を実現した．

キーワード：都市空間モデリング，都市人流データ，深層学習

1. はじめに

都市空間における各場所の意味や役割は，訪問者が持つ背景や生活スタイルによって異なる．ある場所を日常生活の拠点として利用する人もいれば，余暇の場として訪れる人もいる．こうしたユーザごとの場所の捉え方の違いを理解することは，移動場所のレコメンドや施設配置の最適化などの都市サービスの高度化において重要であり，図 1 に示すような，各場所の意味の多様性を考慮したベクトル表現を学習する都市空間モデリングが求められる．近年，スマートフォンなどのモバイル端末の普及に伴い，都市における大規模な人流データの収集が可能となった．これらのデータはそこに住む人々の行動特性を理解する上で重要な手掛かりとなり，都市計画・交通計画・災害対策など様々な都市課題の解決への活用が進められている [1], [2], [3]．

[1] 名古屋大学大学院 工学研究科
Graduate School of Engineering, Nagoya University
[2] LINE ヤフー株式会社
LY Corporation
[3] 名古屋大学 未来社会創造機構
Institutes of Innovation for Future Society, Nagoya University
[a)] haru@ucl.nuee.nagoya-u.ac.jp

このような背景から，都市空間モデリングの研究においても，人流データを活用した手法が数多く提案されている．

人流データを用いた都市空間モデリングの代表的な手法として，都市内の各グリッドへの訪問パターンや滞在情報を活用したモデリング手法 [4], [5], [6] や，個々のユーザの移動軌跡に基づき，移動の順序や時間的特性を考慮したモデリング手法 [7], [8], [9], [10] が提案されている．これらの手法は，人流データに含まれる時空間的な特徴を捉え，場所間の関係性や移動行動の理解を目的としている．しかし，一般的な人流データは移動した場所の座標や時間といった数値情報から構成されており，移動目的など意味的な情報が直接含まれていない．そのため，既存の都市空間モデリング手法の多くは，統計的な訪問特徴や時間的な移動パターンのモデリングにとどまり，ユーザごとの移動文脈に基づく場所の意味の多様性を十分に捉えられていない．

本研究では，ユーザ固有の移動文脈に基づき，各場所の意味の違いを捉える都市空間モデリング手法を提案する．提案手法では，移動の時間的特性を Transformer [11] により表現するとともに，移動を訪問頻度や時間情報に基づく簡易なテキストで表現し，言語空間に埋め込むことで，数値

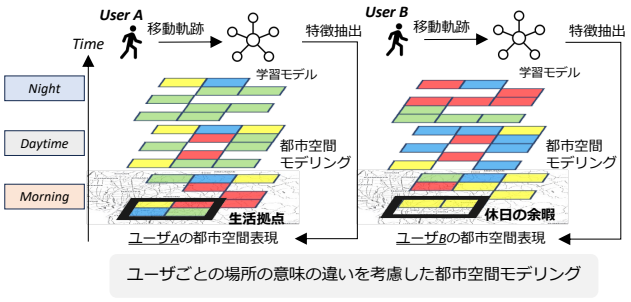

図 1 ユーザ固有の都市空間モデリング

ベースの移動データだけでは表現が難しい場所の意味や役割に関する特徴を捉える．最終的にこれらの表現を Cross Attention により動的に統合し，各場所の埋め込み表現を構築する．構築された場所表現は，ユーザ固有の移動文脈を考慮した豊富な情報を持ち，様々な下流タスクの精度向上への寄与が期待される．

実世界の人流データセット”LYMob-4Cities [12]”を用い，提案手法で構築される埋め込み表現の性能を，移動場所予測，移動軌跡検索，移動軌跡補間の 3 種類の下流タスクを通じて既存の都市空間モデリング手法と比較した．実験の結果，提案手法はすべての下流タスクにおいて既存手法を上回る性能を示し，下流モデルを BERT とした場合に 2 都市の平均スコアで最も競合的な既存手法に対して，Top-1 Accuracy でそれぞれ，移動場所予測では 4.3%，軌跡検索では 3.6%，軌跡補間では 3.9%の性能向上を確認した．これらの結果より，ユーザ固有の移動文脈を考慮した都市空間モデリングによる埋め込み表現の有効性を示した．

本研究の貢献は以下の 3 つである．

- 数値情報から成る人流データの制約を踏まえ，ユーザ固有の移動文脈に基づき場所の意味の多様性を捉える都市空間モデリングの枠組みを提案．
- 移動の時間的特徴と，移動をテキストで表現し言語空間に埋め込んで得た意味的特徴を Cross Attention により統合する都市空間モデリング手法を提案．
- 実世界の都市人流データセットを用いた評価実験により，移動場所予測，軌跡検索，軌跡補間の 3 つの下流タスクを通じて提案手法の有効性を実証．

2. 関連研究

本章では，都市空間モデリングに関する既存研究を整理する．まず，地理情報や統計情報などに基づく統計的な都市空間モデリング手法の特徴と限界を述べる．次に，個々のユーザの移動軌跡を利用した動的なモデリング手法について整理し，本研究が取り組む課題を明確にする．

