News Recommendation via Jointly Modeling Event Matching and Style Matching

Pengyu Zhao[1⋆], Shoujin Wang[2⋆], Wenpeng Lu[1(✉)], Xueping Peng[3], Weiyu Zhang[1], Chaoqun Zheng[1], and Yonggang Huang[4]

[1] School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
[2] Data Science Institute, University of Technology Sydney, Sydney, Australia
[3] Australian Artificial Intelligence Institute, University of Technology Sydney, Sydney, Australia
[4] School of Computer, Beijing Institute of Technology, Beijing, China
{pengyuzhao.h,yonggang.h}@gmail.com, {wenpeng.lu,zwy}@qlu.edu.cn, cqzhengWORK@163.com, {shoujin.wang,xueping.peng}@uts.edu.au

Abstract. News recommendation is a valuable technology that helps users effectively and efficiently find news articles that interest them. However, most of existing approaches for news recommendation often model users' preferences by simply mixing all different information from news content together without in-depth analysis on news content. Such a practice often leads to significant information loss and thus impedes the recommendation performance. In practice, two factors which may significantly determine users' preferences towards news are news event and news style since users tend to read news articles that report events they are interested in, and they also prefer articles that are written in their preferred style. Such two factors are often overlooked by existing approaches.To address this issue, we propose a novel Event and Style Matching (ESM) model for improving the performance of news recommendation. The ESM model first uses an event-style disentangler to extract event and style information from news articles respectively. Then, a novel event matching module and a novel style matching module are designed to match the candidate news with users' preference from the event perspective and style perspective respectively. Finally, a unified score is calculated by aggregating the event matching score and style matching score for next news recommendation. Extensive experiments on real-world datasets demonstrate the superiority of ESM model and the rationality of our design[5].

Keywords: News Recommendation · News Event · News Style.

⋆ Equal contribution. This paper was partially supported by Nature Science Foundation of Shandong under Grant No.ZR2022MF243, National Nature Science Foundation of China under Grant No.61502259, Key Program of Science and Technology of Shandong under Grant No.2020CXGC010901, Program of Science and Technology of Qilu University of Technology under Grant No.2021JC02010.

[5] The source code and the splitted datasets are publicly available at https://github.com/ZQpengyu/ESM.

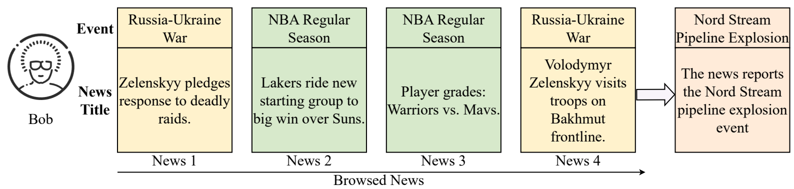
1 Introduction

Recent years have witnessed the increasing popularity of online news websites as the primary source of news of most individuals [6, 27]. A variety of news websites such as BBC News, CNN, Yahoo! News have been well established to provide massive news of various topics from politics to entertainment to millions of users around the world every day. While news websites bring convenience to people, they also bring challenges for individuals to navigate through the vast amount of available content and find articles that align with their interests [11, 33, 31]. As a dominated solution, news recommendation has playing an increasing important role [4, 28].

There are some existing works on news recommendation in the literature. For example, some works first encode news content with deep neural network for generating news representation, and then encode news sequence into a unified user representation by using sequence models such as RNN [1] or attention networks [34–36]. However, since user's reading interests are usually diverse, the aforementioned unified user representation fails to accurately imitate the complicated interests of user. In order to alleviate this problem, some works attempt to model each user's multi-interest reading behaviors by capturing the news content interactions with complex matching modules [17, 18, 21], or generating multiple user representations with attention networks [8, 23] or hierarchy modeling [16].

