Do Not Read the Same News! Enhancing Diversity and Personalization of News Recommendation

Seonghwan Choi*
Seoul National University
Seoul, South Korea
csh3695@snu.ac.kr

Hyeondey Kim*
Bigpearl Inc.
Seoul, South Korea
hdkimaa@connect.ust.hk

Manjun Gim
Bigpearl Inc.
Seoul, South Korea
robin@bigpearl.io

ABSTRACT

Personalized news recommendation by machine is one of the widely studied areas. As the production of news articles increases and topics are diversified, it is impractical to read all the articles available to users. Therefore, the purpose of the news recommendation system should be to provide relevant news based on the user's interest. Unlike other recommendation systems, explicit feedback from users on each item such as ratings is rarely provided in news recommendation systems. Most news recommendation systems use implicit feedback such as click histories to profile user interest, which leads to biased recommendation results towards generally popular articles. In this paper, we suggest a novel news recommendation model for more personalized recommendations. If a user reads news not widely clicked by others, the news reflects the user's personal interest rather than other popular news clicked. We implement two user encoders, one to encode the general interest of the set of users and another one to encode the user's individual interest. We also propose regularization methods that induce two encoders to encode different types of user interest. The experiment on real-world data shows that our proposed method improves the diversity and the quality of recommendations for different click histories without any significant performance drops.

CCS CONCEPTS

• **Information systems** → **Personalization**; **Recommender systems**.

KEYWORDS

Recommender systems, Personalization, News Recommendation

ACM Reference Format:
Seonghwan Choi, Hyeondey Kim, and Manjun Gim. 2022. Do Not Read the Same News! Enhancing Diversity and Personalization of News Recommendation. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3487553.3524936

*Both authors contributed equally to this research.

WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France

ACM ISBN 978-1-4503-9130-6/22/04…$15.00
https://doi.org/10.1145/3487553.3524936

1 INTRODUCTION

Nowadays, online news platforms are ubiquitous [3]. They provide recommended news to billions of users. A myriad of news articles covering various topics and fields are generated every day. It is not realistic for users to read all the articles provided by the news platform they use. Therefore, it is important to build a personalized news recommendation system to users. Many methods are proposed to recommend personalized news[3, 5]. However, there are inherent problems in news recommendations.

Most news recommendation systems interact with users by recommending news based on their click histories [7]. Unlike other recommendation systems such as movies or e-commerce products, however, users' explicit ratings on each news item are rarely provided. Lack of detailed interaction data leads to significant problems: less diversity and less personalization in recommendation results. The click distribution is generally concentrated on politics, gossip, and attention-grabbing clickbait news that is intentionally designed to be clicked on by people. Since almost the only input available for them is the click history, the news recommendation systems tend to recommend clickbait news and trendy issues rather than less-popular ones in which each user has personal interests.

It is unrealistic that all the click behaviors made by a user in a news service reflect her *personal* interests to a similar level. For instance, users may click some articles because they deal with big issues such as political events or international sports games. Users may also click some articles which provide information about the concert schedules of their favorite musician, which is far more about personal interests that distinguish them from others than the former. Therefore, a user-centered news recommendation system may have to distinguish types of user choices and separately process different types of user choices.

Intuitively, if a user reads news that is not widely chosen by others, then the click behavior can be regarded as a highly discriminating feature to help identify the user's interest. The click behaviors on generally popular items, generally clickbait news or articles on trendy issues, are less likely to reflect her personal interests. Processing popular and rare click behaviors equally on the recommendation system can lead to the model shading out the user's unique preferences under public popularity.

To cope with this, we propose a novel model for more personalized news recommendations. We use two user encoders to process generally popular and user-centered interests. Then, we apply regularization methods during the training to encode these general and specialized interests separately.

We evaluate our model on a real-world dataset. The experiment result shows that our model not only improves the personalization and the diversity but also improves the performance on every metric only except the AUC score.

Our main contribution can be summarized as follows:

- We present a novel model and regularization methods for personalized news recommendation.
- Our proposed model shows effectiveness on real-world data in terms of both performance and diversity.

2 METHODOLOGY

Our proposed model leverages NRMS model as baseline structure, [6] which aggregates user interest from multiple user click histories with the multi-head attention mechanism. We present a novel model to learn generally popular and personal interests separately.