2.1 静的特徴に基づく都市空間モデリング

人流データは都市空間における人々の行動パターンを理解する上で重要な情報源であり，様々な都市課題の解決への活用が進んでいる．こうした応用の性能は，各場所をどのように表現するかに大きく依存するため，都市空間の効果的なモデリング手法の構築が求められている．初期の都市空間分析手法では，地理情報や施設情報などの静的な特徴を用いて，各場所の機能を表現する研究が多く行われてきた．例えば，POI (Point of Interest) データを用い，施設カテゴリの分布などに基づいた都市空間分析手法が提案されている [13]．しかし，POI データは主に郊外で情報が少なく，時間的な変化を伴う都市活動を十分に表現できないという課題がある．また，集計型モバイルネットワークデータから得られる時間集計統計量を用いた手法も提案されているが [14], [15], [16]，統計量に基づく表現は特定の分析目的に依存しやすく，多様な下流タスクに対して汎用的に利用可能な表現を得ることが難しい．

こうした限界を踏まえ，近年では人流データや POI データを用いて，各場所をベクトル表現として学習する都市空間モデリング手法が提案されている [4], [5], [6]．さらに，衛星画像や風景画像といった，複数の異なるモダリティを統合したマルチモーダルな都市表現学習も研究されている [17], [18], [19], [20]．しかし，これらの手法は，各場所に対して単一の固定ベクトルを割り当てるため，ユーザごとの場所の意味の違いを十分に表現できない．

2.2 移動文脈に基づく都市空間モデリング

固定ベクトルによる静的な場所表現の限界を補うため，個々のユーザの移動軌跡に含まれる時間的・順序的な移動文脈を考慮した都市空間モデリング手法が提案されている．これらの研究では，自然言語処理における単語埋め込みの考え方を応用し，移動軌跡を単語列に類似した系列として捉え，RNN や Transformer を用いて，軌跡内で共起する場所や移動の順序関係から各場所の埋め込み表現を学習する [7], [8], [9], [10]．これにより，固定的な統計量では捉えられなかった場所間の動的な関係性や移動の時間的パターンを表現できるようになった．しかし，これらの手法の多くは，移動の順序や時間的特徴のモデリングに重点が置かれており，ユーザが各場所をどのような意味や目的で利用しているかといった意味的な文脈を十分に考慮できていない．これは，人流データが主に座標や時間といった数値情報から構成されているため，高次の意味情報の直接的な表現が難しいためである．この課題に対し，大規模言語モデル (LLM) を用いて，各 POI の言語的説明から表現を補強する研究も提案されているが [21]，場所の説明は固定的であり，ユーザごとの移動文脈に応じた場所の意味の多様性を捉えるには至っていない．本研究では，この課題に対して，ユーザ固有の移動文脈に基づく意味的表現を取り込むアプローチを提案する．

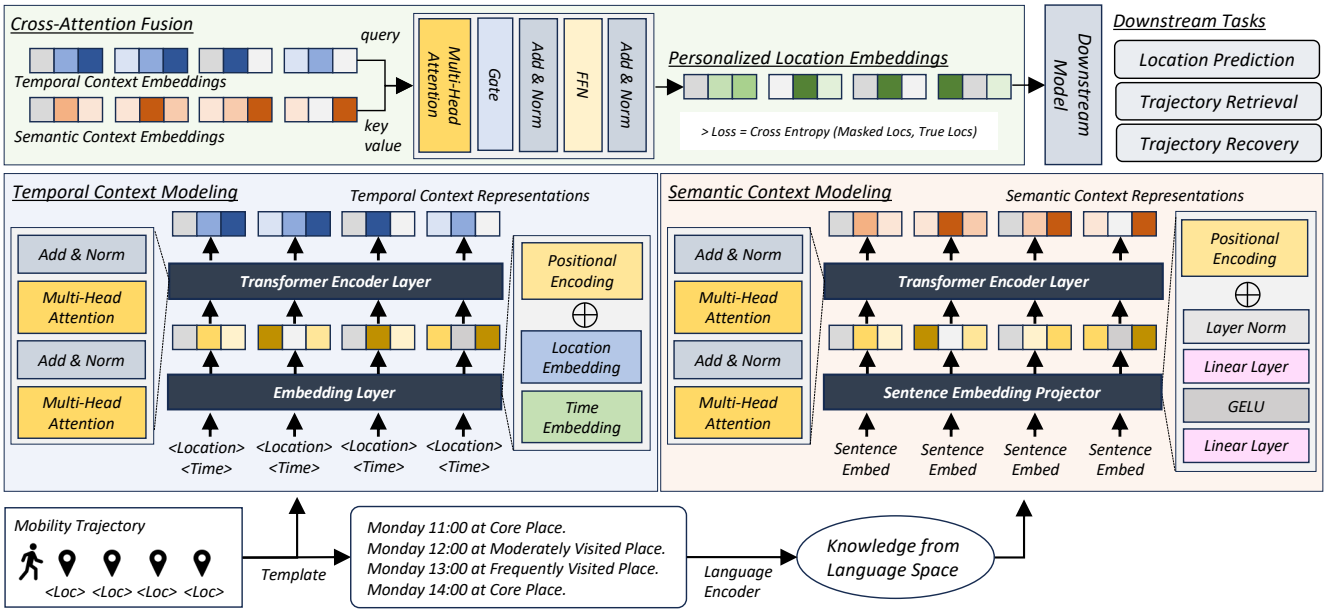

図 2 提案手法：移動の意味的文脈表現と時間的文脈表現を統合した都市空間モデリング

3. 提案手法

本章では，提案手法の詳細を説明する．提案手法の概要を図 2 に示す．まず，移動の意味的な文脈モデリングおよび時間的な文脈モデリングについて述べ，その後二つの表現の統合方法およびモデルの学習方法について説明する．

