

# NSP-BERT：滞在時間を考慮に入れた 効率的な Next Stay 予測手法

寺島 青<sup>1</sup> 田村 直樹<sup>1</sup> 庄子和之<sup>1</sup> 片山 晋<sup>1</sup> 浦野 健太<sup>1</sup> 米澤 拓郎<sup>1</sup> 河口 信夫<sup>1</sup>

**概要：**スマートフォンの普及に伴い、大規模な位置情報データの収集が容易となり、位置情報データを用いた研究が盛んに行われるようになった。大規模な位置情報データを用いて人間の移動傾向の分析・理解をし、その結果を都市計画、交通計画、災害対策等に活用する取り組みが行われている。滞在履歴から滞在の周期性、高頻度で滞在する場所等を見つけ、滞在傾向の特徴を捉えると、ユーザ毎に将来の滞在エリアとそのエリアでの滞在時間が予測でき、各エリアでのマーケティング戦略等に活用できる。週単位や月単位といった様々な滞在の周期性が混在する滞在履歴から、滞在傾向の特徴を詳細に捉えるためには深層学習モデルが有効であると考えられる。しかし、RNN や LSTM は滞在履歴の長期の依存関係の考慮を不得意とし、逐次的な予測手法で多くの時間を要する。そこで、滞在履歴の長期の依存関係を考慮し、かつ効率的に将来の滞재를予測するために NSP-BERT(Next Stay Prediction BERT) を提案する。NSP-BERT は Transformer の Attention 機構により滞在履歴の長期の依存関係を考慮でき、逐次的な予測を行わず Encoder から出力されるベクトルと FC Layer で滞在エリアと滞在時間の予測を行うため効率性が高い。また、同じエリアでの連続する滞재를滞在時間としてまとめて予測し、従来の時間刻みの予測で必要であった計算コストを削減した。実世界の位置情報データを用いて NSP-BERT と LSTM の滞在予測精度および効率性を比較した。滞在エリア一致率、滞在時間の一致率、GEOBLEU、DTW のスコア全てにおいて NSP-BERT が LSTM を上回り、かつ学習時間も少なかったことから NSP-BERT の予測性能および効率における優位性を示した。

## NSP-BERT: An Efficient Method for Next Stay Prediction Considering Duration Time

HARU TERASHIMA<sup>1</sup> NAOKI TAMURA<sup>1</sup> KAZUYUKI SHOJI<sup>1</sup> SHIN KATAYAMA<sup>1</sup>  
KENTA URANO<sup>1</sup> TAKURO YONEZAWA<sup>1</sup> NOBUO KAWAGUCHI<sup>1</sup>

### 1. はじめに

スマートフォンの普及に伴い、大規模な位置情報データの収集が容易となり、位置情報データを用いた研究が盛んに行われるようになった。大規模な位置情報データを用いて人間の移動傾向の分析・理解をし、その結果を都市計画、交通計画、災害対策等に活用する取り組みが行われている [1][2][3]。

人間の日常生活での滞には周期性がある場合が多く、

滞在先の候補を絞ることができる [4][5]。図 1 のように平日に会社員は、自宅→駅→会社→駅→自宅のように滞場所を移し、学生は自宅→学校→塾→自宅のように滞場所を移すと考えられる。このような滞場所の推移から滞傾向の特徴を捉えると、将来どのエリアにどのくらいの時間滞するか予測する手掛かりとなる。将来の滞を予測は、各エリアでのマーケティング戦略等で活用できる。例えば、あるエリアにおいて昼を跨いで滞するユーザが複数予測できた場合はそのエリアでのランチ営業の需要が高まると考えられる。加えて、ユーザの属性(男性、女性、子供、高齢者など)がわかっている場合は、訪れるであろう属性に合った供給が可能となる。

人間の滞には週単位や月単位といった様々な周期性が

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

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

混在しており、混在する様々な周期性を予測に反映させるためには、滞在履歴の長期の依存関係の考慮が必要となる。また、例に挙げたマーケティング戦略に予測結果を用いる場合、最新の滞在傾向の特徴をモデルに反映させることが重要であり、学習データに滞在履歴を順次加えて、モデルを更新し続ける必要があると考える。モデルを高頻度で再学習させる場合、滞在履歴から滞在傾向の特徴を効率的に学習できるモデルが必要である。