2.1 Model Structure

The existing NRMS structure consists of 3 modules: a news encoder, a user encoder, and a click predictor [6]. However, our model consists of 4 modules, two user encoders, a news encoder, and the click predictor. Our two user encoders encode general popularity and people's specific interests separately. We apply two regularization strategies in the training process so that each user encoder can encode the required information. One is *similarity regularization* which pushes one user encoder to encode similar information across each user. The other is *orthogonality regularization* which induces the other user encoder to encode unpopular user behaviors. Figure 1 illustrates the model structure of NRMS model [6] and our proposed model.

2.1.1 *Similarity Regularization.*

To ensure that the general user encoder encodes public popularity, the training process is regularized to increase similarities between user embeddings within the batch. During the training, we regulate our general user encoder's cosine similarity to be 1 in every batch. The regularization loss function is formulated as follows:

$$\mathcal{L}_S = -\frac{2}{B(B-1)} \sum_{i=1}^{B} \sum_{j=i+1}^{B} \frac{\mathbf{u}_i^g \cdot \mathbf{u}_j^g}{||\mathbf{u}_i^g|| \cdot ||\mathbf{u}_j^g||}, \tag{1}$$

where $\mathbf{u}_i^g$ and $\mathbf{u}_j^g$ are respectively the general user embedding of the i-th and j-th user in batch B.

2.1.2 *Orthogonality Regularization.*

Orthogonality regularization is applied to induce the specialized user embeddings to encode information different from the general information. During the training, we regulate our specialized user encoder's cosine similarity to be 0 in every batch. The regularization loss function is formulated as follows:

$$\mathcal{L}_O = \frac{1}{B} \sum_{i=1}^{B} |\frac{\mathbf{u}_i^g \cdot \mathbf{u}_i^s}{||\mathbf{u}_i^g|| \cdot ||\mathbf{u}_i^s||}|, \tag{2}$$

where $\mathbf{u}_i^g$ and $\mathbf{u}_i^s$ are respectively the general and specialized user embedding of the i-th user in batch B.

2.2 Model Training

The final loss function $\mathcal{L}$ is formulated as a linear combination of the recommendation and two regularization loss functions, and is formulated as follows:

$$\mathcal{L} = \mathcal{L}_R + \lambda_S \mathcal{L}_S + \lambda_O \mathcal{L}_O, \tag{3}$$

where $\mathcal{L}_R$ is the recommendation loss function and λ_S and λ_O are coefficients on the impact of each regularization. Let $\hat{y}$ be the predictions of the news recommendation system, and let y be the labels for clicking.

$$\mathcal{L}_R = -\sum y \log \hat{y}, \tag{4}$$

Finally the final user embedding of the i-th user $\mathbf{u}_i$ is defined as the sum of two separated user embeddings as follows:

$$\mathbf{u}_i = \mathbf{u}_i^g + \mathbf{u}_i^s, \tag{5}$$

with which the final logit is defined as the sum of the general and specialized logits.

3 DATASET AND EXPERIMENTAL SETTINGS

Our offline experiments are conducted on a real-world news recommendation dataset collected from our own Korean news platform service in one month(Jun. 23, 2021 to Jul. 24, 2021). The detailed statistics are shown in Table 1. The structure of our dataset is same as MIND [7], each labeled sample being formatted as [*userID*, *timestamp*, *ClickHist*, *ImpLog*] where *ClickHist* is an ID list of news articles previously clicked by this user and *ImpLog* is a list of tuple containing article ids and the boolean labels indicating user click behavior, i.e., [($newsID_1$, $label_1$), ..., ($newsID_n$, $label_n$)]. The logs in the last date are used for the test, and the rest are used for training. The logs on the last date of the training data were used for validation. Figure 3 shows the distribution of CTR (click-through rate) of some categories in the dataset. We notice that most of the user's clicks are concentrated on gossip and political issues. Figure 2 illustrates the distribution of the title sequence length.

