Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.052600

ARTICLE

Improving Diversity with Multi-Loss Adversarial Training in Personalized News Recommendation

Ruijin Xue[1,2], Shuang Feng[1,2,*] and Qi Wang[1,2]

[1]School of Computer and Cyber Sciences, Communication University of China, Beijing, 100024, China

[2]State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing, 100024, China

*Corresponding Author: Shuang Feng. Email: fengshuang@cuc.edu.cn

Received: 08 April 2024 Accepted: 18 July 2024 Published: 15 August 2024

ABSTRACT

Users' interests are often diverse and multi-grained, with their underlying intents even more so. Effectively capturing users' interests and uncovering the relationships between diverse interests are key to news recommendation. Meanwhile, diversity is an important metric for evaluating news recommendation algorithms, as users tend to reject excessive homogeneous information in their recommendation lists. However, recommendation models themselves lack diversity awareness, making it challenging to achieve a good balance between the accuracy and diversity of news recommendations. In this paper, we propose a news recommendation algorithm that achieves good performance in both accuracy and diversity. Unlike most existing works that solely optimize accuracy or employ more features to meet diversity, the proposed algorithm leverages the diversity-aware capability of the model. First, we introduce an augmented user model to fully capture user intent and the behavioral guidance they might undergo as a result. Specifically, we focus on the relationship between the original clicked news and the augmented clicked news. Moreover, we propose an effective adversarial training method for diversity (AT4D), which is a pluggable component that can enhance both the accuracy and diversity of news recommendation results. Extensive experiments on real-world datasets confirm the efficacy of the proposed algorithm in improving both the accuracy and diversity of news recommendations.

KEYWORDS

News recommendation; diversity; accuracy; data augmentation

1 Introduction

In recent years, significant advancements have been made in content recommendation technologies used by online news platforms such as Microsoft News and applications like "Toutiao". These developments have substantially enhanced the browsing experience for readers and attracted a large user base to digital news content [1,2]. Despite these advancements, the overwhelming volume of daily news articles presents a significant challenge for users trying to find content that matches their interests [3]. Excessive information can dilute user attention. The filtering and distribution mechanisms of news recommendation (NR) systems can alleviate information overload, helping users navigate through the vast array of articles to discover those most relevant to their interests [4].

Most existing news recommendation methods prioritize accuracy as their primary optimization goal [5]. They rank candidate news by computing the similarity between user interests and candidate news to generate the final recommendation list [6–9]. For example, Okura et al. [6] use gate recurrent unit (GRU) networks to learn user representations from historical news browsed by users and denoising autoencoders to learn news representations. Wu et al. [7] apply multi-head self-attention networks to learn user and news representations, ranking candidate news based on their correlation. An et al. [8] utilize user ID embeddings for long-term user representations and GRU networks for short-term user representations. Zhang et al. [9] employ the pre-trained model to improve news textual representations. However, these methods often neglect recommendation diversity in favor of optimizing recommendation accuracy, resulting in a failure to comprehensively cover different user interests. Additionally, the lack of diversity awareness in news recommendation models can lead to low levels of recommendation diversity. Therefore, enhancing recommendation diversity is crucial to improving user experience and engagement [10]. Users are dissatisfied with the homogeneous news presented in the recommendation list. Accuracy-based news recommendation algorithms are prone to inducing individuals into filter bubbles [11] and echo chambers [12].

In this paper, we propose a news recommendation algorithm that can effectively improve both news recommendation diversity and accuracy. First, we introduce an augmented user model that focuses on user intent. It infers potential behavioral guidance rather than merely considering the relevance between user-clicked news and candidate news. Specifically, we apply data augmentation techniques on user-clicked news. Then, we incorporate both original and augmented instances within different attention heads to capture user intent, infer potential behavioral guidance, and learn diverse representations for users with similar interests. Furthermore, we introduce an effective adversarial training method for diversity (AT4D) to incentivize the model to avoid selecting homogeneous news information. AT4D is a pluggable component that can enhance both accuracy and diversity in news recommendations. Extensive experiments on real-world news recommendation datasets show that the proposed algorithm outperforms baseline methods in accuracy and diversity, as evidenced by metrics including AUC, MRR, nDCG@5, nDCG@10, and ILAD@10 values.

The contributions of this paper are as follows:

- We propose an augmented user model to capture user intent and potential behavioral guidance comprehensively.
- We introduce an effective adversarial training method for diversity (AT4D), a pluggable component that enhances both accuracy and diversity in news recommendations.
- Extensive experiments conducted on real-world news recommendation datasets demonstrate that the proposed algorithm achieves better performance in terms of both accuracy and diversity of recommendation compared to seven baseline methods.

2 Related Work

Numerous scholars have researched news recommendations in detail. A core task of news recommendation is to compute the relevance between candidate news and user interests to generate a recommendation list. For instance, Okura et al. [6] use GRU networks to learn user representations from historical news browsed by users and denoising autoencoders to learn news representations. Wu et al. [7] apply multi-head self-attention networks to learn user and news representations, ranking candidate news based on their representations. An et al. [8] learn long-term user representations using user ID embeddings and short-term user representations through GRU networks. Wang et al. [13] propose a fine-grained interest-matching method that models the correlation between candidate news and clicked news to compute ranking scores. Wang et al. [14] introduce knowledge graphs into news recommendation, integrating semantic and knowledge-level representations of news. The effectiveness of using knowledge is demonstrated through a content-based deep recommendation framework for click-through rate prediction. These news recommendation methods primarily focus on matching users' personalized interests by optimizing recommendation accuracy, often neglecting recommendation diversity. However, a lack of recommendation diversity can negatively impact users' long-term experience and engagement. Some news recommendation methods explicitly consider recommendation diversity. For example, Qi et al. [11] employ a hierarchical user interest matching framework that matches candidate news with user interests at various levels to enhance the modeling capability of personalized news recommendation. Gharahighehi et al. [15] use a neighborhood-based session recommendation system in anonymous sessions to improve the diversity issue in news recommendation, which demonstrates effective diversity improvements across four datasets. Qi et al. [16] merge news popularity information to address the cold start and diversity problems in tailored news recommendation. These works aim to improve the diversity of recommendation results by more comprehensively mining users' historical interaction information and addressing diverse user interests. In contrast, our proposed algorithm incorporates diversity awareness, thereby enhancing both the accuracy and diversity of news recommendations.

