

Intention-aware for Session-based Recommendation with Multi-channel Network*

WANG Jing-jing, Oliver Tat Sheung Au, Lap-Kei Lee

School of Science and Technology, The Open University of Hong Kong, Hong Kong, China

Session-based recommendation predicts the user's next action by exploring the item dependencies in an anonymous session. Most of the existing methods are based on the assumption that each session has a single intention, items irrelevant to the single intention will be regarded as noises. However, in real-life scenarios, sessions often contain multiple intentions. This paper designs a multi-channel Intention-aware Recurrent Unit (TARU) network to further mining these noises. The multi-channel TARU explicitly group items into the different channels by filtering items irrelevant to the current intention with the intention control unit. Furthermore, we propose to use the attention mechanism to adaptively generate an effective representation of the session's final preference for the recommendation. The experimental results on two real-world datasets denote that our method performs well in session recommendation tasks and achieves improvement against several baselines on the general metrics.

Keywords: Intention-aware network; Session-based recommendation; Recommendation

1 Introduction

Recommender systems play an essential role in real-world applications, to assist users in finding their needs or interests from a large-scale candidate pool based on the sequential records of user-item interactions. In some scenarios, personal attributes are not provided, only anonymous and chronological behavior orders collected from the sessions are available. Most of the existing session recommendation methods such as utilize attention mechanism (STAMP) (Qiao et al., 2018) or Recurrent Neural Network (RNN) (GRE4REC) (Hidasi, et al., 2015), model the item dependencies based on the assumption that each session has a single intention. Items irrelevant to the single intention will be regarded as noises. Taking the session illustrated in Figure 1 as an example, the user first added the phone into the cart, then she searched the phone case, during this shopping, it occurred to her that she needed to buy a new teakettle and a new vacuum cup. In this case, methods based on attention (STAMP) take the most appeared attributes as the long preference and the last item as the short preference, thus a new phone case will be ranked higher in the recommended list since the user viewed phone case as last, and attributes about phone will be taken as the main purpose while “cup” will be considered as noises. However, it is not a simple task to identify noises, usually these noise item effects will be limited by assigning a smaller weight by the attention mechanism or a narrower resetting gate in Recurrent Neural

* **Acknowledgments:** This work is supported by Katie Shu Sui Pui Charitable Trust-Academic Publication Fellowship (Project Reference No. KSPF 2020-01).

WANG Jing-jing, Ph.D. candidate, School of Science and Technology, The Open University of Hong Kong.

Oliver Tat Sheung Au, Ph.D., Assistant Professor, School of Science and Technology, The Open University of Hong Kong.

Lap-Kei Lee, Ph.D., Assistant Professor, School of Science and Technology, The Open University of Hong Kong.

Network (RNN).



Figure 1. An example of session with two intentions.

Obviously, such approaches fail to differentiate the purpose-specific item dependencies (e.g., items for phone vs. items for cup) when the items in a session serve multiple purposes. Based on our observations, there are three shortcomings in methods based on attention mechanism and RNN. In the Attention-based method, the limited effectiveness of the irrelevant items can still dilute the final session target generation significantly, especially sessions containing multi-intention or diverse items. In the RNN-based method, it is easy to generate false dependencies as not all items depend on each other in a multi-purpose session. To both methods, it is a challenging task to detect the main intention accurately. However, in real-life scenarios, sessions containing multiple intentions is commonly occurred. Thereby, it has great potential to improve the recommend accuracy by exploring these noise item dependencies and adaptively generate an effective representation of the session’s final preferences.

In this paper, we propose to design a multi-channel Intention-aware Recurrent Unit (TARU) network to further mining these noises. The multi-channel TARU explicitly group items into the different channels by filtering items irrelevant to the current layer purpose with the intention control unit. Furthermore, we propose to use the attention mechanism to adaptively generate an effective representation of the session’s final preference for recommendation. Our contributions include:

- We propose an Intention-aware Recurrent Unit (TARU) to jointly identify and learn users’ multiple intentions from the different channels. Each channel will be fed into the different subsequence which reflects the same intention.
- In TARU, we propose an intention control unit to judge whether the current timestamp item need to be filtered out and avoid the repeated intention independences learning simultaneously. Attention mechanism will be employed to fuse the multi-channel intentions.
- We carried out experiments on two public datasets, the results show the effectiveness of TARU in terms of Recall and MRR.