3.1 意味的な移動文脈モデリング

人流データは主に，移動場所の座標や時間などの数値情報から構成されており，各場所がどのような意味や役割で利用されているかといった意味的な情報は明示的に含まれない．そこで本研究では，ユーザの移動をテキストで表現し，それらを言語空間に埋め込むことで，移動に含まれる意味的文脈をモデルに取り込む．具体的には，まず各ステップの移動を，移動時刻とユーザ固有の訪問頻度に基づくテンプレートテキストで記述する．テキスト作成に用いる訪問頻度として，ユーザ u の場所 l に対する相対的な訪問頻度 $F_u(l)$ を次式で定義する．

$$F_u(l) = \frac{c_u(l)}{\sum_{l' \in \mathcal{L}_u} c_u(l')} \tag{1}$$

$c_u(l)$ はユーザ u の場所 l への訪問回数，$\mathcal{L}_u$ はユーザ u が訪問したすべての場所の集合を表す．$F_u(l)$ はユーザにとっての場所の相対的な重要度の代理指標として機能する．なお，$F_u(l)$ の計算には学習データのみを用い，評価データとの分離を厳密に保つ．この訪問頻度をユーザ内のパーセンタイル順位 $P_u(l)$ に基づいて表 1 のように 5 段階に離散化し，各場所の重要度レベルを定める．

この重要度レベルと移動時刻を組み合わせ，各ステップの移動を「[曜日・時刻] に [重要度レベル] の場所へ」という形式のテンプレートテキストとして記述する (例：「月曜 18:00 によく行く場所へ」)．次に，生成されたテキストを事前学習済みのテキスト埋め込みモデル (Sentence-BERT [22]) により，埋め込み表現 $\mathbf{s}_i$ を得る．$\mathbf{s}_i$ に，層正規化とゲート付き非線形変換から成る射影関数 f_{sa} を適用して埋め込み次元を調整し，さらに系列内の順序情報を考慮するため位置エンコーディング p_i を加算し，各ステップの表現 $\mathbf{z}_i$ を以下の式で得る．

$$\mathbf{z}_i = f_{sa}(\mathbf{s}_i) + p_i \tag{2}$$

次に，表現系列 $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_T\}$ を Transformer Encoder (f^s_{te}) に入力し，移動軌跡全体の意味的文脈を反映した表現 $\mathbf{G}$ を次式で得る．

$$\mathbf{G} = f^s_{te}(\mathbf{Z}) = \{\mathbf{g}_1, \ldots, \mathbf{g}_T\} \tag{3}$$

最終的に得られた $\mathbf{G}$ は，数値ベースの人流データのみでは表現が難しい移動の意味的文脈を捉えた表現であり，次節で述べる時間的な文脈表現と統合される．このように移動をテキストで記述し，言語空間に写像することで，平日の夕方に頻繁に訪れる場所への移動を「帰宅」といった言語概念に対応する意味表現として捉える．

表 1 訪問頻度の離散化基準

重要度レベル	パーセンタイル条件
Core Place	$P_u(l) \geq 0.99$
Subcore Place	$0.95 \leq P_u(l) < 0.99$
Frequently Visited	$0.85 \leq P_u(l) < 0.95$
Moderately Visited	$0.50 \leq P_u(l) < 0.85$
Rarely Visited	$P_u(l) < 0.50$

3.2 時間的な移動文脈モデリング

都市における人々の移動は，生活スタイルなどに由来する時間的な周期性を持つ．このような時間的文脈を表現するため，本研究では移動軌跡を時系列データとして扱い，Transformer Encoder により時間的な移動パターンをモデリングする．具体的には，ユーザの移動軌跡を位置系列 $\{l_1, l_2, \ldots, l_T\}$ と，対応する時間系列 $\{t_1, t_2, \ldots, t_T\}$ で表現し，各ステップの位置 l_i は位置埋め込み $\mathbf{e}_i^l$ に，時刻 t_i は時間埋め込み $\mathbf{e}_i^t$ に変換する．これに，3.1 節で得た意味的文脈表現 $\mathbf{g}_i$ を線形変換 f_{si} により変換した $\tilde{\mathbf{g}}_i = f_{si}(\mathbf{g}_i)$ および系列内の順序情報を考慮する位置エンコーディング $\mathbf{p}_i$ を加算し，各ステップの入力表現 $\mathbf{x}_i$ を構築する．$\tilde{\mathbf{g}}_i$ の加算は意味的文脈の弱い事前情報として機能し，より明示的な統合は次節の Cross Attention により行われる．

$$\mathbf{x}_i = \mathbf{e}_i^l + \mathbf{e}_i^t + \tilde{\mathbf{g}}_i + \mathbf{p}_i \tag{4}$$

次に，得られた系列表現 $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$ を Transformer Encoder (f_{te}^t) に入力し，時間的文脈表現 $\mathbf{H}$ を得る．

$$\mathbf{H} = f_{te}^t(\mathbf{X}) = \{\mathbf{h}_1, \ldots, \mathbf{h}_T\} \tag{5}$$

最終的に得られた $\mathbf{H}$ は，自己注意機構を通じて移動軌跡全体の時間的依存関係を捉えた文脈表現であり，次節で述べる意味的文脈表現 $\mathbf{G}$ との統合に用いる．