複雑な時系列データの傾向分析や予測問題には深層学習を用いる研究が多く [6][7]、移動予測の既存研究にも、RNN, LSTM といった深層学習モデルを用いた手法が存在する [8][9]。週単位や月単位といった様々な周期性が混在する滞在履歴から特徴を詳細に捉えるためには深層学習モデルが有効であり、他のユーザの移動傾向も考慮した予測により、予測精度は向上した [10]。しかし、深層学習モデルは予測精度を向上させる一方で、学習および予測に多くの時間を要するという課題が存在する。移動予測の研究では、時間刻みで滞在エリアを予測するものが多く、予測頻度に応じて学習および予測時間が影響を受け、図 2 のような逐次的な予測を行う RNN や LSTM 等では多くの時間を要する。また、RNN, LSTM には移動履歴の長期の依存関係の考慮が不得意という課題も挙げられ、月単位等の長期の周期性を捉えることが困難であると考えられる。

本研究では、滞在履歴の長期の依存関係を考慮した効率的な予測を行うため、BERT(Bidirectional Encoder Representations from Transformers)[11] のアーキテクチャを参考にした滞在予測モデルを提案する。BERT は Transformer Encoder[12] を用いた深層学習モデルであり、自然言語処理の分野において翻訳タスク等で活躍する。提案モデルは BERT を参考に将来の滞在予測に応用したため Next Stay Prediction BERT(NSP-BERT) と命名する。NSP-BERT は逐次的な予測を行わず、Encoder から出力される各ベクトルを FC Layer により、滞在エリアおよび滞在時間の確率分布に変換し予測を行う。また、Transformer の Attention 機構により滞在履歴の長期の依存関係を考慮した予測が可能となる。加えて、NSP-BERT は滞在エリアと滞在時間を予測するため、同じエリアでの連続する滞在を滞在時間としてまとめて予測し、従来の時間刻みのエリア予測で必要であった計算コストを削減した。

NSP-BERT の滞在予測精度および効率性の評価を行うための実験を行った。名古屋市における全 5 万人の 3 ヶ月間に及ぶ長期滞在履歴のデータセットを用いて、全 5 万人のユーザのうち 1 万人の最後の 15 日間の滞在を予測した。実験は、半月単位でエリア毎にどんなユーザがいつどれくらい来るか予測し、各エリアにおける飲食や物販等の需要を推定するシチュエーションを想定した。まず、NSP-BERT と LSTM で予測精度および学習時間を比較した。NSP-BERT は滞在エリアの一致率、滞在時間の一致率、

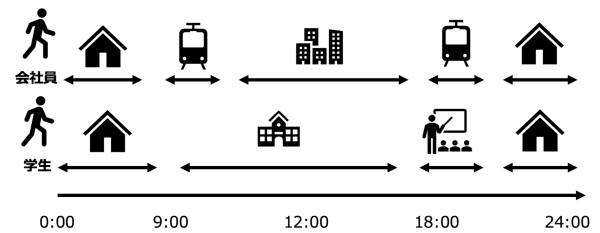


図 1 日常の滞在パターン

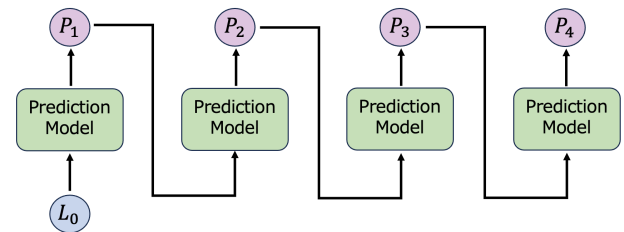


図 2 逐次的予測

GEOBLEU[13], DTW[14] 全てにおいて LSTM より高いスコアを示し、1epoch あたりの学習時間も NSP-BERT の方が少なかったため、予測精度および効率における優位性を示した。次に、各滞在を事前に 30 分刻みのタイムスロットに分割し学習およびエリア予測した場合と、滞在エリアと滞在時間を合わせて予測し、予測滞在時間に応じて予測滞在エリアをタイムスロットに変換した場合を比較した。滞在を事前にタイムスロットに分割した場合の方が、滞在エリアの一致率が約 4%高かったものの、学習時間が 20 倍以上となり、滞在エリアと滞在時間を合わせた予測の効率性を示した。