2 Related Work

There are two types of work that is closely related to ours: session-based recommendation and intention-aware recommendation.

2.1 Session-based Recommendation

Early research about session-based recommendation is based on Markov chains, aiming to obtain the transition matrix from the given sequential data. The transition matrix provides the probabilities from the current state to the next timestamp state, then according to the transition matrix to predict. However, this method cannot capture the long-term dependency and are not suitable for the personalized recommendation. Rendle (2010) proposes incorporating the matrix factorization with Markov chains, named FPMC, to make the transition matrix with personalized features. To solve the limit ability in capturing the long-term dependency

about the Markov chains, Recurrent neural networks (RNNs) are proposed. GRU4Rec (Hidasi et al., 2015) was the first attempt to model the session-based recommendation with RNN and achieved an encouragement result. After that, more works about RNN have been proposed, for example Liu (2016) introduced RNN to predict the next location, Yu (2016) applied RNN in the basket prediction tasks. Compared to the conventional methods, RNN-based methods have achieved the promising improvements, however it is easy to generate false dependency between items, because RNN totally rely on the sequential data and cannot tackle noises, for example item clicked accidentally or out of curiosity. To overcome this limitation, Li (2017) proposes that both user's main purpose and the sequential data are important to the recommendation accuracy, they design a hybrid encoder with attention mechanism to emphasize the main purpose in the current active session. Due to the success of Li, most of the following works incorporate the attention mechanism to the sequential recommendations to avoid the effectiveness of noise. For example, Liu (2018) applied attention mechanism to assign different importance to the sequential data to jointly obtain the long-term intention and the short-term intention in the predict task. Inspired by the Transformer (Vaswani et al., 2017), Kang and McAuley (2018) modeled the sequential recommendation task with the self-attention mechanism to distinguish the importance from the others, then applied a multi-layer feed forward network to predict the next action. Other works, such as Wang (2019) proposed that collaborative information from the neighborhood sessions was useful in improving the predict accuracy, thus they designed two parallel memory networks to capture the current session and its neighborhood separately. Wu (2019) modeled each sequential data as a graph, and applied the gated graph neural network to exploit the complex item dependencies.

2.2 Intention-aware Recommendation

Methods based on attention or RNN will perform poorly as the effects of the contributing target item are diluted by diverse historical records. Zhang (2019) adopted the reinforcement learning algorithm to solve the attention contributing problem, they revised user profiles by removing the noises instead of assigning a smaller attention coefficient to each of them. Guo (2020) proposed a Leap Recurrent Unit (LRU) to identify whether to skip the item to ensure the current channel preference was new-learned. However, item irrelevant to the current purpose can not be filtered out. What's more, the preference manager cannot distinguish the importance between different channels. Wang (2019) proposed a mixture channel purpose for the session next item recommendation, in their method, routing network was used to calculate the relation between items and each preference, then items with different importance coefficient will be fed into each RNN channel, separately. However, this method also suffered the same problem as methods based on attention. Chen (2020) integrated the intent-aware diversity promoting (IDP) module and implicit intent mining (IIM) components into a unified recommender system, the IIM exploited users' multiple intentions and IDP balanced the recommendation accuracy and diversity. Cen (2020) introduced a multi-purpose module to capture the multiple intentions from the sequential data and a controllable factor to generate the prediction. Pan (2020) modeled a basket preference sensitive neural network for item recommendation. The basket preference is a composition of the recent preference, the main preference from the current basket and the historical sequential baskets separately.