We leverage the pre-trained language model KoELECTRA - small [2, 4] for the underlying language model. The dimension of the hidden representations is 128, and the additive attention query vector is 100. The number of self-attention heads is 16. We select a positive-negative sampling method with three positive and six negative samples per each labeled log. This ratio is set empirically and may vary depending on the dataset — 1-to-4 for MIND[6] case. To train the model, we use binary cross-entropy loss for the objective function and Adam for the model optimization, with a 1e-5 learning rate for the language model parameters and 1e-4 for the other parameters. λ_S and λ_O are both set to 0.1. These hyperparameters are tuned on the validation set.

Table 1: Detailed statistics of our dataset.

# Users	11,553	**Avg. # words per title**	7.681
# News	31,546	**# Positive Samples**	386,912
# Impressions	90,402	**# Negative Samples**	3,780,716

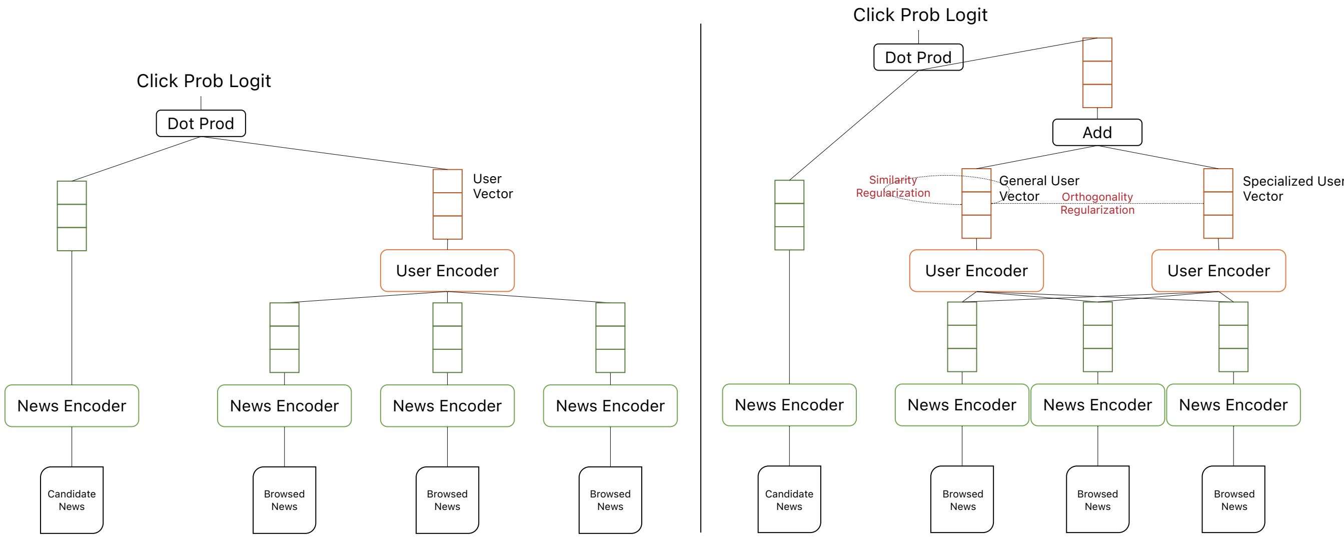

Figure 1: The comparison of the model structures of NRMS model [6] and our proposed model.

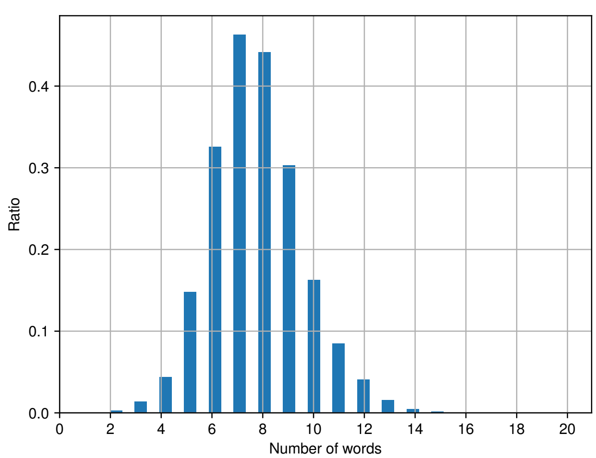

Figure 2: The distribution of title sequence length.

