



# Sequence-based Recommendation with Bidirectional LSTM Network



Hailin Fu , Jianguo Li , Jiemin Chen , Yong Tang , Jia Zhu  
hailin@m.scnu.edu.cn - 510000,  
School of Computer Science, South China Normal University

## Abstract

In modern recommendation systems, most methods often neglect the sequential relationship between items. So we propose a novel Sequence-based Recommendation model with Bidirectional Long Short-Term Memory neural network (BiLSTM4Rec) which can capture the sequential feature of items to predict what a user will choose next. By collecting consumed items of a user in a sequence with time ascending order, fitting the model with the last item as the label, the rest items as the features, we regard this recommendation assignment as a super multiple classification task. Once trained well, the output layer of our model will export the probabilities of the next items with given sequence. In the experiments, we compare our approach with several commonly used recommendation methods on a real-world dataset. Experimental results indicate that our sequence-based recommender can perform well for short-term interest prediction on a sparse, large dataset.

## Introduction

The primary assignments of recommendation systems faced with consist of two parts: ratings predicting and products recommendation. So to predict what a user will choose next given his consumed history is one of the crucial mission [2] in recommendation area. In many websites and applications, such as online electronic business, news/videos website, music/radio station, they need an excellent service for users to recommend what they will like in future. Existing recommenders mainly concentrate on finding the neighbor sets for users or items, or leveraging other explicit/implicit information (such as tags, reviews, item contents and user profiles) for neighborhood-aware. However, to the best of our knowledge, few works use the sequential feature of data to build recommender. We find that the sequence of data implicates much exciting and relevant information, for example in a video website, the user who watched "Winter is coming" (S01, E02 of Game of Thrones) will be more likely to watch "The Kingsroad" (S01, E02 of Game of Thrones). Even at the 2011 Recsys conference, Pandora's researchers gave a speech about music recommendation and said they found many users consumed music in sequences. Our work was inspired by the previous study of Siwei Lai et al. [1], where a neural network is proposed to capture the sequence of words in a sentence. We took a similar approach by considering one item as a word, the catalog of items as a vocabulary, and the historical consumed items of one user as a sentence, to capture the sequence of the user consumed items.

## Architecture

Let us assume that  $\mathbb{U} = \{u_1, u_2, \dots, u_N\}$  is the user set and  $\mathbb{I} = \{i_1, i_2, \dots, i_M\}$  is the item set. For each user  $u$ , there is an observed consumed items sequence  $\mathbb{S}_u = \{s_u^1, s_u^2, \dots, s_u^{t-1}, s_u^t\}$  in ascending order of time, where  $s_u^t$  is the item consumed by user  $u$  at time  $t$ . The sequential prediction problem is to predict  $s_u^{t+1}$  for each user  $u$ .

