A Hierarchical and Disentangling Interest Learning Framework for Unbiased and True News Recommendation

Shoujin Wang
shoujin.wang@uts.edu.au
University of Technology Sydney
Sydney, Australia

Wentao Wang
wangwentao0915@gmail.com
University of Technology Sydney
Sydney, Australia

Xiuzhen Zhang
xiuzhen.zhang@rmit.edu.au
RMIT University
Melbourne, Australia

Yan Wang
yan.wang@mq.edu.au
Macquarie University
Sydney, Australia

Huan Liu
huanliu@asu.edu
Arizona State University
Tempe, United States

Fang Chen
fang.chen@uts.edu.au
University of Technology Sydney
Sydney, Australia

ABSTRACT

In the era of information explosion, news recommender systems are crucial for users to effectively and efficiently discover their interested news. However, most of the existing news recommender systems face two major issues, hampering recommendation quality. Firstly, they often oversimplify users' reading interests, neglecting their hierarchical nature, spanning from high-level event (e.g., US Election) related interests to low-level news article-specific interests. Secondly, existing work often assumes a simplistic context, disregarding the prevalence of fake news and political bias under the real-world context. This oversight leads to recommendations of biased or fake news, posing risks to individuals and society. To this end, this paper addresses these gaps by introducing a novel framework, the Hierarchical and Disentangling Interest learning framework (HDInt). HDInt incorporates a hierarchical interest learning module and a disentangling interest learning module. The former captures users' high- and low-level interests, enhancing next-news recommendation accuracy. The latter effectively separates polarity and veracity information from news contents and model them more specifically, promoting fairness- and truth-aware reading interest learning for unbiased and true news recommendations. Extensive experiments on two real-world datasets demonstrate HDInt's superiority over state-of-the-art news recommender systems in delivering accurate, unbiased, and true news recommendations.

CCS CONCEPTS

• **Information systems** → **Retrieval tasks and goals**.

KEYWORDS

News recommendation, Fake news, Bias

ACM Reference Format:
Shoujin Wang, Wentao Wang, Xiuzhen Zhang, Yan Wang, Huan Liu, and Fang Chen. 2024. A Hierarchical and Disentangling Interest Learning Framework for Unbiased and True News Recommendation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24), August 25–29, 2024, Barcelona, Spain.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3637528.3671944

KDD '24, August 25–29, 2024, Barcelona, Spain

ACM ISBN 979-8-4007-0490-1/24/08.
https://doi.org/10.1145/3637528.3671944

1 INTRODUCTION

News recommender systems (RSs) aim at helping users efficiently discover news items aligned with their specific interests from a vast array of candidates [14, 30, 31]. News RSs have assumed an increasingly vital role in the contemporary era of big data and digital economy. Various news RSs have been widely applied on news portals such as CNN, BBC, as well as social media platforms like Facebook and TikTok [21]. In practical terms, news RSs can influence or even change users' reading behaviors to a great extent [6].

A series of studies on news RSs have been documented in the literature, spanning from conventional machine learning-based methods to advanced deep learning-based approaches. For instance, Capelle et al. [3] introduced a method that represents news using Synset Frequency-Inverse Document Frequency for news recommendations. More recently, deep learning models have been employed to enhance the understanding of news content for more accurate recommendations. For example, Wu et al. [38] used self-attention and additive attention to represent words in each news item and news sequences, respectively, for news recommendations. Wang et al. [27] utilized dilated convolutions to capture fine-grained reading interest matching degrees. Mao et al. [16] utilized graph attention networks to improve user and news representations by incorporating additional semantic information.

Although great advancements have been achieved, most of existing news RSs face two major gaps, significantly hampering the recommendation quality. **Gap 1:** *they often oversimplify users' reading interests into a single level, neglecting the hierarchical nature (i.e., high- and low-level) of them, leading to limited recommendation accuracy*. In practice, a user's reading interests towards news often span from high-level interests in events (e.g., Russia-Ukraine conflict) to low-level interests in specific news articles. Furthermore, these interests often change along with time and low-level interests usually change faster compared with high-level ones, and thus such two levels of interests cannot be well modelled in a single level. For instance, a user Alice has kept reading a sequence of different news articles for the event "Russia-Ukraine conflict" for several months and then changed to reading news articles for another relevant event "Nord Stream pipeline bombed". In this case, Alice's change on news articles is much faster than that on news events since each event often has a series of news articles of various styles associated with it. **Gap 2:** *more importantly, existing news RSs often assume a simplistic context, disregarding the prevalence of political bias (i.e., left-leaning or right-leaning) and fake news under the real-world context* [2]. This oversight easily leads to biased and/or fake news recommendations, posing risks to individuals and society [33, 35]. In the real world, it is not uncommon that there may exist politically biased news and fake news in the cyber space, bringing significantly extra challenges for news recommendations.

These challenges cannot be well addressed by existing news RSs even some of them have taken news bias or veracity into account. On the one hand, existing fairness-aware news RSs mainly focus on sentiment bias on news side [39], exposure bias on provider side [20], or attribute (e.g., gender) bias on user side [40]. However, the commonly existing political polarity bias on news side is less studied though it often generates significant impact on the public opinion. More importantly, they often sacrifice recommendation accuracy and overlook the existence of fake news. On the other hand, those very limited studies [33] which take news veracity into account tend to overlook other important aspects like fairness and accuracy of recommendations. All in all, there lacks of a unified framework to comprehensively address all the aforementioned three important aspects for generating high-quality news recommendations: **accuracy, news bias and veracity**.

To this end, we aim at bridging those aforementioned two significant gaps in the literature. Specifically, we focus on the concrete task of how to make accurate, unbiased and true news recommendations toward a healthy news reading environment under the complex context where biased news and fake news commonly exist.

Accordingly, we propose a novel unified Hierarchical and Disentangling Interest learning framework (HDInt) for accurate, unbiased and true news recommendations. HDInt is composed of two core modules: *(1) a hierarchical interest learning module* for bridging the first gap, and *(2) a disentangling interest learning module* for bridging the second gap. In particular, the first module is equipped with two relatively independent dynamic interest learning components to well learn each user's high-level and low-level reading interests respectively from their historical news articles, leading to accurate recommendations. The second module is equipped with two disentanglers which successively disentangle the political polarity bias and veracity information from news contents and model them in a relatively independent way. In this way, the reading interest learned from users' reading history are more politically unbiased and true news-oriented, contributing to recommending unbiased and true news. Furthermore, the disentangled veracity information will be input into a veracity classifier to accurately predict the veracity label (i.e., true or fake) of each candidate news item so that only those predicted true news will be recommended to end users.

Main contributions of this work are summarized below:

- We propose a novel, practical and emerging research problem, namely how to make accurate, unbiased, and true news recommendations with a unified framework under the real-world complex context where biased and fake news commonly exist?
- We propose a novel and unified HDInt framework as one of the early solutions to this challenging problem. HDInt is more robust to the real-world complex context than existing news RSs.
- In HDInt, a novel hierarchical interest learning module is devised to accurately learn each user's high-level and low-level reading interests respectively, together with another novel disentangling interest learning module to effectively learn each user's unbiased and true news-oriented reading interests.

We instantiate this framework into a specific model. Extensive experimental results on two real-world datasets demonstrate the consistent and significant superiority of our model over the representative and state-of-the-art methods in terms of recommending accurate, unbiased and true news to users.

2 RELATED WORK

2.1 Accuracy-oriented News Recommendation

News RSs generally recommend the next piece of news to users via learning users' reading interest towards news from their reading history, i.e., a sequence of news items. During the past decades, a variety of methods have been proposed to continually improve news recommendation accuracy. For instance, Capelle et al. [3] introduced a method which utilises Synset Frequency-Inverse Document Frequency to represent news contents for news recommendations. In recent years, deep neural networks have been widely employed to improve the accuracy of news recommendations. Various deep neural models from recurrent neural networks (RNNs) [1, 17], attention mechanism [37, 47], dilated convolution [27], graph neural networks [8], knowledge distillation [28], to deep reinforcement learning [46] are explored for news recommendations. For example, Wu et al. [38] introduced self-attention and additive attention models to represent words within each news article and a sequence of historical news for news recommendations. Wang et al. [27] utilised dilated convolution model to capture fine-grained interest matching signals for news recommendations. Ma et al. [15] proposed an unsupervised pre-training paradigm to effectively model user behaviors for accurate new recommendations. Although great success has been achieved, all of these methods are accuracy-oriented only. They overlook other important aspects, such as bias and fake information, which commonly exist in the real-world context.