3 Problem Formulation

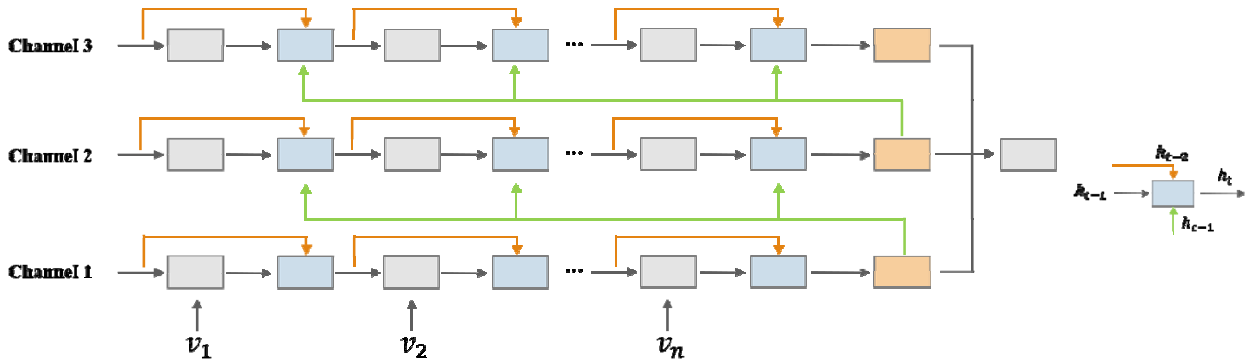


Figure 2. The architecture of TARU. Each blue rectangle represents the intention control unit, and the structure of the intention control unit is shown in the right. Orange rectangle is the current channel intention. TARU recurrently feed the sequential data into each channel, then the intention control unit will determine whether it is item or noise, at last aggregate all intentions to generate the session’s representation.

We first formulate the task of intention-aware for session-based recommendation as follows. Given $S = [s_1, s_2, \dots, s_N]$ denotes the set of all anonymous session sequences, the set of items I denotes all unique items collected from S , and $I = \{i_1, i_2, \dots, i_M\}$, N, M is the total number of sessions and items, separately. An anonymous session sequence s can be represented as $s = \{i_1, i_2, \dots, i_t\}$, where i_t denotes the last timestamp t clicked item in s . Given previous clicked item sequences, the task of our method is to compute the probability distribution of the next clicked happened at the timestamp $t + 1$ from the candidates.

4. Multi-channel Networks

The architecture of TARU is illustrated in Figure 2. Each channel has the same structure but different input. In specifically, TARU cell integrates the intention control unit into the GRU cell to filter out the noise from the whole input when generating the specific channel purpose, the detailed intention control unit is shown in the right part of Figure 2. To ensure each channel purpose’s independent, the preference control unit need identify and pop items, which closely related to the other learned preferences or unrelated to the current channel preference. Then each output from the TARU will be regarded as the specific channel intention, aggregate all the channel outputs to generate the final session representation.

4.1 Preference-Aware Recurrent Unit

Memory networks like long short-term network (LSTM) and gated recurrent unit (GRU) can reserve the long-term purpose from the sequential clicks, and jointly combine the short-term preference. These long short-term memory networks achieve comparable results in the field of sequential recommendations. However, methods based on the memory networks assume that (a) users have a single purpose when searching needs, (b) the records in the session (click or purchase) play the same role when predicting the next item. This hypothesis limited representation of models with multi-purpose session. Besides, these models cannot distinguish the noise in the sequence when click with carelessness or out of curiosity.

We propose to add an intention control unit into the traditional GRU cell. The intention control unit is used to judge whether the current timestamp node is a noise. As shown in Figure 2, it is defined as:

$$flag_{c,t} = (s(h_{c,t}, h_{c,t-1}) > \varepsilon_1) \text{ and } \left(\prod_{i=0}^{c-1} (s(h_{c,t}, p_i) < \varepsilon_2) \right) \quad (1)$$

$$p_c = GRU(x_c) \quad (2)$$

Where c is the index of the TARU channel, $h_{c,t}$ is the output of channel c when $timestamp = t$, p_i is the intention of the channel i , calculated by the equation 2. There are two factors to determine the value of indicator $flag_{c,t}$. One is the similarity between the output of GRU when $timestamp = t$ and the output of GRU when $timestamp = t - 1$, the other one is the similarity between the current channel learned intention $h_{c,t}$ when $timestamp = t$ and the other channel learned intentions p_i . If the similarity between $h_{c,t}$ and p_i larger than the threshold parameter ε_2 , or extremely changes happened makes the similarity between $h_{c,t}$ and $h_{c,t-1}$ smaller than the threshold parameter ε_1 , $flag_{c,t} = 0$, thus x_t will be regarded as the noise, the GRU will pop the timestamp t item from the subsequence. It is defined as:

$$h_{c,t} = \begin{cases} GRU(h_{c,t-2}), & flag_{c,t} = 0 \\ GRU(h_{c,t-1}), & flag_{c,t} = 1 \end{cases} \quad (3)$$

