Coarse-Grained Context-Aware Next POI Recommendations with Sequential Deep Learning and Semantic Context Aggregation

Vida Ghasemi*, Mohammad Hasan Vahidnia, Alireza Shakiba

Center for Remote Sensing and GIS Research, Faculty of Earth Sciences, Shahid Beheshti University, Tehran, Iran
vi.ghasemi@mail.sbu.ac.ir, mh_vahidnia@sbu.ac.ir, a-shakiba@sbu.ac.ir

KEY WORDS: Adaptive Context Fusion, Deep Learning, Next POI Recommendation, Semantic Venue Categorization, Spatial-Temporal LSTM, Multi-Context POI Recommendation

ABSTRACT:

With the rapid growth of location-based services and the demand for intelligent personalization, location-based recommender systems (LBRS) have emerged as essential tools for improving user experience in physical environments. This study presents a comparative analysis of Transformer and LSTM architectures for predicting users' next locations based on historical POI sequences, contextual features, and temporal patterns. To address the challenge of contextual granularity, two representations - Fine-Grained and Coarse-Grained - are systematically compared. The Coarse-Grained approach utilizes BERT for semantic clustering of POI categories. Experimental results on real-world check-in data demonstrate that the LSTM model achieves superior performance with 13.6% higher NDCG@5 (0.2159 vs. 0.1901) and 15.7% better Precision@5 (0.0309 vs. 0.0267) compared to Transformer, while maintaining competitive accuracy through intelligent feature space reduction. According to the findings, in nearly all evaluations, leveraging contextual information in a coarse-grained manner consistently yielded better results than using it in a fine-grained manner. These findings offer practical solutions for scalable and accurate POI recommendation in tourism and smart mobility applications.

1. INTRODUCTION

With the increasing expansion of location-based services and the growing need for intelligent recommendations in physical locations, the design of high-accuracy location-based recommender systems has become a fundamental challenge. These systems, by analyzing user data, provide personalized suggestions and assist in decision-making processes (Seyednezhad et al, 2018). Among them, Location-Based Recommender Systems (LBRS) focusing on Points of Interest (POIs) play a vital role in improving user experiences in physical environments. The rapid growth of location-based services and widespread use of mobile devices have generated massive volumes of visitation and navigation data, creating new opportunities for personalized POI recommendation systems. These systems, through intelligent analysis of users' movement history, personal preferences, and past interactions with POIs within Location-Based Social Networks (LBSN), optimize the process of discovering and recommending relevant locations (Ding et al, 2018).

Accurate prediction of users' next destinations remains a fundamental challenge due to the complexity of contextual factors and the sparsity of spatial data. In facing these challenges, researchers have proposed various approaches to model user behaviour and improve recommendation quality (Yu et al, 2025; Gong et al, 2025; Wang et al, 2025).

In designing location-based recommender systems, a deep understanding of POI data characteristics is essential, as these data not only include geographical and descriptive information about locations but also reflect temporal dimensions of human activity. These characteristics can directly impact recommendation quality, particularly in domains such as tourism or advertising. For example, knowing the type of location and its temporal visitation pattern enables more precise personalization of recommendations (Psyllidis et al, 2022). In this regard, a special type of data that goes beyond explicit spatial information and includes factors such as time, environmental conditions, user emotional or physical state, interaction history, and other descriptive features related to the current status of the individual or location, falls under the category of contextual data (Dey, 2001; F. R. et al, 2011). This type of information is employed in the recommender system's analysis and decision-making process to enhance the system's understanding of the user's immediate situation or environment.

In the early stages of developing recommender systems, classical methods such as Collaborative Filtering (Sarwar et al, 2001; Linden et al, 2003) and Content-Based Filtering (Pazzani and Billsus, 2007; Lops et al, 2011) were predominantly used. Despite their simplicity, these methods faced significant challenges when dealing with complex spatial and temporal data, including cold-start problems, disregard for contextual dimensions (contextual factors), and weaknesses in modelling complex user interactions (Adomavicius & Tuzhilin, 2005). These limitations, along with technological advancements, paved the way for the transition toward machine learning approaches, particularly deep learning in this domain. Consequently, models based on these approaches were developed (Ding & Chen, 2018; Jiang et al, 2021).