2.2 Unbiased and True News Recommendation

Recently, researchers have realized the negative effect of bias in news recommendations, and thus developed various news RSs for unbiased recommendations [12]. They can be generally categorized into different classes according to the type of bias (e.g., user side bias, item side bias, provider side bias). For instance, Wu et al. [40] proposed a decomposed adversarial learning approach to alleviate the biases of sensitive user attributes for fair news recommendations. Wu et al. [39] proposed a sentiment diversity-aware news RS to remove the sentiment bias embedded in news contents. Qi et al. [20] proposed an orthogonal regularization to better learn provider-fair representations in news representations. However, all those studies did not touch the commonly exiting political bias. Only quite limited studies explored political effect in news recommendations. They either proposed an attention-based neural network model to reduce political homogenization in content-based news recommendations [23], or modelled the interaction between political typology and filter bubbles to diversify recommended news contents [13]. However, these studies are just to diversify the political polarities or news contents, totally different from the political bias we focused on in this work. Moreover, they overlooked the veracity information and may unavoidably recommend fake news to end users. Very few researchers take news veracity into account [26, 33]. However, they are just in the early stage and more importantly, they overlook the commonly existing bias issue.

3 PROBLEM FORMULATION

A user-news interaction dataset records each user's sequence of historical interactions (e.g., clicks or reading) with news in a certain time period. Specifically, $\mathcal{D} = \{\mathcal{S}_1, \cdots, \mathcal{S}_u, \cdots, \mathcal{S}_{|\mathcal{U}|}\}$ denotes a collection of news sequences interacted by all users (indexed by $u \in \mathcal{U}$), where $\mathcal{S}_u = \{v_1, \cdots, v_t\} (v \in \mathcal{V})$ consists of t pieces of news sequentially interacted by a user u. Each news is indexed by its interaction timestamp where $\mathcal{V} = \mathcal{V}_{left} \cup \mathcal{V}_{nuetral} \cup \mathcal{V}_{right}$ or $\mathcal{V} = \mathcal{V}_{true} \cup \mathcal{V}_{fake}$. $\mathcal{V}$ denotes the whole news set with polarity and veracity labels. In addition, a news information table $\mathcal{N}$ records the meta information (i.e., news title) of each news piece. For each user u, given their $(t-1)$ historical true and/or fake news pieces with polarity labels, denoted as context $C_u = \{v_1, \cdots, v_{t-1}\}$ together with the meta information, we build a recommendation framework $\mathcal{F}$ (i.e., HDInt) with two-fold goals: (1) learning users' high- and low-level interests from C_u for accurate recommendations, and (2) disentangling the polarity bias and veracity information from news contents for unbiased and true news recommendations.

4 THE HDINT FRAMEWORK

As shown in Figure 1 (a), the HDInt framework contains three main components: (1) *hierarchical interest learning module*, (2) *disentangling interest learning module*, and (3) *next-news prediction*. First, given a context consisting of a sequence of $(t-1)$ news pieces interacted by a user where each news piece is associated with news contents and keywords, we map the contents and keywords of each news item into a latent embedding vector respectively. Afterwards, disentangling interest learning module takes the content embedding as input and successively disentangles the potential political polarity information and veracity information with a polarity disentangler and a veracity disentangler respectively. Subsequently, the resultant polarity- and veracity-free content embedding will be input into the low-level interest learner within the hierarchical interest learning module to learn user's low-level interest towards specific news articles. At the same time, the keyword embedding is input into the high-level interest learner to learn high-level interest towards certain news events. Finally, both levels of interests are aggregated for predicting the next piece of news. In addition, prediction module takes each candidate news' veracity embedding as the input to predict its veracity label to filter out those predicted fake news. Although each module has been instantiated to a specific model structure, they can be easily instantiated to other structures in the future. Hence, our proposed HDInt is a general framework with great potential for expansion and further optimization.

4.1 Disentangling Interest Learning Module

As illustrated in Figure 1 (b), the disentangling interest learning module comprises two serially connected disentanglers to disentangle the polarity information including political bias and veracity information from news contents and model them in a relatively independent manner. Both disentanglers are built on an adversarial auto-encoder framework [5, 33, 44], which encompasses *an encoder*, *two decoders*, and a specifically designed *loss function module*. Here, we delve into the specifics of the polarity disentangler, while the veracity disentangler has a similar structure.

4.1.1 Encoder. The encoder consists of a three-layer network structure that takes the news content embedding as input. For a news v_j, we learn a content embedding $\mathbf{c}_j$ based on its title and description:

$$\mathbf{c}_j = FC(\mathbf{c}_j^m), \tag{1}$$

where FC indicates a fully connection layer with a sigmoid activation function. $\mathbf{c}_j^m$ is a concatenation of both news title embedding and news description embedding. Following the established conventions in text embedding [9, 43], both types of embeddings are initially generated using a widely adopted pre-trained BERT model [10]. Then, we undergo fine-tuning to align with the specifics of our task. We empirically set the dimension of $\mathbf{c}_j$ to 256, employing a grid search method. The dimensions of the title embedding and description embedding remain at the default 768 within the BERT model. Once content embedding $\mathbf{c}_j$ is ready, it is inputted into the encoder to derive the hidden state $\mathbf{h}_j^p$ in the polarity disentangler:

$$\mathbf{z}_j^1 = LeakyReLU(Dense(\mathbf{c}_j)), \tag{2}$$

$$\mathbf{z}_j^2 = LeakyReLU(Dense([\mathbf{c}_j; \mathbf{z}_j^1])), \tag{3}$$

$$\mathbf{h}_j^p = LeakyReLU(Dense([\mathbf{c}_j; \mathbf{z}_j^2])), \tag{4}$$

where the parameter α is empirically configured to 0.1 in the LeakyReLU activation function and *Dense* indicates a fully connection network. Besides, we employ residual connections to enhance the efficiency of feature extraction.

Upon obtaining the hidden state $\mathbf{h}_j^p$ for news v_j, this state serves as the input for both the polarity-free decoder and polarity decoder. From the standpoint of representation learning, $\mathbf{h}_j^p$ can be viewed as the high-level abstract representation of the content of news v_j.

4.1.2 Polarity-free Decoder. A polarity-free decoder extracts the polarity-free information from the news content hidden representation $\mathbf{h}_j^p$. Similar to encoder, the structure of this decoder is also a three-layer network with residual connections. Specifically,

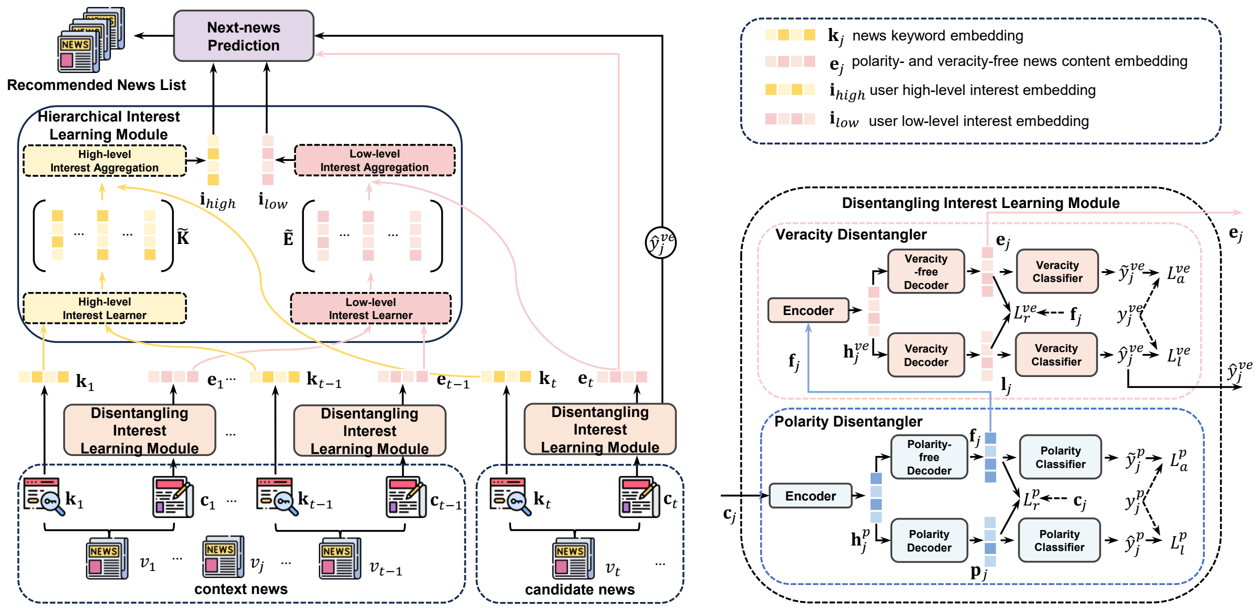
$$\mathbf{d}_j^1 = LeakyReLU(Dense(\mathbf{h}_j^p)), \tag{5}$$

$$\mathbf{d}_j^2 = LeakyReLU(Dense([\mathbf{d}_j^1; \mathbf{h}_j^p])), \tag{6}$$

$$\mathbf{f}_j = LeakyReLU(Dense([\mathbf{d}_j^2; \mathbf{h}_j^p])), \tag{7}$$

where the parameter α in LeakyReLU is empirically set to 0.1. The output $\mathbf{f}_j$ of polarity-free decoder is the latent representation of content(s) without polarity information for news v_j.