While great success has been achieved, most of existing approaches for news recommendations simply model each piece of news by learning a unified news representation [2, 21, 24, 34, 36]. They generally embed all relevant information (e.g., news article content, news topic, news title) together into the news representation without in-depth analysis [9, 11, 12]. However, such a practice may lose or weaken some implicit information which is significant for capturing users' preference towards news and thus news recommendation. In practice, users' preferences are usually driven by two underlying factors, i.e., news event and news style [3, 32]. Specifically, on one hand, a user often read news with the purpose to obtain the information relevant to some events which interest her/him, e.g., US election. On the other hand, even for reporting the same event, there are usually a variety of news pieces with different writing styles from different sources (e.g., ABC news, Yahoo! news). In such a case, users often tend to read news with their preferred styles. Let us take the toy examples shown in Fig. 1 to illustrate such characteristics. In Fig. 1(a), a user Bob has read a series of news pieces which report the event of Russia-Ukraine War or NBA Regular Session, implying he may interest in such events. Then, he choose to read some news reporting the Nord Stream pipeline explosion which is a follow-up event closely related to the Russa-Ukraine War event. In Fig. 1(b), there are three news pieces reporting the Nord Stream pipeline explosion event. However, Bob prone to read the News 5 only which is written in a statement style, similar to the News 1 and News 4 which have been already browsed by Bob before as shown in Fig. 1(a).

Although news event and style are critical for learning users' preferences towards news as well as news recommendation, such two significant driving factors are mostly overlooked in the literature. To this end, in this paper, we aim to build



(a)

Event	Candidate News ID	News Title	Read
Nord Stream Pipeline Explosion	5	Ukraine denies involvement in Nord Stream pipeline blasts.	✓
	6	Who Blew Up the Nord Stream Pipeline?	✗
	7	What do we know about the Nord Stream pipeline explosions?	✗

(b)

Fig. 1. (a) After reading a sequence of news, Bob preferred to read news about the Nord Stream pipeline explosion, which is related to the Russia-Ukraine War event. (b) Among the three news articles reporting the same event, Bob only read the News 5 that is written in his preferred style, similar to that of News 1 and 4.

a more powerful and accurate news recommender system to provide more accurate recommendations via well modeling news event and style respectively. To accommodate this idea, we propose a novel Event and Style Matching model for news recommendation, called ESM model for short. In ESM model, we first propose a novel event-style disentangler to effectively extract the event information and style information of news respectively from the input data (e.g., news article content, news title, category). Then, we design a novel event matching module to comprehensively model the matching degree between a set of a user's browsed historical news pieces and a given candidate news piece by taking the extracted event information as the input. To be specific, a novel fine-grained multi-channel matching module is proposed to model the possible multiple events covered by a set of historical news pieces browsed by each user, e.g., Russia-Ukraine war event and NBA Regular Session event shown in Fig. 1(a). At the same time, a novel style matching module is designed to measure the matching score between the user preferred news style revealed from her/his browsed news and the candidate news' style. Finally, both the event matching score and style matching score are well aggregated to a unified matching score between a user's preferred news and the candidate news for predicting the next news to the user. Extensive experiments on real-world datasets show that our proposed ESM model significantly improve the performance of news recommendation. The main contributions of this paper are summarized as follows:

– We propose modeling the event matching and style matching respectively for accurate news recommendation. A novel Event and Style Matching (ESM) model has been designed for implementing this idea. To the best of our knowledge, ESM is the first work to simultaneously model both event and style matching for news recommendation.

- We propose a novel event-style disentangler for effectively extracting event and style information respectively from the input news.
- We propose a novel and complicated event matching module and a style matching module for effectively identifying those candidate news of the user's interest w.r.t both news event and style.

2 Related work

2.1 News Recommendation

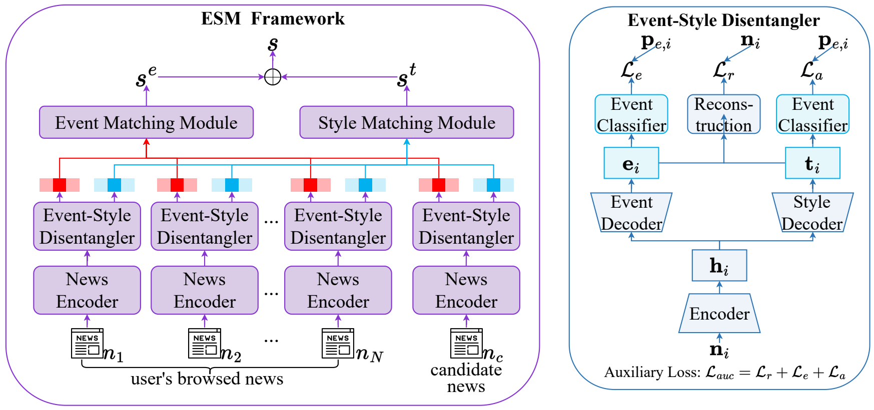
Methods modeling a single interest usually first encode each of a user's browsed news pieces content into a news representation, and then generate a unified user representation based on these news representations for recommending the appropriate candidate news [1, 34–36]. For example, Wu et al. [36] employed multi-head attention to encode news title and generate user representation for modeling the interaction among news. Although these methods achieve satisfactory performance, they encode user behavior into a unified user representation, which fails to capture the multiple interests of the user [10, 22].

