Is It Enough Just Looking at the Title?: Leveraging Body Text To Enrich Title Words Towards Accurate News Recommendation

Taeho Kim*
Hanyang University
Seoul, Korea
hirooms2@hanyang.ac.kr

Yungi Kim*
Hanyang University
Seoul, Korea
gozj3319@hanyang.ac.kr

Yeon-Chang Lee
Hanyang University
Seoul, Korea
lyc0324@hanyang.ac.kr

Won-Yong Shin
Yonsei University
Seoul, Korea
wy.shin@yonsei.ac.kr

Sang-Wook Kim†
Hanyang University
Seoul, Korea
wook@hanyang.ac.kr

ABSTRACT

In a news recommender system, a user tends to click on a news article if she is interested in its topic understood by looking at *its title*. Such a behavior is possible since, when viewing the title, humans naturally think of the contextual meaning of each title word by leveraging their own *background knowledge*. Motivated by this, we propose a novel personalized news recommendation framework **CAST** (Context-aware Attention network with a Selection module for Title word representation), which is capable of *enriching title words* by leveraging *body text* that fully provides the whole content of a given article as the context. Through extensive experiments, we demonstrate (1) the effectiveness of core modules in CAST, (2) the superiority of CAST over 9 state-of-the-art news recommendation methods, and (3) the interpretability with CAST.

CCS CONCEPTS

• **Information systems** → **Recommender systems**.

KEYWORDS

news recommendation; context-aware attention network

ACM Reference Format:
Taeho Kim, Yungi Kim, Yeon-Chang Lee, Won-Yong Shin, and Sang-Wook Kim. 2022. Is It Enough Just Looking at the Title?: Leveraging Body Text To Enrich Title Words Towards Accurate News Recommendation. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3511808.3557619

1 INTRODUCTION

News recommender systems have emerged as a promising solution to the problem of information overload on online news platforms, aiming to enhance the user experience in reading articles [4, 8, 24]. In online news platforms, users tend to read the *title* of news articles recommended to them and then make a click decision if they are interested in the topic that they understood *through the title*. Thus, existing studies have mostly focused on analyzing the *title text* of news articles to predict a user's preference on a news article [1, 17, 20, 21, 25]. Users readily grasp the *contextual meaning of each word in the title* by *leveraging their own background knowledge*; however, it is difficult for the news recommender system to comprehensively understand the contextual meaning of title words only with the title text, which is *a short sentence* composed of only a few words. In this sense, we claim that the *news encoder* generating a news representation should capture correctly the *context* of title words, although not explicitly shown in the title.

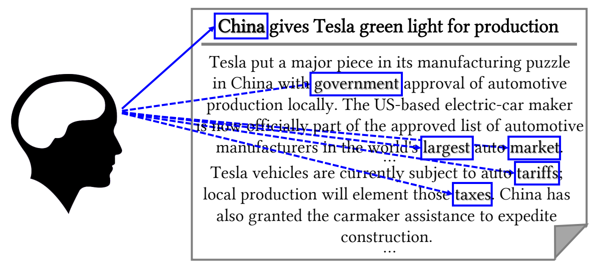

Figure 1: The contextual meaning of the title word 'China' includes 'the world's largest market' and 'country having strong restraints against foreign automakers,' which are described by body words such as 'largest,' 'market,' and 'tariffs'.

Although there have been attempts to leverage the entities in *knowledge graphs* (KGs) as the source of context [14, 17, 25], such methods pose a difficulty in *adaptively* capturing the context of title words since entities in KGs rather contain diverse perspectives [9, 18]. As an alternative to addressing this limitation, there have been a few studies on leveraging the body text which contains the plentiful description of the content addressed in the article. Studies in [10, 19] generated two types of news representations (*i.e.*, one for the body text and the other for the title text) and then fused them. *Multi-field matching* between a user's clicked news and other candidate news (*e.g.*, the title of a user's clicked news and the body of candidate news or vice versa) was presented in [23]. Unlike the aforementioned studies, our work is motivated by the fact that, when viewing the title only, humans naturally capture the contextual meaning of each title word owing to their *background knowledge* (see Figure 1). Thus,

*Two first authors have contributed equally to this work.
†Corresponding author.