3 Methodology

In this section, we introduce our proposed news recommendation algorithm, which aims to enhance both accuracy and diversity. We begin by elucidating the problem formulation addressed in this study. Then, we present the model architecture for the augmented user model and news model separately. Finally, we describe our pluggable component, AT4D. The structures of these components are illustrated in Fig. 1.

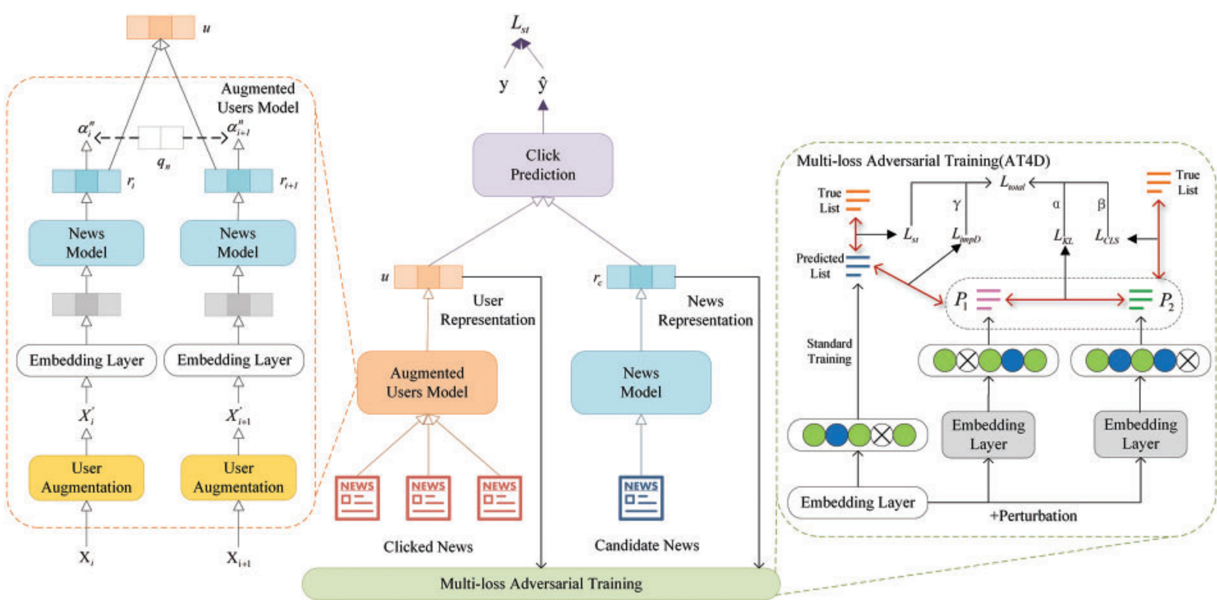

Figure 1: Architecture of the proposed algorithm

3.1 Problem Formulation

Given a user u with a history of clicked news, denoted as $L_u = \left\{l_1, l_2, \ldots, l_{|L_u|}\right\}$, and a list of candidate news to be ranked, denoted as $L_u^c = \left\{l_1^c, l_2^c, \ldots, l_{|L_u^c|}\right\}$, where $u \in U$, $l \in L$, U represents the set of users, and L represents the set of news articles. The model computes the inner product between the user representation vector u and the i^{th} candidate news representation vector l_i^c, denoted as $\hat{y} = u^T l_i^c$, obtaining the click score for user u, denoted as $Y_u = \left\{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_{|Y_u|}\right\}$. The final recommendation list is generated by ranking the candidate news based on their scores.

3.2 Augmented User Model

The augmented user model of the proposed algorithm, as illustrated in Fig. 1, aims to extract user intent from both the original clicked news and the augmented clicked news, inferring potential behavioral guidance to enhance recommendation performance. Data augmentation is initially widely applied in the field of computer vision [17], typically involving random operations such as cropping and rotation to enhance data efficiency. Following this intuition, we apply data augmentation strategies to news text in the proposed algorithm. Firstly, we denote the original text as $S_u = \left\{s_1, s_2, \ldots, s_{|S_u|}\right\}$. Then we utilize random swapping (RS) and random deletion (RD) following the method proposed by Wei et al. [18] to obtain augmented text denoted as $S_u' = \left\{s_1', s_2', \ldots, s_{|S_u'|}'\right\}$. The calculation process is as in Eq. (1).

$$S_u' = \omega_1 \mathrm{RS}\,(S_u) + \omega_2 \mathrm{RD}\,(\omega_1 \mathrm{RS}\,(S_u)) \tag{1}$$

where ω_1 and ω_2 are augmentation weights. Subsequently, we apply multi-head self-attention networks at the news level. We inject the original clicked news and the augmented clicked news into different attention heads simultaneously. The user model perceives additional information that cannot be captured solely by using the original news text. Experimentally, we find that this information can help to capture the intent behind user click behavior and potential behavioral guidance. The representation of the i^{th} original or augmented news learned by the k^{th} attention head is as follows:

$$\omega_{i,j}^k = \frac{\exp\,(s_i'^T Q_k^n s_j')}{\sum_{m=1}^{M} \exp\,(s_i'^T Q_k^n s_m')} \tag{2}$$

$$h_{i,k}^n = V_k^n \left(\sum_{j=1}^{M} \omega_{i,j}^k s_j' \right) \tag{3}$$

where Q_k^n and V_k^n represent the parameters of the k^{th} self-attention head for news. $\omega_{i,j}^k$ is the relative significance of interaction between the j^{th} original news and the k^{th} augmented news. The overall representation of the i^{th} news is a concatenation of vectors outputted by h individual self-attention heads, denoted as $h_i^n = [h_{i,1}^n; h_{i,2}^n; \ldots; h_{i,h}^n]$. When describing user profiles based on historical news, not every news article holds the same value. Hence, we apply the additive attention mechanism to select news articles that better represent user interests and learn more information. The attention weight of the i^{th} news is given by:

$$a_i^n = q_n^T \tanh\,(V_n \times h_i^n + v_n) \tag{4}$$

$$\alpha_i^n = \frac{\exp\,(a_i^n)}{\sum_{j=1}^{N} \exp\,(a_j^n)} \tag{5}$$

where q_n, V_n and v_n are all parameters of the attention network, and N represents how many news items the user has clicked. The final user representation is the weighted sum of representations of news clicked by the user, which is formulated as in Eq. (6).

$$u = \sum_{i=1}^{N} \alpha_i^n h_i^n \tag{6}$$

3.3 News Model

The news model is used to learn representations of news across varying semantic scales from textual data. In the proposed algorithm, the news model is employed to learn representations of historical news including those users have clicked and news that users have not clicked, and candidate news from the recommended list. We utilize the news information in the dataset, including the title, abstract, category, and subcategory. Typically, the title provides a succinct summary of the news content, and the abstract delivers a concise overview of the event [19]. Categories and subcategories help readers quickly locate news and inform them about the news's attributes. Given that news information varies in semantic scale and characteristic level, we propose a multi-level attention learning framework. It treats news information at distinct levels of granularity as separate features. Instead of simply concatenating these features into a single continuous text for representation, our approach aims to achieve a unified representation of news. The news model consists of three major components: text encoder, element encoder, and attention-based pooling layer.