3.3 統合モデリングおよび学習

前節までに得られた意味的文脈表現 $\mathbf{G} = \{\mathbf{g}_1, \ldots, \mathbf{g}_T\}$ と時間的文脈表現 $\mathbf{H} = \{\mathbf{h}_1, \ldots, \mathbf{h}_T\}$ を統合し，最終的な各場所の埋め込み表現を構築する．統合には Cross Attention (f_{ca}) を用い，時間的な移動パターンを基盤としつつ，必要な意味的文脈を動的に参照するため，時間表現 $\mathbf{H}$ をクエリ，意味表現 $\mathbf{G}$ をキーおよびバリューとして，注意表現 $\mathbf{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_T\}$ を得る．

$$\mathbf{a}_i = f_{ca}(\mathbf{h}_i, \mathbf{G}, \mathbf{G}) \tag{6}$$

次に，$\mathbf{a}_i$ をチャンネルごとの学習可能なゲート $\boldsymbol{\beta} \in \mathbb{R}^d$ で重み付けして時間表現へ残差加算し，統合表現 $\mathbf{o}_i$ を得る．

$$\mathbf{o}_i = \mathbf{h}_i + \sigma(\boldsymbol{\beta}) \odot \mathbf{a}_i \tag{7}$$

ここで $\sigma(\cdot)$ はシグモイド関数，$\odot$ は要素積を表す．ゲート $\boldsymbol{\beta}$ はチャンネルごとに意味的文脈の寄与を調整し，時間表現を基盤としつつ必要な意味情報を選択的に統合する．最終的な統合表現 $\mathbf{O} = \{\mathbf{o}_1, \ldots, \mathbf{o}_T\}$ は，時間的文脈と意味的文脈の双方を反映した場所埋め込み表現となる．

モデルの学習には *Masked Location Modeling* を用いる．移動系列の一部のステップをランダムに選択し，時間的文脈モデリングへ入力される場所情報と，意味的文脈モデリングへ入力されるテキストを同時にマスクする．モデルは周辺ステップの時間的・意味的文脈を手がかりに，マスクされた場所を復元するよう学習される．マスクされた位置集合を $\mathcal{M}$ とすると，学習損失は次式で表される．

$$\mathcal{L} = -\sum_{i \in \mathcal{M}} \log P(l_i | \mathbf{O}) \tag{8}$$

両モダリティを同時にマスクした学習により，モデルが単一の情報源に依存することを防ぎ，時間的・意味的文脈を統合的に利用した場所埋め込み表現の獲得を促す．

4. 評価実験

本章では，提案手法により構築される場所埋め込み表現の有効性を検証する．提案手法は下流タスクに依存しない汎用的な埋め込み表現の獲得を目的としているため，移動場所予測，移動軌跡検索，移動軌跡補間の 3 種類の下流タスクを設定し，既存の都市空間モデリング手法と比較する．以下では，データセットおよび下流タスクの設定，学習設定を説明した上で，実験結果を述べる．

4.1 実験設定

4.1.1 データセット

本研究では，日本国内の複数都市で収集された GPS ベースの移動軌跡データから構成されるオープンデータセット”LYMob-4Cities” [12][*1]を用いる．なお，プライバシー保護のためデータ取得都市名および期間は公開されていない．データセットに含まれる 2 都市 (都市 A，都市 B) を対象とし，各都市において各週の移動記録数が 50 件以上のユーザから 10,000 人を無作為に抽出した．データセットの統計情報を表 2，グリッドごとのデータ量を図 3 に示す．各移動は $\langle uid, date, time, x, y \rangle$ の形式で表現されており，移動場所は 500 m × 500 m の空間グリッドに離散化され，30 分刻みで記録されている．曜日は日ごとのデータ量の周期性に基づいて推定し，追加の時間属性として用いる．曜日推定の詳細は付録 A.1 に示す．評価には都市ごとにユーザ単位の 5 分割交差検証を採用し，各 fold を 70%学習用，10%検証用，20%評価用に分割する．埋め込み表現の学習には各ユーザの最初の 3 週間分の移動データを，下流タスクの評価には残り 1 週間分のデータを追加で用い，埋め込み学習と下流タスク評価を時間的に分離する．

4.1.2 下流タスク設定

提案手法により構築される埋め込み表現の有効性を検証するため，移動場所予測，移動軌跡検索，移動軌跡補間の 3 種類の下流タスクを設定する．各タスクでは，提案手法および既存の都市空間モデリング手法 (以下，Location Encoder) により場所埋め込みを構築し，下流モデルとして LSTM [23] および BERT [24] ベースのモデルを用いる．下流モデルの構成は表 3 の通りであり，構成および学習設

[*1] https://zenodo.org/records/14219563

表 2　データセット統計

	都市 A	都市 B
ユーザ数	10,000	10,000
グリッド数	24,206	18,076
レコード数	4,443,978	4,420,943
平均移動レコード数	444	442
平均移動グリッド数	108	92

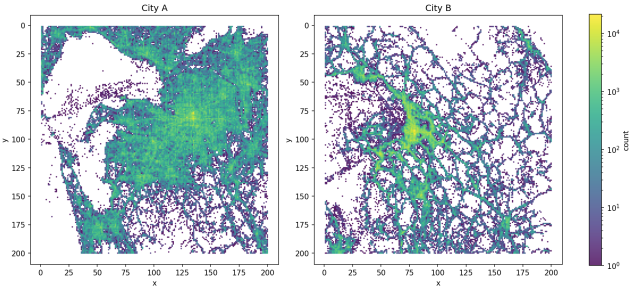

図 3　グリッドごとのデータ量

定はすべての Location Encoder で統一し，埋め込み表現そのものの性能差を検証する．詳細なパラメータ設定は付録 A.2 に示す．以下に各タスクの詳細を示す．