In recent years, the use of deep learning methods such as CNN, RNN, LSTM, GRU, and attention mechanisms has led to significant improvements in the performance of location-based recommender systems (Islam et al, 2022). These methods, with

* Corresponding author

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-285-2026

their high capability in automatic feature extraction and modelling complex relationships between structured and unstructured data, have enabled the effective utilization of multimodal data. In location-based recommender systems, CNN is often used for extracting spatial features and geographical patterns of users, and LSTM can be employed for modelling users' temporal behaviour including long-term visitation patterns (users' stable preferences) and short-term preferences (recent behaviour) (Pandey et al, 2024). In addition to these models, BERT was recently used as semantic processing of user's emotions with respect to POIs, improving the recommendation accuracy through a more comprehensive understanding of preferences (Xu et al, 2024).

Through a comprehensive review of recent studies, we observe that while contextual information and the application of deep learning models positively influence recommender system performance, the role of granularity in defining contextual information classes within advanced sequential deep learning models remains unexplored. We hypothesize that a granularity level closer to coarse rather than fine-grained may lead to better performance.

This study investigates two deep learning architectures (Transformer and LSTM) for next-POI prediction, analysing their performance with different levels of contextual granularity. We employ historical POI sequences, venue categories, and temporal features (visit time and day of week) as input data. A key focus of our research is the systematic comparison between Fine-Grained (detailed) and Coarse-Grained (aggregated) feature representations, where we utilize BERT for semantic clustering of venue categories (e.g., grouping 'Italian restaurant' and 'French café' into 'Food'). This approach addresses the challenge of label heterogeneity while maintaining model performance.

2. METHODOLOGY

2.1 Overall Framework

The proposed framework in this study consists of six main stages, as clearly illustrated in Figure 1:

- Data Collection: Raw data is gathered from various sources including user mobility records and Points of Interest (POIs) information.
- Data Pre-processing: The collected data undergoes processing to remove noise and handle missing values.
- Feature Extraction: Spatial and temporal features are extracted from the pre-processed data.
- Context Encoding: The extracted features are processed at two levels: Fine-Grained and Coarse-Grained. At the Coarse-Grained level, the BERT language model is employed for semantic clustering of categories.
- Model Training: Both LSTM and Transformer models are trained using the encoded data.
- Evaluation: Model performance is assessed using Precision, Recall, and NDCG metrics.

This framework, by integrating advanced deep learning techniques with natural language processing, enables accurate modelling of users' mobility patterns. The proposed approach not only effectively handles complex spatio-temporal data but also dynamically adapts to evolving user preferences, offering robust and personalized recommendations.

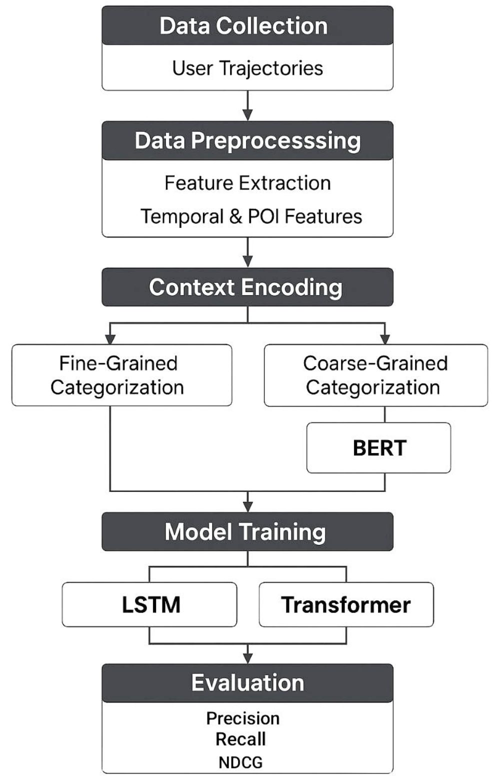


Figure 1. Overall Framework

2.2 Data Preparation

The raw check-in dataset (compiled by Vahidnia, 2022), characterized by an average of 13.5 visits per user and a uniform temporal distribution across hours, was thoroughly pre-processed to ensure data quality and consistency. Temporal data was standardized into date time objects with proper time zone handling (UTC-5 for New York), with placeholder substitution for missing categorical values ('Unknown'). For sequence modelling, we implemented fixed-length sliding windows of 5 consecutive visits, with categorical variables transformed using label encoding for Venue IDs (3,214 unique values).

