Multi-level News Recommendation via Modeling Candidate Interactions

Ying Sun
Institute of Automation, Chinese Academy of Sciences
School of Artificial Intelligence University of Chinese Academy of Sciences
Beijing, China
sunying2019@ia.ac.cn

Qingchao Kong*
Institute of Automation, Chinese Academy of Sciences
School of Artificial Intelligence University of Chinese Academy of Sciences
Beijing, China
qingchao.kong@ia.ac.cn

Wenji Mao
Institute of Automation, Chinese Academy of Sciences
School of Artificial Intelligence University of Chinese Academy of Sciences
Beijing, China
wenji.mao@ia.ac.cn

Shaoqiang Tang
Department of Mechanics and Engineering Science College of Engineering Peking University
Beijing, China
maotang@pku.edu.cn

***Abstract*—Due to the information explosion on the Internet, news recommendation, which helps users quickly find the news they are interested in, has become an essential issue for online news services. Previous research work usually adopts collaborative filtering or content-based methods which extract features and measure the similarities between users and each candidate news independently. However, candidate news often competes with each other for user attention, and modeling the interactions of multiple candidate news helps distinguish them better for news recommendation. In this paper, we propose a multi-level news recommendation method via modeling the interactions of multiple candidate news explicitly. Specifically, we design a Candidate Interaction Module (CIM) to generate interaction-enhanced candidate news representations. For each candidate news, the interaction-enhanced news representation contains information from other candidate news displayed to the user at the same time. Furthermore, in order to identify the connections between candidate news and user preferences at different semantic levels, we add a Multi-level Prediction Module (MPM) to exploit the category and subcategory information of news. Experimental results demonstrate that our proposed model achieves the state-of-the-art performance on two real-world benchmark datasets.**

Keywords—news recommendation, candidate news interaction, multi-level prediction, user modeling

I. INTRODUCTION

With the rapid development of the Internet, reading news online has become an important way for the public to acquire information worldwide. However, the unprecedented increase of information makes it difficult for users to find the content they are interested in. News recommendation, which aims to proactively provide users with news that is more consistent with their reading preferences, has become a key technology to alleviate information overload.

Different from recommendations in other scenarios, news recommendation has two unique characteristics which pose specific technical challenges to this task. Firstly, online news updates with high frequency [20], making news recommendations susceptible to the sparsity of user-news interactions. Secondly, a piece of news usually contains various information, such as the category and related entities, which requires a deep understanding of the embedded semantic information.

A variety of methods have been proposed for news recommendation, including collaborative filtering (CF) based methods [5] [16], content-based methods [1] [2] [12] [15] [17] [20], and the hybrid of the above two types of methods [10] [11]. The CF-based methods analyze the interactive relationships between users and news by matrix factorization to make recommendations. However, traditional CF-based methods suffer from the data sparsity problem in news recommendation. Considering the abundant textual information in news, the content-based methods, which exploit the news text and other attributes, are shown to be more effective and have become the mainstream of news recommendation researches. The content-based methods usually extract semantic features from news, capture user interests based on their historical behaviors, and predict the probability of the user clicking candidate news by matching user interests and candidate news [24].

Existing research work mainly focuses on extracting semantic information from each candidate news and measuring the relatedness between user interests and this piece of candidate news without considering other candidate news displayed to the user at the same time. However, when faced with multiple candidate news, users may only select a few pieces of news to read through comparison, meaning there is a competitive relationship among candidate news [21]. Therefore, extracting semantic information from each candidate news independently ignores the interactions of multiple candidate news caused by competition, and may be inferior for news recommendation.