- **移動場所予測** (*Next Location Prediction*)
  3 週間分の移動軌跡を Location Encoder で埋め込み，その表現を用いて続く 1 週間の移動場所系列を予測するタスク．予測対象期間の移動の日付および時間は既知とし，場所のみを予測する．精度は Top-1 Accuracy および Top-5 Accuracy で評価する．
- **移動軌跡検索** (*Trajectory Retrieval*)
  移動軌跡を 1 週間単位に分割して各軌跡を Location Encoder で埋め込み，クエリ軌跡に対して同一ユーザの軌跡を検索するタスク．負例には，他ユーザの軌跡に加え，自身の軌跡にノイズを付与した軌跡も含める．精度は Top-1 Accuracy および MRR で評価する．
- **移動軌跡補間** (*Trajectory Recovery*)
  移動軌跡の一部を欠損させた状態で Location Encoder で埋め込み，周辺の文脈から欠損箇所の移動場所を予測するタスク．欠損箇所の移動の日付および時間は既知とし，場所のみを予測対象とする．精度は Top-1 Accuracy および Top-5 Accuracy で評価する．

4.2　学習設定

提案手法はタスク非依存な場所埋め込み表現の獲得を目的とし，*Masked Location Modeling* により自己教師あり学習を行う．移動系列のうちランダムに選択した 20%のステップをマスク対象とし，時間的文脈モデリングへ入力される場所情報を専用のマスクトークンに置き換えるとともに，意味的文脈モデリングへ入力される対応テキストも同時にマスクする．最適化には AdamW (学習率 1×10^{-4}) を用い，勾配爆発を防ぐため勾配クリッピングを適用する．

表 3　下流モデルの構成

タスク	モデル	アーキテクチャ
移動場所予測	LSTM	単方向 LSTM + 線形出力層
	BERT	Transformer Encoder + 線形出力層
移動軌跡検索	LSTM	双方向 LSTM AutoEncoder
	BERT	Transformer AutoEncoder (CLS)
移動軌跡補間	LSTM	単方向 LSTM + 線形出力層
	BERT	Transformer Encoder + 線形出力層

バッチサイズ 16 で 100 エポック学習した．また，モデルは PyTorch で実装し，NVIDIA RTX6000 Ada GPU (48GB) 上で学習を行った．意味的文脈モデリングには事前学習済み Sentence-BERT (埋め込み次元 768)[*2] を用い，その出力を射影関数 f_{sa} により埋め込み次元 128 の空間へ射影する．Sentence-BERT のパラメータは固定し，事前学習済みの言語知識を保持しつつ人流データへ適応させる．時間的文脈モデリングの Transformer Encoder は 4 層・8 ヘッド，意味的文脈モデリングの Transformer Encoder は 2 層・8 ヘッドで構成する．後者を軽量にしているのは，入力が Sentence-BERT により既に文脈情報を含んだ表現であるためである．

4.3　実験結果

本節では，提案手法により学習された場所埋め込み表現の有効性を，下流タスクにおける性能比較を通じて評価する．比較手法として，移動軌跡から場所埋め込みを学習する代表的手法である Geo-Teaser [7]，TALE [8]，HIER [9]，CTLE [10] を採用した．各比較手法は原論文の設定に従って実装・学習した．詳細なモデル構成およびパラメータ設定は付録 A.3 に示す．以下に，各比較手法の概要を示す．

- **Geo-Teaser** [7]：Skip-gram に基づく場所埋め込み学習手法であり，地理的近接性と時間情報を考慮した共起関係から空間的・時間的特徴を表現する．
- **TALE** [8]：CBOW に基づく場所埋め込み学習手法であり，複数の時間区間に分割した時間木構造を導入し，各場所が訪問される時間帯の特徴を捉える．
- **HIER** [9]：RNN に基づく場所埋め込み学習手法であり，次移動場所予測を通じて，複数の空間解像度を考慮した階層的な場所表現を構築する．
- **CTLE** [10]：Transformer Encoder に基づく場所埋め込み学習手法であり，マスク学習を通じて，移動軌跡の系列依存関係と時間情報を同時に考慮する．

表 4 に，各下流タスクにおける実験結果を示す．評価値はユーザ単位の 5 分割交差検証の平均値であり，統計的有意性は fold ごとの結果に対する両側対応 t 検定により確認した．検定の結果，有意差が確認できた項目を*で示

[*2] https://huggingface.co/sentence-transformers/all-mpnet-base-v2

表 4 各下流タスクにおけるベースラインモデルとの性能比較結果 (5 分割交差検証の平均値).
* は提案手法と比較手法間の両側対応 t 検定における有意差あり（$p < 0.05$）を示す.

