단방향 및 양방향 순환 신경망의 성능 평가

Sammy Yap Xiang Bang[1], 정경희 [1], 추현승 [2]
[1] 성균관대학교 수퍼인텔리전스학과 석사과정
[2] 성균관대학교 전기컴퓨터공학과 교수

sammyyap98@skku.edu, datakira@g.skku.edu, choo@skku.edu

Performance Evaluation of Unidirectional and Bidirectional Recurrent Neural Networks

Sammy Yap Xiang Bang[1], Kyunghee Jung[1], Hyunseung Choo[2]
[1]Dept. of Superintelligence, Sungkyunkwan University
[2]Dept. of Electrical and Computer Engineering, Sungkyunkwan University

Abstract

The accurate prediction of User Equipment (UE) paths in wireless networks is crucial for improving handover mechanisms and optimizing network performance, particularly in the context of Beyond 5G and 6G networks. This paper presents a comprehensive evaluation of unidirectional and bidirectional recurrent neural network (RNN) architectures for UE path prediction. The study employs a sequence-to-sequence model designed to forecast user paths in a wireless network environment, comparing the performance of unidirectional and bidirectional RNNs. Through extensive experimentation, the paper highlights the strengths and weaknesses of each RNN architecture in terms of prediction accuracy and computational efficiency. These insights contribute to the development of more effective predictive path-based mobility management strategies, capable of addressing the challenges posed by ultra-dense cell deployments and complex network dynamics.

1. Introduction

Reactive signal strength-based mobility management has been effective in microcell architectures with limited overlapping cells. However, the ultra-dense femtocell deployments in Beyond 5G (B5G) and 6G networks, which aim to facilitate hyper-connectivity, present new challenges to existing mobility management approaches. These challenges include increased erroneous and unnecessary handovers due to a higher likelihood of multiple overlapping cells at any given position [1]. To address these challenges and optimize network performance, multi-step ahead User Equipment (UE) path prediction is crucial for effective mobility management in B5G/6G networks.

Machine Learning (ML) and Deep Learning (DL) techniques have recently gained attention for their superior performance over traditional prediction-based proactive mobility management approaches [2]. These traditional methods, such as Hidden Markov models and data mining-based matching strategies, suffer from limitations in prediction accuracy and extensive computational time, making them ill-suited to tackle the emerging challenges of 6G networks [3]. Consequently, an increasing number of studies have focused on developing ML and DL-driven mobility prediction techniques to overcome these limitations.

For instance, some studies have introduced novel predictive handover mechanisms that combine Long Short-Term Memory (LSTM) networks and Support Vector Machines (SVM) for next location prediction [4]. Similarly, other investigations have proposed Recurrent Neural Network (RNN)-based sequence-to-sequence models, which use historical UE associations to predict the next Point of Attachment (PoA) of UE [5]. While these investigations have achieved a certain level of PoA prediction accuracy, they have not yet fully addressed the mobility management challenges specific to 6G networks, nor have they thoroughly investigated the performance of unidirectional and bidirectional recurrent neural network architectures in user path prediction, leaving potential improvements unexplored.

In this paper, we aim to provide a comprehensive evaluation of unidirectional and bidirectional recurrent neural networks for UE path prediction in the context of B5G/6G networks. Through a comprehensive evaluation, our study highlights the strengths and weaknesses of each RNN architecture in terms of prediction accuracy and computational efficiency. By shedding light on these aspects, our findings offer valuable insights for the development of more effective predictive path-based mobility management strategies in the emerging B5G/6G network landscape.

2. Data

The dataset utilized in this study was gathered in a wireless network environment consisting of 12 Access Points (APs) at Sungkyunkwan University's Pangyo Campus. The network controller records handover logs based on user path within the campus. These logs contain mobile device MAC (Media Access Control) addresses and the IDs of the source and destination APs during handover events. Movement sequences for mobile devices are generated using these handover logs. For instance, if a handover occurs for a device with MAC address 0333.0222.0111 from AP number 5 to 1, 1 to 9, 9 to 8, and 8 to 4, a mobile trajectory sequence of {5, 1, 9, 8, 4} is created. To ensure unbiased treatment of all 12 APs, the mobile trajectory sequences are further preprocessed using one-hot encoding. In the resulting vector, the index corresponding to the AP ID is assigned a value of one, while all other values are set to zero. For example, AP number 6 is represented as {0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0}, with the sixth value being one and the remaining 11 values being zero. In this study, we focus on predicting 7-step mobile trajectories using input data from 26 previous APs. A total of 2,160 mobility sequences from 289 users are split into 80% for training data and 20% for testing data.