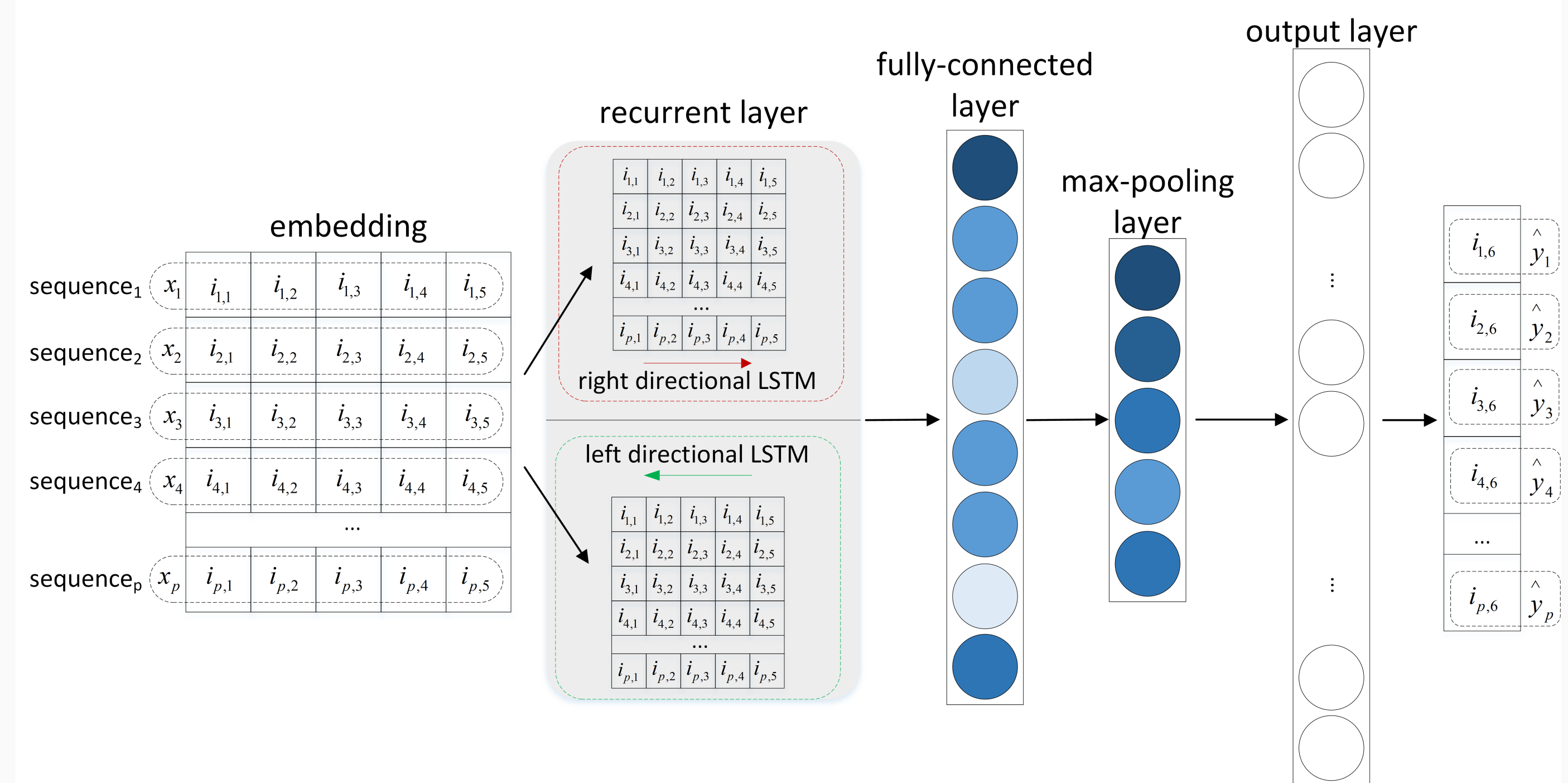


Fig. 1: Framework of sequence-based recommendation

We propose a novel deep neural model to capture the sequence feature of the user's consumed data. Our model mainly consists of five layers: embedding, recurrent structure, fully-connected layer, pooling layer and output layers. Figure 1 shows the structure of our sequence-based recommender.

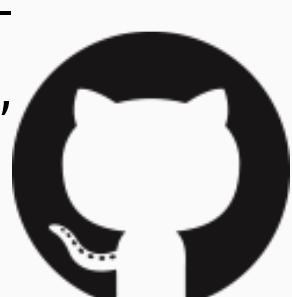
## References

- [1] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. Recurrent convolutional neural networks for text classification. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, pages 2267–2273. AAAI Press, 2015.
- [2] Shengxian Wan, Yanyan Lan, Pengfei Wang, Jiafeng Guo, Jun Xu, and Xueqi Cheng. Next basket recommendation with neural networks. In *RecSys Posters*, 2015.

## Get contacted!

BiLSTM4Rec is currently hosted on GitHub and I am working as a recommendation algorithm engineer intern at *CloudBrain.ai*. Experience exchanging, academic discussion and contributions are strongly encouraged!