		Next Location Prediction				*Trajectory Retrieval*				*Trajectory Recovery*			
		City A (10,000)		*City B (10,000)*		*City A (10,000)*		*City B (10,000)*		*City A (10,000)*		*City B (10,000)*	
		Top-1	*Top-5*	*Top-1*	*Top-5*	*Top-1*	*MRR*	*Top-1*	*MRR*	*Top-1*	*Top-5*	*Top-1*	*Top-5*
LSTM	*Geo-Teaser*	0.0022	0.0081	0.0294	0.0954	0.2070	0.3056	0.1231	0.1809	0.0031	0.0098	0.0265	0.0855
	TALE	0.0112	0.0375	0.0714	0.1963	0.1521	0.2202	0.1913	0.3078	0.0112	0.0370	0.0693	0.1925
	HIER	0.0206	0.0735	0.0723	0.2103	0.2819	0.4822	0.2298	0.4003	0.0166	0.0623	0.0622	0.1886
	CTLE	0.0354	0.0970	0.1212	0.2684	0.2875	0.5133	0.2509	0.4405	0.0275	0.0791	0.1047	0.2445
	Proposal	0.0717*	0.1927*	0.1491*	0.3385*	0.3044	0.5099	0.2879*	0.4558*	0.0793*	0.2069*	0.1632*	0.3583*
BERT	*Geo-Teaser*	0.0751	0.1608	0.1848	0.3753	0.3015	0.5880	0.2908	0.5736	0.0772	0.1634	0.1891	0.3795
	TALE	0.0883	0.1879	0.2010	0.3949	0.2933	0.5909	0.2832	0.5780	0.0991	0.2040	0.2089	0.4072
	HIER	0.0724	0.1966	0.1607	0.3592	0.3418	0.5755	0.3263	0.5558	0.0722	0.1976	0.1691	0.3724
	CTLE	0.1090	0.2367	0.1937	0.3937	0.2920	0.5787	0.2902	0.5728	0.1047	0.2285	0.1949	0.3945
	Proposal	0.1650*	0.3686*	0.2236*	0.4563*	0.3544	0.6066	0.3856*	0.6203*	0.1598*	0.3582*	0.2267*	0.4610*

している．実験の結果，提案手法はほぼ全ての評価項目において，最も競合的な比較手法を上回る性能を示した．各タスクにおける最も競合的な既存手法との比較では，2 都市の平均 Top-1 Accuracy において，下流モデル LSTM では移動場所予測・移動軌跡検索・移動軌跡補間で精度がそれぞれ 3.2%・2.7%・5.5%向上し，BERT で 4.3%・3.6%・3.9%向上した．都市間のスコア差は，都市 A と都市 B のグリッド数の違い (表 2) に起因するものと考えられる．都市 A はグリッド数が多く，ユーザごとの訪問グリッドの多様性が高いため，埋め込み表現の学習および下流タスクの予測がより困難になっていると考えられる．また，下流モデル間で絶対的な性能差は見られるものの，提案手法は LSTM・BERT いずれを用いた場合でも既存手法を上回っており，埋め込み表現の優位性が下流モデルの選択に依存しないことを示している．これらの結果は，意味的文脈と時間的文脈を統合したモデリングが，タスクや下流モデルに依存しない汎用的な場所埋め込み表現の構築に有効であることを示している．

4.4 追加検証

本節では，提案手法の設計選択の有効性を検証するため，2 種類の変種との比較実験を行う．1 つ目は意味的文脈モデリングの寄与を検証するための変種 (*w/o Semantics*)，2 つ目は移動の言語的表現に代えて時間情報と訪問頻度の離散化値の積を意味表現として用いる変種 (*Num Semantics*) である．その他の条件は 4.2 節の学習設定と同一である．

実験結果を図 4 に示す．図中のスコアは提案手法を 1 とした相対値で示している．*w/o Semantics* は提案手法と比較して全タスクでスコアが低下しており，意味的文脈モデリングの場所埋め込み表現の精度向上への寄与が確認できる．特に移動軌跡検索タスクでの低下が顕著であり，LSTM では Top-1 Accuracy が約 16%，BERT では約 23% 低下した．これは，軌跡検索においてユーザ固有の意味的文脈が識別に重要な役割を果たすことを示唆している．*Num Semantics* にも同様の傾向が確認され，ユーザ固有の訪問頻度と時間情報からなるテキストを言語空間に写像することで得られる意味的表現の豊かさが場所埋め込みの質向上に有効であることが示された．

さらに，提案手法により構築される場所埋め込み表現の特性を分析するため，テストユーザ 100 人分の軌跡中の移動場所の埋め込みを PCA により 2 次元に圧縮した結果を図 5 に示す．比較として，*Num Semantics* および *w/o Semantics* の埋め込みも同時に可視化した．提案手法 (Baseline) の埋め込みは，両変種と比較して空間的な広がりが大きく，ユーザごとの移動パターンがより分散した形で分布している．この傾向は，意味的文脈の統合により各場所の表現がユーザ固有の特徴をより豊かに反映していることを示唆しており，下流タスクにおける精度向上に寄与したと考えられる．ただし，本分析は視覚的な観察に基づくものであり，埋め込みの分布特性 (分散・クラスタ間距離など) と下流タスク精度の関係の定量的な分析は今後の課題である．

5. まとめ

本研究では，ユーザ固有の移動文脈に基づいて場所の意味の違いを捉える都市空間モデリング手法を提案した．提案手法は，移動軌跡の時間的文脈を Transformer Encoder でモデル化するとともに，移動をテキストで表現して言語空間に埋め込むことで意味的文脈を取り込み，Cross Attention によりこれらを統合し，各場所の埋め込み表現を構築する．実世界の人流データセットを用いた評価実験において，提案手法は移動場所予測，移動軌跡検索，移動軌跡補間の 3 つの下流タスクすべてで既存手法を上回る性能を示した．また，アブレーションスタディにより，意味的文脈モデリングの寄与が特に移動軌跡検索タスクにおいて顕著であることを確認し，提案手法の有効性を示した．