3. Unidirectional and Bidirectional RNN Models

In this section, we introduce our sequence-to-sequence-based user path prediction model, which incorporates four types of RNN cells for performance evaluation. Our model predicts a mobile trajectory sequence of 7 APs using a historical trajectory sequence of 26 APs through an encoder network and a decoder network. The encoder network processes the input sequence via a fully connected neural network and single sublayers of RNN cells, creating an encoder state containing a compressed representation of all the elements. This encoder state is then passed to the decoder network, which simultaneously predicts 7 elements of the target sequence by employing two fully connected neural networks and single sublayers of RNN cells.

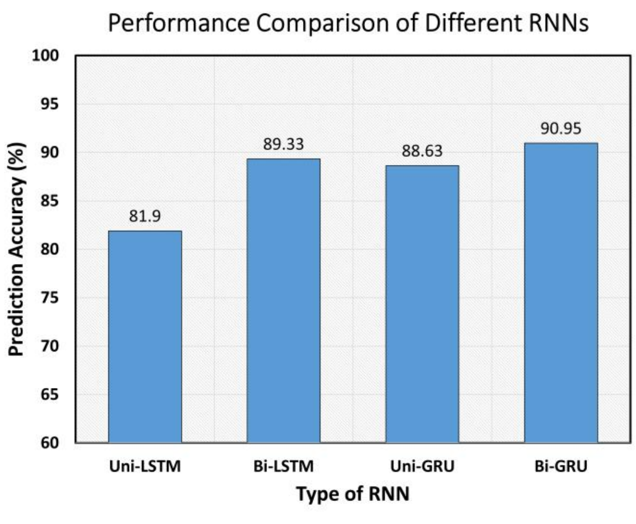
The prediction of the target sequence of 7 APs from the decoder network is enhanced through the use of teacher forcing during training. Teacher forcing is a technique employed in training sequence-to-sequence models, wherein the ground truth is used as input rather than the model output from the previous time step. This approach allows for more accurate trajectory predictions, as it corrects the trajectory prediction at each step.

In our model, we utilize four types of RNN cells within the encoder and decoder networks, including unidirectional Long Short-Term Memory (Uni-LSTM), bidirectional Long Short-Term Memory (Bi-LSTM), unidirectional Gated Recurrent Unit (Uni-GRU), and bidirectional Gated Recurrent Unit (Bi-GRU). A performance comparison of these four types of RNN cells will be presented in the subsequent section.

4. Performance Evaluation

Each result presented in this paper is an average of ten experiments which are conducted on a system with Intel Core i7-11700KF CPU, NVIDIA GeForce RTX 3070 Ti graphic cards, and 32GB RAM. Outright accuracy is used as the evaluation metric for this paper. Here, outright accuracy denotes that each step in the output sequence is correctly predicted which implies that prediction is considered as correct if and only if each step in the output sequence is correctly predicted. Outright accuracy returns an overall measure of how much the model is correctly predicting on the entire set of data. Each sequence in the testing set has the same weight and contributes equally to the accuracy value.

From the experiment results as shown in Figure 1, it can be observed that the Bi-GRU model achieves the highest accuracy of 90.95%, followed by the Bi-LSTM model with 89.33%, the GRU model with 88.63%, and lastly, the LSTM model with 81.90%. The bidirectional models (Bi-LSTM and Bi-GRU) demonstrate better performance in terms of accuracy compared to their unidirectional counterparts (LSTM and GRU). This can be attributed to the way bidirectional models process the input data. They capture information from both past and future time steps, allowing them to understand the context more effectively, leading to improved predictions. In contrast, unidirectional models only process input data from past time steps, limiting their ability to capture future context.



(Figure 1) Performance Evaluation of Different RNNs

In terms of training time, the GRU model has the shortest duration of 398.18s, while the Bi-LSTM model takes the longest with 445.53s. The difference in training time among the models can be attributed to the number of parameters in each model where Bi-LSTM model has the highest number of parameters as shown in Table 1. The increased number of parameters in bidirectional models is due to the additional forward and backward layers, which process the input data from two directions, thus requiring more time for training.