<https://github.com/fuhailin/BiLSTM4Rec>



## Mathematical Section

Word Embedding shines in the field of Natural Language Processing, instead of ending up with huge One-hot encoded vectors we also can use an embedding matrix to keep the size of each vector much smaller:

$$e(l_i) = E l_i \quad (1)$$

$$h_b(l_i) = \sigma(W^{(b)} h_b(l_{i-1}) + W^{(cb)} e(l_{i-1})) \quad (2)$$

$$h_a(l_i) = \sigma(W^{(a)} h_a(l_{i+1}) + W^{(ca)} e(l_{i+1})) \quad (3)$$

$$x_i = [h_b(l_i); e(l_i); h_a(l_i)] \quad (4)$$

$$y_i^{(2)} = \tanh(W^{(2)} x_i + b^{(2)}) \quad (5)$$

$$y_i^{(3)} = \max_{i=1}^n y_i^{(2)} \quad (6)$$

$$y_i^{(4)} = W^{(4)} y_i^{(3)} + b^{(4)} \quad (7)$$

Finally, a softmax activation function applied to  $y_i^{(4)}$ , which can convert the output values to the probabilities of next items.

$$p_i = \frac{e^{y_i^{(4)}}}{\sum_{k=1}^n e^{y_k^{(4)}}} \quad (8)$$

We define all of the parameters to be trained as  $\theta$ .

$$\theta = \{E, b^{(2)}, b^{(4)}, h_b(B_1), h_a(B_n), W^{(2)}, W^{(4)}, W^{(b)}, W^{(a)}, W^{(cb)}, W^{(ca)}\} \quad (9)$$

The training target of the network is to minimize the categorical cross entropy loss:

$$\mathcal{L}(y, S, \theta) = - \sum_{u \in \mathbb{U}} [y_u \log p(y_u | S_u, \theta) + (1 - y_u) \log(1 - p(y_u | S_u, \theta))] \quad (10)$$

## Results

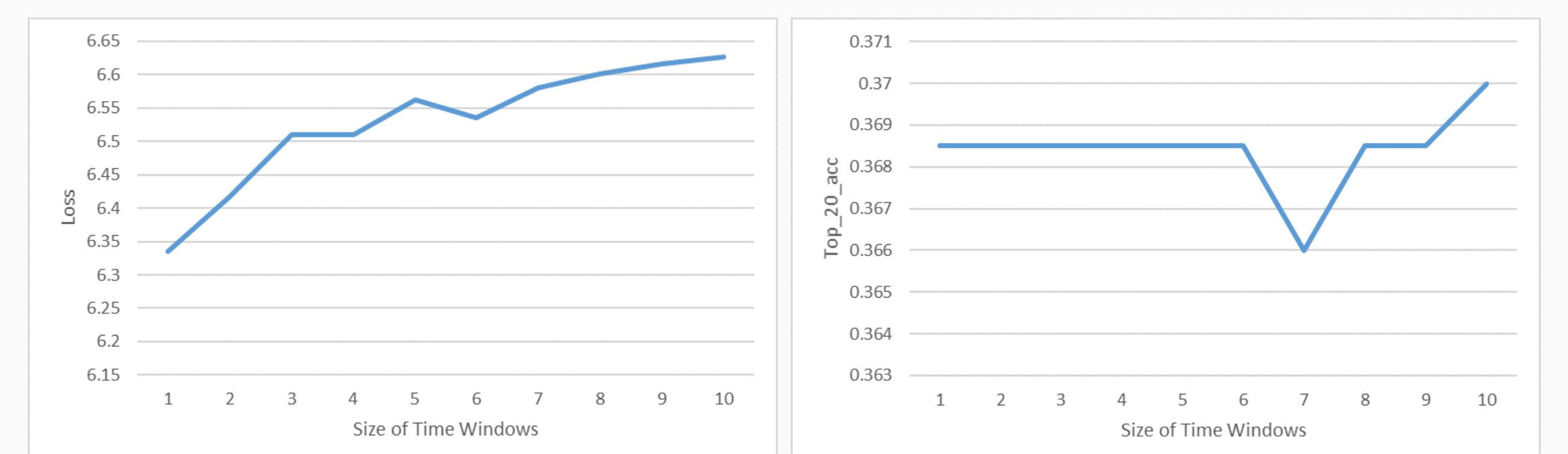


Fig. 2: The training results of BiLSTM4Rec model on 80% LiveStreaming-10M with different Time Windows