To address the above issue, we propose a **M**ulti-**L**evel news recommendation model with **C**andidate **I**nteractions(**MLCI**) in this paper. Firstly, we design a Candidate Interaction Module (CIM) to model the interactions of candidate news explicitly, which could help distinguish the clicked candidate news from unclicked ones. CIM takes candidate news representations extracted from news text as input to generate interaction-enhanced news representations, which are then matched with user preferences to predict click probabilities. To further take advantage of the category and subcategory information, we also design a Multi-level Prediction Module (MPM) and make recommendations in three levels, namely the text level, category

* Corresponding Author

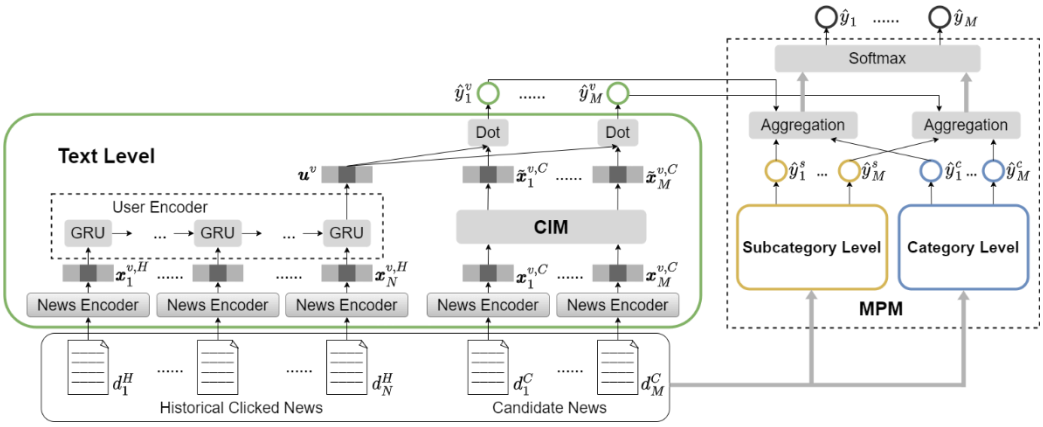


Fig. 1. The framework of MLCI for news recommendation.

level, and subcategory level. Lastly, we aggregate click probabilities of the above three levels as the final prediction.

Our work has made the following contributions.

- We propose a multi-level news recommendation method that models the interactions of candidate news and predicts click probabilities at three different semantic levels.
- We design a candidate interaction module to capture the competitive information between multiple candidate news and extract distinctive features for each candidate news.
- Experimental results on two real-world benchmark datasets show that our proposed model achieves the state-of-the-art performance.

II. Related Work

News recommendation is an essential issue for online news services and has been widely studied [24]. Traditional CF-based methods use the IDs of users and news and analyze the interactive relationships between them [5] [16]. As massive news articles are published on online platforms every day, the performance of CF-based methods is usually suboptimal due to the lack of historical interaction between users and the newly published news.

Since a piece of news contains abundant contents, including title, abstract, entity, and so on, making full use of these information can help learn better representations of news. Thus, the content-based methods [1] [2] [15] [17] are proposed to extract semantic features from news and build user interests from their historical clicked news, which are proven more effective for news recommendation. Early related work relies on feature engineering to represent news and users [2] [3] [6], which requires much effort and domain knowledge.

In recent years, deep learning techniques have been widely used in news recommendation [12] [17] [27]. Some content-based methods employ neural networks to learn news and user representations in an end-to-end manner. For example, An et al. [1] build long-term user representations from user ID and short-term user representations from their recently browsed news via the GRU network to model users more precisely. Wu et al. [22] exploit the embedding of user ID to attend to the important words in each news and important news in user click history. Jia et al. [7] propose the recurrent reasoning memory network to dynamically determine the user representation specific to the candidate news. Qi et al. [15] model user overall interest and user interest in different topics or subtopics to match candidate news with different levels of user interests. In addition, since news content is usually full of entities and commonsense, some methods introduce the knowledge graph to provide extra meaningful information and enhance news representations[14] [18] [20].

However, all of the above methods learn news representations for each candidate news independently and ignore the interactions of multiple candidate news displayed to the user simultaneously. Different from these methods, our MLCI method models the candidate interactions to extract distinctive features for candidate news and predicts the click probabilities, which could achieve performance improvement.