本研究の貢献は以下のようにまとめられる。

- 滞在履歴の長期の依存関係を考慮した効率的な滞在エリアおよび滞在時間の予測が可能な NSP-BERT を提案
- NSP-BERT と LSTM を比較し、NSP-BERT の予測精度および効率における優位性を示した
- 滞在エリアおよび滞在時間を予測し、同じエリアでの連続する滞在を滞在時間としてまとめて予測し、時間刻みのエリア予測に必要な計算コストを削減

本論文の構成は次に示す通りである。まず 2 章で既存の移動先予測に関する研究における課題を整理する。3 章では提案するモデルについて説明し、4 章では提案モデルの評価を行う。最後に 5 章でまとめと今後の展望について述べる。

## 2. 関連研究

### 2.1 移動履歴からの特徴抽出

人間の移動傾向は同じ属性(会社員, 学生等)において共通点はあるものの, 個人に依存している部分が多い. Suzuki らの研究では, 個々人の移動履歴からそのユーザの移動傾向の特徴を抽出し, 将来の移動を予測している [15]. Jiaxin らは GNN(Graph Neural Network) を用いてユーザ毎にグラフを作成した移動予測モデルを提案している [16]. これらの研究ではユーザ毎の移動履歴に焦点を当てて移動傾向の特徴を抽出し, 他ユーザの移動を考慮する計算コストを抑えている. しかし, ユーザ毎に移動傾向の特徴抽出を試みた場合, ユーザ毎のデータ量に予測精度が依存し, データ量が少ないユーザの予測精度が低くなるという課題もある. そこで, 他のユーザの移動履歴を考慮し, 個々人の移動履歴では不足する情報を補った予測がなされてきている [17].

### 2.2 深層学習を用いた移動予測

多数のユーザの様々な移動履歴から移動傾向の特徴を詳細に抽出するために, 移動先予測における研究でも RNN や LSTM といった深層学習モデルを活用したものも多く存在する [8][9]. 深層学習の活用により, 予測するユーザの移動傾向のみでなく, 他のユーザの移動傾向を考慮した予測が可能となる. また, 週単位や月単位といった複雑な周期性を捉えることも可能となった. 一方で, RNN, LSTM では移動履歴の長期の依存関係を考慮した予測が不得意であり月単位等の長期の周期性を捉えることが困難であると考えられる. また, 逐次的な予測方法で多くの時間を要するという課題も挙げられる.

移動履歴の長期の依存関係を考慮した予測を可能にするため, 近年 Transformer が移動予測に活用されつつある. Transformer は Attention 機構により, 時系列データをより詳細に捉えることが可能となり, 移動履歴の長期の依存関係を考慮した予測が可能となる [18][19][20]. 移動予測において Yan らの研究では, Transformer の Encoder を活用した CTLE(Context and Time Aware Location Embeddings) モデルを提案している [21]. CTLE モデルは移動履歴を Embedding するモデルであり, CTLE モデルで Embedding した移動履歴を LSTM 等の下流モデルに入力し, 移動予測を行いモデルの性能を評価している. しかし, 長期の移動履歴に対して高い能力を発揮する一方で, Transformer は RNN や LSTM といった他の深層学習モデルに比べて使用するパラメータ数も多く, 計算時間の増加やメモリの使用量の増加といった懸念も存在する. そこで, 長期時系列予測問題に対して, 内部計算を効率的に行う Transformer ベースのモデルを提案する研究も多くなされてきている [22][23].

このように, 深層学習モデルの使用に際して, 効率性は重要であると考える.

我々は以前の研究で, Transformer Encoder と FC Layer で将来の移動を予測し, Decoder や下流モデルを用いない予測手法を提案した [24]. この手法では, 移動履歴の長期の依存関係の考慮を可能とし, 逐次的でない効率的な予測を可能としている. しかし, この研究では時間刻みで滞在エリアを予測しており, 同じエリアでの連続する滞在を別々に予測している. また, 他のモデルとの比較検証が実施できていない. そこで, 本研究では我々の既存の手法をもとに, 将来の滞在エリアと滞在時間を予測し, 同じエリアでの連続する滞在を滞在時間としてまとめて予測する手法を提案した. また, 提案モデルと逐次的な予測を行う LSTM の予測精度および効率性を比較検証した.