Methods modeling multiple interests aim to capture a user's diverse interests from browsed news sequences [18, 16, 19, 21, 23, 26]. For example, Qi et al. [18] first encoded news piece into embedding representation, and then designed three kinds of candidate-aware components to generate the user representation for recommending diverse news. This type of methods well model the multiple interests of each user, and thus can achieve a significant improvement compared with the methods modeling the single interest only [30]. However, they embed the relevant information of news (e.g., title, content, topic) together into a simple unified news representation. Therefore, they lack the in-depth analysis of some important factors, such as news event and news style, which are important for driving users' preferences towards news.

2.2 News Event and Style Modeling

Since events are the essential content of news pieces, extracting event information from news content is beneficial to model news pieces. Some researchers have attempted to utilize news event information to enhance the performance of downstream tasks [3, 32, 40]. For example, Wang et al. [32] first utilized the event-veracity disentangler to obtain the event information and veracity information from news content, and then designed an event detection and transition module to predict which news event a user may prefer to know about. However, such study not only overlooked the multi-granularity nature of event information, but also ignored the significance of news style in modeling users' preference. Moreover, the method is specially built for fake news detection task, which relies on annotation of news veracity. Hence, it cannot be applied to general news recommendation tasks where news veracity label is not available.

In addition to events, news style is also important for modeling news pieces. However, limited work has been done on exploring news style information for



(a) The overall framework of ESM model.

(b) The structure of Event-Style Disentangler in ESM.

Fig. 2. (a) ESM consists of three main modules: Event-Style Disentangler, Event Matching Module, and Style Matching Module; (b) The Event-Style Distangler is built on the Encoder, Event and Style Decoder, and three auxiliary loss functions.

news representation [15, 39]. For example, Zhou et al. [39] constructed a hierarchical linguistic tree and a recursive neural network for modeling the linguistic style of news and predicting their veracity. However, here the style information does not involve the content of the news. Therefore, merely relying on style information to model the user's representation makes it difficult to accurately capture the user's interests. Additionally, as far as we konw, no research has been conducted to utilize news style for modeling news recommendation tasks. Different from the aforementioned methods, we attempt to jointly model new events and news style for news recommendation in this paper.

3 News Event and Style Matching Model

In this paper, the recommendation task is formalized as the next news prediction, which is similar to the next-item recommendation [20, 25, 29]. Specifically, given a sequence of N news pieces which has been read by a given user, we aim to predict which news piece the user may like to read in the next by considering the matching degree between read news and candidate news in terms of both news event and news style.

Fig. 2(a) presents the overall framework of our model, which consists of three main modules: an event-style disentangler, an event matching module, and a style matching module. First, given a news sequence which user browsed, the event-style disentangler disentangles the each news representation into event information and style information. Then, the event matching module generates the event-based user representation by modeling the event transition information, and matches it with events of candidate news to obtain the event matching score (i.e., fine- and coarse-granularity event matching score). Meanwhile, the style matching module aggregates the style information of the browsed news pieces into a style-based user representation, and matches it with styles of candidate news to obtain the style matching score. Finally, the recommendation score is achieved for predicting the next news by aggregating the event and style matching scores together.

3.1 News Encoder

In ESM, we encode each news piece into news representation using the news encoder proposed by [34]. Specifically, it first uses a CNN and attention network to encode the news title and content into corresponding text representations, and encodes the categories into feature representations using a dense network. Then, a view-level attention network is used to aggregate the representations of the title, content, and categories into a news representation $\mathbf{n}$.

3.2 Event-Style Disentangler

To extract the event and style information from news piece, and avoid interference between them [32], we propose an event-style disentangler to separate news representation into two kinds of information (i.e., event information and style information). As shown in Fig.2(b), event-style disentangler consists of three modules: an encoder, an event and style decoder, and an auxiliary loss module.