III. Problem Definition

The news recommendation problem is defined as follows. Given a user with N historical clicked news $\{d_1^H, d_2^H, \dots, d_N^H\}$ sorted by click time where "H" means "History". We aim to predict the click probabilities of this user on M candidate news $\{d_1^C, d_2^C, \dots, d_M^C\}$ where "C" means "Candidate". A piece of news d_i consists of a category w^c, a subcategory w^s, a title $[w_1^t, \dots, w_{L_t}^t]$ including L_t words, and L_e related entities $\{w_1^e, \dots, w_{L_e}^e\}$.

IV. Proposed Method

Fig. 1 shows the framework of our proposed model MLCI. We build the news representation from news text (including news title and entities) via the news encoder and aggregate the representations of historical clicked news as the user representation. Then we utilize the CIM module to generate interaction-enhanced candidate news representations and combine them together with the user representation to calculate click probabilities. Furthermore, we add the MPM module to obtain click probabilities at the category level and subcategory level in a way similar to the text level, but with the category and

subcategory contents as input. Finally, we make the click probability prediction for candidate news based on the click probabilities at the above three levels (i.e., text, category, subcategory).

A. News Encoder

At the text level, the news representation is learned from the title and entities. The news title $[w_1^t, \dots, w_{L_t}^t]$ is first converted into word embeddings $[\boldsymbol{e}_1^t, \dots, \boldsymbol{e}_{L_t}^t]$. Then, a convolutional neural network (CNN) [8] is used to generate the contextual vector $\boldsymbol{v}_k^t (k \in [1, L_t])$ for each word.

$$\boldsymbol{v}_k^t = ReLU\left(\boldsymbol{W}_F \otimes \boldsymbol{e}_{[k-r:k+r]}^t + \boldsymbol{b}_F\right) \tag{1}$$

where $\boldsymbol{W}_F$ and $\boldsymbol{b}_F$ are parameters of CNN, and the width of convolution kernels is $2r+1$. Lastly, we adopt the attention mechanism to calculate the importance of each word, i.e., the attention weight $\alpha_k^t (k \in [1, L_t])$.

$$\alpha_k^t = \frac{exp(tanh(\boldsymbol{W}^t \boldsymbol{v}_k^t + b^t))}{\sum_{i=1}^{L_t} exp(tanh(\boldsymbol{W}^t \boldsymbol{v}_i^t + b^t))} \tag{2}$$

where $\boldsymbol{W}^t$ and b^t are parameters of the attention mechanism. The title vector $\boldsymbol{v}^t$ is obtained by

$$\boldsymbol{v}^t = \sum_{k=1}^{L_t} \alpha_k^t \boldsymbol{v}_k^t \tag{3}$$

For entities in the title and abstract of news, we first convert them into entity embeddings $[\boldsymbol{e}_1^e, \dots, \boldsymbol{e}_{L_e}^e]$. Since the same entity may have different semantic relevance to the news in different categories, its contribution to the news representation is also different. For example, "Kobe Bryant" is more related to sports news than financial news. Therefore, we design an entity-level attention layer guided by category embedding $\boldsymbol{e}^c$ to generate the entity vector $\boldsymbol{v}^e$.

$$\boldsymbol{q}^e = ReLU(\boldsymbol{W}^q \boldsymbol{e}^c + \boldsymbol{b}^q) \tag{4}$$

$$\alpha_k^e = \frac{\exp\left(\boldsymbol{e}_k^{e^T} \tanh(\boldsymbol{W}^e \boldsymbol{q}^e + \boldsymbol{b}^e)\right)}{\sum_{i=1}^{L_e} \exp\left(\boldsymbol{e}_i^{e^T} \tanh(\boldsymbol{W}^e \boldsymbol{q}^e + \boldsymbol{b}^e)\right)} \tag{5}$$