## 3. 提案手法

### 3.1 モデルの概要

提案する NSP-BERT の概要を図 3 に示す. NSP-BERT は Embedding Layer, Transformer Encoder, FC Layer で構成されている. まず, 滞在履歴 (Stay: Stay\_Area, Stay\_Day, Stay\_Time, Duration\_Time, Time\_diff) のうち一部を実在しないエリアと滞在時間によりマスクし, 滞在履歴が一部マスクされた状態で Embedding Layer に入力し, 滞在毎の滞在ベクトルを生成する. 次に生成された時系列滞在ベクトルを Transformer Encoder に入力し, 滞在ベクトル同士の関係が学習された滞在毎の滞在特徴ベクトルを生成する. 最後に, 滞在履歴のうちマスクされた滞在の滞在特徴ベクトルのみを FC Layer に入力し, それらの滞在特徴ベクトルを滞在エリアと滞在時間の確率分布にそれぞれ変換する. このように, Transformer の Attention 機構を活用し滞在履歴の長期の依存関係の考慮を可能とし, また Transformer Encoder から出力される滞在特徴ベクトルと FC Layer から滞在エリアと滞在時間を予測し, Decoder や下流モデルを用いず効率化を図っている.

### 3.2 Embedding Layer

Embedding Layer では図 3 の右側のように各滞在からその滞在の特徴を表す滞在ベクトルを生成する. 滞在履歴から各滞在毎に滞在エリア (Stay\_Area), 日付 (Stay\_Day), 滞在開始時間 (Stay\_Time), エリアでの滞在時間 (Duration\_Time), 前の滞在との時間差 (Time\_diff) をそれぞれ Embedding によりベクトル化し, 足し合わせることで各滞在における滞在ベクトルを生成する. ここで, ベクトル同士は横方向に結合するのではなく, ベクトルの加算を行い, 次元数の増加に伴う計算量の増加を抑えている. また, 各滞在中の時系列における相対的な位置関係を示すため, 前の滞在との時間差を滞在ベクトルに加えている.