There are many methods to calculate the similarity, in this paper, we choose cosine to compute the similarity between two vectors, as cosine is the commonly used and easily calculate method. Before calculating similarity, x_t needed to translate into the GRU space, since the calculation of similarity will be meaningless if the two vectors are in different spaces.

4.2 Intention Fusion

TARU feeds the subsequences into different channel and generate each unique representation under the intention. Here, note that our method can deal with the multi-intention sessions, which doesn't mean all the sessions have more than one intention, especially the short sessions. It is dangerous to take all the channel intentions as the same important. In our method, the first channel intention p_0 will be regarded as the main intention, as it is the newly learned presentation for the others. In order to effectively take advantage of these intentions, we propose to utilize the attention mechanism to dynamically select different channel intentions to generate the final one.

$$p = \sum_{i=0}^n \alpha_{i0} p_i \quad (4)$$

$$\alpha_{i0} = \text{attn}(p_0, p_i) = v_1 \sigma(v_2 p_0 + v_3 p_i) \quad (5)$$

Here σ denotes an activate function, for example *Leakrelu* function, matrix v_2, v_3 play the same role as the linear transition, to transform p_0 into the same space.

4.3 Prediction and Training

After obtained the session representation p , we compute the matching score vector by multiplying the candidate item sx_i with the session representation p^T , as follows.

$$\hat{z}_i = p^T x_i \quad (6)$$

In order to choose top-k items as the final recommend ones from the score vector \hat{z} , where $\hat{z} \in \mathbb{R}^{|V|}$. We apply a softmax function to calculate the probability distribution over all the candidates with follow equation:

$$\hat{y} = \text{softmax}(\hat{z}) \quad (7)$$

For any given session prefix $S_t \in S$ ($t \in [1, \dots, N]$), we train and optimize the cross-entropy loss of the

prediction results \hat{y} as follows:

$$L(\hat{y}) = - \sum_{i=1}^{|M|} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (8)$$

5 Experiments

5.1 Datasets

We evaluate our proposed method on two publicly available datasets: (1) Yoochoose¹: Yoochoose dataset is released by RecSys Challenge 2015, which collected from an e-commerce website to predict whether the user will buy the item based on the clicked sequences. (2) Diginetica²: this dataset is also obtained from the e-commerce website, including the anonymous search and purchase logs of Diginetica. In our experiment, only the transaction data of Diginetica is used.

For the Yoochoose dataset, following Li (Li et al., 2017) and Liu (Liu et al., 2018), we evaluate our model on the sub-dataset Yoochoose1/64 and Yoochoose1/4, which are the more recent sequences from the sorted whole dataset. To both datasets, sessions with only one item and items appear less than 5 times in the whole dataset will be filtered out. Furthermore, same to NARM, we truncate the long session to $length = 19$. The statistics of the two datasets is shown in the Table 1.

Table 1

Statistics of the used datasets in our experiments

Datasets	YOOCHOOSE 1/64	YOOCHOOSE 1/4	DIGINETICA
# clicks	557,248	8,326,407	982,961
# train	369,859	5,917,746	719,470
# test	55,898	55,898	60,858
# items	16,766	29,618	43,097
#Avg. length	6.16	5.71	5.12

5.2 Experimental Settings

5.2.1 Baselines

We compare and analyze the proposed model with eight representative baselines, including conventional methods and deep learning algorithms. The conventional methods include: (a) recommend the next-click based on item popularity of the current session (S-POP); (b) predict the next item based on matrix factorization (BPR-MF, FPMC), BPR-MF used a pair-wise objective function to optimize the matrix factorization, FPMC incorporates the Markov chain into matrix factorization to solve personalized problem. Deeping learning method can further divide into methods based on GRU (NARM), methods base on graph neural network (SR-GNN), methods based on MLP (STAMP). We also compared our results with a multi-channel method, named HLN.