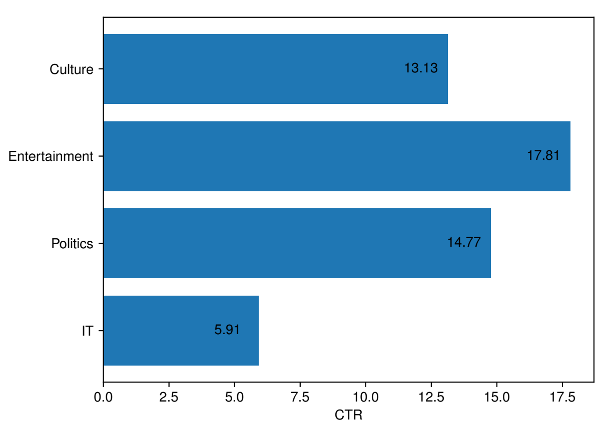

Figure 3: The distribution of CTR by categories.

Table 2: Experiment results.

Method	AUC	MRR	nDCG@5	JacSim@20
NRMS	**0.6518**	0.3978	0.5567	0.3094
NRMS-Reg	0.6488	**0.3983**	**0.5572**	**0.2213**

4 RESULTS AND ANALYSIS

We used NRMS with pre-trained Korean language model as the baseline [2, 6]. We compare both the baseline and the proposed model in Table 2 with our data under the same conditions. Jaccard similarity is one of the widely used metrics to calculate similarity in recommendation system [1]. To evaluate the diversity of our recommendation system, we leverage JacSim@20 as the numeric score on the diversity of recommendation results. JacSim@20, which provides a numeric score on the diversity of recommendation results for each session, is the average of all pair Jaccard similarities between the top 20 predicted articles for each of 3,576 sessions based on click histories for the 3,019 articles in the test set. Higher JacSim@20 indicates more similar recommendation results for different click histories.

Table 4 shows the recommendation performance with the use of the two user embeddings in NRMS-Reg. We improve the diversity of recommendations for different click histories with little performance drop. Results with the specialized user embedding contain much more diverse recommendations(lower JacSim@20 score) for different click histories than the vanilla NRMS. Summing the two user embeddings shows a higher AUC score than separate usage of each embedding, which shows the necessity of reflecting both personalization and public popularity in a recommendation system.

We randomly select two users with highly different user embeddings and predict their scores based on their click histories to analyze recommendation results. Table 3 demonstrates the recommendation result of the two users. We can infer that user 1 may

Table 3: Recommendation examples, showing click histories and recommendation results of sampled users. The numbers on each title represent different categories; [1]: politics, [2]: entertainment, [3]: IT/tech.

User 1 Clicked News		
How much for the presidency? The war of money has begun for presidential election[1]		
Lee says "Yoon is inexperienced, also I respect president Park"[1]		
Boosting up Choi; Warming up Kim; Dark day for Yoon[1]		
Ahn: the balanced national development should be the main topic of president election[1]		
Candidate News	**Score/Rank (Baseline)**	**Score/Rank (Ours)**
Rulling party: Yoon is a political prosecutor[1]	0.7620 / 10	0.8555 / 2
Hong: Yoon and Choi are terrible[1]	0.7546 / 13	0.8484 / 4
Lee presents a house to his gf 17 years younger than him[2]	0.8226 / 5	0.6006 / 125
Investigation reveals the falsehood of Chris[2]	0.7547 / 12	0.6821 / 55
User 2 Clicked News		
MS Windows Terminal's convenience and readability are improved[3]		
iPad 12.9" is over spec if you have a laptop[3]		
TMax wins 11B$ of the next-generation DBMS business[3]		
Offline certificate of Covid-19 immunization with an Android cellphone is available[3]		
Candidate News	**Score/Rank (Baseline)**	**Score/Rank (Ours)**
iPhone has been hacked; Virus without clicking a link[3]	0.8741 / 20	0.9440 / 1
Netflix embarking video game business[3]	0.8837 / 16	0.9148 / 5
KAIST collects 500K of Korean hairstyle pictures[3]	0.7893 / 101	0.9010 / 30
Crazy! hot bikini photo of Ms. Shin[2]	0.9147 / 5	0.7656 / 236

Table 4: Experiment results of NRMS-Reg.