Encoder. For the i-th browsed news, the encoder takes the news representation $\mathbf{n}_i$ as input, followed by a three-layer dense network with residual connections. Specifically,

$$\mathbf{n}_i^1{=}ReLU(Dense(\mathbf{n}_i)), \mathbf{n}_i^2{=}ReLU(Dense([\mathbf{n}_i;\mathbf{n}_i^1])), \mathbf{h}_i{=}ReLU(Dense([\mathbf{n}_i;\mathbf{n}_i^2])), \quad (1)$$

where $[;]$ stands for the concatenate operation, $\mathbf{n}_i$ is the representation of i-th browsed news, and $\mathbf{h}_i$ is the high-level representation extracted from $\mathbf{n}_i$.

Event and Style Decoder. The event and style decoder includes two decoders: an event decoder and a style decoder. The event decoder first takes the high-level representation of news $\mathbf{h}_i$ as input, and then utilizes a three-layer dense network with residual connections to capture the event information of news. Specifically,

$$\mathbf{h}_i^1{=}ReLU(Dense(\mathbf{h}_i)), \mathbf{h}_i^2{=}ReLU(Dense([\mathbf{h}_i;\mathbf{h}_i^1])), \mathbf{e}_i{=}ReLU(Dense([\mathbf{h}_i;\mathbf{h}_i^2])), \quad (2)$$

where $\mathbf{e}_i$ refers to the event information of news. Similar to event decoder, the style decoder also employs this structure to capture the style information of news, i.e., $\mathbf{t}_i$. Here, $\mathbf{e}_i$ and $\mathbf{t}_i$ could be viewed as event-based and style-based news representation, respectively.

In order to predict the event distribution of news, we further employ an event classifier to encode the event information $\mathbf{e}_i$. Specifically,

$$\widehat{\mathbf{p}}_{e,i} = softmax(Dense(\mathbf{e}_i)), \quad (3)$$

where $\widehat{\mathbf{p}}_{e,i}$ stands for the event distribution of the i-th browsed news.

Auxiliary Loss Module. To well train the event-style disentangler, we utilize three kinds of loss function to optimize it, i.e. reconstruction loss, event prediction loss and adversarial loss. The reconstruction loss is used to alleviate the information loss problem during the disentangling process. We first concatenate the event and style information, and then employ a dense layer to encode it for reconstructing the new representation. Formally,

$$\mathcal{L}_r = MSE(Dense([\mathbf{e}_i; \mathbf{t}_i]), \mathbf{n}_i). \tag{4}$$

Event prediction loss is to help the event classifier to correctly predict the event distribution of news. Formally,

$$\mathcal{L}_e = MSE(\widehat{\mathbf{p}}_{e,i}, \mathbf{p}_{e,i}), \tag{5}$$

where $\widehat{\mathbf{p}}_{e,i}$ is the output of event classifier, $\mathbf{p}_{e,i}$ is the real event distribution of i-th browsed news.

The aim of the adversarial loss is to ensure that the disentangled style information does not contain the event information. Specifically, we first take the style information as the input of event classifier to predict the event distribution, and then maximize the loss between it and the real event distribution of news,

$$\mathcal{L}_a = \frac{1}{K} \sum\nolimits_{j=1}^{K} \frac{1}{1 + MSE(\widehat{\mathbf{p}}_{t,ij}, \mathbf{p}_{e,ij})}, \tag{6}$$

where K represents the number of event channels, $\widehat{\mathbf{p}}_{t,ij}$ is the distribution probability of j-th event channel, which is predicted by style information and event classifier, and $\mathbf{p}_{e,ij}$ stands for the real distribution probability of j-th event channel for i-th browsed news.

Finally, the auxiliary loss function is the summation of the three loss function,

$$\mathcal{L}_{auc} = \mathcal{L}_r + \mathcal{L}_e + \mathcal{L}_a. \tag{7}$$

In this auxiliary loss module, we utilize the event prediction loss and adversarial loss to optimize the event-style disentangler. However, due to the dataset limitation, it is hard for us to obtain the real event distribution $\mathbf{p}_{e,i}$. Therefore, we propose a method to calculate the label of the news event distribution. Specifically, we first utilize TF-IDF to extract keywords of each piece of news, and then employ BERT to encode keywords into embeddings, followed by a K-means algorithm to cluster all keywords into K event channels. For the i-th news, the distribution of j-th event channel $\mathbf{p}_{e,ij}$ is calculated as:

$$\mathbf{p}_{e,ij} = \frac{|N_{i,j}|}{|N_i|}, \tag{8}$$

where $|N_i|$ represents the number of keywords in the i-th news, and $|N_{i,j}|$ represents the number of keywords which belong to i-th news and j-th event channel.