CIKM '22, October 17–21, 2022, Atlanta, GA, USA.

ACM ISBN 978-1-4503-9236-5/22/10…$15.00
https://doi.org/10.1145/3511808.3557619

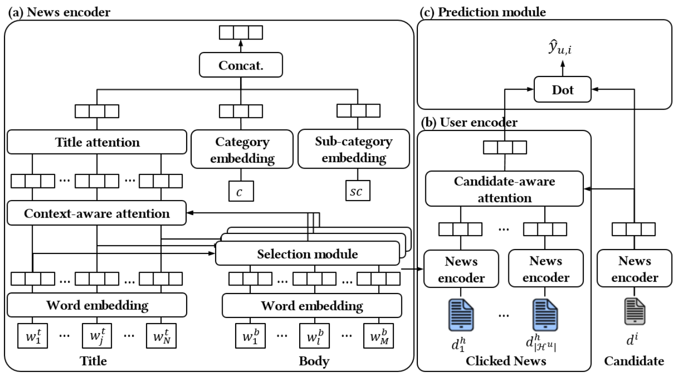

Figure 2: The schematic overview of our CAST framework.

to more precisely grasp the contextual meaning of title words, we aim at leveraging the *body text* as the context to enrich each title word in news recommendation.

In this paper, we propose a novel framework **CAST** (Context-aware Attention network with a Selection module for Title word representation) for personalized news recommendation. Towards this end, as the key component of CAST, we design a sophisticated news encoder making use of the *body text* as the source of context. Specifically, CAST first selects only some body words relevant to a target title word as the context, which corresponds to the selection module; then, CAST calculates the attention weights of the words in the context, which corresponds to the context-aware attention network. By learning different degrees of importance for the body words in terms of the context of each title word, CAST is capable of more effectively understanding the contextual meaning of each title word. Our empirical findings demonstrate that CAST (1) effectively captures the context related to title words by leveraging the body text, (2) enhances the quality of news representations, resulting in improving the recommendation accuracy over 9 state-of-the-art news recommender systems, and (3) offers an interpretation for the context related to a title word via visualization.

2 CAST FRAMEWORK

As mentioned in Section 1, we design a sophisticated news encoder making use of the body text as the context to enrich each title word. Now, we elaborate on our CAST framework shown in Figure 2.

2.1 News Encoder

As depicted in Figure 2-(a), our news encoder is composed of the five modules: word representation, body-word selection, context-aware attention, title representation, and news representation.

Word representation. We convert title text $\mathbf{t} = [w_1^t, w_2^t, \cdots, w_N^t]$ and body text $\mathbf{b} = [w_1^b, w_2^b, \cdots, w_M^b]$ in a news article d to the word representation matrices $\mathbf{E}^t = [\mathbf{e}_1^t, \mathbf{e}_2^t, \cdots, \mathbf{e}_N^t]$ and $\mathbf{E}^b = [\mathbf{e}_1^b, \mathbf{e}_2^b, \cdots, \mathbf{e}_M^b]$, respectively, where N and M denote the lengths of the title and body text. To this end, we use a learnable representation matrix $\mathbf{W}^w \in \mathbb{R}^{|\mathcal{V}| \times d_e}$, which is *initialized* by pre-trained GloVe [13]: each row is the embedding vector corresponding to the index of the word; $|\mathcal{V}|$ and d_e denote the vocabulary size and the dimensionality of the word representation, respectively.

Body-word selection. Some body words may be irrelevant to a given title word w_j^t. To filter out such irrelevant words, we introduce the selection module in our news encoder. Given the j-th word w_j^t in the title, this module first calculates the similarity between its title-word embedding $\mathbf{e}_j^t$ and every body-word embedding $\mathbf{e}_l^b$. Then, it selects the top-K body words having highest similarities to $\mathbf{e}_j^t$, where K is a hyperparameter to be determined empirically. Then, we build a matrix $\mathbf{E}^K \in \mathbb{R}^{K \times d_e}$, which consists of word representations corresponding to only the selected top-K body words.