The text encoder is employed to learn representations of news text. The text encoder consists of three layers. The first layer is the word embedding layer, which uses GloVe [20] to convert the text into low-dimensional word vectors. Note the word sequence of the text as $\left[w_1^t, w_2^t, \ldots, w_M^t\right]$, and M is the length of the text. Through a mapping in the $A \times B$ dimensional word vector space, where A is the number of words and B is the dimensionality of the word vectors, it is transformed into a sequence of word vectors $\left[e_1^t, e_2^t, \ldots, e_M^t\right]$. The second layer is the convolutional neural network (CNN) layer. In the process of learning news representations, focusing on the local context of words is essential. The same entity may hold different perspectives on different events. Hence, we employ CNN to attend to the local context so that we can capture the relationship between the entity in events and their perspectives. The representation of the context of the i^{th} word in the text is denoted as c_i^t, computed as in Eq. (7).

$$c_i^t = \mathrm{ReLU}(K_t \times e_{[-I:+I]}^t + b_t) \tag{7}$$

where ReLU is a non-linear activation function [21]. K_t represents the kernel parameters of the CNN filters. $e_{[-I:+I]}^t$ indicates that the position of word embedding vectors is selected within the range of $-I$ to $+I$. b_t denotes the bias parameters of the CNN filters. The output of this layer is a sequence of context word representations, denoted as $\left[c_1^t, c_2^t, \ldots, c_M^t\right]$. The third layer comprises an additive word attention network. Compared to convolutional neural networks, the attention mechanism exhibits strong parallelism. It focuses on more important information from a global perspective. Therefore, identifying crucial words in news text facilitates the learning of richer semantic representations. The attention weight of the i^{th} word in the news text is denoted as α_i^w, computed as follows:

$$a_i^w = q_t^T \tanh(V_t \times c_i^t + v_t) \tag{8}$$

$$\alpha_i^w = \frac{\exp(a_i^w)}{\sum_{j=1}^{M} \exp(a_j^w)} \tag{9}$$

where q_t is the query vector. V_t and v_t are projection parameters. The final representation of the news is computed by weighting and summing the context word representations based on their attention weights. The calculation formula is as in Eq. (10).

$$r = \sum_{i=1}^{M} \alpha_i^w c_i^t \tag{10}$$

The second component of the news model is the element encoder, which is utilized to learn representations of news from the categories and subcategories. Many users, when using online news services, first select a category such as "politics" before engaging in extended reading sessions. Clearly, the categories and subcategories are crucial for modeling news. They enable us to obtain the information conveyed by the news more accurately. Therefore, we simultaneously consider category and subcategory information in learning news representations. The input of the element encoder comprises the category ID denoted as d_c and the subcategory ID denoted as d_{sc}. The element encoder consists of two layers. The first layer is a category and subcategory ID embedding layer, which converts discrete IDs of categories and subcategories into low-dimensional dense representations, denoted as g^c and g^{sc}, respectively. The second layer is a dense layer, which transforms the dense representations of categories and subcategories to learn hidden representations of them. The transformation calculation formulas are:

$$r^c = \mathrm{ReLU}(D_c \times g^c + d_c) \tag{11}$$

$$r^{sc} = \mathrm{ReLU}(D_{sc} \times g^{sc} + d_{sc}) \tag{12}$$

where D_c, d_c, D_{sc}, and d_{sc} are parameters in dense layers.

The third component of the news model is the attention-based pooling layer. Different types of news texts contain varying levels of semantic information, leading to differences in the quality of learned news representations. Some news texts are ambiguous and brief, which is detrimental to learning news representations. On the other hand, some news texts have clear perspectives and detailed content. In these cases, there are many news expressing different attitudes towards the same event. We aim to take note of these aspects in the implementation of news recommendation algorithms to prevent users from being confined within information bubbles. Therefore, words with explicit attitudes when representing the news carry high weight. For short and ambiguous news, categories are weighted more heavily than the text in representing such news. Inspired by these observations, we use an attention-based pooling layer to capture semantic information of news with varying qualities. The attention weights for the title, abstract, category, and subcategory are denoted as α_t, α_{ab}, α_c, and α_{sc}, respectively. The calculation formula for the attention weight of the abstract is as follows:

$$a_{ab} = q_v^T \tanh(U_v \times r^{ab} + u_v) \tag{13}$$

$$\alpha_{ab} = \frac{\exp(a_{ab})}{\exp(a_t) + \exp(a_{ab}) + \exp(a_c) + \exp(a_{sc})} \tag{14}$$

where U_v and u_v are projection parameters, q_v is the attention query vector. A similar method can also be used to calculate the attention weights for other news items, such as the title, category, and subcategory.

