5690

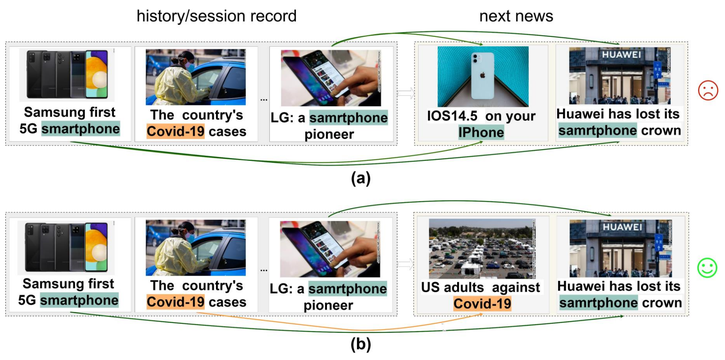
NEWS RECOMMENDATION VIA MULTI-INTEREST NEWS SEQUENCE MODELLING

Rongyao Wang， Shoujin Wang， Wenpeng Lu， Xueping Peng
School of Computer, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
School of Computing Technologies, RMIT University, Melbourne, Australia
Australian Artificial Intelligence Institute, University of Technology Sydney, Sydney, Australia

Introduction

A session-based news recommender system recommends the next news to a user by modeling the potential interests embedded in a sequence of news read/clicked by her/him in a session. Generally, a user's interests are diverse, namely there are multiple interests corresponding to different types of news, e.g., news of distinct topics, within a session. However, most of existing methods typically overlook such important characteristic and thus fail to distinguish and model the potential multiple interests of a user, impeding accurate recommendation of the next piece of news. Aiming at the above problem, we propose a novel *multi-interest news sequence* (MINS) model for news recommendation. In MINS, a news encoder is designed to learn an informative representation of each news. Then a novel *parallel channel interest network* (PIN) is designed to first detect the potential interest embedded in each news, and then models the multiple interests while each channel models one interest by taking those news with the same interest as the input. As a result, an informative news session representation is obtained by aggregating the interest representations from all channels. Finally, the dot production is employed to predict the next news by taking the session representation as the input.

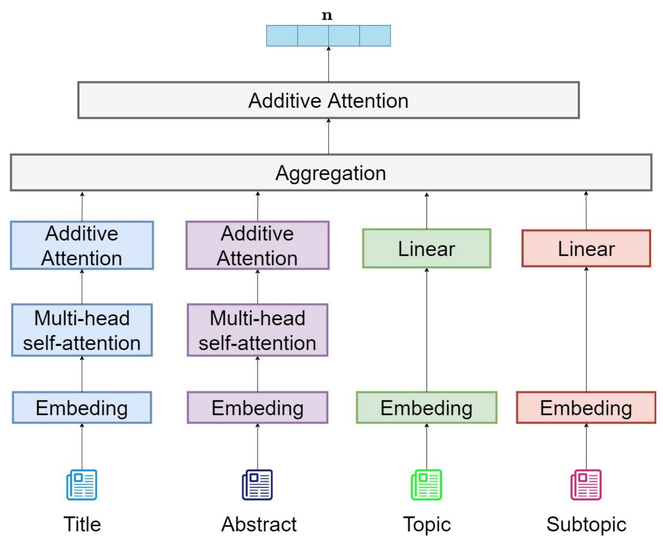
Motivation



Tom first clicked the news about *Samsung's 5G smartphone* and then he was attracted to the news on *Covid-19*, afterwards, he clicked another news about *smartphone*. In this session, Tom's main reading interest is *smartphone* revealed by the first and third piece of news and the secondary interest is *Covid-19* revealed by the second piece of news. However, most of the existing methods for news recommendation work as the upper row shown in figure (a). which are referred as single-interest methods. They only model the user's main interest in the session while ignoring the secondary interest. As a result, they only recommend the news on *smartphone* and neglect the other news that may also be of interest to the user. Different from the single-interest methods, the bottom row in Fig. 1(b) captures the multiple interests, and recommend diverse news on both *smart phone* and *Covid-19* for users. Obviously, the single-interest based method fails to satisfy multiple potential interests of users, which is inferior to the multi-interest method.