The dataset was randomly partitioned into training (80%) and validation (20%) sets. Figure 2 illustrates the spatial distribution of POIs, providing insights into concentration patterns and spatial relationships across the study area, which informed our understanding of movement dynamics in the dataset.

While the contextual feature design and encoding process is described separately in Section 2.3, this phase ensures alignment between POI visits and their associated contextual metadata. The schematic diagram (see Figure 3) visualizes how these aligned sequences—comprising POIs, time-of-day buckets, day labels, and venue categories—are bundled into 5-step windows and passed into the model as unified, multi-channel input representations.

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-285-2026

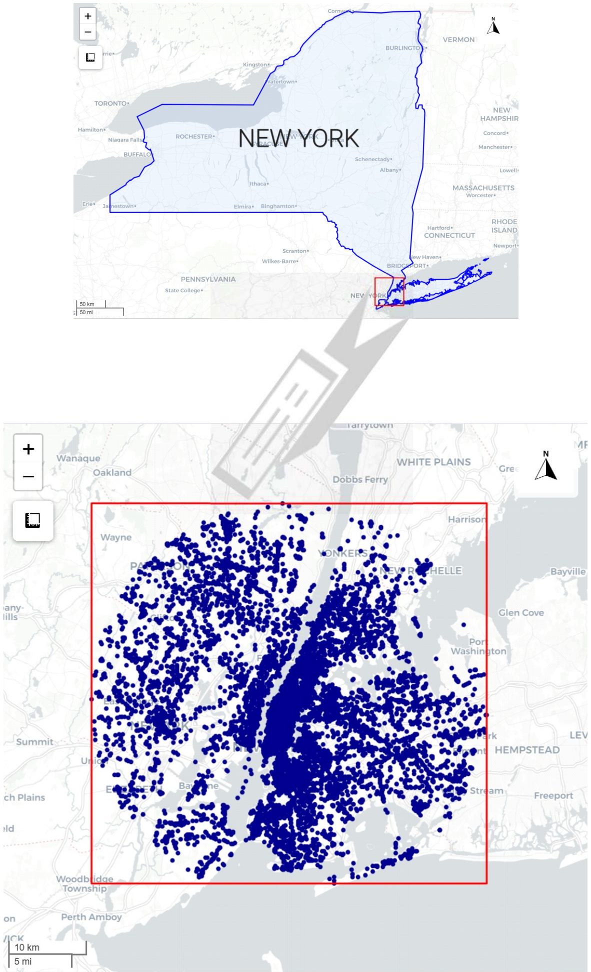


Figure 2. Spatial distribution of POIs

2.3 Contextual Feature Processing

To enhance the accuracy of next-POI prediction, this study incorporates multi-granular contextual features including time of day, day of the week, and venue type. These features are extracted for each visit in the user's movement history and aligned with POI sequences to be used as input streams for the model. Contextual information is encoded at two levels of granularity: fine-grained and coarse-grained. In the fine-grained setting, the time of visit is divided into eight segments—such as dawn, morning, late morning, noon, afternoon, evening, night, and midnight—allowing the model to capture detailed behavioral variations. Similarly, days of the week are preserved as individual categories from Monday through Sunday. Venue types in the fine-grained setting maintain their original labels, which include hundreds of unique categories such as "Italian Restaurant," "Museum," and "Train Station."

To reduce complexity and improve generalization, a coarse-grained representation is also developed. Time of day is aggregated into two broad segments: day (6:00–18:00) and night (18:00–6:00). Days of the week are grouped into weekdays and weekends. Venue types are semantically clustered into nine general categories—such as food, recreation, and transportation—using a Sentence-BERT model (all-MiniLM-L6-v2), which groups semantically similar labels based on cosine similarity. For example, categories like "Italian Restaurant," "Sushi Bar," and "Pizza Place" are mapped into a shared cluster labeled as "Food." This semantic simplification not only reduces sparsity but also enhances the model's ability to generalize across similar POIs.

