LEMU: Personalized Urban Location Modeling through Mobility-Derived Temporal and Semantic Contexts

Haru Terashima
haru@ucl.nuee.nagoya-u.ac.jp
Nagoya University
Nagoya, Aichi, Japan

Kota Tsubouchi
ktsubouc@lycorp.co.jp
LY Corporation
Chiyoda, Tokyo, Japan

Shuji Yamaguchi
shyamagu@lycorp.co.jp
LY Corporation
Chiyoda, Tokyo, Japan

Naoki Tamura
tam@ucl.nuee.nagoya-u.ac.jp
Nagoya University
Nagoya, Aichi, Japan

Kazuyuki Shoji
shoji@ucl.nuee.nagoya-u.ac.jp
Nagoya University
Nagoya, Aichi, Japan

Nobuo Kawaguchi
kawaguti@nagoya-u.jp
Nagoya University
Nagoya, Aichi, Japan

Takuro Yonezawa
takuro@nagoya-u.jp
Nagoya University
Nagoya, Aichi, Japan

Abstract

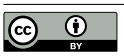
The same street corner can serve as a familiar anchor for one person while being just a passing stop for another; hence, personalized urban services would benefit from representing locations through the lens of user-specific semantics. However, most existing location modeling methods capture primarily general spatiotemporal mobility patterns from GPS trajectories and fail to represent user-specific meanings due to the absence of semantic supervision. We propose LEMU, an urban location modeling method that integrates mobility-derived temporal and semantic contexts. LEMU translates raw movements into concise natural language sentences, thereby leveraging general linguistic priors from pretrained language encoders to supply the semantics that numerical trajectories lack. These semantic representations are fused with Transformer-based temporal contexts through a cross-attention mechanism to construct personalized location embeddings. On a large-scale human mobility dataset, LEMU outperforms competitive baselines on the downstream task of *Next Location Prediction*, improving Top-1 Accuracy by 4.3 percentage points over the best-performing one. This result suggests that incorporating user-specific semantics is beneficial for learning expressive location embeddings.

CCS Concepts

• **Information systems** → **Spatial-temporal systems**; • **Human-centered computing** → **Ubiquitous and mobile computing**.

Keywords

Location Modeling, Human Mobility, Urban Computing



This work is licensed under a Creative Commons Attribution 4.0 International License.
SIGSPATIAL '26, Riverside, CA, USA
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2950-8/2026/11
https://doi.org/10.1145/3841645.3843058

ACM Reference Format:
Haru Terashima, Kota Tsubouchi, Shuji Yamaguchi, Naoki Tamura, Kazuyuki Shoji, Nobuo Kawaguchi, and Takuro Yonezawa. 2026. LEMU: Personalized Urban Location Modeling through Mobility-Derived Temporal and Semantic Contexts. In *The 34th ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '26), November 03–06, 2026, Riverside, CA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3841645.3843058

1 Introduction

The same location can be a daily anchor for one person, whereas it is merely a passing stop for another; hence, personalized urban services would benefit from representing locations through the lens of individual experience. Numerous studies in urban computing have proposed methods for learning location embeddings from human mobility data, supporting a wide range of urban applications [3, 4, 6, 8, 10, 11]. However, existing methods primarily capture spatiotemporal patterns shared across users, and thus fail to reflect the role a location plays for each individual user. This limitation stems from the fact that large-scale GPS datasets consist of numerical records (coordinates and timestamps) and contain no movement-purpose labels, making direct semantic supervision unavailable. While recent LLM-based methods incorporate textual features into mobility modeling, they anchor these features at the location level and overlook the user-specific roles that emerge from individual mobility behaviors [1, 2]. This motivates projecting user-level mobility signals into a language-induced semantic space, where pretrained linguistic priors can supplement the semantic signals absent from raw numeric features.

We propose **LEMU** (Location Embedding via Mobility-derived User-specific Contexts), an urban location modeling method that constructs personalized location embeddings by jointly capturing mobility-derived temporal and semantic contexts. The *semantic context* models the user-specific roles of locations by converting each movement into a concise sentence describing its temporal attributes and relative location importance, and encoding it with a pretrained

language encoder, without relying on external semantic annotations. The *temporal context* characterizes periodic mobility patterns by modeling sequences of visited locations and timestamps with a Transformer [7]. These representations are integrated through a cross-attention fusion mechanism, producing expressive location embeddings for a downstream mobility task.