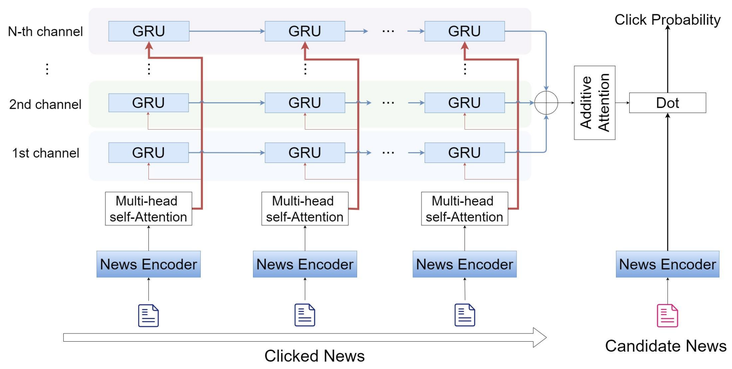
Method

News Encoder



For each piece of news, the title, abstract, topic and subtopic are inputted into the encoder and a unified news representation n is outputted. For title and abstract, we first utilize multi-head self attention to learn the rich semantic meaning from the sentences and then employ additive attention to integrate the outputs from multi-head self attention into a unified embedding vector respectively. We use linear transformation to encode the topic and subtopic. Finally, another additive attention module is employed to effectively aggregate the embedding vectors of all the four parts to build the final news representation.

Parallel-channel Interest Network



First, we utilize a multi-head self attention as an interest detector to detect the potential interests embedded in each news piece clicked at each time step, while each self-attention head vector represents one interest.

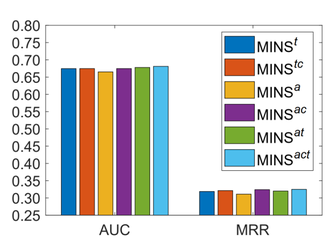
Then, we devise a multi-channel GRU-based recurrent network where each channel models the sequential dependencies over news within each interest. Finally we take the hidden state at the final step as the representation of each interest.

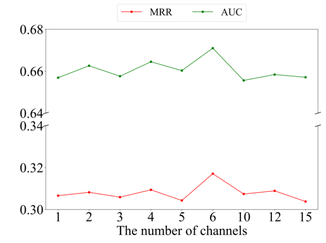
Third, we integrate all channels together to obtain the user's final compound interest representation. Specifically, additive attention is employed to emphasize the crucial information from the multiple interest representations.

Experiment

Model	MIND-small				MIND-large			
	AUC	MRR	nDCG@5	nDCG@10	AUC	MRR	nDCG@5	nDCG@10
BiasMF	0.5108	0.2258	0.2318	0.2952	0.5111	0.2257	0.2346	0.2963
DKN	0.5726	0.2339	0.2418	0.3033	0.6329	0.2902	0.3163	0.3930
LSTUR	0.6021	0.2659	0.2873	0.3529	0.5633	0.2454	0.2583	0.3252
NRMS	0.6391	0.3017	0.3282	0.3937	0.6701	0.3185	0.3534	0.4175
HiFi-Ark	0.6403	0.2996	0.3272	0.3925	0.6394	0.2969	0.3221	0.3888
TANR	0.6455	0.3107	0.3367	0.4017	0.6611	0.3148	0.3467	0.4114
MINS	**0.6710**	**0.3171**	**0.3525**	**0.4150**	**0.6811**	**0.3249**	**0.3601**	**0.4242**
Improvement[1] (%)	3.95	2.06	4.69	3.31	1.64	2.01	1.90	1.60

The experimental comparison results on MIND dataset are shown in Table 1. There are several observations. First, compared with neural network methods, the method based on statistical machine learning such as BiasMF consistently shows the worse performance on two versions of MIND dataset. This is because that BiasMF fails to capture nonlinear features and complex semantic representations for news recommendation. Second, the method applying multi-head self-attention such as NRMS outperforms the methods without multi-head self-attention (i.e., DKN, LSTUR, HiFi-Ark, TANR) on the MIND-large dataset. This may because multi-head self-attention can efficiently model news content representations. Third, our model significantly outperforms other baselines on all metrics and datasets, which improves about 4% in terms of AUC on MIND-small dataset than TANR. The reason may be that MINS tends to learn multiple interests contained in news session, which can satisfy the diverse requirements of users.





Impact of inputs: According to the figure, we found MINS^*act* achieve the best performance, which demonstrates that it is reasonable and necessary for our model to integrate the four kinds of news information.

Impact of interest channel: According to the figure, we find that when the number of channels is set to 6, our model can achieve the best performance. The less channels may fail to capture the diverse interests while the more channel may lead to overfitting issue.

Conclusion

In this paper, we propose a *multi-interest news sequence* (MINS) model for news recommendation. A parallel-interest network is devised to detect the potential interest of each news and assign it into the corresponding interest-specific channel, followed by GRU based network to generate the multi-interest representation for the session. Besides, a news encoder is devised to learn the accurate news representation with multi-head self-attentions. Our extensive experiments demonstrate that our MINS model can outperform the state-of-the-art compared baselines for news recommendation.