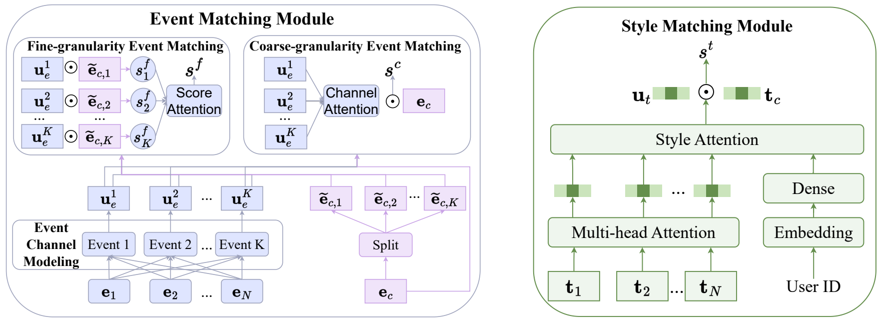


Fig. 3. The structure of Event Matching Module and Style Matching Module.

3.3 Event Matching Module

Users read a piece of news because the reported event may be one that they are interested in or has a potential relation to their interests. To measure the matching degree between user interest and candidate news event, we propose an event matching module. As shown in the left side of Fig. 3, the event matching module contains three modules: an event channel modeling module, a fine-granularity event matching module, and a coarse-granularity event matching module.

Event Channel Modeling Module. For generating the event-based user representation for each event channel, we employ attention networks to model the event information within the same event channel. Each event channel aggregates event information with potential connections. Specifically, according to the event distribution $\widehat{\mathbf{p}}_{e,i}$ calculated in Equ. (3), we first split the event information into different event channels. Formally,

$$\widetilde{\mathbf{e}}_i = \widehat{\mathbf{p}}_{e,i}\mathbf{e}_i, \tag{9}$$

where the event information $\mathbf{e}_i \in \mathbb{R}^{\text{dim}}$ is split to K event channels, the splitted event representation is $\widetilde{\mathbf{e}}_i \in \mathbb{R}^{K,\text{dim}}$.

Since browsed news event may have different influences for generating event-based user representation, we separately utilize attention network to generate event-based user representation for each event channel. For the j-th event channel, the event-based user representation $\mathbf{u}_e^j$ is computed as below,

$$\alpha_{i,j}^e = \mathbf{q}_j^\top \tanh(\mathbf{V}_j \times \widetilde{\mathbf{e}}_{i,j} + \mathbf{v}_j), \quad \alpha_{i,j}^e = \frac{\exp(\alpha_{i,j}^e)}{\sum_{k=1}^N \exp(\alpha_{k,j}^e)}, \quad \mathbf{u}_e^j = \sum_{i=1}^N \alpha_{i,j}^e \widetilde{\mathbf{e}}_{i,j}, \tag{10}$$

where $\mathbf{V}_j$, $\mathbf{v}_j$ and $\mathbf{q}_j$ are trainable parameters, N is the number of user browsed news, $\widetilde{\mathbf{e}}_{i,j}$ is the event information of i-th news on j-th event channels, and $\mathbf{u}_e^j$ is the event-based user representation on j-th event channel.

Fine-granularity Event Matching Module. To model the possible multiple events covered by a set of historical news pieces browsed by each user, we first calculate the matching score on different event channels, and then aggregate them into a fine-granularity matching score by a score attention network.

The matching score of j-th event channel is computed by the event-based user representation and the event representation of candidate news on the j-th event channel,

$$s_j^f = \mathbf{u}_e^j \widetilde{\mathbf{e}}_{c,j}. \tag{11}$$

Afterward, we design a score attention network to aggregate all matching score on different event channels. Considering the event distribution of candidate news may influence the score aggregation, we incorporate it into score attention network. Specifically, we map the event distribution of candidate news $\widehat{\mathbf{p}}_{e,c}$ into a discrete value vector $\widetilde{\mathbf{p}}_{e,c}$ with the function $\widetilde{\mathbf{p}}_{e,c} = \text{round}(10 \times \widehat{\mathbf{p}}_{e,c})$. Then, we obtain the representation of $\widehat{\mathbf{p}}'_{e,c}$ via a dense layer:

$$\widehat{\mathbf{P}}''_{e,c} = Dense(\widehat{\mathbf{P}}'_{e,c}), \tag{12}$$

where $\widehat{\mathbf{P}}'_{e,c}$ is the embedding of $\widetilde{\mathbf{p}}_{e,c}$.