The unified news representation learned by the news model is obtained as in Eq. (15).

$$r = \alpha_t r^t + \alpha_{ab} r^{ab} + \alpha_c r^c + \alpha_{sc} r^{sc} \tag{15}$$

3.4 Click Predictor

In this section, we explain how to calculate click prediction scores for candidate news to obtain the recommended list. The click predictor is utilized to forecast the probability of users clicking on candidate news. The representation of candidate news is denoted as r_c, and the representation of the user as u. The click probability score is obtained through inner product, as shown in Eq. (16).

$$\widehat{y} = u^T r_c \tag{16}$$

3.5 Multi-Loss Adversarial Training

In this section, we introduce our pluggable component, AT4D. During the training process, most news recommendation methods are for accuracy [5]. An enormous amount of news is generated worldwide every day, containing rich semantics. However, it is undeniable that a considerable amount of homogeneous information is also produced, which is common in news recommendation. Inspired by Wu et al. [22], we recognize that adversarial training can enhance the precision and fairness of news recommendation, yet its capability to improve both accuracy and diversity remains unexplored. To address this issue, we propose multi-loss adversarial training, which effectively facilitates learning on hard samples. Adversarial training is a training technique that enhances model robustness by incorporating adversarial samples during the training process. It mitigates the effect of redundant information from past user click activities on the model's performance and improves the model's generalization capacity. Additionally, we aim to prevent a decrease in model accuracy due to inconsistencies between the training and testing phases. We adopt a variant of the regularization training method R-AT proposed by Ni et al. [23]. Specifically, through the perturbation method of FreeLB [24], we dynamically perturb the input news embedding vectors in multiple steps and minimize adversarial risks in different regions around input samples. The output samples are passed through two sub-modules of dropout, where neural units are randomly discarded to generate two adversarial samples with other distributions. Next, by lowering the bidirectional Kullback-Leibler (KL) divergence between two distinct output probability distributions, we regularize the model prediction, which is denoted as $\mathcal{L}_{KL}$. Meanwhile, to further enhance the accuracy of model recommendations, we abstract the news recommendation task into a classification task that corresponds to two levels of click action: clicked and non-clicked. Therefore, we introduce classification loss to improve recommendation accuracy, denoted as $\mathcal{L}_{CLS}$. Finally, we design a diversity-aware loss function, denoted as $\mathcal{L}_{impD}$, to enforce the model to not only satisfy accuracy but also possess the ability to enhance diversity. Notably, in each training step, we first perform forward propagation and backward computation on the original samples, utilizing the cross-entropy loss function, denoted as $\mathcal{L}_{st}$ in Eq. (17), to ensure model learning.

$$\mathcal{L}_{st} = -\sum_{i=1}^{P} \log \frac{\exp(\widehat{y}_i^+)}{\exp(\widehat{y}_i^+) + \sum_{i=1}^{K} \exp(\widehat{y}_i^-)} \tag{17}$$

where P represents the number of positive samples, K represents the number of negative samples. The calculation process involves computing the KL divergence of corresponding adversarial samples as in Eq. (20), the classification loss as in Eq. (21), and the diversity improvement loss as in Eq. (22).

$$\mathcal{L}_{adv_1} = -\sum_{i=1}^{P} \log P_1(y_i|x_i + \delta_{adv_i}) \tag{18}$$

$$\mathcal{L}_{adv_2} = -\sum_{i=1}^{P} \log P_2(y_i|x_i + \delta_{adv_i}) \tag{19}$$

$$\mathcal{L}_{KL} = \frac{D_{KL}[P_1(y_i|x_i + \delta_{adv_i}) + P_2(y_i|x_i + \delta_{adv_i})]}{2} \tag{20}$$

$$\mathcal{L}_{CLS} = \frac{\mathcal{L}_{adv_1} + \mathcal{L}_{adv_2}}{2} \tag{21}$$

$$\mathcal{L}_{impD} = \frac{D_{KL}[P_1(y_i|x_i + \delta_{adv_i}) + P(y_i|x_i)] + D_{KL}[P_2(y_i|x_i + \delta_{adv_i}) + P(y_i|x_i)]}{2} \tag{22}$$

where δ_{adv_i} is the perturbation on the news embedding vector, $P_1(y_i|x_i+\delta_{adv_i})$ and $P_2(y_i|x_i+\delta_{adv_i})$ denote the two distributions of the perturbed samples, and D_{KL} is the method for calculating KL divergence. Ultimately, the second backward propagation loss is obtained as in Eq. (23).

$$\mathcal{L}_{total} = \mathcal{L}_{st} + \alpha\mathcal{L}_{KL} + \beta\mathcal{L}_{CLS} + \gamma\mathcal{L}_{impD} \tag{23}$$

where α, β and γ are adaptive parameters, and $\alpha + \beta + \gamma = 1$, which are used to adjust the weights of multiple different loss functions during the second forward propagation process.

4 Experiments

4.1 Datasets

We conduct experiments on the real-world news recommendation dataset MIND [1], which was collected from MSN News logs. The MIND dataset is available in two versions: MIND-large and MIND-small. MIND-large is constructed from Microsoft News user behavior data spanning six weeks, from 12 October 2019, to 22 November 2019. The user data from the first four weeks is utilized for building user reading histories, data from the second-to-last week is used for model training, and data from the final week is reserved for evaluation. A uniformly random sample of MIND-large's daily behavioral records was used to create MIND-small, a scaled-down version of MIND-large. Detailed statistics of the dataset are presented in Table 1.

Table 1: Detailed statistics of the datasets

	MIND-large	MIND-small
# User	1,000,000	94,057
# News	161,013	65,238
# Impressions	15,777,377	230,117
# Click behaviors	24,155,470	347,727
# Number of words	101,233	70,973
# Number of categories	296	275
# Average of news	596.86	629.51

4.2 Experimental Settings

We tune parameters in our work on the MIND-small dataset, followed by training and evaluation on both MIND-large and MIND-small datasets. We initialize the word embedding using pre-trained 300-dimensional GloVe [20]. For consistency with other baselines and training efficiency, we treated each news article clicked by the user as a positive sample and randomly selected four articles from the recommendation list that were not clicked as negative samples. In the multi-loss adversarial training, hyperparameters α, β, and γ can be set to 0.1, 0.5, and 0.4, respectively, for a quick start. We conduct model optimization using Adam [25]. Moreover, following Qi et al. [26], we train the discriminator and adversarial task iteratively for a single step instead of seeking the optimal discriminator. We adopt AUC, MRR, nDCG@5, and nDCG@10 as measures of recommendation accuracy. Consistent with Qi et al. [16], we employ the Intra-List Average Distance (ILAD) of the top 10 ranked news articles as a measure of recommendation diversity. The experimental results represent the averages of these metrics over five random experiments conducted on all interaction logs from MIND-large and MIND-small datasets. All experiments are conducted on the RTX 2080 Ti.