$$\boldsymbol{v}^e = \sum_{k=1}^{L_e} \alpha_k^e \boldsymbol{e}_k^e \tag{6}$$

where $\boldsymbol{W}^q$, $\boldsymbol{b}^q$, $\boldsymbol{W}^e$, $\boldsymbol{b}^e$ are parameters of the entity-level attention layer. Finally, we concatenate the title vector $\boldsymbol{v}^t$ and the entity vector $\boldsymbol{v}^e$ as the news representation $\boldsymbol{x}^v$ at text level, i.e., $\boldsymbol{x}^v = [\boldsymbol{v}^t; \boldsymbol{v}^e]$, which contains semantic information from both title and entities.

B. User Encoder

To track the evolution of user preferences, we exploit the gated recurrent unit (GRU) [4] to capture the sequence information in historical user click behaviors. Specifically, the user encoder takes the sequence of clicked news representations $[\boldsymbol{x}_1^{v,H}, \dots, \boldsymbol{x}_i^{v,H}, \dots, \boldsymbol{x}_N^{v,H}]$ as input, and the last hidden state of the GRU network is taken as the user representation $\boldsymbol{u}^v$.

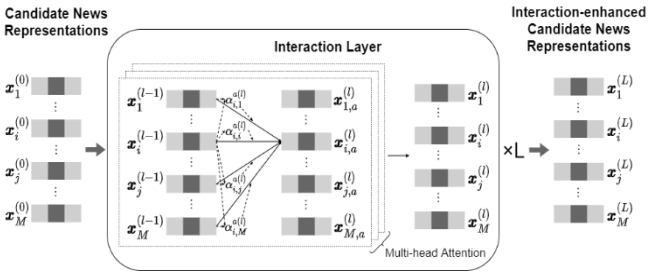


Fig. 2. The structure of candidate interaction module.

C. Candidate Interaction Module (CIM)

To model the competitive relationship among multiple candidate news, we propose a candidate interaction module to generate the interaction-enhanced candidate news representation. As shown in Fig. 2, the candidate interaction module consists of L interaction layers, which are implemented using the multi-head attention mechanism [19] [23]. Each layer has A attention heads. Given M candidate news, the i-th news vector $\boldsymbol{x}_i^{(l)} (i \in [1, M], l \in [1, L])$ calculated by the l-th interaction layer can be written as follows:

$$\alpha_{i,j}^{a(l)} = \frac{\exp\left(\boldsymbol{x}_i^{(l-1)^T} \boldsymbol{Q}_a^{(l)} \boldsymbol{x}_j^{(l-1)}\right)}{\sum_{m=1}^{M} \exp\left(\boldsymbol{x}_i^{(l-1)^T} \boldsymbol{Q}_a^{(l)} \boldsymbol{x}_m^{(l-1)}\right)} \tag{7}$$

$$\boldsymbol{x}_{i,a}^{(l)} = \boldsymbol{W}_a^{(l)} \left(\sum_{j=1}^{M} \alpha_{i,j}^{a(l)} \boldsymbol{x}_j^{(l-1)}\right) \tag{8}$$

$$\boldsymbol{x}_i^{(l)} = \left[\boldsymbol{x}_{i,1}^{(l)}, \dots, \boldsymbol{x}_{i,A}^{(l)}\right] \tag{9}$$

where $\alpha_{i,j}^{a(l)} (i, j \in [1, M], a \in [1, A], l \in [1, L])$ can be seen as the interaction effect between the i-th and j-th candidate news modeled by the a-th attention head in the l-th layer, $\boldsymbol{Q}_a^{(l)}$ and $\boldsymbol{W}_a^{(l)}$ are parameters of the a-th attention head in the l-th layer. The news vector $\boldsymbol{x}_i^{(l)}$ is the concatenation of $\boldsymbol{x}_{i,a}^{(l)} (a \in [1, A])$.

