Deep Learning Based User's Demand Predictor

Kyi Thar, Nguyen H. Tran, Choong Seon Hong

Department of Computer Science and Engineering, Kyung Hee University.

{kyithar, nguyenth, cshong}@khu.ac.kr.

Abstract

Edge network caching and network virtualization are the most promising technologies among the key technologies to fulfill the requirements of next-generation cellular networks. Caching popular contents at edge nodes (base stations) becomes a promising solution to improve the user's quality of services, as well as to reduce the network traffic for each Mobile Virtual Network Operator (MVNO). However, a decision to buy cache space to efficiently store popular contents is the challenging issue for each MVNO. So, in this paper, we proposed deep learning based user's demand predictor to support cache space buying decision of each MVNO. We develop, train and test the prediction models using Keras libraries.

1. Introduction

According to the Cisco Visual Networking Index, watching videos from wireless devices has been generating most of the Internet traffic and is forecast to continue to grow exponentially. In order to handle the overwhelming Internet traffic, several future Internet network architectures have been proposed to support in-network caching capability. Also, Wireless Network Virtualization (WNV) support scheme to share resources (e.g. cache space, computation, spectrum, the core network) among Infrastructure Providers (InPs) and Mobile Virtual Network Operators (MVNOs), in order to reduce the capital expenses (CapEx) and operation expenses (OpEx). Thus, with the help of in-network caching and virtualized cache space sharing, edge nodes (Base Stations (BSs) and Small-cell Base Stations (SBSs)) temporarily store video contents in their virtual cache space to satisfy user requests in the near future. Thus, the cache space buying decision to store efficiently and effectively becoming challenging issues, where the future user demand is unknown. Therefore, in this paper, we focus on the user's demand prediction scheme to support the cache storage space buying decision of MVNOs, where user demand information is time series data. Thus, we utilize the deep learning model such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU)), which are suitable for time series predicting.

Our contributions are summarized as follows:

- We utilize Deep Recurrent Neural Network model such as (LSTM, GRU) to predict the future user's demand.
- We utilize the Keras [1] machine learning libraries to construct the prediction model and train the prediction model in centralize server.
- We utilize Movie Lens [2] 1M dataset to train and validate the performance of proposed prediction model.

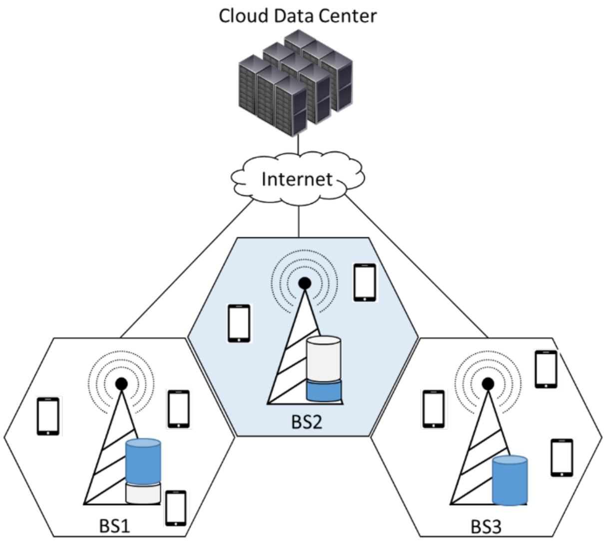

Figure 1 System Model

2. System Model

The system model of the proposed scheme is shown in Fig.1, where the infrastructures such as BSs/SBSs are owned by one InP, and those BSs/SBSs are attached to the cache storage. With the help of virtualization technology, physical cache storage at each BS/SBS can share among MVNOs as virtual cache storage space, where each MVNO can request required virtual cache space from InP. Also, MVNO requests the required virtual cache space depending on its user's demand. MVNO collects user's information from each BS/SBS at the cloud data center/centralized server. Then, MVNO train the prediction model and predict the future user's demand by utilizing the collected data at cloud data center. Then, the predicted user demand will be used as one of the input parameters to make a cache space buying decision at every time t. In this paper, the cache space buying decision process is out of scope and we will only focus on predicting user's demand because of pages limitation.

2. Demand Prediction

For the user's demand prediction, we utilize the Recurrent Neural Network (RNN) because user's requests are in a sequential manner. We use many to one RNN model which has 50 cells in each layer. Among the variants of RNN, we utilize Long Short-Term Memory (LSTM) [3] and Gated Recurrent Unit (GRU) [4] and test the performance. Then, the optimized trained model is stored at the cloud data center. The user's demand prediction models are shown in Fig. 2, where we use Mean Squared Error (MSE) as loss function and ADAM [5] optimizer to minimize training loss. We choose "tanh" function as the activation function.