4.3 Compare Models

We employ seven comparative models as baseline models for this task, including:

- NAML [7], using attention networks and CNN to obtain news and user modeling;
- LSTUR [8], utilizing user ID embeddings to model long-term user interests and GRU networks to model short-term user interests;
- UNBERT [9], using the pre-trained model BERT to get multi-grained user-news interactions at news-level and word-level;
- DKN [14], using knowledge-aware convolutional neural networks for news modeling and candidate-aware attention for user modeling;
- TANR [27], a neural news recommendation method with an auxiliary topic classification task to get topic-aware news representation;
- NRMS [28], employing multi-head self-attention mechanisms to learn user and news representations;
- MINS [29], embedding topics and textual features to encode news and using a parallel interest network to encode users.

5 Results and Discussions

In this section, we evaluate the performance of our proposed algorithm on the MIND-large and MIND-small datasets, focusing on both accuracy and diversity in news recommendation. Specifically, we pose the following four research questions to guide the experiments:

RQ1: Does our proposed algorithm effectively improve diversity while enhancing the accuracy of recommendation results?

RQ2: Is AT4D effective in our proposed algorithm?

RQ3: How does each module of our proposed algorithm effectively enhance its performance?

RQ4: Can the pluggable component AT4D boost the performance of other methods in terms of accuracy and diversity?

5.1 Performance Comparison (RQ1)

In the preceding chapters, we introduce the proposed news recommendation algorithm, two open-source news recommendation datasets, and four accuracy-based evaluation metrics, along with one diversity-based evaluation metric. To tackle RQ1, we carry out extensive experiments on the MIND-small and MIND-large datasets. We evaluated the performance of our method by comparing it with the baseline models. The average results and the variance of these methods on the two datasets are summarized in Tables 2 and 3, as well as Fig. 2. Considering the size of the page, we have shifted the decimal point of the result to the right by two places in Tables 2 and 3. *Improv*. refers to the improvement of the best performance over the next best performance.

Table 2: Performance comparison of the proposed algorithm and other baselines on the MIND-small datasets. In each row, the best performance is bolded, and the second best is underlined

Model	MIND-small			
	AUC	MRR	nDCG@5	nDCG@10
DKN	64.27 ± 0.19	30.22 ± 0.11	33.39 ± 0.23	39.68 ± 0.25
TANR	64.66 ± 0.17	30.05 ± 0.08	33.23 ± 0.06	39.57 ± 0.11
NAML	65.45 ± 0.17	31.00 ± 0.16	34.44 ± 0.23	40.63 ± 0.27
LSTUR	63.12 ± 0.19	28.83 ± 0.30	31.29 ± 0.28	38.00 ± 0.27
UNBERT	66.56 ± 0.21	31.35 ± 0.12	35.17 ± 0.18	41.39 ± 0.17
NRMS	63.32 ± 0.20	27.91 ± 0.17	30.51 ± 0.12	37.26 ± 0.14
MINS	66.86 ± 0.15	31.81 ± 0.15	35.45 ± 0.16	41.70 ± 0.14
Ours	**67.50 ± 0.13**	**32.77 ± 0.09**	**36.27 ± 0.14**	**42.33 ± 0.10**
Improv.	0.96%	3.02%	2.31%	1.51%

Table 3: Performance comparison of the proposed algorithm and other baselines on the MIND-large datasets. In each row, the best performance is bolded, and the second best is underlined

Model	MIND-large			
	AUC	MRR	nDCG@5	nDCG@10
DKN	64.92 ± 0.23	30.34 ± 0.20	33.65 ± 0.21	39.98 ± 0.19
TANR	65.36 ± 0.11	30.42 ± 0.18	33.72 ± 0.19	40.26 ± 0.19
NAML	66.32 ± 0.12	31.46 ± 0.16	34.87 ± 0.17	41.20 ± 0.20
LSTUR	65.65 ± 0.21	30.53 ± 0.18	33.96 ± 0.17	40.37 ± 0.20
UNBERT	66.53 ± 0.10	32.27 ± 0.09	35.26 ± 0.07	42.16 ± 0.10
NRMS	66.01 ± 0.05	31.28 ± 0.17	34.49 ± 0.16	40.89 ± 0.09
MINS	66.79 ± 0.17	32.57 ± 0.25	34.97 ± 0.09	41.72 ± 0.11
Ours	**68.62 ± 0.12**	**33.57 ± 0.14**	**37.54 ± 0.15**	**43.61 ± 0.18**
Improv.	2.74%	3.07%	6.47%	3.44%

We obtain several observations from Tables 2 and 3. Firstly, on the publicly available datasets MIND-large and MIND-small, the proposed algorithm outperforms the current baseline models, DKN, TANR, NAML, LSTUR, UNBERT, NRMS, and MINS, in terms of four accuracy-based metrics. This indicates the effectiveness of the proposed algorithm. The proposed algorithm can improve the accuracy of user interest matching, so as to attract more users' attention to the online news platforms.

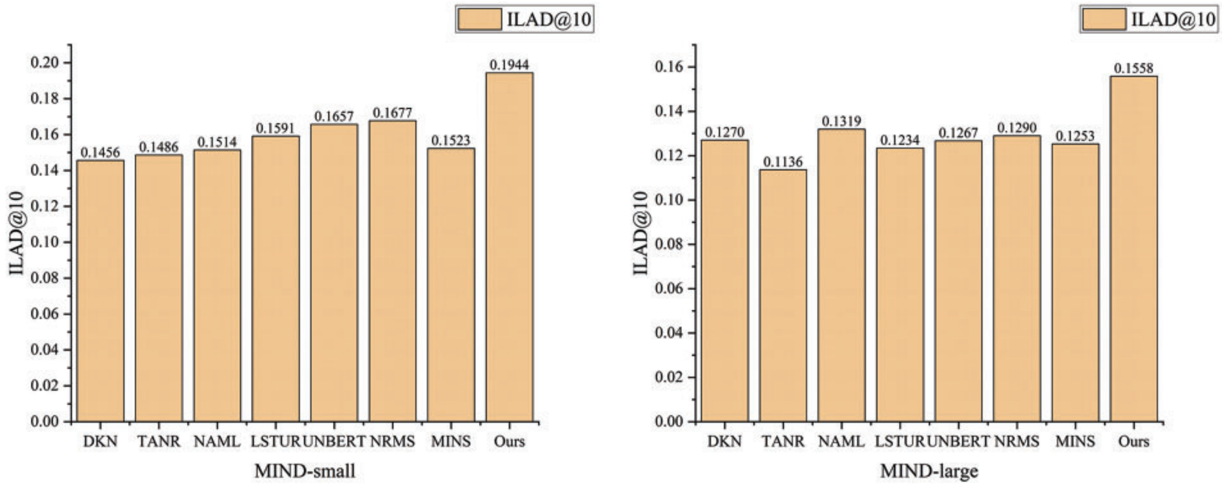

Figure 2: Performance of different methods on recommendation diversity on the MIND-small datasets and the MIND-large datasets, respectively