今後の課題として以下の 2 点を挙げる．第 1 に，本研究ではテンプレートベースのテキストによる意味表現を用い

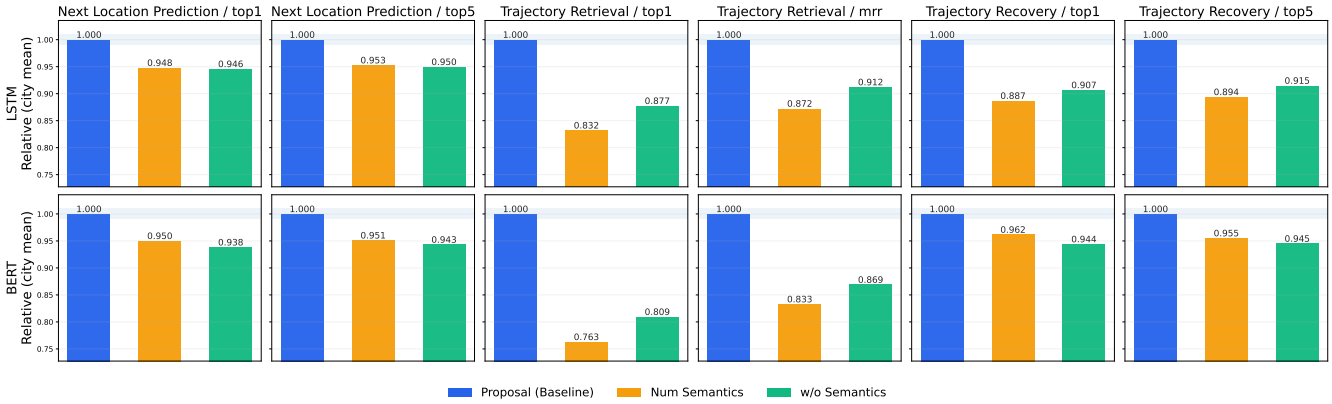

図 4 意味的文脈モデリングを数値表現に置き換えた変種 (*Num Semantics*)，意味的文脈モデリングを除いた変種 (*w/o Semantics*) との比較．スコアは提案手法を 1 とした相対値．

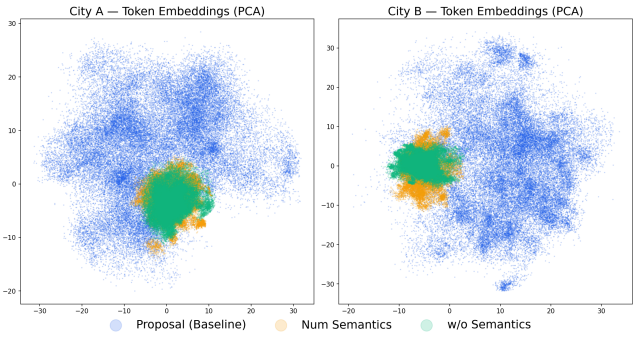

図 5 各場所の埋め込み表現の可視化 (PCA)

たが，LLM 等を用いてテンプレートを多様なパラフレーズへ変換する仕組みを構築する．これにより，言語空間における表現の連続性を高め，より豊かな意味的文脈の獲得が期待される．第 2 に，本研究の意味表現は移動ステップ単位で生成されるが，ユーザの長期的な行動傾向を抽象化した表現を取り込むことで，場所の意味をより安定的かつ包括的にモデル化できる仕組みの構築を目指す．

謝辞 本研究は，JST CREST(JPMJCR22M4)，JST RISTEX (JPMJRS23K)，内閣府 SIP3(JPJ012495) に支援いただいています．

参考文献

[1] Carlo Ratti, Dennis Frenchman, Riccardo Maria Pulselli, and Sarah Williams. Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis. *Environment and Planning B: Planning and Design*, 33:727–748, 2006.

[2] Shan Jiang, Joseph Ferreira, and Marta C Gonzalez. Activity-based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore. *IEEE Transactions on Big Data*, 3:208–219, 2017.

[3] Takahiro Yabe, Nicholas KW Jones, P Suresh C Rao, Marta C Gonzalez, and Satish V Ukkusuri. Mobile Phone Location Data for Disasters: A Review from Natural Hazards and Epidemics. *Computers, Environment and Urban Systems*, 94, 2022.

[4] Kazuyuki Shoji, Shunsuke Aoki, Takuro Yonezawa, and Nobuo Kawaguchi. Area Modeling Using Stay Information for Large-Scale Users and Analysis for Influence of COVID-19. *arXiv preprint arXiv:2401.10648*, 2024.

[5] 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 International Joint Conference on Artificial Intelligence (IJCAI)*, 2018.

[6] Shanshan Feng, Gao Cong, Bo An, and Yeow Meng Chee. POI2Vec: Geographical Latent Representation for Predicting Future Visitors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 102–108, 2017.

[7] Shenglin Zhao, Tong Zhao, Irwin King, and Michael R. Lyu. Geo-Teaser: Geo-Temporal Sequential Embedding Rank for Point-of-Interest Recommendation. In *Proceedings of the World Wide Web Conference (WWW)*, pages 153–162, 2017.

[8] Huaiyu Wan, Fuchen Li, Shengnan Guo, Zhong Cao, and Youfang Lin. Learning Time-Aware Distributed Representations of Locations from Spatio-Temporal Trajectories. In *International Conference on Database Systems for Advanced Applications*, pages 268–272. Springer, 2019.

[9] Toru Shimizu, Takahiro Yabe, and Kota Tsubouchi. Enabling Finer Grained Place Embeddings Using Spatial Hierarchy from Human Mobility Trajectories. In *Proceedings of the International Conference on Advances in Geographic Information Systems (SIGSPATIAL)*, pages 187–190, 2020.