Context-aware attention network. Next, we generate the representation matrix of context $\mathbf{X}(w_j^t) \in \mathbb{R}^{(K+1) \times d_e}$, defined as $\mathbf{X}(w_j^t) = [\,\mathbf{e}_j^t || \mathbf{E}^K]$ that helps understand the contextual meaning of w_j^t by concatenating $\mathbf{e}_j^t$ and $\mathbf{E}^K$. We note that each representation vector $\mathbf{x}_l \in \mathbf{X}(w_j^t)$ may have different degrees of importance as the context of w_j^t. This motivates us to design a context-aware attention network that generates a contextual word representation $\mathbf{c}_j^t$ of w_j^t by considering the attention weight of $\mathbf{x}_l \in \mathbf{X}(w_j^t)$ as follows:

$$\mathbf{c}_j^t = \mathbf{W}_v \left(\sum_{l=1}^{K+1} \alpha_l^w \mathbf{x}_l \right), \tag{1}$$

where $\alpha_l^w = \frac{\exp(z_l^w)}{\sum_{k=1}^{K+1} \exp(z_k^w)}$ indicates the attention weight of $\mathbf{x}_l$; $z_l^w = \frac{(\mathbf{W}_q \mathbf{e}_j^t)^{\mathrm{T}} (\mathbf{W}_k \mathbf{x}_l)}{\sqrt{d_t}}$; $\mathbf{W}_v \in \mathbb{R}^{d_t \times d_e}$, $\mathbf{W}_q \in \mathbb{R}^{d_t \times d_e}$, and $\mathbf{W}_k \in \mathbb{R}^{d_t \times d_e}$ are the learnable weights; and d_t denotes the dimensionality of the contextual word representation. We perform this process for all the words in the title text $\mathbf{t}$ to obtain the matrix of contextual word representations $\mathbf{C}^t = [\mathbf{c}_1^t, \mathbf{c}_2^t, \cdots, \mathbf{c}_N^t]$ for $\mathbf{t}$.

Title representation. Different words in the title text $\mathbf{t}$ may have different degrees of informativeness for representing the title [19, 20]. Thus, we generate the title representation $\mathbf{r}^t$ by selecting important contextual word representations as follows [1, 10, 19, 21]:

$$\mathbf{r}^t = \sum_{j=1}^{N} \beta_j^t \mathbf{c}_j^t, \tag{2}$$

where $\beta_j^t = \frac{\exp(z_j^t)}{\sum_{k=1}^{N} \exp(z_k^t)}$ indicates the attention weight of $\mathbf{c}_j^t$; $z_j^t = \mathbf{W}_p^{\mathrm{T}} \tanh(\mathbf{W}_t \mathbf{c}_j^t + b_t)$; $\mathbf{W}_p \in \mathbb{R}^f, \mathbf{W}_t \in \mathbb{R}^{f \times d_t}$, and $b_t \in \mathbb{R}^f$ are learnable weights; f denotes the dimensionality of the hidden layer.

News representation. Basically, the news encoder uses title representation $\mathbf{r}^t$ to represent the news article d. Since it is well known that the categorical features (*e.g.*, category c and sub-category sc) are informative clues to understand the topic of the article [1, 19], we finally obtain $\mathbf{r}^d \in \mathbb{R}^{d_n}$ as follows:

$$\mathbf{r}^d = [\,\mathbf{r}^t || \mathbf{r}^c || \mathbf{r}^{sc}], \tag{3}$$

where $d_n = d_t + d_c + d_{sc}$, $\mathbf{r}^c \in \mathbb{R}^{d_c}$ (resp. $\mathbf{r}^{sc} \in \mathbb{R}^{d_{sc}}$) indicates the category (resp. sub-category) representation, and d_c (resp. d_{sc}) denotes the dimensionality of the category (resp. sub-category).