We evaluate LEMU on LYMob-4Cities [9], a large-scale open dataset of human mobility trajectories in Japan, comparing it against existing location embedding methods on the downstream task of *Next Location Prediction*. LEMU improves Top-1 Accuracy over the best-performing baseline by 4.3 percentage points with the BERT-based downstream model and by 3.2 percentage points with the LSTM-based one, demonstrating that personalized location representations improve predictive performance.

2 Methodology

We propose **LEMU** (Location Embedding via Mobility-derived User-specific Contexts), with an overview illustrated in Figure 1. LEMU constructs personalized location embeddings by modeling user mobility trajectories from two perspectives: (i) *Semantic Mobility Context*, which captures the user-relative role of a location, characterized by temporal visit attributes and relative location importance rather than functional meaning; and (ii) *Temporal Mobility Context*, which captures periodic and sequential patterns over the sequences of visited locations and timestamps. These two representations are learned separately and then integrated. We denote the mobility trajectory of user u as $M_u = [(l_i, t_i) \mid i = 1, \ldots, K]$, a sequence of K movement steps, where l_i is the location visited at step i and t_i the time at which the movement occurs.

2.1 Semantic Context Modeling

LEMU captures semantic context by transforming raw mobility trajectories into natural language sentences and encoding them with a pretrained language encoder, thereby leveraging linguistic priors embedded in the language space. Each movement step i is mapped to a sentence s_i following a template that combines temporal attributes with the importance of the visited location to the user:

"[Day of the week/Time] at [Importance level]."

For example, a movement at 8:00 on Monday to a location the user visits almost every day yields "Monday 8:00 at Core Place." The importance level is derived from the user's relative visit frequency to the location, defined for a location l and user u as:

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

where $c_u(l)$ is the number of visits by user u to location l, and $\mathcal{L}_u$ is the set of all locations visited by user u. Each location is then assigned one of five importance levels, *Core Place*, *Subcore Place*, *Frequently Visited*, *Moderately Visited*, and *Rarely Visited*, according to its percentile rank within the user's distribution of $F_u(l)$, at thresholds 0.99, 0.95, 0.85, and 0.50. These thresholds were set through preliminary analysis of visit-frequency distributions to avoid extreme imbalance across levels.

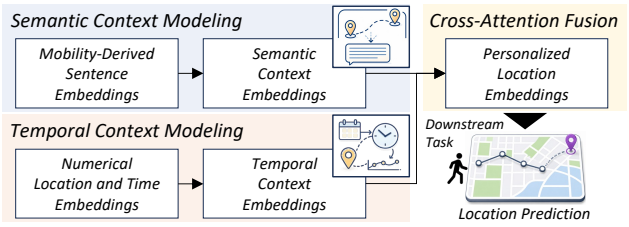


Figure 1: Overview of LEMU. Mobility trajectories are encoded into semantic and temporal contexts, which are integrated by cross-attention fusion to construct personalized location embeddings.

For a trajectory of K movement steps, each sentence is encoded and projected into the semantic representation of step i as:

$$\tilde{\mathbf{s}}_i = f_{sa}\big(f_{le}(s_i)\big) + \mathbf{p}_i, \qquad \tilde{\mathbf{s}}_i \in \mathbb{R}^d, \tag{2}$$

where f_{le} is a frozen pretrained language encoder [5] that maps s_i to a d^{text}-dimensional vector, f_{sa} is a learnable adaptor that projects it to d dimensions, consisting of layer normalization, a SwiGLU-based gated transformation, and a residual connection, and $\mathbf{p}_i$ is a sinusoidal positional encoding. The generation of sentences and their encoding are performed offline as a preprocessing step. Since each sentence is determined by the combination of day of the week, time of day, and importance level, the number of distinct sentences is bounded, and their embeddings are precomputed once and reused across all movements. The resulting sequence is fed into a Transformer encoder (f_{te}^{sem}) to model semantic dependencies:

$$\mathbf{G}_u = f_{te}^{sem}\big([\tilde{\mathbf{s}}_1, \ldots, \tilde{\mathbf{s}}_K]\big) = [\mathbf{g}_1, \ldots, \mathbf{g}_K], \tag{3}$$

where $\mathbf{g}_i \in \mathbb{R}^d$ represents the semantic embedding at movement step i, which is integrated with the temporal context in Section 2.3.