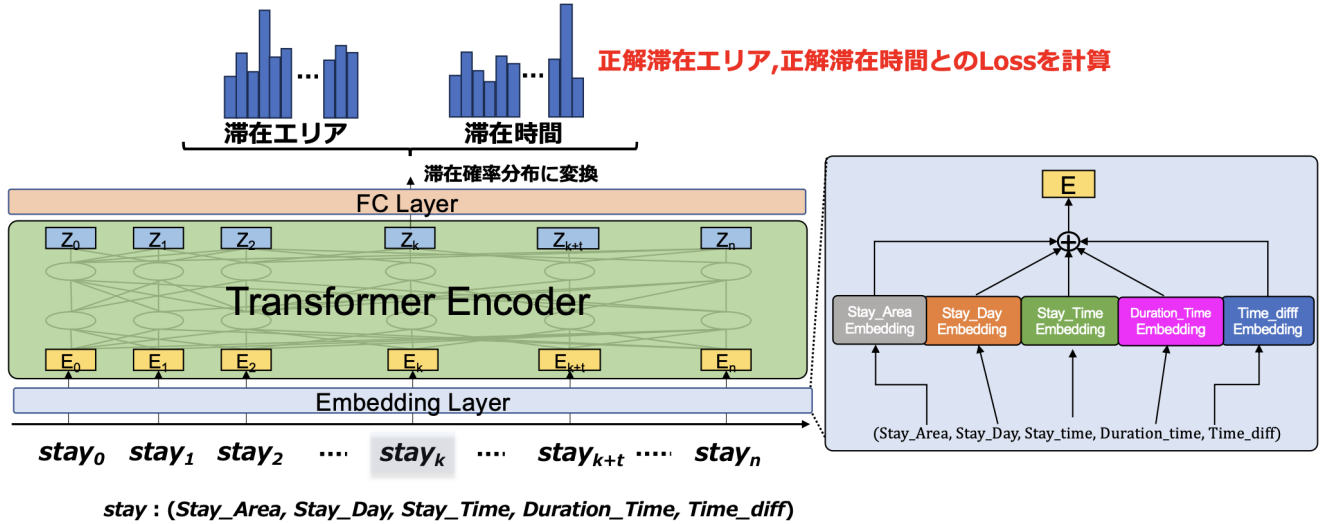


図 3 NSP-BERT 全体像

### 3.3 学習および予測手法

学習時には、バッチ毎にユーザの滞在履歴のうちランダムな位置の連続する複数回の滞在をマスクし、予測時には予測期間の滞在をマスクと見なし学習時と同じ状況を作る。学習時、予測時ともに Transformer Encoder から出力される滞在毎の滞在特徴ベクトルのうちマスク部分のベクトルのみを FC Layer に入力し、滞在エリア確率分布 (Probability\_Area) と滞在時間確率分布 (Probability\_Time) にそれぞれ変換する。学習時には、以下の式のようにそれぞれの確率分布と正解エリア (True\_Area) および正解滞在時間 (True\_Time) のクロスエントロピーロスを計算し、滞在エリアのロスと滞在時間のロスを足し合わせたものを最小化するように学習を進める。

$$Loss_{Area} = CrossEntropy(Probability_{Area}, True_{Area})$$

$$Loss_{Time} = CrossEntropy(Probability_{Time}, True_{Time})$$

$$Loss = \frac{Loss_{Area} + Loss_{Time}}{2}$$

予測時には、以下の式のように変換された確率分布のうち滞在確率が最大のエリアを予測エリアとして出力し、同様に確率が最大の滞在時間を出力する。

$$Predict\_Stay\_Area = \arg \max(Probability_{Area})$$

$$Predict\_Duration\_Time = \arg \max(Probability_{Time})$$

## 4. 評価実験

### 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。本研究ではエリア毎での滞在予測を行うため、[25] の手法を用い、元データからユーザ ID、緯度、経度、滞在開始時間、滞在経過時間を取得し、各滞在を緯度経度から 100m × 100m メッシュのエリアでの滞在中に変換した。また、滞在開始時間は 1 日を 30 分刻みの 48 区分に分割しており、滞在経過時間も 30 分刻みとなっている。実験では名古屋市の 2019/04/01 から 2019/06/30 の期間において、30 分未満および 24 時間以上の滞在を除外し、30 分から 24 時間以内の滞在履歴を 100 レコード以上持つ全 5 万人分のデータを用いる。全データで 10,913,954 レコードあり、エリア数は 53,508 である。

本研究では、各エリアでの将来訪れるであろうユーザに対するマーケティング戦略のシチュエーションを想定し、全 5 万人のうち 1 万人の最後の 15 日間の滞在予測を行う。学習には予測対象でない 4 万人の 91 日分の滞在履歴と予測対象 1 万人の予測期間でない 76 日分の滞在履歴を用いる。滞在エリアは全 53,508 エリアから予測し、滞在時間は 30 分刻みで予測する。実験は RTX 6000 Ada (メモリ 48GB) 上で実施した。

### 4.2 評価指標

本実験において、滞在予測の精度を評価する指標として Area\_Accuracy, Time\_Accuracy, GEOBLEU, DTW の 4 つを用いる。評価指標のうち GEOBLEU, DTW は予測エリアと正解エリアと空間的距離もスコアに考慮されている。

– 滞在エリアの一致率 (Area\_Accuracy)

$$Area\_Accuracy = \left( \frac{\text{正解予測エリア数}}{\text{予測エリア数}} \right) \times 100$$

表 1 NSP-BERT, LSTM 予測結果

Model	Area Accuracy (%)	Time Accuracy (%)	GEOBLEU	DTW	Train Time/Epoch (s)
NSP-BERT	48.31	37.04	0.5330	2.5484	42
LSTM	34.39	34.50	0.3640	4.7041	121

表 2 ハイパーパラメータの設定

パラメータ	NSP-BERT	LSTM
batch_size	16	16
embed_size	128	128
epoch	100	100
Layer	4	4
Transformer_heads	8	None

表 3 滞在予測と時間刻みでの予測の比較

Model	Area_timeslot Accuracy (%)	Train Time/Epoch (s)
NSP-BERT	58.68	42
時間刻みでの予測	62.21	1024

#### – 滞在時間の一致率 (Time Accuracy)

$$\text{Time Accuracy} = \left( \frac{\text{正解予測滞在時間数}}{\text{予測滞在時間数}} \right) \times 100$$