User Embedding	General	Specialized	General + Specialized
AUC	0.5995	0.5569	0.6488
JacSim@20	0.8838	0.0666	0.2213

be highly interested in politics, especially the presidential election. Our model detects this user interest and recommends political issues over gossip in the entertainment category, which the baseline model does not. The baseline model recommends entertainment news with a high recommendation score — the rank of the recommendation score among all candidate news is 12th. However, our proposed model recommends political news with higher recommendation scores. User 2 seems to have an interest in technology news such as news about new technology, news from a software company, and about mobiles, which are not a relatively common taste. The baseline model recommends provocative — generally high click-through rate — articles over tech ones, while our model recommends tech-related articles with high scores. The baseline model recommends entertainment news even though the user's obvious interest in technology. The prediction scores of our model on the technology news are much higher than the entertainment one. These examples show the lowered JacSim@20 score leads to more personalized recommendation results, by lowering recommendation scores for generally popular items and increasing users' personal interests.

5 CONCLUSION

In this paper, we propose a user-centered news recommendation model that uses separate user encoders for general and user-specific interests with regularization. Our proposed method improves recommendation performance in every metric except AUC. Moreover, Jaccard dissimilarity is highly improved, which shows the effectiveness of our model on recommendation diversity. The recommendation example indicates that our model provides more personalized and user-interest-centered recommendations. For future work, we will conduct experiments on other datasets such as MIND [7] to ensure our model's performance on the other languages and data. We are also interested in detecting and removing other unintended biases from the news recommendation system.

ETHICAL CONSIDERATIONS

Our dataset consists of article information and article consumption records of multiple users on our own Korean news platform service. Only the title of the article is used as article information, which does not infringe on the copyright of the creator. The article consumption records were collected with the user's consent for product development and customized service provision, and are stored after anonymization.

Our proposed model does not explicitly reflect the user's identity-based inference bias because it does not take any explicit information about the user's identity as input. However, the recommendation results reflect not only individual but also overall user patterns, which may lead to unintended bias.

The recommendation system is designed to help people make efficient choices. It may be socially positive to help people efficiently consume information and expand their knowledge. Furthermore, this study aims to improve the diversity of individual recommendation results, which may be helpful in terms of social pluralism. However, the problem of preemptive item censorship, so-called *filter bubbles*, can cause confirmation bias and selective attention.

The impact of possible side effects may be great since the news articles tend to be perceived as reliable media compared to other media.

ACKNOWLEDGMENTS

This research was results of a study on the "HPC Support" Project, supported by the 'Ministry of Science and ICT' and NIPA. The authors would like to thank Kyungwoo Song and Won Ik Cho for helpful support.

REFERENCES

[1] Sujoy Bag, Sri Krishna Kumar, and Manoj Kumar Tiwari. 2019. An efficient recommendation generation using relevant Jaccard similarity. *Information Sciences* 483 (2019), 53–64.

[2] Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. In *ICLR*. https://openreview.net/pdf?id=r1xMH1BtvB

[3] Abhinandan S. Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. 2007. Google News Personalization: Scalable Online Collaborative Filtering. In *Proceedings of the 16th International Conference on World Wide Web* (Banff, Alberta, Canada) *(WWW '07)*. Association for Computing Machinery, New York, NY, USA, 271–280. https://doi.org/10.1145/1242572.1242610

[4] Jangwon Park. 2020. KoELECTRA: Pretrained ELECTRA Model for Korean. https://github.com/monologg/KoELECTRA.

[5] Chuhan Wu, Fangzhao Wu, Suyu Ge, Tao Qi, Yongfeng Huang, and Xing Xie. 2019. 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 (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 6389–6394. https://doi.org/10.18653/v1/D19-1671

[6] Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2021. Empowering News Recommendation with Pre-trained Language Models. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, Fernando Diaz, Chirag Shah, Torsten Suel, Pablo Castells, Rosie Jones, and Tetsuya Sakai (Eds.). ACM, 1652–1656. https://doi.org/10.1145/3404835.3463069

[7] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, and Ming Zhou. 2020. MIND: A Large-scale Dataset for News Recommendation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 3597–3606. https://doi.org/10.18653/v1/2020.acl-main.331