2.2 User Encoder

We describe how to encode a target user $u \in \mathcal{U}$ with her click history $\mathcal{H}^u$ for the user representation (see Figure 2-(b)). We first obtain a news representation matrix $\mathbf{H} = [\mathbf{r}^{d_1^h}, \mathbf{r}^{d_2^h}, \cdots, \mathbf{r}^{d_{|\mathcal{H}^u|}^h}]$ of history $\mathcal{H}^u = \{d_1^h, d_2^h, \cdots, d_{|\mathcal{H}^u|}^h\}$ from the news encoder. Then,

Table 1: Dataset statistics

Datasets	# Users	# News	# Click behavior	# Impression
MIND-small	94,057	65,238	347,727	230,117
MIND-large	1,000,000	161,013	24,155,470	15,777,377

we generate the user representation $\mathbf{r}^u$ of u by considering the diversity of u's interest. To this end, we aggregate all news representations $\mathbf{r}^{d^h_m}$ in $\mathbf{H}$ based on their attention weights calculated by the candidate-aware attention network [10, 17, 25] as follows:

$$\mathbf{r}^u = \sum_{m=1}^{|\mathcal{H}^u|} \gamma^u_m \mathbf{r}^{d^h_m}, \tag{4}$$

where $\gamma^u_m = \frac{\exp(z^u_m)}{\sum_{k=1}^{|\mathcal{H}^u|}\exp(z^u_k)}$ indicates the attention weight of $\mathbf{r}^{d^h_m}$; $z^u_m = \mathbf{W}_o^{\mathrm{T}}\tanh(\mathbf{W}_{u1}\mathbf{r}^{d^h_m} + \mathbf{W}_{u2}\mathbf{r}^{d^i})$; $\mathbf{W}_o \in \mathbb{R}^f$, $\mathbf{W}_{u1} \in \mathbb{R}^{f\times d_n}$, and $\mathbf{W}_{u2} \in \mathbb{R}^{f\times d_n}$ are learnable weights.

2.3 Prediction and Model Training

Given a user u and impression [20, 22] $\mathcal{I}^u = \{d^i_1, d^i_2, \cdots, d^i_{|\mathcal{I}^u|}\}$, we compute the matching score of a user representation $\mathbf{r}^u$ and a news representation $\mathbf{r}^{d^i}$, *i.e.*, $\hat{y}_{u,i} = \mathbf{r}^u \cdot \mathbf{r}^{d^i}$ (see Figure 2-(c)).

Now, we explain how to train the entire model parameters based on the output of the prediction module. we randomly sample Q *negative samples* $[d^i_{u,n_1}, d^i_{u,n_2}, \cdots, d^i_{u,n_Q}]$ that user u did not click in impression $\mathcal{I}^u$ for a *positive sample* $d^i_{u,p}$ clicked in $\mathcal{I}^u$. Then, we make a training tuple $s = (d^i_{u,p}, d^i_{u,n_1}, d^i_{u,n_2}, \cdots, d^i_{u,n_Q})$. Finally, we minimize our loss $\mathcal{L}$ so that the positive sample has a higher matching score than that of negative samples in a training tuple s:

$$\mathcal{L} = -\sum_{\mathcal{U}}\sum_{\mathcal{I}^u}\sum_{s}\log\left(\frac{\exp(\hat{y}_{u,p})}{\exp(\hat{y}_{u,p}) + \sum_{q=1}^{Q}\exp(\hat{y}_{u,n_q})}\right), \tag{5}$$

where $\hat{y}_{u,p}$ and $\hat{y}_{u,n_q}$ denote the matching score for the positive sample and its associated q-th negative sample, respectively.

3 EVALUATION

3.1 Experimental Setup

Datasets. We used two versions of a real-world dataset MIND [22], which are publicly available[1] and widely used in previous studies: MIND-small and MIND-large (see Table 1).