#### – GEOBLEU[13]

自然言語処理の分野で用いられる BLEU を参考にした評価指標。局所的な予測精度を評価するためのもので、値が大きい方がスコアは良く、予測結果と正解が完全に一致すると値は 1 になる。

#### – DTW[14]

予測軌跡と正解軌跡の時系列間の類似度を評価する指標。値が小さい方がスコアは良く、予測軌跡と正解軌跡が完全に一致すると値は 0 になる。

### 4.3 NSP-BERT の予測精度および効率性の評価

NSP-BERT および LSTM を用いて、予測対象ユーザー 1 万人の 15 日間の滞在エリアおよび滞在時間予測を行った。学習時において、NSP-BERT は滞在履歴のうちランダムな日にちから連続する 15 日間の滞在をマスクし、LSTM ではバッチ内でデータが最小のユーザーに応じて滞在履歴の最後尾からランダムな数の滞在をマスクして学習した。パラメータを表 2 に、スコアを表 1 に示す。

表 1 より、NSP-BERT は予測精度に関する全ての評価指標において LSTM より高いスコアを達成し、1epoch あたりの学習時間も少なかったことから、LSTM が不得意とする滞在履歴の長期の依存関係の考慮を可能とし、高い効率性を示した。図 4 に、NSP-BERT と LSTM が 100 回以上予測したエリアにおけるそのエリアでの全予測回数に対する予測間違え率と、学習期間における各エリアのデータ量をそれぞれ可視化している。バーの高さが各エリアにおける予測間違え率を示しており、高いほど予測間違えが多かったことを示す。底面のヒートマップは学習時における各エリアのデータ量を示しており、色が濃いエリア程データ量が多いことを示している。両モデルともに、学習時に

多く出現したエリアを予測しやすい傾向にあり、そのエリアでの予測間違え率が高くなっている。LSTM の方が高頻度で予測されるエリアでの予測間違え率が高く、学習時におけるエリア毎のデータ量に影響を受け、ユーザー毎の滞在傾向の特徴を NSP-BERT に比べて予測に反映できなかったため、滞在エリアの一致率において NSP-BERT のスコアが 10%以上よくなったと考えられる。また、図 5 より、滞在時間の予測分布には大きな差はなく、Time Accuracy が同程度の値であることにも矛盾はなかった。

### 4.4 滞在予測の効率性評価

NSP-BERT を用いて滞在エリアと滞在時間を予測した場合と、事前に各滞在を 30 分刻みのタイムスロット毎に分割し学習を行い、30 分毎の滞在エリア予測を行った場合を比較した。滞在履歴を 30 分刻みのタイムスロット毎に分割した時、データセットの総レコード数は 128,360,745 となった。また、滞在を分割する前の入力データの最大データ長は 811 であったが、分割後の最大データ長は 4,120 となった。NSP-BERT では予測滞在時間に応じて予測滞在エリアを 30 分刻みの各タイムスロットに当てはめた。比較結果を表 3 に示す。Area\_timeslot Accuracy はタイムスロットに分割後の予測エリアと正解エリアの一致率を示している。使用したパラメータは表 2 と同様である。滞在をタイムスロット毎に分割して学習および予測の方が滞在エリアの一致率が高いが、学習データ量の増加により学習時間が 20 倍以上となっている。この結果から滞在エリアと滞在時間を合わせた予測は効率性の観点において優位であると考えられる。一方で、タイムスロットで予測した方が予測精度が高くなる理由は以下の 2 つが挙げられる

- 学習データ数の増加により滞在履歴が密になり、モデルが滞在エリア同士の関係をより詳細に捉えられたこと。
- 滞在エリアと滞在時間を合わせて予測する場合、Transformer Encoder からの出力されたベクトルに滞在エリアと滞在時間の特徴が混在し、予測タスクが複雑化したこと。