4.1.3 Polarity Decoder. The polarity decoder comprises a decoder with an identical structure to that of the polarity-free decoder, alongside a news polarity classifier. This classifier is designed to predict the polarity label of the news article v_j by utilizing the output of the decoder as input. During the training phase, we minimize the discrepancy between the predicted polarity label and the ground-truth label in the training set. This process encourages the polarity decoder to extract polarity-specific information effectively.



(a) The overall structure of HDInt

(b) The detailed structure of disentangling interest learning module

Figure 1: (a) HDInt framework is bulit on three main components: hierarchical interest learning module, disentangling interest learning module and next-news prediction, (b) The disentangling interest learning module is equipped with two serially connected disentanlgers: polarity disentangler and veracity disentangler.

Specifically, with the news hidden representation $\mathbf{h}_j^p$ as input, the decoder performs operations akin to those outlined in Eqs. (5) to (7) to generate the polarity representation $\mathbf{p}_j$ for news v_j. Then, the classifier takes $\mathbf{p}_j$ as the input to predict the polarity label $\hat{y}_j^p$,

$$\hat{y}_j^p = softmax(Dense(\mathbf{p}_j)), \tag{8}$$

where we employ the softmax function here, which is a widely accepted choice for classification task. However, exploring more powerful classifiers remains a potential avenue for future research.

4.1.4 Loss Function Module. In order to effectively train the polarity disentangler, we have devised three specific loss functions for its optimization: *reconstruction loss*, *label prediction loss*, *adversarial loss*. Throughout the training phase, these losses collectively contribute to the overall loss. The reconstruction loss is a fundamental component inherent to any auto-encoder structure. Here, we concatenate the disentangled polarity-free content representation $\mathbf{f}_j$ and the polarity representation $\mathbf{p}_j$ to facilitate the reconstruction of the initial news content embedding $\mathbf{c}_j$. This ensures no information loss during the disentanglement process. Formally,

$$L_r^p = \frac{1}{2}\left([\mathbf{f}_j; \mathbf{p}_j] - \mathbf{c}_j\right)^2. \tag{9}$$

The polarity label prediction loss aims at ensuring that the disentangled polarity representation $\mathbf{p}_j$ accurately reflects the polarity information (left-leaning, neutral or right-leaning) of the news content. In this context, we employ the widely accepted binary cross-entropy loss [45] to quantify this loss for each news item v_j:

$$L_l^p = -y_j^p \log(\hat{y}_j^p), \tag{10}$$

y_j^p and $\hat{y}_j^p$ are the true polarity label and predicted one respectively.

The adversarial loss plays a pivotal role in ensuring that the disentangled polarity-free content representation remains devoid of any polarity-related information. To attain this goal, we utilize the disentangled polarity-free content representation $\mathbf{f}_j$ of news v_j to predict the polarity label $\tilde{y}_j^p$ and subsequently aim to maximize the prediction error (or minimize the reciprocal of the error),

$$L_a^p = -\frac{1}{y_j^p \log(\tilde{y}_j^p)}, \tag{11}$$

where $\tilde{y}_j^p = softmax(Dense(\mathbf{f}_j))$.

Finally, the overall loss of the polarity disentangler is the sum of all the aforementioned three losses:

$$L_p = L_r^p + L_l^p + L_a^p. \tag{12}$$

In a similar vein, by adhering to the framework illustrated in Figure 1(b), the overall loss of the veracity disentangler can be calculated:

$$L_{ve} = L_r^{ve} + L_l^{ve} + L_a^{ve}. \tag{13}$$

4.2 Hierarchical Interest Learning Module

The hierarchical interest learning module consists of two components: (1) the *high- and low-level interest learners* for assimilating the user's reading interests. One takes news keyword embedding as input to learn high-level interests towards certain news events/topics, while the other takes polarity- and veracity-free content embedding for learning unbiased and true news-oriented low level interests towards news articles for certain events/topics; and (2) the *a high- and low-level interest aggregation module* to effectively aggregate both levels of interests while well modeling the intricate relationships between the context news and the candidate news.

4.2.1 High-level & Low-level Interest Learner. Given a piece of news v_j, its three keywords are extracted and their embeddings $\mathbf{k}_j^1, \mathbf{k}_j^2, \mathbf{k}_j^3$ are pre-trained using BERT aforementioned in 4.1.1. Then, we input the concatenation of them into a fully connected layer:

$$\mathbf{k}_j = FC([\mathbf{k}_j^1; \mathbf{k}_j^2; \mathbf{k}_j^3]), \tag{14}$$

where $\mathbf{k}_j$ is a column vector. Once $\mathbf{k}_j$ is ready, the keyword embedding matrix $\mathbf{K} = [\mathbf{k}_1, \mathbf{k}_2, ..., \mathbf{k}_{t-1}]$ of the context can be fed into the high-level interest learner, specified as a one-layer transformer:

$$\tilde{\mathbf{K}} = Transformer(\mathbf{K}), \tag{15}$$

where $\tilde{\mathbf{K}} \in \mathbb{R}^{\dim(\tilde{\mathbf{k}}_j)\times(t-1)}$. Similarly, all the polarity- and veracity-free content embeddings, denoted as $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{t-1}]$, where $dim(\mathbf{e}_j) = dim(\mathbf{k}_j)$, can be seamlessly integrated into the low-level interest learner, which is also specified as a one-layer transformer:

$$\tilde{\mathbf{E}} = Transformer(\mathbf{E}), \tag{16}$$

where $\tilde{\mathbf{E}} \in \mathbb{R}^{\dim(\tilde{\mathbf{e}}_j)\times(t-1)}$ and these two transformers share an identical structure.

4.2.2 High-level & Low-level Interest Aggregation. In order to generate a unified and informative interest embedding for next news prediction, we aggregate interests learned from each interest learner with the guidance of candidate news information. Concretely, we utilize the keyword representation $\mathbf{k}_t$ and disentangled polarity- and veracity-free content representation $\mathbf{e}_t$ of each candidate news v_t as aggregation cues. Accordingly, a specialized attention mechanism is devised to selectively extract relevant information at each level to form the high- and low-level interest representations $\mathbf{i}_{high}$ and $\mathbf{i}_{low}$ respectively:

$$\mathbf{i}_{high} = \sum_{k=1}^{t-1} \alpha_k \tilde{\mathbf{k}}_k, \mathbf{i}_{low} = \sum_{k=1}^{t-1} \beta_k \tilde{\mathbf{e}}_k, \tag{17}$$

$$\alpha_k = softmax(\mathbf{k}_t^{\mathrm{T}} \mathbf{W}_1 \tilde{\mathbf{k}}_k), \beta_k = softmax(\mathbf{e}_t^{\mathrm{T}} \mathbf{W}_2 \tilde{\mathbf{e}}_k), \tag{18}$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are trainable parameters.

4.3 Next-news Prediction Module

In addition to the news context C, user ID information is also incorporated as an input to enhance the personalization of recommendations. Therefore, we combine the high- and low-level interests learned from context news and user ID together to build an informative user embedding with a dense layer:

$$\mathbf{u} = FC([\mathbf{i}_{high}; \mathbf{i}_{low}; \mathbf{u}^{id}]). \tag{19}$$

Once $\mathbf{u}$ is prepared, it is fed into the next-news prediction module to forecast the next news item for user u. First, the polarity- and veracity-free representation $\mathbf{e}_t$ and the veracity representation $\mathbf{l}_t$ are disentangled for each candidate news article v_t, a task handled by the disentangling interest learning module. Subsequently, a standard inner product operation is executed between $\mathbf{u}$ and $\mathbf{e}_t$ to compute a score that quantifies the relevance between the given context C_u and v_t. Finally, a sigmoid function is employed to transform the computed score into a probability value, which signifies the likelihood that the news article v_t will be appreciated by u:

$$p_t = sigmoid(\mathbf{u}^{\mathrm{T}} \cdot \mathbf{e}_t). \tag{20}$$

Simultaneously, aiming at recommending true news only to users, the veracity label $\hat{y}_t^{ve}$ for each candidate news v_t is predicted via the veracity classifier with $\mathbf{l}_t$ as the input. As a result, from the predicted true candidate news, the top-K news items with the highest probabilities are selected as the recommendation list.

4.4 Model Optimization and Training

4.4.1 The Loss Function. As we are jointly optimizing multiple tasks, the total loss L comprises three components: (1) the loss arising from the next-item prediction task, denoted as L_{next}, (2) the disentangling losses incurred during the processing of context news, specifically L_p^{con} and L_{ve}^{con}, (3) the disentangling losses incurred during the processing of candidate news, namely L_p^{can} and L_{ve}^{can}.

$$L = \lambda L_{next} + \left(\gamma L_p^{con} + L_{ve}^{con}\right) + \left(\gamma L_p^{can} + L_{ve}^{can}\right), \tag{21}$$

where λ and γ are hyper-parameters.