Specifically, the candidate interaction module takes candidate news representations from news encoder as input, i.e., $\boldsymbol{x}_i^{(0)} = \boldsymbol{x}_i^{v,C} (i \in [1, M])$. The output of the last interaction layer is the interaction-enhanced candidate news representation, i.e., $\tilde{\boldsymbol{x}}_i^{v,C} = \boldsymbol{x}_i^{(L)} (i \in [1, M])$, which is used to compute the click probability for the i-th candidate news with the user representation $\boldsymbol{u}^v$, i.e., $\hat{y}_i^v = \boldsymbol{u}^v \cdot \tilde{\boldsymbol{x}}_i^{v,C} (i \in [1, M])$.

D. Multi-level Prediction Module (MPM)

To identify the connections between user interests and candidate news at the category and subcategory level, we design a multi-level prediction module to compute click probabilities at these two levels separately. The model structure of the category level and subcategory level in MPM is similar to that of the text level without CIM, including the news encoder, the user encoder, and click probability prediction.

In MPM, news representations are learned respectively from the category embedding $\boldsymbol{e}^c$ and subcategory embedding $\boldsymbol{e}^s$ of news, which are initialized using word embedding and fine-

tuned during model training. These two embeddings are passed through a fully connected layer to obtain the category vector $\boldsymbol{x}^c$ and the subcategory vector $\boldsymbol{x}^s$ as the news representation in the corresponding semantic level. Then the calculation in the user encoder and click probability prediction are the same as those at the text level.

Given the above multi-level prediction results, we aggregate them to obtain the final click probability $\hat{y}_i (i \in [1, M])$.

$$\hat{y}_i = \frac{e^{\gamma_i^v}}{\sum_{*\in\{v,s,c\}} e^{\gamma_i^*}} \cdot \hat{y}_i^v + \frac{e^{\gamma_i^s}}{\sum_{*\in\{v,s,c\}} e^{\gamma_i^*}} \cdot \hat{y}_i^s + \frac{e^{\gamma_i^c}}{\sum_{*\in\{v,s,c\}} e^{\gamma_i^*}} \cdot \hat{y}_i^c \quad (10)$$

where the superscripts v, c, s represent the text level, category level and subcategory level, $\hat{y}_i^*$ is the predicted clicked probability for the i-th candidate news at the corresponding semantic level. The weight coefficient γ_i^* is computed as follow:

$$\gamma_i^* = \tanh\left(\boldsymbol{W}_{sum}^* \left[\boldsymbol{u}^*; \boldsymbol{x}_i^{*,C}\right] + b_{sum}^*\right), \ *\in \{v, c, s\} \quad (11)$$

where $\boldsymbol{u}^*$ and $\boldsymbol{x}_i^{*,C}$ are the user representation and candidate news representations, and $\boldsymbol{W}_{sum}^*$, b_{sum}^* are trainable parameters.

E. Model Training

Since most users only click a few news, the number of positive samples (i.e., candidate news clicked by the user) and negative samples (i.e., candidate news unclicked by the user) are very imbalanced. Following [22], we adopt the negative sampling technique to construct the training dataset. For each positive sample, K negative samples in the same impression log are randomly selected as a group. The loss function is

$$\mathcal{L} = -\sum_{p=1}^{P} \log \frac{\exp(\hat{y}_p)}{\exp(\hat{y}_p) + \sum_{k=1}^{K} \exp(\hat{y}_k)} \quad (12)$$

where P is the number of positive training samples.

V. Experiments

A. Experimental Settingss

1) Dataset: We conduct experiments on the public large-scale English news dataset MIND [26], which is widely used in recent representative work [7] [14] [15] [18] [25]. This dataset is collected from the user impression logs of Microsoft News in a week. A log records the historical clicked news of a user, the candidate news articles displayed to him when he visits the news website at a specific time, and his click behaviors on these news articles.

There are two versions of this dataset, namely MIND-large and MIND-small. The statistics of the two datasets are summarized in Table I. The MIND-large dataset contains 750,434 users with 2,609,219 impression logs, and the MIND-small dataset randomly samples 94,057 users with 230,111 impression logs. For both of these two datasets, we use the logs in the first six days for training and split the logs in the last day for validation and test with split ratio 2:5.