Competitors. We compare CAST with 9 state-of-the-art news recommendation methods built upon DNNs, *i.e.*, DKN [17], LSTUR [1], NPA [20], NRMS [21], NAML [19], FIM [16], HieRec [14], AMM [23], and CNE-SUE [10]. For DKN, LSTUR, NPA, NRMS, and NAML, we used the implementations in the popular open-source library.[2] For FIM, HieRec, and CNE-SUE, we used the source code provided by the authors. For AMM, as its source code is *not available*, we quote the accuracy of AMM reported in [23].

Evaluation task. Following [3, 6, 7, 10, 15, 19, 21], we performed top-n recommendation with each method and then evaluated their accuracy by using area under the curve (AUC), mean reciprocal rank (MRR), and normalized discounted cumulative gain (nDCG). We report the average and the standard deviation of values obtained by the five independent evaluations for each measure [2, 5, 11, 12].

[1] https://msnews.github.io/
[2] https://github.com/microsoft/recommenders

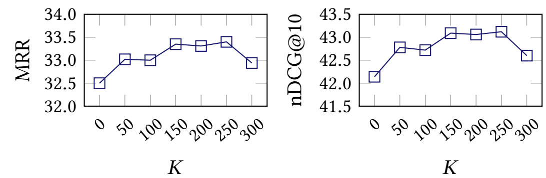

Figure 3: The effect of K on the accuracies of CAST.

Table 2: Accuracies of CAST variants with different sources used as context

Variants	AUC	MRR	nDCG@5	nDCG@10
CAST(T)	66.63±0.13	32.18±0.17	35.51±0.20	41.82±0.12
CAST(KG)	67.49±0.31	32.43±0.20	35.98±0.25	42.21±0.21
CAST(B)	**68.55**±0.20	**33.40**±0.11	**36.85**±0.18	**43.12**±0.14

Implementation details. We carefully tuned the hyperparameters of competitors and CAST. For CAST, we set its hyperparameters as follows: d_e=300, d_t=400, d_c=d_{sc}=50, f=200, K=250, batch size=32, dropout rate=0.2, learning rate=1e-4, and Q=4. If the body length is smaller than K, then K is set to the body length.

3.2 Results and Analysis

We designed our experiments, aiming at answering the following five key research questions (RQs):

- **(RQ1)** Does our selection module in CAST accurately select body words relevant to the title word?
- **(RQ2)** Which source of context is most influential in terms of the accuracy?
- **(RQ3)** Does our news encoder more effectively exploit the body text than other news encoders using the body text?
- **(RQ4)** Does CAST provide more-accurate recommendations than state-of-the-art news recommnedation methods?
- **(RQ5)** How does our CAST provide an interpretation for the context related to each title word?

For RQs except RQ4, we omit the results of CAST on MIND-large because they showed tendencies similar to those on MIND-small.

(RQ1) Effectiveness of the selection module. We judiciously analyze how the accuracy of CAST depends on the values of K, the number of relevant body words in the selection module.[3] Figure 3 illustrates the accuracy versus the value of K for each measure. Note that setting $K = 0$ is equivalent to the case of CAST that does *not* utilize the body words *at all*, whereas setting $K = 300$ is equivalent to the case of CAST using *all* the body words as the context. We see that the accuracy steadily increases up to 250 and then gradually decreases. Specifically, CAST yields gains up to 2.78% and 2.32% in terms of MRR and nDCG@10, respectively, when $K = 250$, compared to the case of $K = 0$. The results indicate that carefully selecting relevant body words as the context of a title word is indeed effective in offering more-accurate news recommendations.

(RQ2) Impact of each source used as context. We evaluate the performance of CAST with modifications by using the following different sources: (1) **T**: using all words only in the *title* text; (2) **KG**: using the entity corresponding to each word in the title text and

[3] We only used the first 30 (resp. 300) words in the title text (resp. body text).