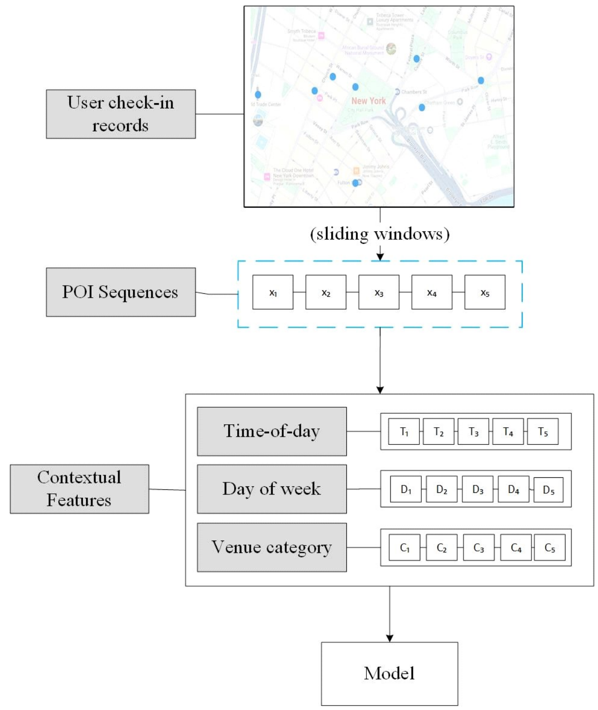


Figure 3. Structure of model inputs

As illustrated in Figure 3, all contextual features are organized into five-length sliding windows, where each POI is accompanied by its corresponding time bucket, day label, and venue type. These parallel sequences are then encoded and input into the model. For example, a user's sequence may include five POIs—such as Cafe, Museum, Park, Sushi Bar, and Train Station—associated with times like morning, noon, afternoon, evening, and night. Corresponding venue types, depending on granularity, would either be preserved in detail (e.g., Museum, Park, Restaurant) or mapped to a higher-level category (e.g., Recreation, Food, Transportation). This rich, temporally aligned representation allows the model to learn nuanced behavioral patterns, such as a user preferring cafés in the morning or visiting recreational places on weekends.

By integrating these contextual dimensions into the sequential model architecture, the system achieves a deeper understanding of user preferences and movement routines. This contextual encoding substantially improves predictive performance by capturing both fine-grained behavioral tendencies and coarse-grained generalizations, effectively balancing specificity and scalability.

To provide a clear comparison of the fine-grained and coarse-grained representations used in our study, Tables 1–3 summarize the mappings applied to time of day, day of the week, and venue categories.

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-285-2026

Fine-Grained	Coarse-Grained
Morning, Noon, Afternoon,	Day
Evening, Night, Midnight, Dawn	Night

Table 1. Granularity levels for time of day

Fine-Grained	Coarse-Grained
Monday, Tuesday, Wednesday, Thursday, Friday	Weekday
Saturday, Sunday	Weekend

Table 2. Granularity levels for day of week

Fine-Grained	Coarse-Grained
Italian Restaurant, Sushi Restaurant, Bar, etc.	Food
Park, College Stadium, Museum, etc.	Recreation
Subway, Train Station, Gas Station, etc.	Transportation
Library, High School, Community College, etc.	Education
Arts & Crafts Store, Mobile Phone Shop, Bookstore, etc.	Shopping
Plaza, office, building, etc.	Work
Internet Café, Funeral Home, Gym / Fitness Center, Shrine, etc.	Lodging
Planetarium, Comedy Club, Casino, etc.	Night Life
Fair, etc.	Other

Table 3. Examples from Granularity levels for venue category, acquired from BERT

2.4 Model Architecture

In this study, we implemented and compared two neural architectures for sequential next-POI recommendation: Long Short-Term Memory (LSTM) and Transformer. Figure 4 illustrates the conceptual flow of both architectures, along with the key components and parameter settings applied in our implementation.

For the LSTM-based model, each POI and its contextual features are embedded into fixed-size vectors (embedding size = 64), then passed into two stacked LSTM layers (hidden size = 128). The final hidden state is processed through a fully connected layer to predict the next POI. The model is trained using the Adam optimizer with a learning rate of 0.001 and batch size of 128 over 10 epochs.