The calculation of the loss terms L^{con} and L^{can} has been detailed in Section 4.1.4. Therefore, we only focus on explaining how to compute L_{next}. For the task of next-item prediction, we employ the commonly used cross-entropy loss. Specifically, given a news context C, we create a contrastive pair $\langle s^+, S^- \rangle$. Here, s^+ represents the ground truth next news item, denoted as v_s, serving as the positive sample. Additionally, we randomly select n news articles from the news set $\mathcal{V} \setminus v_s$ to form the negative sample set S^-. The loss associated with the positive sample and each negative sample is computed as follows: $-\log(p_s^+)$ for the positive sample and $-\log(1-p_{s^-})$ for each negative sample, where p_s^+ represents the predicted probability (as defined in Eq (20)). Consequently, the loss for one contrastive pair is calculated as:

$$L_{next} = -[\log(p_{s^+}) + \sum_{s^- \in S^-} \log(1 - p_{s^-})]. \tag{22}$$

4.4.2 Model Training. Our model is implemented using TensorFlow 1.14 within a Python 3.6 environment. Model parameters are learned through the minimization of the total loss denoted as L, employing a mini-batch learning procedure. We utilize the Adam optimizer [11] for gradient-based learning. The initial learning rate is determined through a grid search, exploring values within the range [0.0006, 0.0014] with a step size of 0.0002, and it is empirically set to 0.0001. The batch size is also empirically set to 64, and the number of negative samples for training is configured to 4. All model parameters are fine-tuned using the validation set. Our experiments are conducted on a cluster, where each node is equipped with a 32-core CPU running at 2.0GHz and 128GB of RAM.

5 EXPERIMENTS AND EVALUATIONS

5.1 Data Preparation

In our particular setting, we need user-news interaction data, as well as news political polarity and veracity information for accurate, unbiased and true news recommendations. To the best of our knowledge, there is not such an existing public dataset. Existing public datasets for news recommendations do not contain news veracity information while datasets for fake news detection usually lack comprehensive information regarding reading histories of individual users. Fortunately, we found user-news interactions and news veracity information can be extracted from a commonly used dataset FakeNewsNet[1] [24]. In addition, we retrieved the political polarity of each news item according to its source from a commonly used media bias/fact check website[2]. Therefore, we built

[1]https://github.com/KaiDMML/FakeNewsNet
[2]https://mediabiasfactcheck.com/left/

Table 1: The characteristics of experimental datasets

Statistics	PolitiFact	GossipCop
#Users	37,873	22,540
#True news	306	6,792
#Fake news	310	2,737
#User-news interactions	150,350	646,154
#Training instance	38,062	108,802
#Test instance	4,701	13,601
#Validation instance	4,701	13,601
Polarity proportion (left:right)	55%:45%	66%:34%

our dataset based on such information. FakeNewsNet dataset comprises two extensive datasets, namely PolitiFact and GossipCop. In each dataset, for each user, we built their reading sequence $\mathcal{S}$ by arranging all news items (may contain true or/and fake news) consumed by them in chronological order according to their respective reading timestamps. The aggregation of all such sequences constitutes the user-news interaction dataset $\mathcal{D}$. Concurrently, for each news item, we extracted its content information, comprising the title and a concise one-paragraph description, keywords, together with news political polarity and veracity information.

Once users' reading sequences are built, we construct sequence instances for training and testing using each user u's news sequence $\mathcal{S}$. Each instance is a context-target news pair $\langle C, v_t \rangle$, where $C = \{v_1, \cdots, v_{t-1}\}$ is the reading history as the input, and v_t is the target news item to be predicted. For better personalising, the user ID is also taken as part of the input. To counteract recommending fake news, we retain only those instances with a ground-truth true target news piece in experiments. We apply the commonly adopted sliding window technique [25] and padding and masking techniques [4] to process those sequences longer or shorter than the fixed length t respectively. In line with typical practices [32], we maintain a fixed value of $t = 5$ in our work. Subsequently, we create three distinct test sets by randomly selecting 10%, 20% and 30% of the constructed instances. Similarly, we allocate 10% as the corresponding validation set, the remainder for the training set. For each split ratio, we conduct 10-fold cross-validation and report the average results. Notably, our method consistently outperforms all baseline approaches across all three splits. Due to space constraints, we present results exclusively for the 10% split. Detailed characteristics of the experimental datasets are provided in Table 1.

5.2 Experiment Settings

5.2.1 Baseline Methods. Our task is to recommend the next news article, similar to sequential recommendations [22]. In addition, we aim at reducing bias and fake news. To this end, we select two classes of methods as baselines: (1) representative and state-of-the-art sequential RSs [34], encompassing both general (item) sequential RSs and those tailored for next news recommendations; and (2) representative and state-of-the-art RSs for unbiased news recommendations or for mitigating fake news. In total, the selected 11 baseline methods are grounded in various modeling paradigms, including nearest neighbor models, graph neural networks (GNNs), recurrent neural networks (RNNs), convolutional neural networks (CNNs), and attention models, etc. The details are as follows:

- **SKNN**: a sequence and time aware neighborhood model for next-item recommendation [7].
- **SR-GNN**: a representative GNN based model for next-item recommendation [41].
- **NRMS**: a multi-head self-attention based model for news recommendation according to each user's reading history [38].
- **LSTUR**: an RNN and attention based model for next-news recommendation. It learns users' long and short term preferences [1].
- **FedRec**: a decentralized next-news recommendation model based on RNN, CNN and attention mechanism [19].
- **FIM**: a 3D CNN-based next-news recommendation model which preforms fine-grained interest matching [27].
- **ESM**: a news RS that effectively combines event matching and style matching for accurate news recommendations [18].
- **Glory**: an advanced news recommender which combines global representations learned from other users with local representations to enhance personalized news recommendations [42].
- **Rec4Mit**: a personalized news recommendation model that takes into account news veracity to mitigate the spread of fake news [33].
- **SentiRec**: a sentiment diversity-aware neural news RS, which can recommend news with more diverse sentiment [39].
- **ProFairRec**: a news RS focusing on provider fairness to improve the fairness and diversity of recommendations [20].

5.2.2 Evaluation Metrics. In light of our approach's objective of recommending accurate, unbiased and true news to end users, we assess the performance of all the approaches from three distinct angles: *(1) recommendation accuracy*, gauging the ability to precisely recommend news articles that align with a user's reading interests; *(2) recommendation fairness*, evaluating the approach's capacity to mitigate political bias within the recommended news list; and *(3) the ratio of true news* included in the recommendation list, indicating the approach's effectiveness in avoiding recommending fake news. In accordance with established conventions [1], we employ two widely recognized ranking-based metrics: recall (REC) and normalized discounted cumulative gain (NDCG) [36] for measuring recommendation accuracy, commonly used for assessing next-news recommendation performance [19, 27]. They are computed based on the top K ranked news, denoted as REC@K and NDCG@K, respectively. Concerning recommendation fairness, we introduce polarity proportion error (PE@K) and fairness score (FS@K):

$$PE@K = 1 - \left|\frac{\#left\ news}{K} - \frac{\#right\ news}{K}\right|, \tag{23}$$

$$FS@K = \frac{-1 \times (\#left\ news) + 1 \times (\#right\ news)}{K}, \tag{24}$$

where −1, 0 and 1 indicate the polarity label left-leaning, neutral and right-leaning respectively. Hence, the more approaching 0 the FS, the more unbiased the recommendation list.

Drawing inspiration from the conventional F1-score, we design a new metric, i.e., a redefined F1-score@K (F1@K), to aggregate NDCG@K and PE@K into a unified metric:

$$F1@K = \frac{2 \times NDCG@K \times PE@K}{NDCG@K + PE@K}, \tag{25}$$

Moreover, we introduce another metric, the ratio of true news (denoted as RT@K) [33] for each recommendation list, for evaluating the capability of recommending true news:

$$RT@K = \frac{\#true\ news}{K}. \tag{26}$$

Table 2: Comparison of prediction performances with baselines on two datasets, *the improvement is significant at $p < 0.05$.