Table 3: Accuracies from the variants of CAST with different news encoders

Variants	AUC	MRR	nDCG@5	nDCG@10
CAST(NAML)	67.44±0.25	32.18±0.16	35.58±0.23	41.82±0.12
CAST(CNE)	67.33±0.13	31.97±0.27	35.66±0.29	41.86±0.24
CAST	**68.55**±0.20	**33.40**±0.11	**36.85**±0.18	**43.12**±0.14

Table 4: Accuracies of CAST and 9 competitors

(a) MIND-small

Methods	AUC	MRR	nDCG@5	nDCG@5
DKN	64.48 ±0.29	30.09 ±0.23	32.82 ±0.35	39.55 ±0.24
LSTUR	66.52 ±0.27	31.32 ±0.32	34.64 ±0.41	40.93 ±0.29
NPA	65.56 ±0.13	31.12 ±0.23	34.27 ±0.23	40.57 ±0.15
NRMS	65.84 ±0.34	31.23 ±0.13	34.39 ±0.24	40.84 ±0.21
NAML	67.02 ±0.42	31.62 ±0.38	35.00 ±0.48	41.29 ±0.42
FIM	65.93 ±0.32	31.35 ±0.32	34.51 ±0.43	40.98 ±0.32
HieRec	67.55 ±0.22	32.55 ±0.14	36.00 ±0.19	42.24 ±0.14
AMM	<u>67.96</u>	<u>32.98</u>	<u>36.64</u>	<u>42.77</u>
CNE-SUE	67.76 ±0.13	32.10 ±0.25	35.78 ±0.20	42.06 ±0.23
CAST	**68.55** ±0.20	**33.40** ±0.11	**36.85** ±0.18	**43.12** ±0.14

(b) MIND-large

Methods	AUC	MRR	nDCG@5	nDCG@5
DKN	66.28 ±0.24	32.11 ±0.15	34.82 ±0.14	40.54 ±0.14
LSTUR	<u>68.57</u> ±0.14	<u>33.56</u> ±0.09	<u>36.55</u> ±0.08	<u>42.29</u> ±0.08
NPA	67.11 ±0.28	32.49 ±0.20	35.26 ±0.19	40.99 ±0.19
NRMS	67.69 ±0.16	32.82 ±0.08	35.63 ±0.10	41.38 ±0.10
NAML	68.35 ±0.13	33.45 ±0.15	36.41 ±0.17	42.16 ±0.16
FIM	67.39 ±0.20	33.06 ±0.14	35.98 ±0.16	41.67 ±0.16
HieRec	67.93 ±0.13	33.26 ±0.09	36.22 ±0.09	41.91 ±0.10
AMM	-	-	-	-
CNE-SUE	-	-	-	-
CAST	**69.68** ±0.20	**34.54** ±0.10	**37.72** ±0.08	**43.43** ±0.07

its neighboring entities in the *knowledge graph* (KG)[4] [14, 17, 25]; (3) **B**: using the body words relevant to each title word obtained from our selection module. We then made the following 3 variants of CAST depending on which sources are taken into account as the context: CAST(**T**), CAST(**KG**), and CAST(**B**). Table 2 shows that CAST(**B**) consistently outperforms all other variants. This indicates that, since the body text provides the whole content of the news article, it is quite useful in providing the contextual meaning of each title word, resulting in accurate news recommendations. We also examined the performance of CAST(**T+KG**), CAST(**T+B**), CAST(**KG+B**), and CAST(**T+KG+B**), confirming that their accuracy is comparable to that of CAST(**B**). This result implies that using sources other than body words is not indeed beneficial.

(RQ3) Comparative study among the methods exploiting the body text. From the fact that NAML [19] and CNE-SUE [10] exploit the body text toward different perspectives from ours, we compare CAST with its two variants, namely CAST(NAML) and CAST(CNE), that employ the news encoders designed by NAML and CNE-SUE, respectively. From Table 3, we see that CAST consistently outperforms CAST(NAML) and CAST(CNE) with respect to all measures. The results reveal that our design choice is most effective for accurate news recommendation, rather than other designs from NAML and CNE-SUE, in the sense of exploiting the body text.