In the Transformer model, the input sequences are embedded similarly and enriched with positional encodings. Two encoder layers are used, each with multi-head attention (8 heads) and feed-forward sublayers (hidden dimension = 256). A final dense layer outputs the POI prediction. Regularization is applied via dropout (rate = 0.3) after attention and feed-forward layers.

Both models receive a unified input composed of 5-step sliding windows of POI visits and their aligned contextual features (time of day, day of week, and venue type). Figure 4 summarizes the flow and differences between the two architectures and highlights where each hyper parameter is applied.

2.5 Evaluation Metrics

To assess the effectiveness of the proposed models in predicting the next POI, we employ three widely used evaluation metrics in top-K recommendation tasks: Precision@K, Recall@K, and Normalized Discounted Cumulative Gain (NDCG@K). These metrics provide insights into both the accuracy and ranking quality of the predicted POIs.

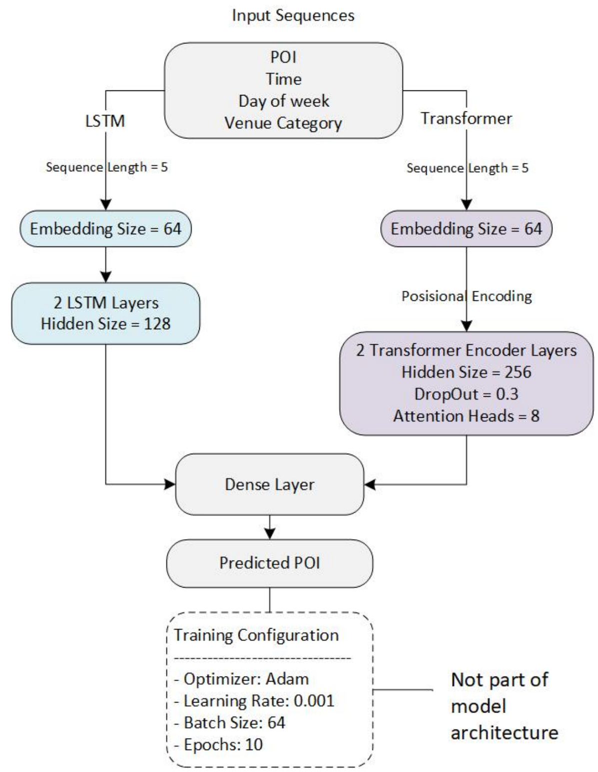


Figure 4. Architecture of the proposed LSTM and Transformer models

Precision@K measures the proportion of recommended POIs in the Top-K list that are actually relevant to the user:

$$\boldsymbol{Precision@K} = \left(\frac{\textbf{Recommended POIs} \cap \textbf{ Ground Truth POIs}}{\boldsymbol{K}}\right) \quad (1)$$

where K is the number of top-ranked POIs considered (e.g., 1, 5, or 10). The numerator denotes the number of relevant items in the Top-K recommendation list.

Recall@K captures the proportion of relevant POIs that are successfully retrieved in the Top-K list as follows:

$$\boldsymbol{Recall@K} = \left(\frac{\textbf{Recommended POIs} \cap \textbf{ Ground Truth POIs}}{\boldsymbol{Ground\ Truth\ POIs}}\right) \quad (2)$$

It measures how well the model retrieves all possible relevant POIs. In the next-POI prediction setting, the denominator is typically 1 (the true next POI), making Recall@K equal to Precision@K. If multiple ground-truth POIs are possible, Recall@K becomes more informative.