Secondly, methods using attention mechanisms (TANR, NAML, LSTUR, UNBERT, NRMS, and MINS) for learning news representations outperform DKN. This is because attention mechanisms can model deep interactions between words, enabling more accurate learning of news representations by capturing the relative importance of these interactions. Particularly, in terms of accuracy-based metrics, MINS outperforms other attention-based methods on MIND-small datasets as it additionally considers the inter-news dependencies. UNBERT outperforms other attention-based methods on MIND-large datasets in terms of the nDCG@5 and nDCG@10, as it additionally leverages the pre-trained model to improve textual understanding.

From Fig. 2, it can be observed that when considering the diversity of recommendation results, previous accuracy-driven works often exhibit mediocre performance in diversity evaluation. This results in low diversity of the final candidate news. It is shown that the experimental results of the same method on the MIND-large dataset are consistently lower than those on the MIND-small dataset. Moreover, the baseline model shows no significant trend in terms of diversity performance on both datasets. The proposed algorithm shows significant improvements in diversity evaluation metrics ILAD@10 over the best-performing baseline models on the MIND-small and MIND-large datasets, with enhancements of 15.92% and 18.12%, respectively. The advancement indicates that the proposed algorithm not only enhances accuracy but also further improves diversity.

5.2 Effectiveness in Adversarial Training (RQ2)

As mentioned in Section 3.5, a large amount of homogeneous news information can lead to insufficient or ambiguous semantic information, rendering news recommendation results unsatisfactory. Additionally, adversarial training proves effective for the news recommendation task, although its effectiveness in learning from hard samples is limited. In this section, we experiment with seven adversarial training methods to assess their positive impact on the news recommendation task.

We derive some observational results from Table 4. Using adversarial training strategies not only serves to prevent overfitting and effectively enhance the model's generalization capability but also, to some extent, can improve the accuracy and diversity of news recommendation. From the experimental results in Table 4, it is evident that employing FGM and PGD adversarial training methods yields good recommendation performance under accuracy-based evaluation metrics. In contrast, the performance under diversity-based evaluation metrics is moderate. However, using the FreeLB adversarial training method does not significantly compromise recommendation performance under accuracy-based evaluation metrics while yielding optimal performance under diversity-based evaluation metrics. Furthermore, FreeAT performs less favorably than FreeLB in terms of diversity metrics for news recommendation tasks, and it does not outperform FGM and PGD in terms of accuracy metrics. Based on this promising preliminary experimentation, we propose AT4D, which adopts the perturbation manner of FreeLB. This choice is informed by AT4D's aim to balance the accuracy and diversity of news recommendation results.

Table 4: Performance comparison of the proposed algorithm without AT4D and with different adversarial training methods on the MIND-small datasets. In each row, the best performance is bolded, and the second best is underlined

Model	MIND-small				
	AUC	MRR	nDCG@5	nDCG@10	ILAD@10
Ours w/o AT	0.6649	0.3204	0.3547	0.4156	0.1813
Ours-FreeLB [24]	0.6649	0.3204	0.3547	0.4156	**0.1619**
Ours-R-Drop [30]	0.6659	0.3217	0.3574	0.4188	0.1313
Ours-FGSM [31]	0.6623	0.3127	0.3443	0.4073	0.1199
Ours-FGM [32]	<u>0.6670</u>	**0.3253**	**0.3612**	**0.4216**	0.1504
Ours-PGD [33]	**0.6682**	<u>0.3227</u>	<u>0.3587</u>	<u>0.4194</u>	0.1538
Ours-FreeAT [34]	0.6632	0.3161	0.3485	0.4118	<u>0.1578</u>
Ours-SMART [35]	0.6629	0.3180	0.3495	0.4116	0.1506

5.3 Ablation Study (RQ3)

In this section, we conduct ablation studies to assess how each module of the proposed algorithm effectively enhances its performance. We validate the effectiveness of the augmented user representation and the AT4D by removing them from the proposed algorithm and evaluating their impact on news recommendation using the MIND-small dataset.

From Fig. 3, we find that the absence of any module can lead to a decrease in the performance of the model. The AT4D contributes the most to the model, primarily because it is a multi-loss adversarial training that can significantly improve both the accuracy and diversity of recommendation results. Following AT4D, the augmented user representation shows the next highest contribution. In summary, each module in the news recommendation process contributes significantly to the model's performance.

5.4 Compatibility with Other NR Models (RQ4)

As previously mentioned, AT4D is a pluggable component, allowing us to integrate it with existing news recommendation methods to enhance their performance. Therefore, to validate AT4D's compatibility with other news recommendation methods, we apply it to several news recommendation models listed in Table 2, namely NAML, DKN, TANR, NRMS, LSTUR, and MINS. Besides adding

AT4D, we do not modify their original training settings. We evaluate their accuracy and diversity on the MIND-small dataset. The results of their performance in terms of recommendation are summarized in Table 5 and Fig. 4. Green font indicates an increase in results, while red font indicates a decrease.

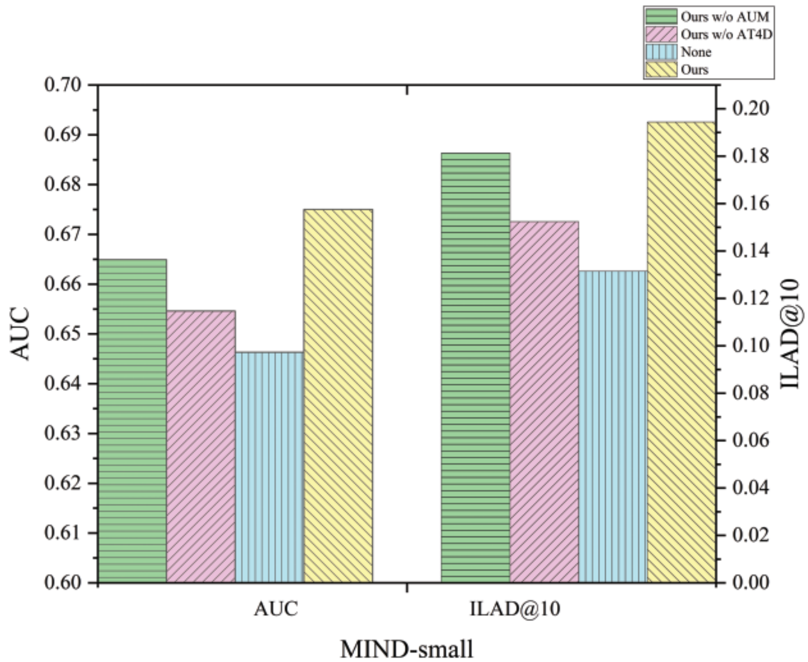

Figure 3: Effectiveness of several core model components on the NIND-small dataset