(RQ4) Comparison with 9 competitors. We conduct comparative experiments on the two datasets to demonstrate the superiority of

[4]These entities were extracted from WikiData and their embeddings were trained via the TransE method [9].

China gives Tesla green light for production

Tesla put a major piece in its manufacturing puzzle in China with government approval of automotive production locally. ⋯ automotive manufacturers in the world's largest auto market. ⋯
Tesla vehicles are currently subject to auto tariffs; local production will element those taxes.

Figure 4: Visualization of relevant body words to 'China'.

CAST over the 9 competing methods. Note that, for MIND-large, the accuracies of CNE-SUE could not be obtained due to its out-of-memory issue. Also, we could not quote the accuracies of AMM on MIND-large because they are not reported in [23]. There have been no papers reporting the performance of CNE-SUE and AMM on the *original* MIND-large dataset. The best and second best performers are highlighted by bold and underline, respectively.

We summarize our empirical findings as follows. First, we see that the methods leveraging the body text (*i.e.*, NAML, AMM, CNE-SUE, and CAST) generally outperforms the ones that do not utilize the body text (*i.e.*, DKN, LSTUR, NPA, NRMS, and FIM). Second and most importantly, CAST consistently outperforms *all* competitors on *all* datasets for *all* metrics. Specifically, on MIND-large, CAST significantly outperforms the *best* competitor (*i.e.*, LSTUR) exhibiting gains of 1.61%/2.92%/3.20%/2.67% in terms of AUC/MRR/nDCG@5/nDCG@10. On MIND-small, CAST outperforms the reported accuracies of the *best* competitor (*i.e.*, AMM) by 0.86%/1.27%/0.56%/0.81% in terms of AUC/MRR/nDCG@5/nDCG@10.

(RQ5) Case study on interpretations. We conduct a case study to show the interpretation of CAST using the context-aware attention network, which calculates the attention weight of each body word for a target title word. Figure 4 shows the visualization result when a title text 'China gives Tesla light for production' of a news article is given, where the colored words indicate the words selected by our selection module (the darker the color, the larger the attention weight) for a title word 'China' and the words with no color mark indicate the irrelevant words, thus not selected by the module. Among all the body words, CAST assigns the highest attention values to the words such as 'government' and 'largest', which can help understand the contextual meaning of the title word 'China'.

4 CONCLUSIONS

In this paper, we aimed to design a high-quality news encoder that effectively understands the contextual meaning of title words. Toward this goal, we proposed CAST, a novel news recommendation framework, which leverages the *body text* as the context to enrich title words. Through extensive experiments, we demonstrated the effectiveness of core modules in CAST as well as the superiority of CAST over 9 state-of-the-art methods.

5 ACKNOWLEDGEMENT

This work was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2022-0-00352) and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.2018R1A5A7059549 and No. 2021R1A2C3004345). Also, we thank the Naver Corporation for their financial support and provisions of the computing facilities, which helped us greatly in performing this research successfully.

REFERENCES

[1] Mingxiao An, Fangzhao Wu, Chuhan Wu, Kun Zhang, Zheng Liu, and Xing Xie. 2019. Neural News Recommendation with Long- and Short-term User Representations. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. 336–345.

[2] Dong-Kyu Chae, Jihoo Kim, Duen Horng Chau, and Sang-Wook Kim. 2020. Ar-cf: Augmenting virtual users and items in collaborative filtering for addressing cold-start problems. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1251–1260.

[3] Won-Seok Hwang, Juan Parc, Sang-Wook Kim, Jongwuk Lee, and Dongwon Lee. 2016. "Told you i didn't like it": Exploiting uninteresting items for effective collaborative filtering. In *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*. IEEE, 349–360.

[4] Ilija Ilievski and Sujoy Roy. 2013. Personalized News Recommendation based on Implicit Feedback. In *Proceedings of the 2013 International News Recommender Systems Workshop and Challenge*. 10–15.