2.2 Temporal Context Modeling

LEMU models temporal context by capturing the periodic mobility patterns within each user trajectory, employing a Transformer encoder that attends to relationships across the entire sequence. Each movement step is mapped to a d-dimensional embedding by summing the location and time embeddings $\mathbf{e}_{l_i}$ and $\mathbf{e}_{t_i}$, obtained from learnable embedding layers over location IDs and discretized time slots. To incorporate a weak semantic prior, we lightly inject the semantic embedding $\mathbf{g}_i$ after aligning it to the temporal representation space via a learnable projection $\mathbf{W}_s \in \mathbb{R}^{d \times d}$; the semantic information is more explicitly integrated through cross-attention. With positional encoding $\mathbf{p}_i$ preserving the sequence order, the augmented input and the resulting representations produced by a Transformer encoder (f_{te}^{tmp}) are:

$$\tilde{\mathbf{m}}_i = \mathbf{e}_{l_i} + \mathbf{e}_{t_i} + \mathbf{p}_i + \mathbf{W}_s \mathbf{g}_i, \tag{4}$$

$$\mathbf{H}_u = f_{te}^{tmp}\big([\tilde{\mathbf{m}}_1, \ldots, \tilde{\mathbf{m}}_K]\big) = [\mathbf{h}_1, \ldots, \mathbf{h}_K], \tag{5}$$

where $\mathbf{h}_i \in \mathbb{R}^d$ represents the temporal embedding at movement step i, which is integrated with the semantic context in Section 2.3.

2.3 Cross-Attention Fusion and Training

We integrate the two contexts using cross-attention, with the temporal representation $\mathbf{H}_u$ (Eq. (5)) as the Query and the semantic

representation $\mathbf{G}_u$ (Eq. (3)) as the Key and Value, yielding attention-enhanced semantic representations $\mathbf{A}_u = [\mathbf{a}_1, \ldots, \mathbf{a}_K]$. To control the contribution of semantic information, $\mathbf{a}_i$ is combined with $\mathbf{h}_i$ through a gated residual fusion:

$$\mathbf{r}_i = \mathbf{h}_i + \sigma(\boldsymbol{\beta}) \odot \mathbf{a}_i, \tag{6}$$

where $\boldsymbol{\beta} \in \mathbb{R}^d$ is a learnable parameter vector, $\sigma(\cdot)$ denotes the sigmoid function, and $\odot$ represents element-wise multiplication, so that the gate adjusts the influence of semantic information on a per-dimension basis. Each cross-attention layer additionally applies a position-wise feed-forward network with a residual connection. The resulting representation $\mathbf{R}_u = [\mathbf{r}_1, \ldots, \mathbf{r}_K]$ serves as the personalized location embedding for the downstream task.

LEMU is trained in a self-supervised manner using a Masked Location Prediction objective. For a user u, we randomly select a subset of movement steps $\mathcal{M}_u \subset \{1, \ldots, K\}$ at a fixed masking ratio, and mask both the location and its semantic representation before feeding them into the model. Each masked $\mathbf{r}_i$ is passed through a linear output layer, and the cross-entropy loss $\mathcal{L} = -\sum_u \sum_{i \in \mathcal{M}_u} \log p(l_i \mid \mathbf{r}_i)$ is computed over the predicted location distribution.

3 Experiments

We evaluate whether the personalized location embeddings learned by LEMU provide expressive representations, through comparative experiments with existing location modeling methods on the downstream task of *Next Location Prediction*. To assess whether the advantage stems from the learned embeddings rather than the downstream model, we evaluate two downstream architectures under an identical training protocol.

3.1 Experimental Settings

Datasets. We use LYMob-4Cities[1], an open dataset of human mobility trajectories collected from multiple cities in Japan, which for privacy protection discloses neither city names nor exact collection periods. We focus on two cities (City A, City B) that differ in the spatial concentration of visits, allowing us to examine whether the advantage of LEMU holds under different urban structures. We sample 10,000 users per city with at least 50 weekly movement records, yielding approximately 4.4M records per city. Dataset statistics are summarized in Table 1, and Figure 2 shows a heatmap of the data volume for each location. Each record is a tuple $\langle uid, date, time, x, y \rangle$, discretized onto a 500 m × 500 m spatial grid at 30-minute intervals. The day of the week is inferred from the periodicity observed in daily data volume. We train the location encoder on weeks 1–3 and predict the locations visited in week 4, which is unseen during embedding training. Training and evaluation follow a user-based 5-fold cross-validation protocol conducted independently for each city, with training, validation, and test ratios of 70%, 10%, and 20%.

Downstream Task. We evaluate the learned embeddings on *Next Location Prediction*, which predicts the locations visited in the following week from three weeks of embedded trajectories, with the date and time of each future movement given as input. Performance is measured by Top-k accuracy with $k \in \{1, 5\}$. We employ two downstream models to assess generalizability across architectures, a unidirectional LSTM and a BERT-style Transformer encoder. The

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

Table 1: Dataset statistics.