5.2.2 Evaluation metrics

In our experiments, we evaluate the quality of the recommendation matters with two commonly used metrics: Recall and Mean Reciprocal Rank (MRR). Recall@k is a well-known metric which represents the proportion of the target item in the top-k ranked recommended items while ignoring the actual rank. MRR@k is

¹ <http://2015.recsyschallenge.com/challenge.html>

² <http://cikm2016.cs.iupui.edu/cikm-cup>

sensitive to the rank of items if the rank $> k$, $MRR=0$, else $MRR = 1/k$. We set $k = 10$ in our reported results.

5.2.3 Experimental setup

The number of the batch size is set to 512, and the dimension of embedding size is set to 50, the hidden size of GRU is set to 50. We choose three channels for the multi-intentions. We train and optimize our method with Adam. The initial learning rate $lr = 0.005$ and decay by 0.1 after each 3 epochs. The threshold of each channel, $\varepsilon_1 = 0.7$, $\varepsilon_2 = 0.7$. We apply dropoutlayer to avoid overfitting and $dropout = 0.4$.

Table 2

Performance comparison of TARU with baseline methods

	YOOCHOOSE 1/64		YOOCHOOSE 1/4		Diginetica	
	Recall@10	MRR@10	Recall@10	MRR@10	Recall@10	MRR@10
S-POP	3.24	1.79	1.52	0.46	0.68	0.16
BPR-MF	31.30	18.89	22.93	16.24	4.21	1.89
FPMC	37.44	20.05	36.12	18.54	15.43	6.20
GRU4REC	52.43	24.53	55.49	26.05	17.93	7.73
NARM	57.83	27.42	57.98	28.51	35.44	15.13
STAMP	52.96	25.17	57.67	28.32	33.98	14.26
SR-GNN	58.01	28.92	61.06	31.08	38.42	16.89
HLN	58.91	29.86	61.47	31.37	38.60	17.02
OUR-METHOD	59.49	30.00	61.98	31.44	39.94	18.56

5.3 Performance Comparison

Table 1 presents the detailed performance of the eight baseline methods and our method on three datasets, the best result has been highlighted in boldface. From Table 1, we have the following findings: (1) TARU outperforms both conventional methods and neural network approaches. This ascertains the ability of our method in distinguishing multi-intention of sessions. Recent neural approaches learn the dynamic preferences from the sequential data by considering the intention as the composition of the main purpose and the current one, which is not sufficient for the diverse intentions. Our method is also better than HLN, although HLN can identify the preference-unrelated items in each subsequence group, it is difficult to cope with the noises such as users accidentally click or click due to curiosity. As to the preferences fusion, HLN directly utilizes the sum pooling of each group preferences as the session representation, it is dangerous to take all the group preferences as the same important, especially when the session is short or only has one intention. (2) SR-GNN achieves better than other neural network methods, demonstrating the implicit item connections are useful in exploring the more complex transition matrix. (3) The neural network methods consistently work better than the conventional methods. This result shows that neural network is more suitable for capturing the complex transition in sequential data. (4) Both RNN-based methods and the personalized matrix factorization method (FPMC) perform well than the simple methods (S-POP, BPR-MF), which demonstrates that the context of the sequential is an essential factor limit the ability in modeling the item transition.

5.4 Analysis

In our method, we assign three TARU channels to learn the final representation, it is worth to compare the performances of different aggregate functions, in this experiment, we choose sum pooling, mean pooling, max pooling, and the attention method reported in the Table 2 to compare the result.

For sum pooling, we take the sum of each channel intention as the whole session intention h_s :

$$h_s = \sum_{c=1}^{N_C} h_c \quad (9)$$

Similar to sum pooling, max pooling takes the max value of every dimension from each channel. Mean pooling calculates the final result used of the mean value instead of max.

Figure 3 shows the performance of different aggregation operations on the Yoochoose 1/64 and Yoochoose 1/4. It can be observed that TARU with attention fusion function outperforms other aggregation operations on Yoochoose in terms of different k@Recall.