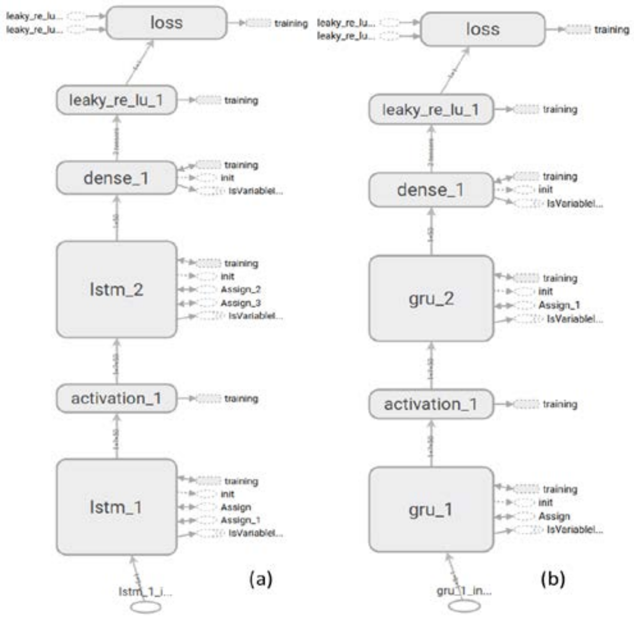

Figure 2 Prediction Model (a) LSTM, (b) GRU

5. Performance Evaluations

The performance comparison of different optimizer (ADAM, Stochastic Gradient Descent (SGD), RMSprop [6]) is shown in Fig. 3, where ADAM optimizer gives minimum MSE error among three optimizers. Thus, we choose ADAM optimizer to train our proposed user's demand prediction model. The training and testing results of proposed model which utilized LSTM is shown in Fig. 4, in which train score by mean of Root Mean Squared Error (RMSE) is 0.01 in scaled output and 368.89 in un-scaled output. As for the test score, 0.04 in scaled output and 5557.63 in un-scaled output. The training and testing results of proposed model which utilized GRU is shown in Fig. 5, in which train score by mean of Root Mean Squared Error (RMSE) is 0.07 in scaled output and 2467.64 in un-scaled output. As for the test score, 0.08 in scaled output and 5892.23 in un-scaled output.

3. Conclusion

In this paper, we proposed the scheme to predict the user's demand for MVNO to support cache storage space buying decision. As results from performance evaluations section, prediction model with LSTM gives the better performance than the GRU. As for the future work, we will merge user's demand prediction into cache storage space buying decision of MVNO.

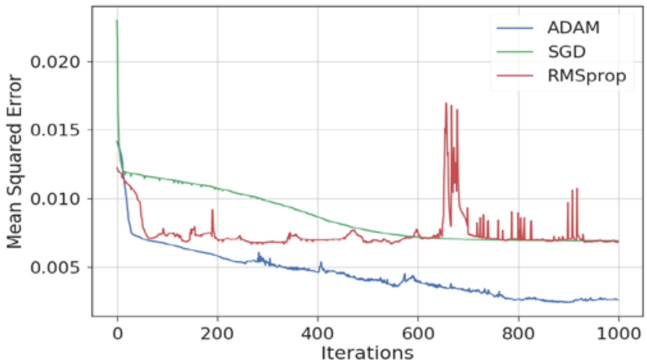

Figure 3 Performance comparison of different optimizer

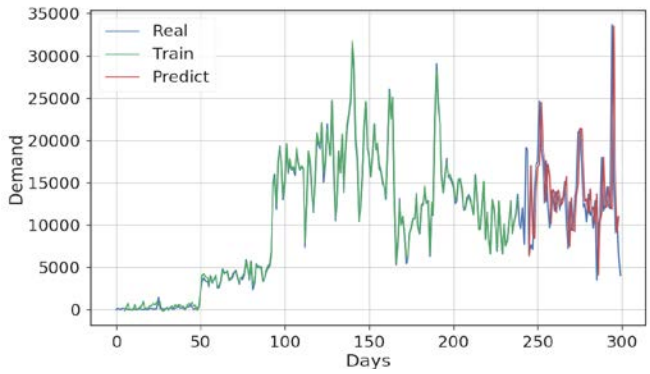

Figure 4 Result of LSTM

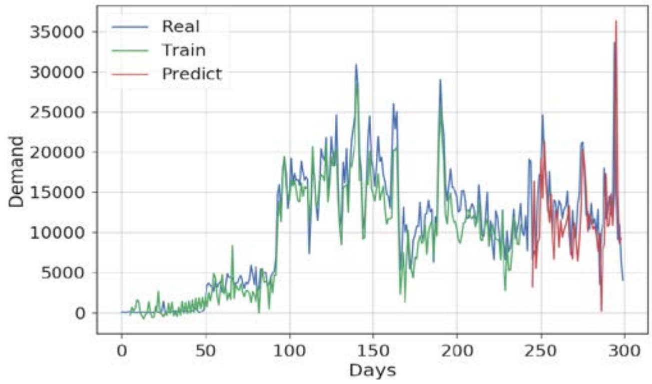

Figure 5 Result of GRU

4. 6. Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2017R1A2A2A05000995). *Dr. CS Hong is the corresponding author

3. References

[1] https://keras.io
[2] https://grouplens.org/datasets/movielens/
[3] Hochreiter, et al., "Long short-term memory" in Neural computation, vol. 9.8, pp. 1735-1780, 1997.
[4]Chung, Junyoung, et al., "Gated feedback recurrent neural networks." in International Conference on Machine Learning, 2015.
[5] Kingma, et al. "Adam: A method for stochastic optimization." in arXiv preprint arXiv:1412.6980 , 2014.
[6] Tieleman, et al., "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude.", pp. 26-31, 2012.