Dataset	Metric	SKNN	SR-GNN	NRMS	LSTUR	FedRec	FIM	ESM	Glory	Rec4Mit	SentiRec	ProFairRec	HDInt	Improvement
PolitiFact	REC@5	0.2183	0.3729	0.4712	0.4725	0.3731	0.3606	0.4608	0.4906	0.4663	0.4156	0.5312	**0.5781***	8.83%
	REC@20	0.6086	0.6703	0.8234	0.8056	0.7715	0.7002	0.7964	0.8182	0.7883	0.7350	0.8454	**0.8594***	1.66%
	NDCG@5	0.1417	0.2956	0.3386	0.3454	0.2547	0.2279	0.3232	0.3527	0.3323	0.3211	0.3598	**0.3675***	2.14%
	NDCG@20	0.2524	0.3971	0.4392	0.4403	0.3528	0.3221	0.4185	**0.4472***	0.4261	0.4237	0.4436	0.4459	-0.29%
	PE@5	0.8438	0.7476	0.5526	0.5818	0.7060	0.6294	0.7082	0.5560	0.6508	0.6054	0.7534	**0.8546***	1.28%
	PE@20	0.8320	0.6956	0.6440	0.6502	0.6982	0.6830	0.7116	0.6524	0.7192	0.7834	0.8450	**0.8656***	2.44%
	FS@5	-0.1469	0.2157	0.3449	0.3347	0.2569	0.3261	0.2392	0.3439	-0.3061	-0.3625	-0.2126	**-0.1273***	15.40%
	FS@20	-0.1491	0.2589	0.2892	0.2860	0.2591	0.2723	0.2384	0.2826	-0.2441	-0.1853	-0.1320	**-0.1211***	9.00%
	F1@5	0.2427	0.4237	0.4199	0.4335	0.3743	0.3346	0.4438	0.4316	0.4400	0.4196	0.4870	**0.5140***	5.54%
	F1@20	0.3873	0.5056	0.5222	0.5250	0.4687	0.4378	0.5270	0.5306	0.5351	0.5500	0.5818	**0.5886***	1.17%
GossipCop	REC@5	0.1698	0.4889	0.6297	0.66	0.2174	0.3661	0.6745	0.6630	0.6931	0.4062	0.663	**0.7188***	3.71%
	REC@20	0.6249	0.6253	0.8158	0.8545	0.4664	0.6135	0.8922	0.8600	0.9341	0.7344	0.9148	**0.9531***	2.03%
	NDCG@5	0.0703	0.4215	0.4953	0.5293	0.1449	0.2665	0.4902	**0.5508***	0.4957	0.3076	0.4845	0.5406	-1.85%
	NDCG@20	0.2073	0.5602	0.5583	0.5560	0.2067	0.3337	0.5894	0.6007	0.5675	0.3995	0.5585	**0.6103***	1.60%
	PE@5	0.5508	0.6256	0.5752	0.5522	0.6138	0.6316	0.6282	0.5468	0.5470	0.6720	0.6566	**0.6968***	3.69%
	PE@20	0.5484	0.6142	0.6286	0.6246	0.6444	0.6468	0.6386	0.6196	0.6194	0.6784	0.7324	**0.7358***	0.46%
	FS@5	-0.3980	-0.3287	-0.3480	-0.3690	-0.3115	-0.2965	-0.2968	-0.3737	-0.4032	-0.2901	-0.2986	**-0.2231***	30.03%
	FS@20	-0.4029	-0.3390	-0.2982	-0.3022	-0.2851	-0.2830	-0.2879	-0.3063	-0.3335	-0.2837	-0.2290	**-0.2263***	1.19%
	F1@5	0.1247	0.5037	0.5323	0.5405	0.2345	0.3748	0.5507	0.5487	0.5201	0.4220	0.5576	**0.6088***	9.18%
	F1@20	0.3009	0.5860	0.5914	0.5883	0.3130	0.4403	0.6130	0.6100	0.5923	0.5029	0.6337	**0.6672***	5.29%

[1] The improvement is over the best-performing baseline methods whose performance is underlined. For FS@K, negative and positive values indicate left biased and right biased respectively, and the smaller its absolute value, the more fair the recommendations.

For all these four metrics, we report them at $K = 5$ and $K = 20$ respectively in our experiments. Following [29], a paired t-test with p<0.05 is used for significance test.

5.2.3 Parameter Settings. To be fair, we meticulously tune all model parameters, including hyper-parameters, for both baseline methods and our proposed approach on a validation dataset. Specifically, for each baseline method, we initialize parameters using values from the original paper and fine-tune them on our datasets to optimize performance. The embedding dimensions for user ID and news ID are set to 32 and 128, respectively, using a grid search with increments of 4 and 16, respectively, across all methods. Further details about crucial parameters for each baseline will be provided in Section 5.3 during the experimental result analysis. In our model, we empirically assign a weight of 1.3 and 10 to the prediction loss for the next news (λ) on the PolitiFact and GossipCop dataset, respectively, based on validation set performance. The polarity loss weight (γ) is set to 3 for both datasets. Empirically, we use 4 heads in the one-layer transformer architecture for both datasets. We conduct training for 15 epochs on both datasets and use 4 negative samples (comprising 2 true news pieces and 2 fake news pieces) during training.

5.3 Performance Comparison with Baselines

5.3.1 Comparisons w.r.t. Recommendation Accuracy and Fairness. In order to achieve the best performance, we carefully tune parameters for each baseline on the validation set. For SKNN, we set the number of neighborhoods to 500. In SR-GNN, we adjust the decay ratio of the learning rate to 0.1 and introduce an L2 penalty of 10^{-5}. NRMS and FedNewsRec both feature 14 heads for self-attention, while LSTUR utilizes 300 filters in the CNN with a filter window size of 3. ESM employs 4 GCN layers and 10 clusters. In Glory, the number of most recently clicked news article of a user is set to 50, the maximum title length is set to 30 and the number of maximum used entities of each news item is set to 5. In Rec4Mit, the number of latent events is set to 10 and 20 on Political and Gossip dataset respectively. SentiRec adjusts two weights within the total loss function, setting them to 0.5 and 15. In ProFairRec, the provider encoder is considered as the polarity encoder, and the weight for the adversarial loss is set to 0.04.

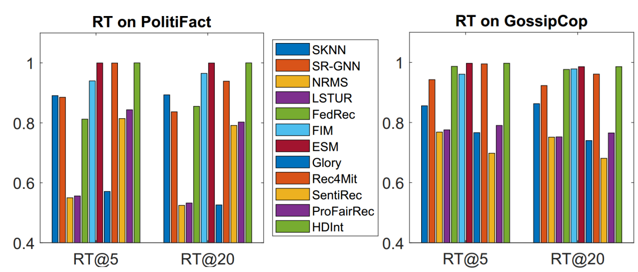


Figure 2: The ratio of true news (RT) in recommendation lists.

The comparison results on accuracy and fairness metrics are presented in Table 2. It is evident that representative next-item recommendation approaches, such as SKNN and SR-GNN, exhibit lower accuracy compared to prominent sequential news recommendation methods like NRMS, LSTUR, ESM and Glory. This emphasizes the superiority of sequential news RSs in effectively handling news data, characterized by distinctive features like news meta-information (e.g., title and description) and the real-time nature of news events. In the context of unbiased news recommendation, SentiRec and ProFairRec stand out as two of the few approaches that strike a better balance between accuracy and fairness, as indicated by F1@K when compared with other accuracy-focused baselines. In contrast, our HDInt not only captures users' high- and low-level reading interests but also disentangles political polarity information from news contents and thus mitigates political bias. This comprehensive approach results in superior performance w.r.t. both accuracy and fairness. HDInt generally outperforms the best-performing baseline(s) by an average of 2.23% across accuracy metrics, with improvements up to 8.83%, and an average of 7.94% across fairness metrics, with improvements up to 30.03% (cf. the last column in Table 2).

5.3.2 Comparisons w.r.t. the Ratio of True News. In Figure 2, we illustrate the ratio of true news (Eq. (26)) within the recommendation lists generated by all compared approaches. Our HDInt model consistently outperforms all baseline methods regarding RT@5 and RT@20 on both datasets. This superiority can be attributed to the fact that most of the baseline approaches ignore the existence of fake news, and thus generate recommendations based on interest matching only, leading to recommendations of both true and fake news. Notably, Rec4Mit demonstrates the second-best performance w.r.t. RT, achieving an impressive RT@5 (resp. RT@20) of 99.85% (resp. 99.76%) on PolitiFact, and 99.59% (resp. 98.31%) on GossipCop respectively. This is attributed to its specific design for recommending true news. In contrast, our HDInt approach incorporates a disentangling interest learning module, which not only mitigates political polarity bias but also classifies each candidate news piece into true or fake ones. Consequently, HDInt can recommend true news to users only. Thanks to the high accuracy of the classifier, it achieves an RT@5 and RT@20 of 100% on PolitiFact and 99.72% and 98.59% on GossipCop, respectively.

5.4 Ablation Study

5.4.1 Settings. To analyze the rationality and the effectiveness of the designed components in our HDInt framework, we conduct an ablation study to compare HDInt with its three variants: (1) **HDInt-KW**: which ignores the high-level interest by removing the high-level interest learner; (2) **HDInt-P**: which removes the polarity disentangler by directly taking the initial news content embedding $\mathbf{c}_j$ as the input of the veracity disentangler; (3) **HDInt-V**: which removes the veracity disentangler by directly replacing its output $\mathbf{e}_j$ with the output of polarity disentangler $\mathbf{f}_j$.