TABLE I. The Statistics of the Two Datasets Used in Our Experiments

Dataset	MIND-small	MIND-large
# Users	94,057	750,434
# Logs	230,111	2,609,219
# News	65,238	104,151
# Categories	18	18
# Subcategories	270	277
Train/Dev/Test	156,959/20,831/52,321	2,232,748/107,563 /268,908

2) Implementation Details: In our experiments, we use the pre-trained GloVe embedding [13] to initialize the word embedding matrix. The dimension of word embedding, entity embedding[1], and query vector in the entity-level attention layer are 300, 100, and 200. The number of filters in CNN is 100 for each of the window sizes {1, 3}, and the vectors generated under these two window sizes are concatenated as the 200-dimensional contextual vector of words. The dimension of category vector $\boldsymbol{x}^c$ and subcategory vector $\boldsymbol{x}^s$ are equal to that of text vector $\boldsymbol{x}^v$, i.e., 300. CIM has 2 candidate interaction layers with 10 heads, and the output dimension of each head is 30. The negative sampling ratio K is 4. The maximum number of user clicked news is 50, the maximum length of news title is 20, and the maximum number of entities associated with a piece of news is 3. We optimize the model by Adam [9] with a learning rate 10^{-4} and set the mini-batch as 400.

The metrics used to evaluate the performance of our model include area under the ROC curve (AUC), mean reciprocal rank (MRR), and normalized discounted cumulative gain (NDCG).

3) Baseline Methods: We compare our proposed model with two groups of recommendation methods[2].

The first group consists of two popular general-purpose recommendation methods, which achieve competitive results in various recommendation scenarios.

- Wide&Deep [3] jointly trains wide linear models and deep neural networks to combine the benefits of memorization and generalization for recommender systems.
- DeepFM [6] combines the architectures of factorization machines and deep neural network to model both low-order and high-order feature interactions.

The second group contains several recently proposed news specific recommendation methods.

[1] The entity embeddings are provided by the MIND dataset, which are learned from the subgraph of the WikiData knowledge graph by the TransE method.

[2] Since the datasets in the literature are different from those in this paper, we implement the baseline methods on the datasets in this paper and report the corresponding experimental results for comparison.

TABLE II. EXPERIMENTAL RESULTS OF DIFFERENT METHODS. THE BEST AND SECOND BEST RESULTS ARE HIGHLIGHTED IN BOLDFACE AND UNDERLINED

Method	MIND-small			MIND-large		
	AUC	MRR	NDCG@5	AUC	MRR	NDCG@5
Wide&Deep [3]	0.5423	0.2392	0.3414	0.5643	0.2979	0.3262
DeepFM [6]	0.5000	0.2279	0.2675	0.5434	0.2871	0.3169
DKN [20]	0.6728	0.2784	0.3067	0.7106	0.3226	0.3553
NPA [22]	0.7315	0.2683	0.2887	0.7327	0.3178	0.3486
LSTUR [1]	0.7485	0.2963	0.3247	0.7635	0.3247	0.3601
MVL [17]	0.7352	0.2805	0.3097	0.7619	0.3231	0.3581
HieRec [15]	0.7579	0.3073	0.3338	0.7668	0.3287	0.3659
RMBERT [7]	0.7484	0.3089	0.3419	0.7653	0.3280	0.3661
MLCI (Ours)	**0.7605**	**0.3137**	**0.3468**	**0.7757**	**0.3306**	**0.3672**

TABLE III. ABLATION RESULTS OF OUR MLCI ON TWO DATASETS. "W/O CIM" MEANS REMOVING CIM FROM MLCI. "W/O CIM, MPM" MEANS REMOVING CIM AND MPM FROM MLCI