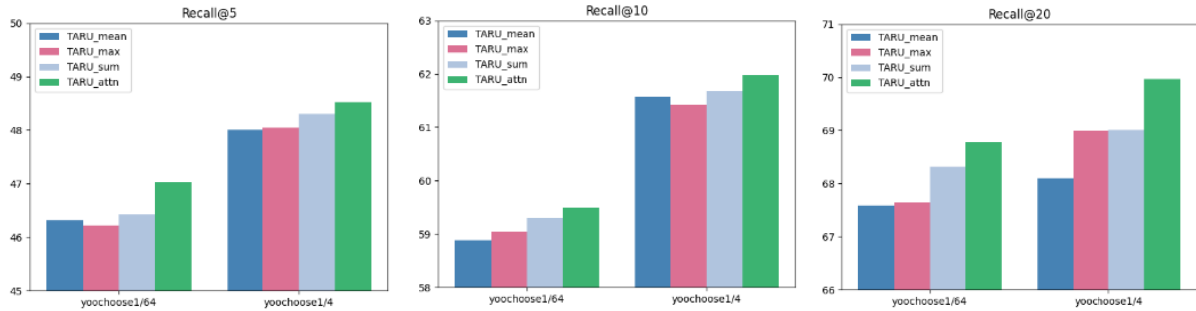


Figure 3. Comparison of the results of different functions.

Conclusions

In this paper, we proposed a multi-channel Intention-aware Recurrent Unit network (TARU) to explicitly filter out the noises and group items into different channels. This proposed method not only avoids the repeated learning of the intention, but also consider alleviate the effective of noises under the current intention. The experimental results on two public datasets show the effectiveness of our method.

References

- Cen, Y., Zhang, J., Zou, X., Zhou, C., Yang, H., & Tang, J. (2020). Controllable multi-interest framework for recommendation. Paper presented at the *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.
- Guo, C., Zhang, M., Fang, J., Jin, J., & Pan, M. (2020). Session-based recommendation with hierarchical leaping networks. Paper presented at the *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*.
- Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. J. A. P. A. (2015). Session-based recommendations with recurrent neural networks.
- Kang, W.-C., & McAuley, J. (2018). Self-attentive sequential recommendation. Paper presented at the *2018 IEEE International Conference on Data Mining (ICDM)*.
- Li, J., Ren, P., Chen, Z., Ren, Z., Lian, T., & Ma, J. (2017). Neural attentive session-based recommendation. Paper presented at the *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*.
- Liu, Q., Wu, S., Wang, L., & Tan, T. (2016). Predicting the next location: A recurrent model with spatial and temporal contexts. Paper presented at the *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Liu, Q., Zeng, Y., Mokhosi, R., & Zhang, H. (2018). STAMP: short-term attention/memory priority model for session-based recommendation. Paper presented at the *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.
- Pan, Z., Chen, W., & Chen, H. J. I. A. (2020). A hybrid-Preference neural model for basket-sensitive item recommendation. 8, 226131-226141.
- Qu, S., Yuan, F., Guo, G., Zhang, L., & Wei, W. J. A. P. A. (2020). CmnRec: Sequential recommendations with

Chunk-accelerated memory network.

- Rendle, S., Freudenthaler, C., & Schmidt-Thieme, L. (2010). Factorizing personalized markov chains for next-basket recommendation. Paper presented at the *Proceedings of the 19th international conference on World wide web*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). *Attention is all you need*. arXiv preprint arXiv:1706.03762.
- Wang, S., Hu, L., Wang, Y., Sheng, Q. Z., Orgun, M. A., & Cao, L. (2019). Modeling multi-purpose sessions for next-item recommendations via mixture-channel purpose routing networks. Paper presented at the *IJCAI*.
- Wu, S., Tang, Y., Zhu, Y., Wang, L., Xie, X., & Tan, T. (2019). Session-based recommendation with graph neural networks. Paper presented at the *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Yu, F., Liu, Q., Wu, S., Wang, L., & Tan, T. (2016). A dynamic recurrent model for next basket recommendation. Paper presented at the *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*.
- Zhang, J., Hao, B., Chen, B., Li, C., Chen, H., & Sun, J. (2019). Hierarchical reinforcement learning for course recommendation in MOOCs. Paper presented at the *Proceedings of the AAAI Conference on Artificial Intelligence*.