Table 5: Compatibility with other baseline news recommendation methods on the MIND-small datasets

Model	MIND-small			
	AUC	MRR	nDCG@5	nDCG@10
DKN	0.6506 + 0.0079	0.3422 + 0.0400	0.3518 + 0.0179	0.4117 + 0.0149
TANR	0.6398 − 0.0068	0.3022 + 0.0017	0.3349 + 0.0026	0.3954 − 0.0003
NAML	0.6652 + 0.0107	0.3471 + 0.0371	0.3539 + 0.0095	0.4158 + 0.0095
LSTUR	0.6458 + 0.0146	0.3215 + 0.0332	0.3231 + 0.0102	0.3891 + 0.0091
UNBERT	0.6725 + 0.0069	0.3258 + 0.0123	0.3610 + 0.0093	0.4225 + 0.0086
NRMS	0.6405 + 0.0073	0.3228 + 0.0437	0.3272 + 0.0221	0.3920 + 0.0194
MINS	0.6716 + 0.0030	0.3269 + 0.0088	0.3616 + 0.0071	0.4222 + 0.0052

We find that our AT4D consistently and effectively improves the accuracy of existing news recommendation models, except for the TANR model. We think TANR first learns important words as topics, which can bring some inevitable errors. Since AT4D acts directly on the model, this error is magnified. Additionally, AT4D efficiently enhances the diversity of these methods, significantly increasing the diversity level of previously accuracy-based recommendation approaches without sacrificing much accuracy. The AT4D performs best in improving diversity for the DKN method, which may be attributed to the additional incorporation of entity information in DKN, enabling AT4D to implement across multiple semantic dimensions effectively.

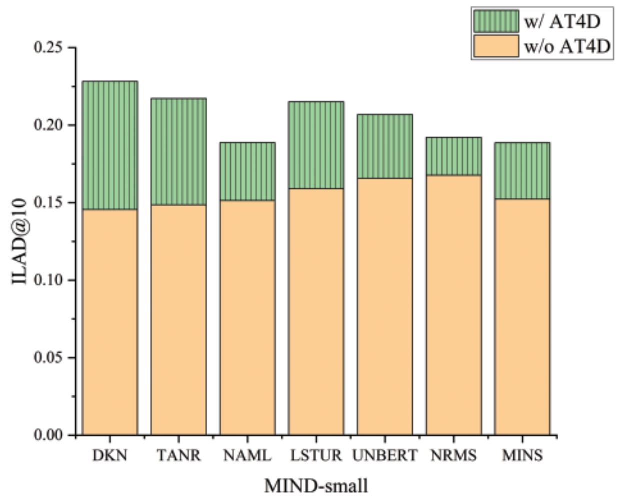

Figure 4: Effectiveness of AT4D under ILAD@10 metric on the MIND-small dataset

6 Conclusion

In this paper, we propose a news recommendation algorithm that achieves high performance in both accuracy and diversity. Unlike most existing approaches that solely optimize accuracy or incorporate more features into the model to enhance diversity, our proposed algorithm attains high performance in both accuracy and diversity through the model's capacity. In the proposed algorithm, we introduce an augmented user model to fully capture user intent and the consequent behavior guidance. Additionally, we propose an effective multi-loss adversarial training method for diversity (AT4D), a pluggable component that improves both the accuracy and diversity of news recommendation results.

However, our proposed algorithm has certain limitations. Due to constraints within the dataset, we only have access to limited news and user information. In future work, we aim to explore the integration of semantic understanding and text generation capabilities powered by large language models to expand datasets and enhance the performance of our news recommendation algorithm.

Acknowledgement: The authors extend their appreciation to the anonymous reviewers for their constructive comments.