Finally, we fuse $\widehat{\mathbf{P}}''_{e,c}$ with the event-based user representation to generate the attention weight of score attention network for aggregating all scores:

$$\alpha_j^s = \mathbf{q}_s^\top \tanh(\mathbf{V_s} \times [\mathbf{u}_e^j; \widehat{\mathbf{P}}''_{e,cj}] + \mathbf{v}_s), \quad \alpha_j^s = \frac{\exp(\alpha_j^s)}{\sum_{i=1}^K \exp(\alpha_i^s)}, \; s^f = \sum_{j=1}^K \alpha_j^s s_j^f, \tag{13}$$

where $\mathbf{q}_s$, $\mathbf{V_s}$, and $\mathbf{v}_s$ are trainable parameters, K is the number of event channels, and s^f represents the fine-granularity event matching score.

Coarse-granularity Event Matching Module. Since the user may focus on all the event information covered by candidate news, we match the overall event-based user representation with the candidate news event.

To generate the overall event-based user representation, we devise a channel attention network that aggregates all user representations across different event channels. More specifically, the number of news articles on an event channel likely reflects its importance in generating the overall user representation. We assume that a news belongs to the j-th event channel when its distribution $\widehat{\mathbf{p}}_{e,ij}$ is greater than $1/K$. We first convert the number of news on each event channel into embedding representation R. Then, we merge it with event-based user representation to construct the channel attention network for generating overall user representation,

$$\alpha_j^c = \mathbf{q}_c^\top \tanh(\mathbf{V}_c \times [\mathbf{u}_e^j; \mathbf{r}_j] + \mathbf{v}_c), \quad \alpha_j^c = \frac{\exp(\alpha_j^c)}{\sum_{i=1}^K \exp(\alpha_i^c)}, \quad \mathbf{u}_e^o = \sum_{j=1}^K \alpha_j^c \mathbf{u}_e^j, \tag{14}$$

where $\mathbf{q}_c$, $\mathbf{V}_c$ and $\mathbf{v}_c$ are trainable parameters, $\mathbf{u}_e^o$ is the overall event-based user representation. The coarse-granularity matching score is calculated by the overall event-based user representation $\mathbf{u}_e^o$ and event-based candidate news representation $\mathbf{e}_c$ from the event-style disentangler,

$$s^c = \mathbf{u}_e^{o\top} \mathbf{e}_c. \tag{15}$$

3.4 Style Matching Module

For each event, there exist abundant news to describe it. However, users only like to read the news pieces that are written in their preferred style. Therefore, we devise a style matching module to measure the matching score between the style of user preferred news and candidate news.

Specifically, we first utilize the multi-head attention to capture the interaction of style information among browsed news $\mathbf{T} = \{\mathbf{t}_i\}_{i=1}^{N}$,

$$[\mathbf{t}_1^{'}, \mathbf{t}_2^{'}, \cdots, \mathbf{t}_N^{'}] = MultiHeadAttention(\mathbf{T}). \tag{16}$$

Next, we aggregate them into a style-based user representation by a style attention network. Since user IDs may contain user preferred style feature, we merge it with style information of browsed news to construct the style attention network:

$$\begin{aligned} &\mathbf{u}_d^{'} = ReLU(Dense(\mathbf{u}_d)), \quad &&\alpha_i^t = \mathbf{q}_t^{\top} \tanh(\mathbf{V}_t \mathbf{t}_i^{'} + \mathbf{V}_d \mathbf{u}_d^{'} + \mathbf{v}), \\ &\alpha_i^t = \frac{\exp(\alpha_i^t)}{\sum_{j=1}^{N} \exp(\alpha_j^t)}, &&\mathbf{u}_t = \sum\nolimits_{i=1}^{N} \alpha_i^t \mathbf{t}_i^{'}, \end{aligned} \tag{17}$$

where $\mathbf{q}_t$, $\mathbf{V}_t$, $\mathbf{V}_d$, $\mathbf{v}$ are trainable parameters, $\mathbf{u}_d$ stands for the embedding of user IDs, $\mathbf{u}_t$ refers to the style-based user representation.