[5] Yeon-Chang Lee, Sang-Wook Kim, and Dongwon Lee. 2018. gOCCF: Graph-theoretic one-class collaborative filtering based on uninteresting items. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

[6] Yeon-Chang Lee, Taeho Kim, Jaeho Choi, Xiangnan He, and Sang-Wook Kim. 2021. M-BPR: A novel approach to improving BPR for recommendation with multi-type pair-wise preferences. *Information Sciences* 547 (2021), 255–270.

[7] Yeon-Chang Lee, Jiwon Son, Taeho Kim, Daeyoung Park, and Sang-Wook Kim. 2021. Exploiting uninteresting items for effective graph-based one-class collaborative filtering. *The Journal of Supercomputing* 77, 7 (2021), 6832–6851.

[8] Hongjun Lim, Yeon-Chang Lee, Jin-Seo Lee, Sanggyu Han, Seunghyeon Kim, Yeongjong Jeong, Changbong Kim, Jaehun Kim, Sunghoon Han, Solbi Choi, et al. 2022. AiRS: A Large-Scale Recommender System at NAVER News. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 3386–3398.

[9] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

[10] Zhiming Mao, Xingshan Zeng, and Kam-Fai Wong. 2021. Neural News Recommendation with Collaborative News Encoding and Structural User Encoding. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 46–55.

[11] Junha Park, Yeon-Chang Lee, and Sang-Wook Kim. 2022. Effective and efficient negative sampling in metric learning based recommendation. *Information Sciences* 605 (2022), 351–365.

[12] Sung-Jun Park, Dong-Kyu Chae, Hong-Kyun Bae, Sumin Park, and Sang-Wook Kim. 2022. Reinforcement Learning over Sentiment-Augmented Knowledge Graphs towards Accurate and Explainable Recommendation. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 784–793.

[13] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1532–1543.

[14] Tao Qi, Fangzhao Wu, Chuhan Wu, Peiru Yang, Yang Yu, Xing Xie, and Yongfeng Huang. 2021. HieRec: Hierarchical User Interest Modeling for Personalized News Recommendation. *arXiv preprint arXiv:2106.04408* (2021).

[15] Cong Tran, Jang-Young Kim, Won-Yong Shin, and Sang-Wook Kim. 2019. Clustering-based collaborative filtering using an incentivized/penalized user model. *IEEE Access* 7 (2019), 62115–62125.

[16] Heyuan Wang, Fangzhao Wu, Zheng Liu, and Xing Xie. 2020. Fine-grained Interest Matching for Neural News Recommendation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. 836–845.

[17] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. DKN: Deep Knowledge-Aware Network for News Recommendation. In *Proceedings of The Web Conference (WWW)*. 1835–1844.

[18] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Transactions on Knowledge and Data Engineering* 29, 12 (2017), 2724–2743.

[19] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. Neural News Recommendation with Attentive Multi-View Learning. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. 3863–3869.

[20] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. NPA: Neural News Recommendation with Personalized Attention. In *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (ACM KDD)*. 2576–2584.

[21] 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 Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 6390–6395.

[22] 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 Annual Meeting of the Association for Computational Linguistics (ACL)*. 3597–3606.

[23] Qi Zhang, Qinglin Jia, Chuyuan Wang, Jingjie Li, Zhaowei Wang, and Xiuqiang He. 2021. AMM: Attentive Multi-field Matching for News Recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1588–1592.

[24] Guanjie Zheng, Fuzheng Zhang, Zihan Zheng, Yang Xiang, Nicholas Jing Yuan, Xing Xie, and Zhenhui Li. 2018. DRN: A Deep Reinforcement Learning Framework for News Recommendation. In *Proceedings of the 2018 World Wide Web Conference*. 167–176.

[25] Qiannan Zhu, Xiaofei Zhou, Zeliang Song, Jianlong Tan, and Li Guo. 2019. DAN: Deep Attention Neural Network for News Recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. 5973–5980.