Funding Statement: This research was funded by Beijing Municipal Social Science Foundation (23YTB031) and the Fundamental Research Funds for the Central Universities (CUC23ZDTJ005).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Shuang Feng, Ruijin Xue; data collection: Ruijin Xue; analysis and interpretation of results: Ruijin Xue; draft manuscript preparation: Ruijin Xue, Shuang Feng, Qi Wang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available at https://msnews.github.io/ (accessed on 18 May 2024).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] F. Wu *et al.*, "MIND: A largescale dataset for news recommendation," in *Proc. 58th Annu. Meet. Assoc. Comput. Linguist.*, Jul. 2020, pp. 3597–3606.
[2] K. Yang, S. Long, W. Zhang, J. Yao, and J. Liu, "Personalized news recommendation based on the text and image integration," *Comput. Mater. Contin.*, vol. 64, no. 1, pp. 557–570, 2020. doi: 10.32604/cmc.2020.09907.
[3] S. Raza and C. Ding, "News recommender system: A review of recent progress, challenges, and opportunities," *Artif. Intell. Review.*, vol. 55, no. 1, pp. 749–800, Jan. 2022. doi: 10.1007/s10462-021-10043-x.
[4] C. Wu, F. Wu, M. An, J. Huang, Y. Huang and X. Xie, "NPA: Neural news recommendation with personalized attention," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. & Data Mining*, Anchorage, AK, USA, Jul. 2019, pp. 2576–2584.
[5] C. Wu, F. Wu, T. Qi, and Y. Huang, "End-to-end learnable diversity-aware news recommendation," arXiv:2204.00539, 2022. doi: 10.48550/arXiv.2204.00539.
[6] S. Okura, Y. Tagami, S. Ono, and A. Tajima, "Embedding-based news recommendation for millions of users," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discov. & Data Mining*, Halifax, NS, Canada, 2017, pp. 1933–1942.
[7] C. Wu, F. Wu, M. An, J. Huang, Y. Huang and X. Xie, "Neural news recommendation with attentive multi-view learning," in *Proc. Twenty-Eighth Int. Joint Conf. on Artif. Intell.*, Macao, China, 2019, pp. 3863–3869.
[8] M. An, F. Wu, C. Wu, K. Zhang, Z. Liu and X. Xie, "Neural news recommendation with long-and short-term user representations," in *Proc. 57th Annu. Meet. Assoc. Comput. Linguist.*, Florence, Italy, Jul. 2019, pp. 336–345.
[9] Q. Zhang *et al.*, "UNBERT: User-news matching BERT for news recommendation," in *Proc. Thirtieth Int. Joint Conf. Artif. Intell.*, Montreal, QC, Canada, 2021, pp. 3356–3362.
[10] C. Wu, F. Wu, Y. Huang, and X. Xie, "Personalized news recommendation: Methods and challenges," *ACM Trans. Inf. Syst.*, vol. 41, no. 1, pp. 1–50, Jan. 2023. doi: 10.1145/3530257.
[11] T. Qi *et al.*, "HieRec: Hierarchical user interest modeling for personalized news recommendation," in *Proc. 59th Annu. Meet. Assoc. Comput. Linguist. 11th Int. Joint Conf. Nat. Lang. Process.*, 2021, pp. 5446–5456.
[12] K. Shivaram, P. Liu, M. Shapiro, M. Bilgic, and A. Culotta, "Reducing cross-topic political homogenization in content-based news recommendation," in *Proc. 16th ACM Conf. Recommender Syst.*, New York, NY, USA, 2022, pp. 220–228.
[13] H. Wang, F. Wu, Z. Liu, and X. Xie, "Fine-grained interest matching for neural news recommendation," in *Proc. 58th Annu. Meet. Assoc. Comput. Linguist.*, Jul. 2020, pp. 836–845.
[14] H. Wang, F. Zhang, X. Xie, and M. Guo, "DKN: Deep knowledge-aware network for news recommendation," in *Proc. 2018 World Wide Web Conf.*, Lyon, France, 2018, pp. 1835–1844.
[15] A. Gharahighehi and C. Vens, "Diversification in session-based news recommender systems," *Pers. Ubiquitous Comput.*, vol. 27, no. 1, pp. 5–15, 2023. doi: 10.1007/s00779-021-01606-4.
[16] T. Qi, F. Z. Wu, C. H. Wu, and Y. F. Huang, "PP-Rec: News recommendation with personalized user interest and time-aware news popularity," in *Proc. 59th Annu. Meet. Assoc. Comput. Linguist.*, 2021, pp. 5457–5467.
[17] L. Zhang, P. Liu, and J. A. Gulla, "A deep joint network for sessionbased news recommendations with contextual augmentation," in *Proc. 29th Hypertext Soc. Media*, Baltimore, MD, USA, 2018, pp. 201–209.

[18] J. Wei and K. Zou, "EDA: Easy data augmentation techniques for boosting performance on text classification tasks," in *Proc. 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Joint Conf. Nat. Lang. Process.*, Hong Kong, China, 2019, pp. 6383–6389.

[19] W. Fan, Y. Wang, and H. Hu, "Mimicking human verification behavior for news media credibility evaluation," *Appl. Sci.*, vol. 13, no. 17, pp. 9553, 2023. doi: 10.3390/app13179553.

[20] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proc. 2014 Conf. Empir. Methods Nat. Lang. Process.*, Doha, Qatar, 2014, pp. 1532–1543.

[21] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. Fourteenth Int. Conf. Artif. Intell. Stat.*, Fort Lauderdale, FL, USA, 2011, pp. 315–323.

[22] C. Wu, F. Wu, X. Wang, Y. Huang, and X. Xie, "Fairness-aware news recommendation with decomposed adversarial learning," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 5, pp. 4462–4469. doi: 10.1609/aaai.v35i5.16573.

[23] S. Ni, J. Li, and H. Kao, "R-AT: Regularized adversarial training for natural language understanding," in *Findings Assoc. Comput. Linguist.: EMNLP 2022; Assoc. Comput. Linguist.*, Abu Dhabi, United Arab Emirates, 2022, pp. 6427–6440.

[24] C. Zhu, Y. Cheng, Z. Gan, S. Sun, T. Goldstein and J. Liu, "FreeLB: Enhanced adversarial training for natural language understanding," 2019. doi: 10.48550/arXiv.1909.11764.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014. doi: 10.48550/arXiv.1412.6980.

[26] T. Qi *et al.*, "ProFairRec: Provider fairness-aware news recommendation," in *Proc. 45th Int. ACM SIGIR Conf. Res Dev. Inf. Retr.*, Madrid, Spain, 2022, pp. 1164–1173.

[27] C. Wu, F. Wu, M. An, Y. Huang, and X. Xie, "Neural news recommendation with topic-aware news representation," in *Proc. 57th Annu. Meet. Assoc. Comput. Linguist.*, Florence, Italy, 2019, pp. 1154–1159.

[28] C. Wu, F. Wu, S. Ge, T. Qi, Y. Huang and X. Xie, "Neural news recommendation with multi-head self-attention," in *Proc. 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Joint Conf. Nat. Lang. Process.*, Hong Kong, China, 2019, pp. 6389–6394.

[29] R. Wang, S. Wang, W. Lu, and X. Peng, "News recommendation via multi-interest news sequence modelling," in *ICASSP 2022-2022 IEEE Int. Conf. Acoust., Speech Signal Process.*, Singapore, 2022, pp. 7942–7946.

[30] L. Wu *et al.*, "R-Drop: Regularized dropout for neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 10890–10905, 2021.

[31] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," 2014. doi: 10.48550/arXiv.1412.6572.

[32] T. Miyato, A. Dai, and I. Goodfellow, "Adversarial training methods for semi-supervised text classification," 2016. doi: 10.48550/arXiv.1605.07725.

[33] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," 2017. doi: 10.48550/arXiv.1706.06083.

[34] A. Shafahi *et al.*, "Adversarial training for free!" in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, Vancouver, Canada, 2019.

[35] H. Jiang, P. He, W. Chen, X. Liu, J. Gao and T. Zhao, "SMART: Robust and efficient fine-tuning for pre-trained natural language models through principled regularized optimization," in *The 58th Annu. Meet. Assoc. Comput. Linguist.*, 2020.