	City A	*City B*
Total Number of Users	10,000	10,000
Total Number of Locations	24,206	18,076
Total Number of Records	4,443,978	4,420,943
Number of Records per User	444	442
Number of Locations per User	108	92

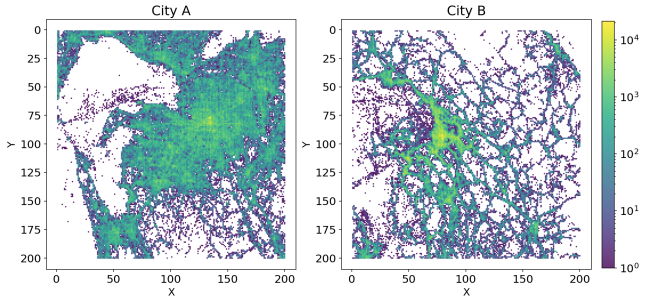


Figure 2: Heatmap illustrating the spatial distribution and density of observed mobility records across locations in the dataset. City A and City B show different spatial patterns.

LSTM-based model adopts a sequence-to-sequence formulation, whereas the BERT-style model feeds the full sequence at once and predicts each target location from its corresponding output position. The location encoder is kept fixed and only downstream parameters are trained, with an identical architecture and training protocol applied to LEMU and all baselines, so that performance differences primarily reflect embedding quality. Downstream models use 2 layers (8 heads for BERT), trained for 50 epochs with AdamW (learning rate 1e-4, weight decay 1e-3, batch size 16).

Training Details. During LEMU training, 20% of movement steps are randomly masked, with the location replaced by a mask token and the corresponding semantic sentence embedding by a zero vector, which prevents the model from relying on a single information source. The temporal-context Transformer (Eq. (5)) has 4 layers and 8 attention heads, while the semantic-context Transformer (Eq. (3)) is lightweight with 2 layers and 8 heads, as it operates on representations derived from a pretrained language encoder. Sentence embeddings are obtained from a frozen Sentence-BERT model[2] with $d^{\text{text}} = 768$ and projected through a hidden dimension of $d^h = 512$. The embedding dimension is $d = 128$ and the batch size is 16, and we train for 100 epochs using AdamW (learning rate 1e-4, weight decay 1e-2) on an NVIDIA RTX6000 Ada GPU.

3.2 Performance on Downstream Task

We compare LEMU with representative context-aware location modeling approaches that learn location embeddings from mobility trajectories, namely Geo-Teaser [11], TALE [8], HIER [6], and CTLE [4]. Each baseline follows its original paper, and is trained

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

		City A		*City B*	
		Top-1	*Top-5*	*Top-1*	*Top-5*
LSTM	*Geo-Teaser*	0.0022	0.0081	0.0294	0.0954
	TALE	0.0112	0.0375	0.0714	0.1963
	HIER	0.0206	0.0735	0.0723	0.2103
	CTLE	0.0354	0.0970	0.1212	0.2684
	LEMU	0.0717*	0.1927*	0.1491*	0.3385*
BERT	*Geo-Teaser*	0.0751	0.1608	0.1848	0.3753
	TALE	0.0883	0.1879	0.2010	0.3949
	HIER	0.0724	0.1966	0.1607	0.3592
	CTLE	0.1090	0.2367	0.1937	0.3937
	LEMU	0.1650*	0.3686*	0.2236*	0.4563*

Table 2: Performance comparison on *Next Location Prediction* over two cities (City A and City B), with two downstream models (LSTM-based and BERT-based). Results are mean values over 5-fold cross-validation. The best score in each column is highlighted in red and the second-best in blue. Asterisks (*) denote statistically significant improvements of LEMU over the best-performing baseline ($p < 0.05$, two-sided paired t-test on fold-level results).

with Adam under a learning rate of 1e-4, a batch size of 16, and an embedding dimension of 128, matching those of LEMU.