Finally, the style matching score is calculated by the style-based user representation $\mathbf{u}_t$ and the style-based candidate news representation $\mathbf{t}_c$,

$$s^t = \mathbf{u}_t^{\top} \mathbf{t}_c. \tag{18}$$

3.5 News Recommendation

For each candidate news, its final recommendation score is obtained by merging the event and style matching scores, i.e., $s = s^e + \beta s^t$, where $s^e = s^f + s^c$. Besides, following [5], we employ the NCE loss $\mathcal{L}_{rec}$ to optimize our model.

Since the $\mathcal{L}_{auc}$ is utilized to optimize the event-style disentangler, we combine it with the $\mathcal{L}_{rec}$ as the total loss:

$$\mathcal{L} = \mathcal{L}_{rec} + \gamma \mathcal{L}_{auc}, \tag{19}$$

where γ is a hyperparmeter.

4 Experiment

4.1 Dataset and Experimental Settings

We carried out extensive experiments on two datasets, namely MIND-large and MIND-500K. The MIND-large[6] is an official benchmark dataset [38], consisting of around 877k users and approximately 130k news articles. Each news article

[6] https://msnews.github.io/

comprises the news title, abstract, entity, and categories. Considering the large scale of MIND-large, following common practice [13, 37], we randomly sampled 500k users and their associated news browsing logs from MIND-large to constitute the MIND-500K dataset. The original test set of MIND-large cannot be used for testing since it did not have any released labels, we submitted our predictions to the official evaluation platform[7] to assess our model's performance. For MIND-500K, we divided the original validation set equally into an experimental validation set and a test set for validation and test.

In our model, the news titles and abstracts were truncated to a maximum length of 20 and 128 words, respectively. The user's browsing history was restricted to a maximum of 50 news articles. Glove embedding [14] with 300 dimensions was used as word embedding. In the auxiliary loss function module, the max number of keywords for each piece of news was set to 5. The number of event channels K was set to 10 and 17 for for MIND-large and MIND-500K respectively. In news recommendation module, the weight of style loss β (resp. auxiliary loss γ) was set to 0.6 (resp. 1) and 1.9 (resp. 1.6) for MIND-large and MIND-500K respectively. The dropout and negative sampling rate was set to 0.2 and 4 respectively. Adam[7] was selected as the optimizer and the learning rate was set to 1e-4. All experiments were ran on one RTX 2080Ti GPU. Three commonly used evaluation metrics, i.e., AUC, MRR, nDCG were used.

4.2 Comparison with Competing Methods

Baselines. Six representative and/or state-of-the-art news recommendation approaches were used as baselines. They can be classified into two categories: single-interest approaches including NPA, NAML, LSTUR, and NRMS, and multiple-interest approaches including FIM and MINS. The details are as follows: (1) NPA [35] employs a personalized attention network to encode news content and user behavior; (2) NAML [34] aggregates multiple kinds of information together into the news representation; (3) LSTUR [1] combines the long- and short-term user interest for news recommendation; (4) NRMS [36] models user interactions on news content and sequence levels respectively by multi-head self-attention networks; (5) FIM [21] extracts fine-granularity interest with 3D-CNN network; (6) MINS [26] applies multi-head self-attention and GRU to model news sequence based on different interest channels.

Baseline settings. To ensure a fair comparison between our ESM model and the baseline models, we fine-tuned all models' parameters consistently on the validation dataset. Specifically, we initialized the parameters of each baseline model with the values reported in the original paper and then fine-tuned them on the target dataset to achieve optimal performance. For NPA and NRMS, we fused category information to improve their performance. Additionally, we set the number of filters in CNN to 400 (resp. 150) and the filter size to 3 (resp. 3) for NAML and LSTUR (resp. FIM). Furthermore, we set the number of heads for multi-head self-attention to 20 and 15 for NRMS and MINS, respectively.

[7] https://codalab.lisn.upsaclay.fr/competitions/420

Table 1. Comparison of our model with baseline methods (%).