[10] Yan Lin, Huaiyu Wan, Shengnan Guo, and Youfang Lin. Pre-training Context and Time Aware Location Embeddings from Spatial-Temporal Trajectories for User Next Location Prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4241–4248, 2021.

[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All You Need. *Advances in Neural Information Processing Systems (NeurIPS)*, pages 5998–6008, 2017.

[12] Takahiro Yabe, Kota Tsubouchi, Toru Shimizu, Yoshihide Sekimoto, Kaoru Sezaki, Esteban Moro, and Alex Pentland. YJMob100K: City-Scale and Longitudinal Dataset of Anonymized Human Mobility Trajectories. *Scientific Data*, 11(1):397, 2024.

[13] Shan Jiang, Ana Alves, Filipe Rodrigues, Joseph Fer-

reira Jr, and Francisco C Pereira. Mining Point-of-Interest Data from Social Networks for Urban Land Use Classification and Disaggregation. *Computers, Environment and Urban Systems*, 53:36–46, 2015.

[14] Víctor Soto and Enrique Frías-Martínez. Automated Land Use Identification Using Cell-phone Records. In *Proceedings of the 3rd ACM International Workshop on MobiArch*, pages 17–22, 2011.

[15] Sébastian Grauwin, Stanislav Sobolevsky, Simon Moritz, István Gódor, and Carlo Ratti. Towards a Comparative Science of Cities: Using Mobile Traffic Records in New York, London, and Hong Kong. In *Computational Approaches for Urban Environments*, pages 363–387. Springer, 2014.

[16] Angelo Furno, Marco Fiore, Razvan Stanica, Cezary Ziemlicki, and Zbigniew Smoreda. A Tale of Ten Cities: Characterizing Signatures of Mobile Traffic in Urban Areas. *IEEE Transactions on Mobile Computing*, 16(10):2682–2696, 2016.

[17] Vicente Vivanco Cepeda, Gaurav Kumar Nayak, and Mubarak Shah. GeoCLIP: Clip-Inspired Alignment between Locations and Images for Effective Worldwide Geo-Localization. *Advances in Neural Information Processing Systems (NeurIPS)*, 36:8690–8701, 2023.

[18] Konstantin Klemmer, Esther Rolf, Caleb Robinson, Lester Mackey, and Marc Rußwurm. SatCLIP: Global, General-purpose Location Embeddings with Satellite Imagery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 4347–4355, 2025.

[19] Xinglei Wang, Tao Cheng, Stephen Law, Zichao Zeng, Ilya Ilyankou, Junyuan Liu, Lu Yin, Weiming Huang, and Natchapon Jongwiriyanurak. Into the Unknown: Applying Inductive Spatial-Semantic Location Embeddings for Predicting Individuals' Mobility Beyond Visited Places. In *Proceedings of the ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL)*, pages 1046–1055, 2025.

[20] Ya Wen, Jixuan Cai, Qiyao Ma, Linyan Li, Xinhuan Chen, Chris Webster, and Yulun Zhou. MoRA: Mobility as the Backbone for Geospatial Representation Learning at Scale. In *International Conference on Learning Representations (ICLR)*, 2026.

[21] Jiawei Cheng, Jingyuan Wang, Yichuan Zhang, Jiahao Ji, Yuanshao Zhu, Zhibo Zhang, and Xiangyu Zhao. POI-Enhancer: An LLM-based Semantic Enhancement Framework for POI Representation Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.

[22] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 3982–3992, 2019.

[23] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Comput.*, 9(8):1735–1780, 1997.

[24] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT*, pages 4171–4186, 2018.

付　　録

A.1　データセット詳細

日ごと，時間ごとのデータ量の推移を図 A·1 に示す．日ごとのデータ量には週周期の変動パターンが観察され，この周期性に基づき曜日を推定した．

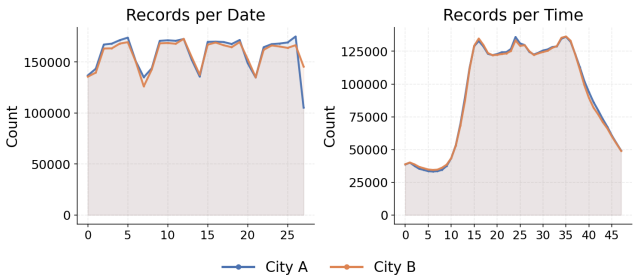

図 A·1　日ごと・時間ごとのデータ量

A.2　下流モデルのパラメータ設定

各下流タスクで用いる LSTM および BERT モデルの構成を表 A·1 に示す．すべての下流モデルは AdamW (学習率 1×10^{-4}，weight decay 1×10^{-3}，50 エポック，バッチサイズ 16) で最適化し，勾配クリッピングを適用する．

表 A·1　下流モデルのパラメータ

タスク	モデル	層数	ヘッド数	埋め込み次元
移動場所予測	LSTM	2	–	128
	BERT	2	8	128
移動軌跡検索	LSTM	2	–	128
	BERT	2	8	128
移動軌跡補間	LSTM	2	–	128
	BERT	2	8	128

A.3　比較手法のパラメータ設定

各比較手法は原論文の設定に従って PyTorch で実装し，提案手法と同一の学習設定を採用した (AdamW，学習率 1×10^{-4}，100 エポック，バッチサイズ 16)．各手法のモデル構成を表 A·2 に示す．

表 A·2　比較手法のパラメータ

モデル	層数	ヘッド数	埋め込み次元
Geo-Teaser [7]	–	–	128
TALE [8]	–	–	128
HIER [9]	4	–	128
CTLE [10]	4	8	128