Table 3: Performance comparison of HDInt with its variants

Dataset	Metric	HDInt	HDInt-KW	HDInt-P	HDInt-V
PolitiFact	REC@5	0.5781	0.4035	0.5625	0.2969
	REC@20	0.8594	0.8288	0.8750	0.8281
	NDCG@5	0.3675	0.2384	0.4032	0.1692
	NDCG@20	0.4459	0.3630	0.4967	0.3275
	PE@5	0.8546	0.7720	0.5166	0.8514
	PE@20	0.8656	0.7866	0.5614	0.8370
	FS@5	-0.1273	-0.1962	-0.4120	-0.1406
	FS@20	-0.1211	-0.1827	-0.3698	-0.1422
	RT@5	1.000	1.000	1.000	0.7992
	RT@20	1.000	1.000	1.000	0.7428
GossipCop	REC@5	0.7188	0.4906	0.7656	0.6538
	REC@20	0.9531	0.8344	0.9062	0.8981
	NDCG@5	0.5406	0.3560	0.5876	0.4729
	NDCG@20	0.6103	0.4489	0.6293	0.5444
	PE@5	0.6968	0.4236	0.5400	0.6696
	PE@20	0.7358	0.3836	0.5636	0.7314
	FS@5	-0.2231	-0.5188	-0.4313	-0.2872
	FS@20	-0.2263	-0.5422	-0.3929	-0.2301
	RT@5	0.9972	0.9891	0.9996	0.7867
	RT@20	0.9859	0.9778	0.9876	0.7763

5.4.2 Findings. The results presented in Table 3 demonstrate the following findings: (1) *The high-level interest learner benefit the next-news recommendations.* We note a notable numerical decline in NDCG@K when removing high-level interest learner. This phenomenon underscores the critical role of high-level interests; (2) *The polarity disentangler can boost the recommendation fairness.* Comparing the values in the fifth column in Table 3, we observe that in the absence of a polarity disentangler, the fairness metrics PE@K and FS@K experience a sharp decline; (3) *The veracity disentangler not only greatly reduces the recommendations of fake news, but also improves recommendation accuracy.* It is evident that the absence of the veracity disentangler leads to a significant numerical decline in RT@K as well as REC@K and NDCG@K.

5.5 Parameter Test

We assess the sensitivity of HDInt on three critical parameters: the weight assigned to predicting the next news loss, i.e., λ, the weight attributed to polarity loss, i.e., γ (cf. Section 4.4.1), and the number of heads in transformers (cf. Section 4.2.1).

5.5.1 Performance w.r.t. the weight of loss for predicting next news (λ). We systematically vary the values of λ across two datasets. For the PolitiFact (resp. GossipCop) dataset, we range λ from 1.1 (resp. 6) to 1.9 (resp. 14) with an increment step of 0.2 (resp. 2). The results presented in Figure 3 show the optimal performance is attained when λ is set to 1.3 and 10 for PolitiFact and GossipCop, respectively. A smaller value fails to provide accurate recommendations, while larger values compromise recommendation fairness.

5.5.2 Performance w.r.t. the weight of polarity loss (γ). We vary the values of γ on both datasets, ranging from 1 to 9 with a step size of 2. Figure 4 presents the results. Our findings indicate that an optimal value of 3, employed on both datasets, yields the best performance. This weight strikes a balance by providing an informative representation for each piece of news. A smaller value fails to capture the rich information from both news occurrence patterns and meta information. Conversely, a larger weight over emphasizes the fairness and thus reduces recommendation accuracy.

5.5.3 Performance w.r.t. the number of heads in transformers within interest learners. We systematically varied the number of heads in both transformers across both the PolitiFact and GossipCop datasets, ranging from 2 to 32 with a step size of 2^n. The results for both datasets are depicted in Figure 5. Our findings indicate that an optimal value, specifically 4 for both datasets, delivers the best performance by providing informative representations for keywords and news contents. Smaller values for the number of heads fall short in capturing the rich information encompassed by both news keywords and news contents. Conversely, larger values introduce unnecessary model parameters, ultimately impairing performance.

5.6 Computation Efficiency Analysis

5.6.1 Complexity Analysis . The computational load of our proposed HDInt model primarily stems from two components: (1) the Disentangling Interest Learning Module, and (2) the Hierarchical Interest Learning Module. On one hand, within the Disentangling Interest Learning Module, the computational resources are consumed

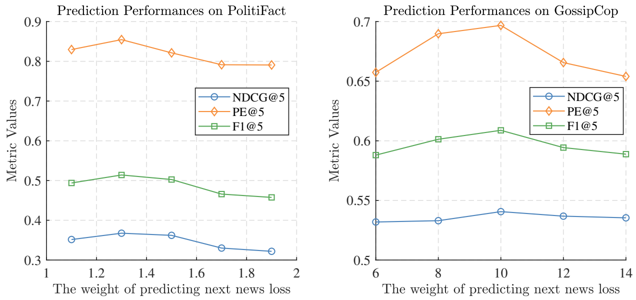


Figure 3: The impact of weight of predicting next news loss.

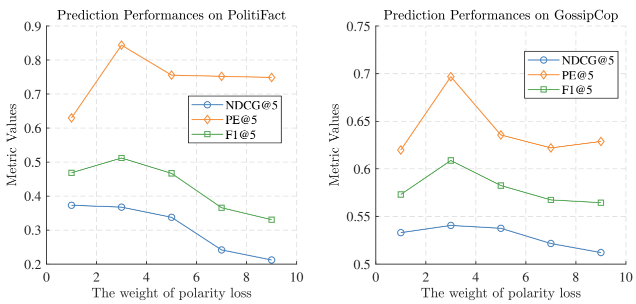


Figure 4: The impact of weight of polarity loss.

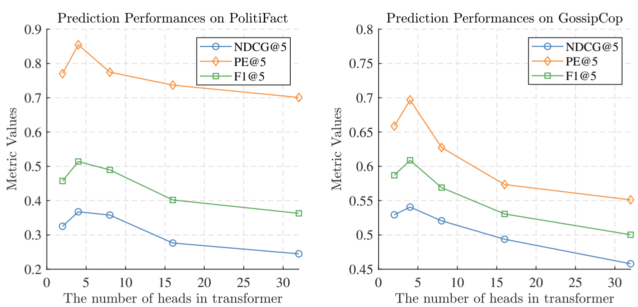


Figure 5: The impact of the number of heads in transformers on the prediction performance.

by two encoder-decoder architectures (consisting of 2 encoders and 4 decoders, each composed of three layers of fully connected neural networks with neuron counts h_1, h_2, and h_3 respectively) and 4 classifiers. The time complexity of encoder-decoder architectures is $O(6 \cdot (d \cdot h_1 + h_1 \cdot h_2))$ and that of classifiers is $O(2 \cdot (2d) + 2 \cdot (3d))$, where d is the input dimension. On the other hand, in the Hierarchical Interest Learning Module, the computational cost primarily arises from 2 Interest Learners (each implemented with a transformer layer) and 2 Interest Aggregation operations, with time complexities of $O(2 \cdot (T^2 + 2Td))$ and $O(2d^2)$, respectively, where T is the length of news sequence read by users. In summary, for each user, the overall model's time complexity approximates cubic complexity.

5.6.2 *Inference Time Comparison.*

We compare the average inference time per user of HDInt with three representative and state-of-the-art baseline methods on the two experiment datasets. Our experiments are ran on a cluster node with an 8-core CPU running at 3.70GHz and 64GB of RAM. As shown in Table 4, the average inference time of HDInt is shorter on test sets of both datasets than ProFairRec but slightly longer than that of ESM and Rec4Mit. This result is acceptable, considering that HDInt can recommend accurate, true and fair news to users, whereas the other three methods can achieve at most two of these goals (cf. Section 5.3).

Table 4: Comparison of the average inference time (second)

Dataset	ESM	Rec4Mit	ProFairRec	HDInt
PolitiFact	1.14	1.19	1.33	**1.21**
GossipCop	1.15	1.18	1.36	**1.22**

5.7 Case Study

To illustrate the practical impact of HDInt, on recommending unbiased and true news, we conducted a case study on the GossipCop dataset, which is presented in Section A in the appendix.

6 CONCLUSIONS

In this paper, we focused on a novel and significant research problem: how to effectively recommend accurate, unbiased and true news in the real-world complex context? To address this problem, we have proposed a novel hierarchical and disentangling interest learning framework (HDInt). Owning to the special design, HDInt is able to not only learn users' high-level and low-level interests towards news events and news articles respectively for improving recommendation accuracy, but also learn unbiased and true news oriented reading interests to reduce political bias and fake news in recommendation lists. Extensive experiments on real-world datasets demonstrated the superiority of HDInt over the state-of-the-art news recommendation methods and verified the rationality and effectiveness of its specific design. Future work includes the exploration of more effective methods to disentangle political bias and news veracity information from news contents for further enhancing the fairness and truth of recommendations.