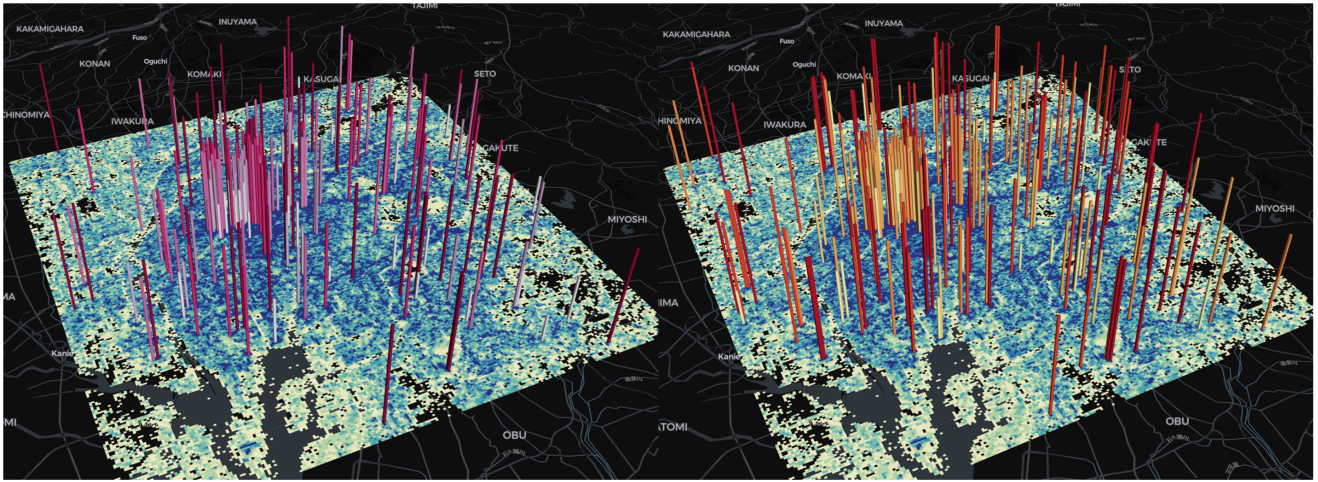


図 4 NSP-BERT(左), LSTM(右) の各エリアにおける予測間違い率 (バーの高さ) および各エリアの学習時のデータ量 (底面のヒートマップ)

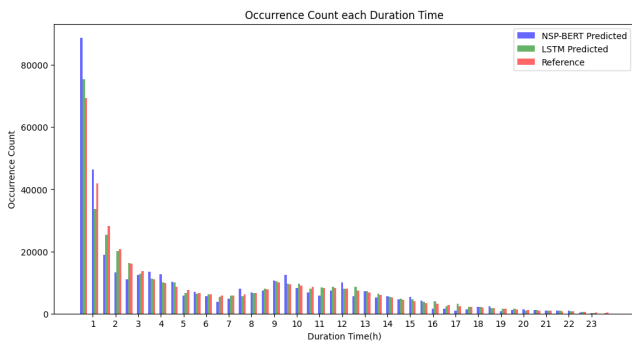


図 5 予測および正解滞在時間分布

NSP-BERT では前の滞在との時間差を入力し、各滞在が滞在履歴のうちどこに位置するかを相対的に示したが、滞在ベクトルに付与する情報量が不十分であったと考える。また、学習時における滞在エリア予測と滞在時間予測のロスの計算方法の検討により更なる精度の向上が可能であると考え。

## 5. まとめと今後の展望

本研究では、大規模な滞在履歴を用いて、人間の周期的な滞在傾向の特徴を捉え、将来の滞在エリアとそのエリアでの滞在時間の予測を行った。週単位や月単位の周期性といった複雑な滞在履歴から滞在傾向の特徴を詳細に捉えるためには、深層学習モデルが有効であると考え NSP-BERT を提案した。NSP-BERT は RNN や LSTM が不得意とする滞在履歴の長期の依存関係の考慮を可能とし、また逐次的でない効率的な予測を可能とした。実世界の位置情報データを用いて、NSP-BERT と LSTM で滞在予測精度および効率性を比較した。滞在エリアの一致率、滞在時間の一致率、GEOBLEU、DTW 全ての予測精度の評価指標において NSP-BERT が LSTM を上回り、学習時間も NSP-BERT の方が短かったことから予測性能および効率における優位

性を示した。また、同じエリアでの連続する滞在を滞在時間としてまとめて予測し、従来の時間刻みのエリア予測で必要であったコストを削減した。今後の展望として、滞在エリアと滞在時間を同時に精度良く予測するための学習方法の更なる検討を行いたいと考えている。また、予測結果を空間的に分析し、予測モデルの性能向上に繋がりたいと考えている。