Variant	MIND-small			MIND-large		
	AUC	MRR	NDCG@5	AUC	MRR	NDCG@5
MLCI	**0.7605**	**0.3137**	**0.3468**	**0.7757**	**0.3306**	**0.3672**
w/o CIM	0.7560	0.2949	0.3263	0.7710	0.3228	0.3573
w/o CIM, MPM	0.7408	0.2892	0.3183	0.7694	0.3207	0.3555

- DKN [20] is a knowledge graph based method that exploit entity embeddings in the news encoder.
- NPA [22] proposes a personalized attention network to represent news and users.
- LSTUR [1] builds long-term user representations and short-term user representations.
- MVL [17] constructs a graph to capture the high-order relatedness between news and users.
- HieRec [15] contains three levels of user interests to model user preferences in different aspects.
- RMBERT [7] designs a recurrent reasoning memory network to represent users and news.

B. *Experimental Results*

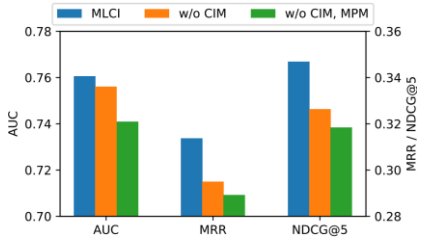
Table II reports the results of our proposed model and all the baseline methods. News specific recommendation methods consistently perform better than general-purpose methods, which shows that recently proposed content-based methods extract more effective features from news articles and user click histories using various deep neural network architectures. Moreover, HieRec and RMBERT perform better than other methods in the second group, because they make fine-grained matching between candidate news and some parts of user interests. Compared with all other methods, our proposed model MLCI achieves the best performance in terms of all metrics on both datasets. We attribute the superiority of MLCI to its two properties. Firstly, MLCI can learn more distinctive features for news recommendation by explicitly modeling the interactions between multiple candidate news. Secondly, MLCI utilizes information at three different semantic levels to identify relatedness between candidate news and user interests from multiple aspects.

C. *Ablation Study*

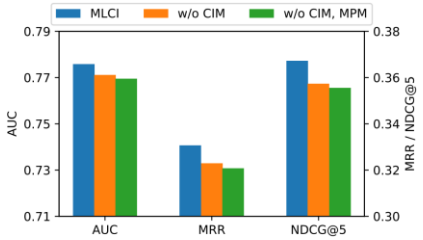
To evaluate the performance of each component used in the MLCI model, we also conduct a detailed ablation study on two model variants of the proposed method. The ablation results are shown in Table III and Fig. 3. Removing the CIM module from MLCI results in average performance degradation of 1.46% on the MIND-small and 0.75% on the MIND-large. This verifies the importance of the interaction information between candidate news, which helps obtain more distinctive news representations. Further removing the MPM component, all metrics continue to drop by an average of 0.97% on the MIND-small and 0.18% on the MIND-large, showing that the multi-level prediction is also important for the news recommendation task. Although most of the candidate news are newly published, their category and subcategory usually already exist in user click history. Therefore, by incorporating information at multiple semantic levels, our MLCI model can discover more direct connections between candidate news and user interests.

VI. CONCLUSIONS

In this paper, we propose a multi-level news recommendation model by modeling the interaction relationship between multiple candidate news. Different from previous work which represents each candidate news individually, our model captures the interactions of multiple candidate news to generate the interaction-enhanced candidate news representation. In addition, we also conduct click probability prediction at different semantic levels, namely the text level, category level, and subcategory level. Extensive experiments on two real-world datasets demonstrate the effectiveness of our proposed method.



(a) Results on the MIND-small dataset



(b) Results on the MIND-large dataset

Fig. 3. Ablation results of our MLCI on two datasets.

ACKNOWLEDGMENT

This work is supported in part by the Ministry of Science and Technology of China under Grant #2020AAA0108405, the National Natural Science Foundation of China under Grants #71621002, and Beijing Nova Program Z201100006820085 from Beijing Municipal Science and Technology Commission.