ACKNOWLEDGMENTS

This work was partially supported by Australian Research Council Discovery Projects DP200101441, DP230100676 and DP240100955 and 2023 UTS Key Technology Partnerships Seed Funding Scheme.

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 57th Annual Meeting of the Association for Computational Linguistics*. 336–345.
[2] Jennifer L Bonnet and Judith E Rosenbaum. 2020. "Fake News", Misinformation, and Political Bias: Teaching News Literacy in the 21st Century. *Communication Teacher* 34, 2 (2020), 103–108.
[3] Michel Capelle, Flavius Frasincar, Marnix Moerland, and Frederik Hogenboom. 2012. Semantics-based News Recommendation. In *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*. 1–9.

[4] John E Collins and et al. 2012. Incorporating RNA-seq Data into the Zebrafish Ensembl Genebuild. *Genome Research* 22, 10 (2012), 2067–2078.

[5] Philip J Feng, Pingjun Pan, Tingting Zhou, Hongxiang Chen, and Chuanjiang Luo. 2021. Zero Shot on the Cold-start Problem: Model-agnostic Interest Learning for Recommender Systems. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*. 474–483.

[6] Florent Garcin, Kai Zhou, Boi Faltings, and Vincent Schickel. 2012. Personalized News Recommendation based on Collaborative Filtering. In *Proceedings of the 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*. 437–441.

[7] Diksha Garg, Priyanka Gupta, Pankaj Malhotra, Lovekesh Vig, and Gautam Shroff. 2019. Sequence and Time Aware Neighborhood for Session-based Recommendations: STAN. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1069–1072.

[8] Linmei Hu, Siyong Xu, Chen Li, Cheng Yang, Chuan Shi, Nan Duan, Xing Xie, and Ming Zhou. 2020. Graph Neural News Recommendation with Unsupervised Preference Disentanglement. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 4255–4264.

[9] Qinglin Jia, Jingjie Li, Qi Zhang, Xiuqiang He, and Jieming Zhu. 2021. RMBERT: News Recommendation via Recurrent Reasoning Memory Network over BERT. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1773–1777.

[10] Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 4171–4186.

[11] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *Proceedings of the 3rd International Conference on Learning Representations*. 1–15.

[12] Jin Li, Shoujin Wang, Qi Zhang, Longbing Cao, Fang Chen, Xiuzhen Zhang, Dietmar Jannach, and Charu C Aggarwal. 2024. Causal Learning for Trustworthy Recommender Systems: A Survey. *arXiv preprint arXiv:2402.08241* (2024).

[13] Ping Liu, Karthik Shivaram, Aron Culotta, Matthew A Shapiro, and Mustafa Bilgic. 2021. The Interaction between Political Typology and Filter Bubbles in News Recommendation Algorithms. In *Proceedings of the Web Conference 2021*. 3791–3801.

[14] Wenpeng Lu, Rongyao Wang, Shoujin Wang, Xueping Peng, Hao Wu, and Qian Zhang. 2022. Aspect-driven User Preference and News Representation Learning for News Recommendation. *IEEE Transactions on Intelligent Transportation Systems* 23, 12 (2022), 25297–25307.

[15] Guangyuan Ma, Hongtao Liu, W Xing, Wanhui Qian, Zhepeng Lv, Qing Yang, and Songlin Hu. 2023. PUNR: Pre-training with User Behavior Modeling for News Recommendation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*. 8338–8347.

[16] Zhiming Mao, Jian Li, Hongru Wang, Xingshan Zeng, and Kam-Fai Wong. 2022. DIGAT: Modeling News Recommendation with Dual-Graph Interaction. In *Findings of the Association for Computational Linguistics: EMNLP*. 6595–6607.

[17] Shumpei Okura, Yukihiro Tagami, Shingo Ono, and Akira Tajima. 2017. Embedding-based News Recommendation for Millions of Users. In *Proceedings of the 23rd SIGKDD Conference on Knowledge Discovery and Data Mining*. 1933–1942.

[18] Zhao Pengyu, Wang Shoujin, Lu Wenpeng, Peng Xueping, Zhang Weiyu, Zheng Chaoqun, and Huang Yonggang. 2023. News Recommendation via Jointly Modeling Event Matching and Style Matching. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. 404–419.

[19] Tao Qi, Fangzhao Wu, Chuhan Wu, Yongfeng Huang, and Xing Xie. 2020. Privacy-preserving News Recommendation Model Learning. In *Findings of the Association for Computational Linguistics: EMNLP*. 1423–1432.

[20] Tao Qi, Fangzhao Wu, Chuhan Wu, Peijie Sun, Le Wu, Xiting Wang, Yongfeng Huang, and Xing Xie. 2022. ProFairRec: Provider Fairness-aware News Recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1164–1173.

[21] Shaina Raza and Chen Ding. 2022. News Recommender System: A Review of Recent Progress, Challenges, and Opportunities. *Artificial Intelligence Review* 55, 1 (2022), 749–800.

[22] Heng-Shiou Sheu, Zhixuan Chu, Daiqing Qi, and Sheng Li. 2022. Knowledge-Guided Article Embedding Refinement for Session-Based News Recommendation. *IEEE Transactions on Neural Networks and Learning Systems* 33, 12 (2022), 7921–7927.

[23] Karthik Shivaram, Ping Liu, Matthew Shapiro, Mustafa Bilgic, and Aron Culotta. 2022. Reducing Cross-Topic Political Homogenization in Content-Based News Recommendation. In *Proceedings of the 16th ACM Conference on Recommender Systems*. 220–228.

[24] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. *Big Data* 8, 3 (2020), 171–188.

[25] Syed Khairuzzaman Tanbeer, Chowdhury Farhan Ahmed, and et al. 2009. Sliding Window-based Frequent Pattern Mining over Data Streams. *Information Sciences* 179, 22 (2009), 3843–3865.

[26] Nguyen Vo and Kyumin Lee. 2018. The Rise of Guardians: Fact-checking URL Recommendation to Combat Fake News. In *The 41st International ACM SIGIR Conference on Research and Development in Information Rtrieval*. 275–284.

[27] Heyuan Wang, Fangzhao Wu, Zheng Liu, and Xing Xie. 2020. Fine-grained Interest Matching for Neural News Recommendation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 836–845.

[28] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. DKN: Deep Knowledge-Aware Network for News Recommendation. In *Proceedings of the 27th International World Wide Web Conference*. 1835–1844.

[29] Meirui Wang, Pengjie Ren, Lei Mei, Zhumin Chen, Jun Ma, and Maarten de Rijke. 2019. A Collaborative Session-based Recommendation Approach with Parallel Memory Modules. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 345–354.

[30] Rongyao Wang, Shoujin Wang, Wenpeng Lu, and Xueping Peng. 2022. News Recommendation via Multi-interest News Sequence Modelling. In *2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 7942–7946.

[31] Rongyao Wang, Shoujin Wang, Wenpeng Lu, Xueping Peng, Weiyu Zhang, Chaoqun Zheng, and Xinxiao Qiao. 2023. Intention-aware User Modeling for Personalized News Recommendation. In *International Conference on Database Systems for Advanced Applications*. Springer, 179–194.

[32] Shoujin Wang, Liang Hu, Yan Wang, Quan Z Sheng, Mehmet Orgun, and Longbing Cao. 2019. Modeling Multi-Purpose Sessions for Next-Item Recommendations via Mixture-Channel Purpose Routing Networks. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. 3771–3777.

[33] Shoujin Wang, Xiaofei Xu, Xiuzhen Zhang, Yan Wang, and Wenzhuo Song. 2022. Veracity-aware and Event-driven Personalized News Recommendation for Fake News Mitigation. In *Proceedings of the ACM Web Conference 2022*. 3673–3684.

[34] Shoujin Wang, Qi Zhang, Liang Hu, Xiuzhen Zhang, Yan Wang, and Charu Aggarwal. 2022. Sequential/session-based Recommendations: Challenges, Approaches, Applications and Opportunities. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3425–3428.

[35] Shoujin Wang, Xiuzhen Zhang, Yan Wang, and Francesco Ricci. 2023. Trustworthy Recommender Systems. *ACM Transactions on Intelligent Systems and Technology* (2023), 1–19. https://doi.org/10.1145/3627826

[36] Markus Weimer, Alexandros Karatzoglou, Quoc Le, and Alex Smola. 2007. Cofirank-maximum Margin Matrix Factorization for Collaborative Ranking. In *Proceedings of the 21st Annual Conference on Neural Information Processing Systems*. 222–230.

[37] 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 28th International Joint Conference on Artificial Intelligence*. 3863–3869.

[38] 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*. 6389–6394.

[39] Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2020. SentiRec: Sentiment Diversity-aware Neural News Recommendation. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*. 44–53.