The differences in testing times are relatively small among the models. GRU model has the shortest testing time of 0.91s, while the Bi-LSTM model has the longest of 1.27s. This can be attributed to the model complexity, as GRU models have fewer gates and are generally less complex than LSTM models. Additionally, bidirectional models require more computation, as they process input data from both directions, which may contribute to slightly longer testing times.

<Table 1> Trainable Parameters for Different RNNs

Type of RNN	*Trainable Parameters*
Unidirectional LSTM	571,393
Bidirectional LSTM	1,107,980
Unidirectional GRU	503,297
Bidirectional GRU	835,596

<Table 2> Training and Testing Time Comparison

Type of RNN	*Training Time(s)*	*Testing Time(s)*
Unidirectional LSTM	407.86	0.97
Bidirectional LSTM	445.53	1.27
Unidirectional GRU	398.18	0.91
Bidirectional GRU	423.43	1.03

In summary, the superior performance of the Bi-GRU model in terms of accuracy can be attributed to its ability to process input data from both past and future time steps, thus capturing the context more effectively. Although the training time is slightly longer than the GRU model, the improved accuracy justifies the trade-off. The Bi-LSTM model also performs well in terms of accuracy, but its longer training and testing times make it less appealing for certain applications where computational efficiency is crucial. The unidirectional models (LSTM and GRU) are less accurate, primarily due to their inability to capture future context. However, they offer faster training and testing times, which may be desirable in some cases.

5. Conclusion

This paper presented a comprehensive evaluation of unidirectional and bidirectional recurrent neural networks for User Equipment (UE) path prediction in the context of Beyond 5G and 6G networks. The experimental results demonstrated the superior performance of bidirectional models, particularly the Bi-GRU model, in terms of prediction accuracy. The bidirectional models' ability to process input data from both past and future time steps enabled them to capture context more effectively, resulting in improved UE path predictions. While the Bi-GRU model exhibited slightly longer training times compared to the unidirectional GRU model, its enhanced accuracy justifies the trade-off. The Bi-LSTM model also performed well in terms of accuracy; however, its longer training and testing times make it less suitable for applications where computational efficiency is critical. In contrast, the unidirectional models (LSTM and GRU) offered faster training and testing times but were less accurate due to their inability to capture future context.

In conclusion, the findings of this study provide valuable insights into the performance of unidirectional and bidirectional recurrent neural network architectures for UE path prediction. These insights can inform the development of more effective predictive path-based mobility management strategies in the emerging B5G/6G network landscape, ultimately improving network performance and reducing erroneous and unnecessary handovers. Additionally, these findings have broader applicability beyond the context of mobility management in wireless networks, as the performance characteristics of unidirectional and bidirectional RNNs can be relevant to other areas, such as natural language processing, and time series forecasting, where sequence-to-sequence predictions are essential.

Acknowledgement

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICT Creative Consilience Program (IITP-2023-2020-0-01821) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation). This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2021-0-02068, Artificial Intelligence Innovation Hub).

Reference

[1] M. U. A. Siddiqui, F. Qamar, M. Tayyab, M. N. Hindia, Q. N. Nguyen, and R. Hassan, "Mobility management issues and solutions in 5g-and-beyond networks: A comprehensive review," *Electronics*, vol. 11, p. 1366, 2022.

[2] M. K. Shehzad, L. Rose, M. M. Butt, I. Z. Kovács, M. Assaad and M. Guizani, "Artificial Intelligence for 6G Networks: Technology Advancement and Standardization," *IEEE Vehicular Technology Magazine*, vol. 17, no. 3, pp. 16-25, 2022.

[3] H. Zhang and L. Dai, "Mobility prediction: A survey on state-of-the-art schemes and future applications," *IEEE Access*, vol. 7, pp. 802-822, 2019.

[4] G. Kaur, R. K. Goyal, and R. Mehta, "An efficient handover mechanism for 5g networks using hybridization of lstm and svm," *Multimedia Tools and Applications*, vol. 81, pp. 37057–37085, 2022.

[5] V. Yajnanarayana, "Proactive mobility management of ues using sequence-to-sequence modeling," 2022 National Conference on Communications (NCC), pp. 320–325.