Table 2 summarizes the results with both LSTM-based and BERT-based downstream models, reported as mean values over 5-fold cross-validation. Statistical significance is assessed using a two-sided paired t-test on the fold-level results, comparing LEMU with the best-performing baseline under identical data splits, and asterisks (*) indicate significant improvements ($p < 0.05$). LEMU outperforms the baselines in both cities and across all evaluated metrics. Averaged over the two cities, LEMU improves Top-1 Accuracy by 4.3 percentage points with the BERT-based downstream model and by 3.2 percentage points with the LSTM-based one. The advantage therefore holds under both downstream architectures, indicating that it stems from the quality of the learned embeddings rather than the choice of downstream model. We also observe a consistent gap between the two cities, which is likely attributable to the differing numbers of locations (Table 1). City A has more locations, leading to greater diversity in visited grids per user and making both embedding learning and prediction more challenging.

These results show that personalized location embeddings yield consistent gains across two cities and two downstream architectures, supporting their expressiveness for mobility modeling.

4 Conclusion

This paper proposed LEMU, an urban location modeling method that constructs personalized location embeddings by integrating temporal mobility context with language-derived semantic representations. Its core design encodes user-relative location importance and temporal attributes into concise sentences and projects them into a language-embedding space, capturing user-relative location roles without requiring explicit semantic labels. Comparative experiments demonstrate that LEMU outperforms existing context-aware location modeling methods on *Next Location Prediction*, improving Top-1 Accuracy over the best-performing baseline by 4.3 percentage points with the BERT-based downstream model and by 3.2 percentage points with the LSTM-based one. These gains confirm that the personalized location representations constructed by LEMU are expressive and effective for a downstream mobility task. As future work, we plan to analyze how the learned representations differ across users for the same location and how such differences relate to the functional characteristics of urban areas, in order to better understand semantically diverse locations.

AI Usage Disclosure

ChatGPT and Claude were used to assist with translation and language editing of the manuscript. The authors reviewed and are fully responsible for all content.

Acknowledgments

This work was supported by JST BOOST (JPMJBS2422), JST CREST (JPMJCR22M4), and JST RISTEX (JPMJRS23K3).

References

[1] Yile Chen, Yicheng Tao, Yue Jiang, Shuai Liu, Han Yu, and Gao Cong. 2025. Enhancing Large Language Models for Mobility Analytics with Semantic Location Tokenization. In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD). 262–273. doi:10.1145/3711896.3736937

[2] Jiawei Cheng, Jingyuan Wang, Yichuan Zhang, Jiahao Ji, Yuanshao Zhu, Zhibo Zhang, and Xiangyu Zhao. 2025. POI-Enhancer: An LLM-based Semantic Enhancement Framework for POI Representation Learning. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). doi:10.1609/aaai.v39i11.33252

[3] Shanshan Feng, Gao Cong, Bo An, and Yeow Meng Chee. 2017. POI2Vec: Geographical Latent Representation for Predicting Future Visitors. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). 102–108. doi:10.1609/aaai.v31i1.10500

[4] Yan Lin, Huaiyu Wan, Shengnan Guo, and Youfang Lin. 2021. 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 (AAAI). 4241–4248. doi:10.1609/aaai.v35i5.16548

[5] Nils Reimers and Iryna Gurevych. 2019. 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). 3982–3992. doi:10.18653/v1/D19-1410

[6] Toru Shimizu, Takahiro Yabe, and Kota Tsubouchi. 2020. 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). 187–190. doi:10.1145/3397536.3422229

[7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. Advances in Neural Information Processing Systems (NeurIPS) (2017), 6000–6010. doi:10.48550/arXiv.1706.03762

[8] Huaiyu Wan, Yan Lin, Shengnan Guo, and Youfang Lin. 2022. Pre-Training Time-Aware Location Embeddings from Spatial-Temporal Trajectories. IEEE Transactions on Knowledge and Data Engineering 34, 11 (2022), 5510–5523. doi:10.1109/TKDE.2021.3057875

[9] Takahiro Yabe, Kota Tsubouchi, Toru Shimizu, Yoshihide Sekimoto, Kaoru Sezaki, Esteban Moro, and Alex Pentland. 2024. YJMob100K: City-Scale and Longitudinal Dataset of Anonymized Human Mobility Trajectories. Scientific Data 11, 397 (2024). doi:10.1038/s41597-024-03237-9

[10] Zijun Yao, Yanjie Fu, Bin Liu, Wangsu Hu, and Hui Xiong. 2018. Representing Urban Functions through Zone Embedding with Human Mobility Patterns. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI). doi:10.24963/ijcai.2018/545

[11] Shenglin Zhao, Tong Zhao, Irwin King, and Michael R. Lyu. 2017. Geo-Teaser: Geo-Temporal Sequential Embedding Rank for Point-of-Interest Recommendation. In Proceedings of the World Wide Web Conference (WWW). 153–162. doi:10.1145/3041021.3054138