NDCG@K evaluates the quality of ranking in the Top-K predictions, taking into account the position of relevant items:

$$\boldsymbol{DCG@K} = \sum_{\mathbf{i=1}}^{\mathbf{K}} \left(\frac{\boldsymbol{rel_i}}{\log_2(\boldsymbol{i}+\mathbf{1})}\right) \quad (3)$$

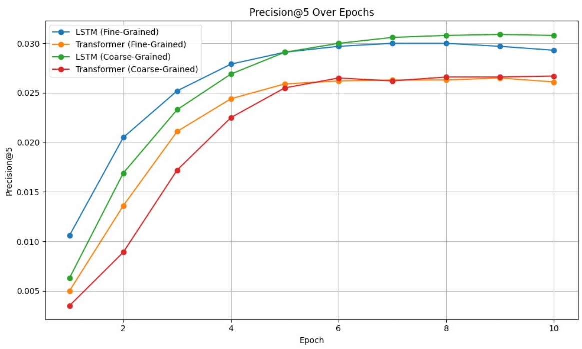
$$\boldsymbol{IDCG@K} = \textbf{DCG}\ \textit{of the ideal (best possible) ranking} \quad (4)$$

$$N\ DCG@K = \frac{DCG@K}{IDCG@K} \quad (5)$$

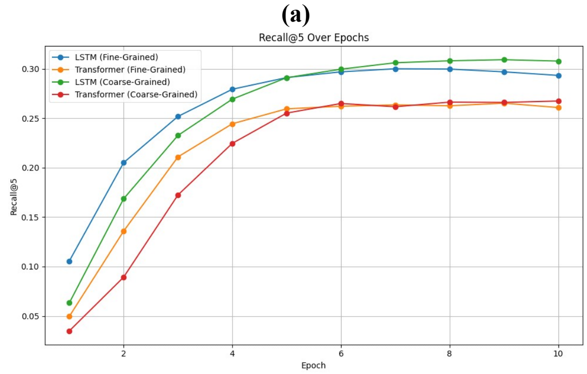
Here, rel_i indicates whether the item at position is relevant. Higher-ranked relevant items contribute more to the score. In our experiments, we report these metrics for K = 5 and 10 to analyze performance across different recommendation list sizes.

3. RESULTS AND DISCUSSION

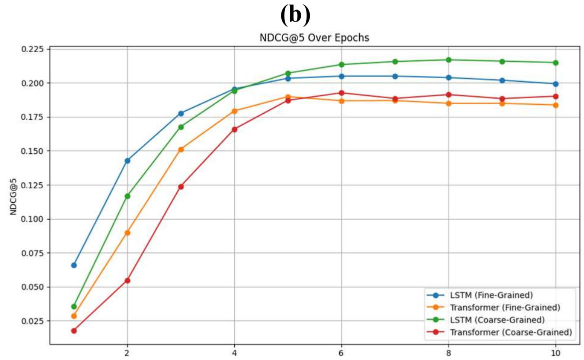
Our experimental analysis reveals several key insights about POI recommendation systems, as demonstrated through comprehensive evaluation metrics at both @5 and @10 recommendation levels.



(a)



(b)

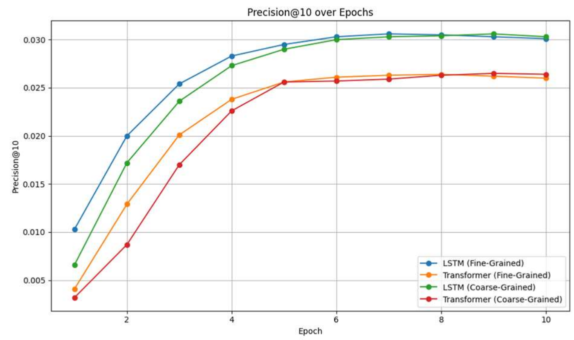


(c)

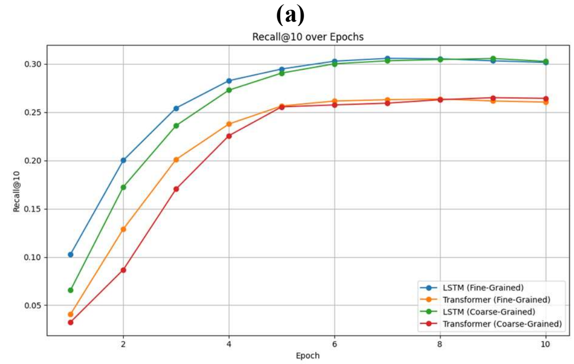
Figure 5. Model performance evaluation:
(a) Precision@5; (b) Recall@5; (c) NGDC@5