[40] Chuhan Wu, Fangzhao Wu, Xiting Wang, Yongfeng Huang, and Xing Xie. 2021. Fairness-aware News Recommendation with Decomposed Adversarial Learning. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. 4462–4469.

[41] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-Based Recommendation with Graph Neural Networks. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. 346–353.

[42] Boming Yang, Dairui Liu, Toyotaro Suzumura, Ruihai Dong, and Irene Li. 2023. Going Beyond Local: Global Graph-Enhanced Personalized News Recommendations. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 24–34.

[43] Qi Zhang, Jingjie Li, Qinglin Jia, Chuyuan Wang, Jieming Zhu, Zhaowei Wang, and Xiuqiang He. 2021. UNBERT: User-News Matching BERT for News Recommendation. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*. 3356–3362.

[44] Zeyu Zhang, Heyang Gao, Hao Yang, and Xu Chen. 2023. Hierarchical Invariant Learning for Domain Generalization Recommendation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3470–3479.

[45] Zhilu Zhang and Mert R Sabuncu. 2018. Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels. In *Proceedings of the 32nd Conference on Neural Information Processing Systems*. 1–11.

[46] 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 27th International World Wide Web Conference*. 167–176.

[47] Qiannan Zhu, Xiaofei Zhou, Zeliang Song, Jianlong Tan, and Li Guo. 2019. DAN: Deep Attention Neural Network for News Recommendation. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. 5973–5980.

A CASE STUDY (CONT.)

To illustrate how our proposed framework HDInt contributes to unbiased and true news recommendations in a more straightforward manner, we conduct a case study on one of the real-world datasets, namely the GossipCop dataset, which we use in our experiments. Specifically, we randomly select 5 users from the GossipCop dataset. For each user, we display their current reading history (referred to as 'context news') alongside the corresponding recommendation list generated by our HDInt model.

The results are presented in Table 5, where each user is represented by two rows: the first row displays the context news read sequentially by the user, and the second row showcases the top-5 news pieces recommended to the user based on their context news. From this table, we observe the following findings:

- *Finding 1: Users may read left leaning, right leaning or neutral news for the same/related events, which may be true or fake.* Table 5 reveals noteworthy patterns. For instance, $user_1$ and $user_3$ mainly read left-leaning news, which encompasses both true news and fake news. For $user_4$ and $user_5$, their reading history consists solely of fake news.
- *Finding 2: For the users who have read news with polarity bias for certain interests, our model is able to recommend the corresponding unbiased news list for the same/related news events/topics to counteract the biased news.* For instance, let's consider $user_1$, who has previously read a piece of left biased news related to the topic "Trump and North Korea", denoted as CN_2. In this case, our model recommended a piece of neutral news also pertaining to the same topic, labeled as RN_5. Similar instances can be readily identified for all the other four users presented in the table. This serves as direct confirmation of the effectiveness of our proposed model in mitigating the polarity bias of news.
- *Finding 3: For the users who have read fake news for certain interests, our model is able to recommend the corresponding true news for the same/related topics to counteract the fake news.* Consider $user_3$ as an example. He/She has previously encountered a piece of fake news related to the event "Chris Pratt divorce", referred to as CN_3. In this instance, our model recommends a piece of true news regarding the same topic, labeled as RN_3, and successfully hit the ground truth. Similar scenarios can be readily identified for all the other four users presented in the table. This serves as direct confirmation of the effectiveness of our proposed model in mitigating fake news by recommending corrective true news.

Table 5: Sampled Recommendation Lists for 5 Users on GossipCop Dataset.

User$_1$	**Context news (CN)**	CN$_1$:[1] FKA Twigs reacts to Robert Pattinson and Kristen Stewart's reunion. She is mad. (L)[2]	CN$_2$: CN2: Trump[3] warns North Korea of 'fire and fury'. (L)	CN$_3$:Kylie Jenner Made a Contestant Cry on Khloé Kardashian's New Reality Show "Happy tears (L)	CN$_4$:All the Details on Prince Harry and Meghan Markle's Wedding Cake (N)	
	Recommended news (RN)	RN$_1$:**All Chicago West Baby Photos Timeline (N)**[4]	RN$_2$:Local Couple Celebrates 50th Wedding Anniversary, Inspiring Love and Longevity (R)	RN$_3$:Community Comes Together to Make Dream Wedding a Reality for Terminally Ill Bride (R)	RN$_4$:FKA Twigs Expresses Displeasure Over Robert Pattinson and Kristen Stewart's Reunion (L)	RN$_5$:Trump's Warning to North Korea Sparks Global Concern: 'Fire and Fury' Remark Draws Attention (N)
User$_2$	**Context news (CN)**	CN$_1$: ulineLady Gaga has found a most distasteful way to lose weight for her leading role in A Star Is Bron (L)	CN$_2$:Princess Kate Stars in New Children's Hospice Video with a Little Help from Ed Sheeran (R)	CN$_3$:The Weeknd Drops $20 Million on Hidden Hills Mega Mansion (N)	CN$_4$:That Twitter-Inspired Rihanna-Lupita Heist Movie Is Actually Happening (R)	
	Recommended news (RN)	RN$_1$:The Weeknd Faces Criticism for Lavish $20 Million Mansion Purchase (L)	RN$_2$:**Here Is the Complete List of Winners From the 2017 Billboard Music Awards (N)**	RN$_3$:ulineLady Gaga's Dedication to Her Craft: Transforming for a Leading Role in A Star Is Born (R)	RN$_4$:Billboard Music Awards 2017 Celebrates Musical Excellence: A Look Back at the Winners (R)	RN$_5$:Controversy Surrounds ulineLady Gaga's Extreme Weight Loss Methods for Film Role (L)
User$_3$	**Context news (CN)**	CN$_1$: Hugh Hefner's Cause of Death Revealed (N)	CN$_2$:Justin Bieber fan arrested for trespassing at his Los Angeles home (L)	CN$_3$:Chris Pratt and Anna Faris Finalize Divorce One Year After Separating (L)	CN$_4$:My Big Fat Greek Wedding's Lainie Kazan Denies She Shoplifted (L)	
	Recommended news (RN)	RN$_1$:Lainie Kazan successfully refutes allegations of shoplifting, clearing her name and putting an end to the controversy surrounding the incident (R)	RN$_2$:Lainie Kazan's Shoplifting Controversy Continues to Haunt Her (L)	RN$_3$:**Chris Pratt Files for Divorce From Anna Faris (L)**	RN$_4$:Hugh Hefner's Life and Legacy Remembered Following the Revelation of His Cause of Death (N)	RN$_5$:Chris Pratt and Anna Faris Maintain Cordial Relationship Post-Divorce (R)
User$_4$	**Context news (CN)**	CN$_1$:Selena Gomez Regrets Fighting With Justin Bieber After Cheating Claims Surface (L)	CN$_2$:Katherine Jackson Resigns as Blanket's Guardian (N)	CN$_3$:Kate Winslet, Allison Janney Kiss at Hollywood Film Awards (N)	CN$_4$:CMA Awards 2017: The Complete Winners List 'From Entertainer of the Year to New Artist' (N)	
	Recommended news (RN)	RN$_1$:CMA Awards 2017: A Night of Music and Honoring Excellence (N)	RN$_2$:International Film Awards Celebrate Global Cinema Talent (N)	RN$_3$:**A Detailed History of Selena Gomez and Justin Bieber (N)**	RN$_4$:Selena Gomez and Justin Bieber have reportedly reconciled after a period of tension, signaling a positive turn in their relationship (R)	RN$_5$: 2017 Film Awards Season Kicks Off with Nominations Announcement (N)
User$_5$	**Context news (CN)**	CN$_1$: Is Beyonce's Demanding Silence In The Delivery Room For Births Of Blueprint 1 And Blueprint 2? (N)	CN$_2$:Beyonce's Birth, Twins Still in the Hospital with 'Minor Issue' "Beyonce had a boy and a girl (L)	CN$_3$:11 Things You Didn't Know About Sofia Richie, Justin Bieber's New Girl (N)	CN$_4$: Madonna Files Emergency Order To Stop Tupac Shakur Letter Auction (N)	
	Recommended news (RN)	RN$_1$:**Justin Bieber Cancels Remaining Purpose Tour Dates (L)**	RN$_2$:Beyonce's twins are reportedly making positive progress in their health as they remain in the hospital with a minor issue. (R)	RN$_3$:Sofia Richie and Justin Bieber's Relationship Facts (N)	RN$_4$:Celebrity Romances Continue to Captivate Public Interest (N)	RN$_5$:Beyonce Welcomes Boy and Girl (N)

[1] Green color indicates true news and red color indicates fake news.
[2] (L) represents a left-leaning polarity, (R) represents a right-leaning polarity, and (N) represents a neutral polarity.
[3] _____ is used to highlight the identical/related topics from the context news and the recommended news of each user.
[4] **Boldline news** indicates the ground truth.