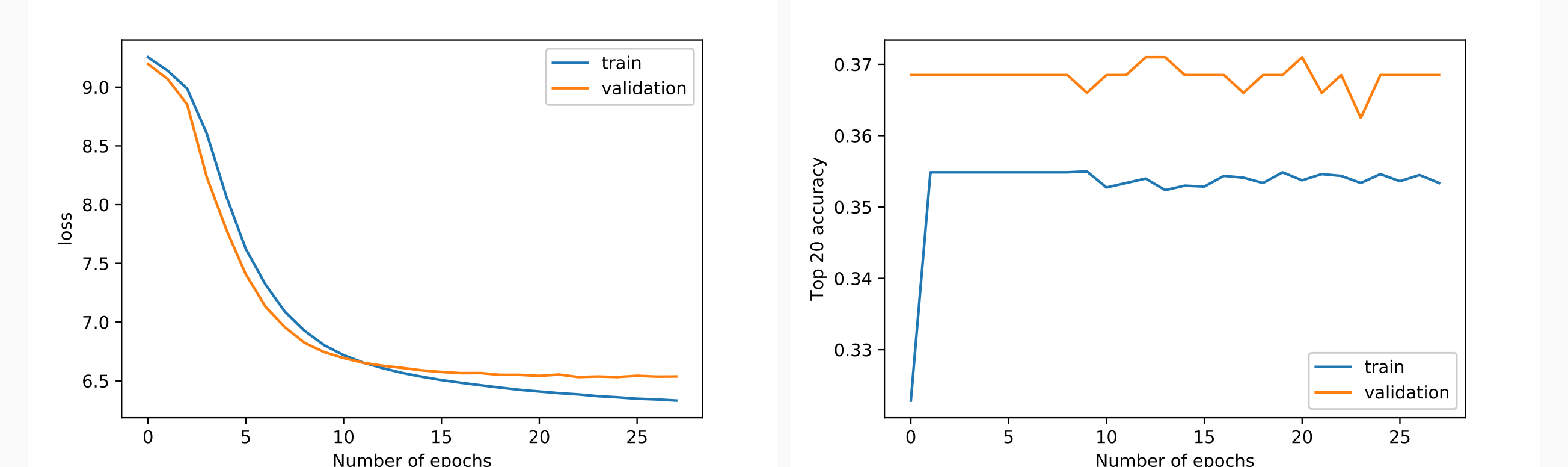


Fig. 3: BiLSTM4Rec training on 80% LiveStreaming-10M with Time Windows as 5, and recommends items with the top 20 probability values

## Comparisons

Metrics	POP	UBCF	IBCF	LSTM	BiLSTM4Rec
Precision@1	0.0056	0.0108	0.0112	0.1135	<b>0.1165</b>
Recall@1	0.0056	0.0108	0.0112	0.1135	<b>0.1165</b>
F1@1	0.0056	0.0108	0.0112	0.1135	<b>0.1165</b>
Precision@5	0.00018	0.0476	0.006223	0.0463	<b>0.0488</b>
Recall@5	0.0009	0.0476	0.0313	0.2315	<b>0.244</b>
F1@5	0.0003	0.0476	0.010433333	0.077167	<b>0.081333</b>
Precision@20	0.00015	<b>0.1154</b>	0.004705	0.01825	0.01895
Recall@20	0.003	0.1154	0.0941	0.365	<b>0.379</b>
F1@20	0.000285714	<b>0.1154</b>	0.008961905	0.034762	0.036095

**Conclusion** BiLSTM4Rec is a novel recommendation by modeling recent consumed items as a "sentence" to predict what users will choose next. We introduced the Bidirectional Long Short-Term Memory neural network to a new application domain: recommendation system. We dealt with consumed items sequences by embedding matrix to save memory cost, and the final model can learn short-term interest of the user. Experimental results show that our approach outperforms existing methods to a great extent, and shows it is suitable to do a short-term prediction. Moreover, it doesn't have an unacceptable time complexity.