REFERENCES

[1] M. An, F. Wu, C. Wu, K. Zhang, Z. Liu, and X. Xie, "Neural news recommendation with long-and short-term user representations," in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 336–345.

[2] M. Capelle, F. Frasincar, M. Moerland, and F. Hogenboom, "Semantics-based news recommendation," in Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics, 2012, pp. 1-9.

[3] H. T. Cheng, et al., "Wide & deep learning for recommender systems," in Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, 2016, pp. 7-10.

[4] K. Cho, et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1724–1734.

[5] A. S. Das, M. Datar, A. Garg, and S. Rajaram, "Google news personalization: Scalable online collaborative filtering," in Proceedings of the 16th International Conference on World Wide Web, 2007, pp. 271-280.

[6] H. Guo, R. Tang, Y. Ye, Z. Li, and X. He, "DeepFM: A factorization-machine based neural network for CTR prediction." in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017, pp. 1725-1731.

[7] Q. Jia, J. Li, Q. Zhang, X. He and J. Zhu, "RMBERT: News recommendation via recurrent reasoning memory network over BERT," in Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 1773-1777.

[8] Y. Kim, "Convolutional neural networks for sentence classification," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1746-1751.

[9] D. P. Kingma, and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[10] A. Lommatzsch, "Real-time news recommendation using context-aware ensembles," in Advances in Information Retrieval, 2014, pp. 51-62.

[11] T. Miranda, et al., "Combining content-based and collaborative filters in an online newspaper," in Proceedings of ACM SIGIR Workshop on Recommender Systems, 1999.

[12] S. Okura, Y. Tagami, S. Ono, and A. Tajima, "Embedding-based news recommendation for millions of users," in Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 1933-1942.

[13] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1532-1543.

[14] T. Qi, F. Wu, C. Wu, and Y. Huang, "Personalized news recommendation with knowledge-aware interactive matching," in Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 61–70.

[15] T. Qi, et al., "HieRec: Hierarchical user interest modeling for personalized news recommendation," in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 5446–5456.

[16] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "Grouplens: An open architecture for collaborative filtering of netnews," in Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, 1994, pp. 175-186.

[17] T. Y. S. S. Santosh, A. Saha, and N. Ganguly, "MVL: Multi-view learning for news recommendation," in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 1873–1876.

[18] Y. Tian, et al., "Joint knowledge pruning and recurrent graph convolution for news recommendation," in Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 51-60.

[19] A. Vaswani, et al., "Attention is all you need," in Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 6000–6010.

[20] H. Wang, F. Zhang, X. Xie, and M. Guo, "DKN: Deep knowledge-aware network for news recommendation," in Proceedings of the 2018 World Wide Web Conference, 2018, pp. 1835-1844.

[21] X. Wang, B. Fang, and H. Zhang, "Predicting the popularity of news based on competitive matrix," in 2017 IEEE Second International Conference on Data Science in Cyberspace, 2017, pp. 151-155.

[22] C. Wu, F. Wu, M. An, J. Huang, Y. Huang, and X. Xie, "NPA: Neural news recommendation with personalized attention," in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 2576-2584.

[23] C. Wu, F. Wu, S. Ge, T. Qi, Y. Huang, and X. Xie, "Neural news recommendation with multi-head self-attention," in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019, pp. 6389-6394.

[24] C. Wu, F. Wu, Y. Huang, and X. Xie, "Personalized news recommendation: A survey," arXiv preprint arXiv:2106.08934, 2021.

[25] C. Wu, F. Wu, T. Qi, and Y. Huang, "Empowering news recommendation with pre-trained language models," in Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 1652–1656.

[26] F. Wu, et al., "MIND: A large-scale dataset for news recommendation," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 3597-3606.

[27] Q. Zhu, X. Zhou, Z. Song, J. Tan, and L. Guo, "DAN: Deep attention neural network for news recommendation," in Proceedings of the AAAI Conference on Artificial Intelligence, 2019, vol. 33, no. 01, pp. 5973-5980.