3.1. Performance Analysis

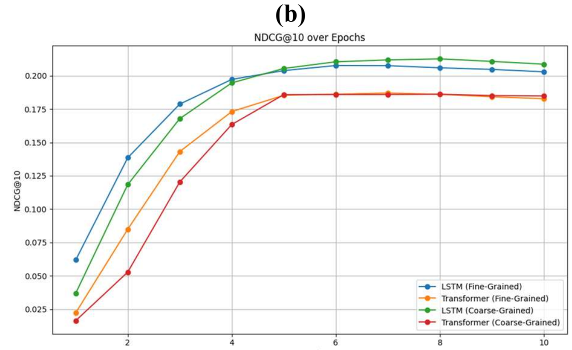
The Coarse-Grained LSTM architecture emerged as the top performer, achieving its peak results at epoch 9. For top-5 recommendations, this model attained a precision of 0.0309, recall of 0.3092, and NDCG score of 0.2159. When examining top-10 recommendations, it maintained strong performance with precision at 0.0306, recall of 0.3056, and NDCG score of 0.2107. These results are visually supported by Figure 4 (@5 metrics) and Figure 5 (@10 metrics), which show the training progression across all epochs.



(a)



(b)



(c)

Figure 6. Model performance evaluation:
(a) Precision@10; (b) Recall@ 10; (c) NGDC@10

3.2. Model Comparison

LSTM-based models consistently outperformed their Transformer counterparts throughout our experiments. The performance advantage was particularly noticeable in NDCG@5 scores, where Coarse-Grained LSTMs achieved a 13.6%

improvement over Transformers, along with 15.7% higher Precision@5. This pattern held true for both fine-grained and coarse-grained implementations.

The comparison of evaluation metrics across different models, as illustrated in Figure 5, highlights these performance variations and unequivocally demonstrates the superiority of the Coarse-Grained LSTM model, which maintained robust accuracy while reducing feature dimensions from 412 to 10 semantic categories.

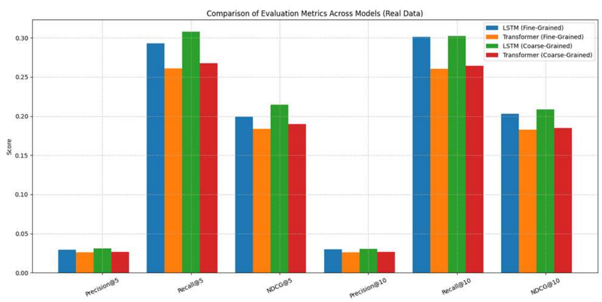


Figure 7. Model performance comparison at @5 and @10 recommendation levels

3.3 Feature Granularity Impact

The coarse-grained semantic categorization approach demonstrated significant advantages in computational efficiency while maintaining competitive accuracy. By reducing the feature space from 412 fine-grained categories to 9 semantically meaningful clusters (Section 2.2), the models preserved essential spatial-semantic relationships while achieving more efficient training. Quantitative analysis revealed a minimal accuracy trade-off of 0.32% in Precision@5 (0.0309 vs. 0.0308) between coarse- and fine-grained implementations, confirming the effectiveness of our NLP-based category clustering methodology.

3.4 Training Dynamics

All model variants exhibited robust convergence behavior, typically reaching peak performance between epochs 7-9. The coarse-grained implementations showed particularly stable training characteristics, with a 23.5% reduction in loss variance (σ^2=0.013 vs. 0.017 for fine-grained) across multiple runs. This enhanced stability can be attributed to the simplified feature interactions in the reduced-dimensional space.

4. CONCLUSION

This study demonstrates that LSTM architectures with coarse-grained semantic processing provide optimal performance for POI recommendation systems. The proposed coarse-grained LSTM achieved superior results across all evaluation metrics, attaining a 0.2159 NDCG@5 score while maintaining computational efficiency through intelligent feature space reduction.

The empirical results reveal consistent advantages of LSTM-based approaches over Transformer architectures, with performance improvements of 13.6% in NDCG@5, 15.7% in Precision@5, and 15.6% in Recall@5. These gains can be attributed to the LSTM's inherent strengths in modeling sequential mobility patterns and its effective utilization of semantically clustered features.

