BERT-M: A Transformer-Based Model for Context-Aware User Mobility Prediction in MEC Environments

Jeonghwa Lee[1], Eunjeong Ju[1], Duksan Ryu[1*], and Jongmoon Baik[2]

[1] Department of Software Engineering, Jeonbuk National University, Jeonju, South Korea
{dlwjdghk133, jeju3146, duksan.ryu}@jbnu.ac.kr

[2] School of Computing, Korea Advanced Institute of Science and Technology, Daejeon, South Korea
jbaik@kaist.ac.kr

Abstract. Mobile Edge Computing (MEC) has emerged as a promising paradigm for reducing latency and enhancing Quality of Service (QoS) by bringing computational resources closer to end users. However user mobility often leads to frequent service handovers and QoS degradation. To address this challenge we propose BERT-M a Transformer-based mobility prediction model encoding sequences of historical edge server connections as natural language. Unlike conventional LSTM-based approaches BERT-M eliminates the need for distance matrices or complex preprocessing pipelines. We evaluate BERT-M on a real-world MEC dataset under diverse hyperparameter configurations. Experimental results demonstrate that BERT-M consistently achieves higher Top-1 accuracy compared to LSTM baselines and exhibits robust generalization. The findings suggest that BERT-M offers a lightweight, scalable, and context-aware solution for enhancing QoS in dynamic MEC environments.

1 Introduction

Mobile Edge Computing (MEC) is a network paradigm that minimizes latency and enhances real-time service responsiveness by deploying computational resources in proximity to end users. This architecture is particularly critical for latency-sensitive applications such as augmented reality, autonomous driving, and smart cities. However frequent user mobility poses significant challenges to maintaining stable Quality of Service (QoS), often resulting in unnecessary handovers and increased network overhead. Accurate user mobility prediction is therefore essential for proactive QoS optimization. One widely adopted approach is the LSTMM which leverages LSTM to predict future user locations based on historical trajectories and spatial distance features. While effective this method relies on complex preprocessing and distance matrix construction limiting its scalability and adaptability in real-time systems. To overcome these limitations we propose BERT-M a novel user mobility prediction model built upon the BERT architecture.

* Corresponding author

2 Methodology

BERT-M interprets sequences of previously accessed edge server locations as natural language sentences and predicts the next likely connection point. Input sequences are constructed from location coordinates in the form of "latitude/longitude" strings excluding additional features such as user identity or spatial distances. During preprocessing the final location in each users record is designated as the prediction target while prior locations form the input sequence. These are tokenized using the `bert-base-uncased` tokenizer with [CLS] and [SEP] tokens appended to match BERTs input format. The model is trained as a multi-class classifier using the AdamW optimizer. Various learning rates, batch sizes, and epoch counts are explored to evaluate sensitivity and optimize performance.

3 Experiment

We evaluate BERT-M using the Shanghai Telecom dataset which consists of cellular access logs from 7,691 users collected over June 2014. LSTMM serves as the baseline model. Performance is assessed using Top-K Accuracy (ACC@1 and ACC@5) as the evaluation metric. Key findings include:

- BERT-M outperforms LSTMM in Top-1 accuracy achieving a maximum of 0.2945 versus 0.2700 for LSTMM.
- In terms of Top-5 accuracy LSTMM yields superior performance (0.5800) while BERT-M achieves 0.4181.
- BERT-M maintains robust performance across various hyperparameter configurations demonstrating consistent Top-1 prediction capability.
- The best configuration for BERT-M was observed at a learning rate of 3×10^{-5}, a batch size of 16, and 100 training epochs.
- Smaller batch sizes generally resulted in better generalization while extreme learning rates adversely affected convergence.

4 Conclusion and Future Work

This study presents BERT-M a Transformer-based user mobility prediction model tailored for MEC environments. BERT-M eliminates the need for spatial or identity-related features and reduces preprocessing overhead. Empirical results confirm that BERT-M delivers accurate Top-1 predictions and generalizes well under diverse conditions.

Future work will explore hybrid architectures that combine LSTM and BERT components to improve Top-K accuracy, introduce multitask learning objectives, and leverage longer-term mobility patterns to further enhance generalization and performance in real-world deployments.

Acknowledgements

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2025-RS-2020-II201795) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).