## 謝辞

本研究の一部は NEDO 委託研究 (JPNP23003, JPNP23025) NICT 委託研究 (22C01), JST CREST(JPMJCR21F2, JPMJCR22M4), JPSP 科学研究費補助金 (22H03580, 22K18422) に支援いただいている。また、データ提供にご協力いただ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] 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*, Vol. 33, No. 5, pp. 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*, Vol. 3, No. 2, pp. 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*, 2022.
- [4] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*, Vol. 453, No. 7196, pp. 779–782, 2008.
- [5] Maria Luisa Damiani, Andrea Acquaviva, Fatima Hachem, and Matteo Rossini. Learning behavioral representations of human mobility. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, pp. 367–376, 2020.

- [6] Mohammad Amin Morid, Olivia R. Liu Sheng, and Joseph Dunbar. Time series prediction using deep learning methods in healthcare. 2023.
- [7] Peipei Wang, Xinqi Zheng, Gang Ai, Dongya Liu, and Bangren Zhu. Time series prediction for the epidemic trends of covid-19 using the improved lstm deep learning method: Case studies in russia, peru and iran. *Chaos, Solitons and Fractals*, 2020.
- [8] Abdulrahman Al-Molegi, Mohammed Jabreel, and Baraq Ghaleb. Stf-rnn: Space time features-based recurrent neural network for predicting people next location. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–7, 2016.
- [9] Dejiang Kong and Fei Wu. Hst-lstm: A hierarchical spatial-temporal long-short term memory network for location prediction. In *IJCAI*, Vol. 18, pp. 2341–2347, 2018.
- [10] Haru Terashima, Naoki Tamura, Kazuyuki Shoji, Shin Katayama, Kenta Urano, Takuro Yonezawa, and Nobuo Kawaguchi. Human mobility prediction challenge: Next location prediction using spatiotemporal bert. In *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge*, pp. 1–6, 2023.
- [11] 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 2018*, pp. 4171–4186, 2018.
- [12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, Vol. 30, pp. 5998–6008, 2017.
- [13] Toru Shimizu, Kota Tsubouchi, and Takahiro Yabe. Geoblu: Similarity measure for geospatial sequences, 2022.
- [14] Pavel Senin. Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, Vol. 855, No. 1-23, p. 40, 2008.
- [15] Masahiro Suzuki, Shomu Furuta, and Yusuke Fukazawa. Personalized human mobility prediction for humob challenge. In *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge*, pp. 22–25, 2023.
- [16] Jiaxin Du and Xinyue Ye. Batch and negative sampling design for human mobility graph neural network training. In *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge*, pp. 47–50, 2023.
- [17] Josh Jia-Ching Ying, Wang-Chien Lee, Tz-Chiao Weng, and Vincent S Tseng. Semantic trajectory mining for location prediction. In *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems*, pp. 34–43, 2011.
- [18] Tomoki Choya, Naoki Tamura, Shin Katayama, Kenta Urano, Takuro Yonezawa, and Nobuo Kawaguchi. Crowdflowtransformer: Capturing spatio-temporal dependence for forecasting human mobility. In *2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, pp. 496–501, 2023.
- [19] He Li, Shiyu Zhang, Xuejiao Li, Liangcai Su, Hongjie Huang, Duo Jin, Linghao Chen, Jianbin Huang, and Jaesoo Yoo. Detectornet: Transformer-enhanced spatial temporal graph neural network for traffic prediction. In *Proceedings of the 29th International conference on advances in geographic information systems*, pp. 133–136, 2021.
- [20] Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. *arXiv preprint arXiv:2310.10688*, 2023.
- [21] 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*, Vol. 35, pp. 4241–4248, 2021.
- [22] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35, pp. 11106–11115, 2021.
- [23] Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*, 2022.
- [24] 寺島青, 田村直樹, 庄子和之, 片山晋, 浦野健太, 米澤拓郎, 河口信夫. Bert に基づく都市における移動エリア予測手法の提案. 研究報告モバイルコンピューティングと新社会システム (MBL), Vol. 2024-MBL-111, No. 8, pp. 1–7, 2024.
- [25] Sae Iwata, Tomoyuki Nitta, Toshinori Takayama, Masao Yanagisawa, and Nozomu Togawa. A stayed location estimation method for sparse gps positioning information. In *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*, pp. 1–5, 2017.