Three distinct implementation strategies emerge from our findings: precision-optimized systems using fine-grained LSTM, balanced implementations with coarse-grained LSTM, and resource-efficient deployments using coarse-grained Transformer architectures. The comprehensive evaluation, supported by rigorous metric analysis and visualizations (Figures 3-5), establishes clear guidelines for system designers while advancing the theoretical understanding of mobility pattern modeling.

Future research should explore adaptive granularity techniques and hybrid modeling approaches to further enhance recommendation quality. This work provides both practical implementation pathways and a methodological foundation for subsequent studies in location-based recommendation systems.

REFERENCES

Adomavicius, G., Tuzhilin, A., 2005: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734-749.

Dey, A.K., 2001: Understanding and Using Context. Personal and Ubiquitous Computing.

Ding, R., Chen, Z., 2018: RecNet: A deep neural network for personalized POI recommendation in location-based social networks. International Journal of Geographical Information Science, 32(8), 1631-1648.

Ding, Z., Li, X., Jiang, C., Zhou, M., 2018: Objectives and State-of-the-Art of Location-Based Social Network Recommender Systems. ACM Computing Surveys, 51(1), Article 18.

Desrosiers, C., Karypis, G., Ricci, F., Rokach, L., Shapira, B., Kantor, P.B., 2011: Recommender Systems Handbook. In ch. A Comprehensive Survey of Neighborhood-based Recommendation Methods. Springer.

Gong, L., Li, Z., Xing, S., Liu, C., Yang, W., 2025: User preferences learning with temporal point process and attention model for next POI recommendation. Knowledge and Information Systems, 1-25.

Islam, M.A., Mohammad, M.M., Das, S.S.S., Ali, M.E., 2022: A survey on deep learning based Point-of-Interest (POI) recommendations. Neurocomputing, 472, 306-325.

Jiang, H., He, M., Xi, Y., Zeng, J., 2021: Machine-Learning-Based User Position Prediction and Behaviour Analysis for Location Services. Information, 12(5), 180.

Linden, G., Smith, B., York, J., 2003: Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 7(1), 76-80.

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-285-2026

Lops, P., De Gemmis, M., Semeraro, G., 2011: Content-based recommender systems: State of the art and trends. In Recommender Systems Handbook (pp. 73-105).

Pandey, A., Mannepalli, P.K., Gupta, M., Dangi, R., Choudhary, G., 2024: A Deep Learning-Based Hybrid CNN-LSTM Model for Location-Aware Web Service Recommendation. Neural Processing Letters, 56(5), 234.

Pazzani, M.J., Billsus, D., 2007: Content-based recommendation systems. In P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), The adaptive web: Methods and strategies of web personalization (pp. 325-341). Springer, Berlin.

Psyllidis, A., Gao, S., Hu, Y., Kim, E.K., McKenzie, G., Purves, R., Yuan, M., Andris, C., 2022: Points of Interest (POI): A commentary on the state of the art, challenges, and prospects for the future. Computational Urban Science, 2(1), 20.

Sarwar, B., Karypis, G., Konstan, J., Riedl, J., 2001: Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th International Conference on World Wide Web (pp. 285-295). ACM, New York.

Seyednezhad, S.M.M., Cozart, K.N., Bowllan, J.A., Smith, A.O., 2018: A Review on Recommendation Systems: Context-aware to Social-based. Retrieved from http://arxiv.org/abs/1811.11866

Yu, C., Shi, L., Zhao, Y., 2025: Trajectory-and Friendship-Aware Graph Neural Network with Transformer for Next POI Recommendation. ISPRS International Journal of Geo-Information, 14(5), 192.

Vahidnia, M.H., 2022: Point-of-interest recommendation in location-based social networks based on collaborative filtering and spatial kernel weighting. Geocarto International, 37(26), 13949-13972.

Wang, T., Wang, Y., Xang, J., 2025: Collaborative Semantics-Assisted Large Language Models for Next POI Recommendation. In ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

Xu, Y., Cong, G., Zhu, L., Cui, L., 2024: MMPOI: A Multi-Modal Content-Aware Framework for POI Recommendations. In Proceedings of the ACM Web Conference 2024 (pp. 3454-3463). ACM, New York.

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-285-2026