Models	MIND-large				MIND-500K			
	AUC	MRR	nDCG@5	nDCG@10	AUC	MRR	nDCG@5	nDCG@10
NPA	67.30	32.75	35.58	41.33	67.64	32.70	36.21	42.47
NAML	68.25	33.53	36.52	42.28	68.65	33.08	36.82	43.18
LSTUR	68.41	33.58	36.58	42.31	68.60	33.48	37.10	43.46
NRMS	68.62	33.69	36.72	42.46	68.58	33.32	36.85	43.35
FIM	68.98	34.18	37.33	43.02	69.03	33.45	37.19	43.50
MINS	69.03	33.90	36.96	42.74	68.98	33.22	36.99	43.41
ESM	**69.38**	**34.23**	**37.38**	**43.11**	**69.87**	**33.63**	**37.52**	**43.90**

The comparison results are shown in Table 1, and the following observations can be made. First, all metrics indicate that the approaches that model multiple interests (i.e. FIM, MINS, and ESM) performed better than those that model a single interest (i.e., NPA, NAML, LSTUR, and NRMS). This outcome may be due to the limited ability of single-interest models to comprehensively representing users via well capturing the wide range of reading interests (cf. Section 2) that a user may have. Second, our proposed ESM method demonstrated superior performance in all metrics compared to the other methods. This can possibly be attributed to the fact that these baseline methods solely embed relevant news information, such as the title, content, and categories, into a unified news representation, without conducting a thorough analysis of news events and styles, which are crucial factors influencing user preferences towards news. Conversely, ESM can effectively extract event and style information from news, and comprehensively calculate the matching score based on both the news event and news style perspectives, resulting in better performance.

4.3 Ablation Study

To show the effectiveness of ESM's core components, we customized ESM into five variants, referred to as ESM^{-dise}, ESM^{-fine}, ESM$^{-coarse}$, ESM^{-event}, and ESM^{-style}. These variants exclude event-style disentangler (cf. Section 3.2), fine-granularity event matching module (cf. Equ. (11)-Equ. (13)), coarse-granularity event matching module (cf. Equ. (14) - Equ. (16)), event matching module (cf. Section 3.3), and style matching module (cf. Section 3.4), respectively.

Table 2 shows the experimental results of ESM and its five variants. Based on these results, we make the following observations. First, the performance of ESM^{-dise} was lower than that of the standard ESM, possibly due to the fact that ESM^{-dise} does not separate the event and style information from the news

Table 2. Comparison of ESM with its variants on MIND-500K dataset (%).

	AUC	MRR	nDCG@5	nDCG@10
ESM^{-dise}	69.11	33.48	37.13	43.58
ESM^{-fine}	69.63	33.38	37.30	43.72
ESM$^{-coarse}$	69.30	33.30	37.13	43.57
ESM^{-event}	69.15	33.06	36.83	43.33
ESM^{-style}	68.17	32.20	35.87	42.42
ESM	**69.87**	**33.63**	**37.52**	**43.90**

representation. Second, both ESM^{-fine} and ESM$^{-coarse}$ achieved worse results than ESM. This is because ESM^{-fine} only considers the coarse-granularity event matching in the event matching module, neglecting the fine-granular multi-event matching. Similarly, ESM$^{-coarse}$ ignores the coarse-granularity event matching score, which also weakens the model's ability. Third, ESM^{-event} was outperformed by ESM, possibly due to the fact that ESM^{-event} only takes into account the style information of news, while ignoring the event information. Last, ESM^{-style} exhibited inferior performance compared to ESM. This is primarily due to the fact that this variant only concentrates on matching news events, disregarding news style matching, resulting in unsatisfactory performance. In summary, the five components of ESM are advantageous in improving the performance of news recommendation tasks.

4.4 Comparison with Different News Encoders

To show the generalization of our model, we replaced the news encoder used in ESM with four news encoders that are respectively utilized in four representative baselines, namely NPA, LSTUR, NRMS, and MINS to obtain four more ESM variants. We then compared their performance with that of their corresponding baseline methods. As shown in Fig. 4, ESM variants (denoted as w/ ESM) significantly outperform the corresponding baseline methods in terms of AUC and nDCG@10. This could be attributed to the fact that ESM emphasizes the modeling of event and style information, which are critical to the news recommendation task. Furthermore, compared with all the four ESM variants here, the standard ESM delivered the best performance, likely because ESM's encoder (i.e., the NAML encoder) can effectively aggregate news content and category information into news representation, laying a robust foundation for extracting news event and style information.

4.5 Hyperparameter Analysis

Fig. 5 provides a detailed analysis of the impact of key hyperparameters on the performance of ESM model. The hyperparameters examined include the number of keywords, the number of event channels, the weight of style matching score,

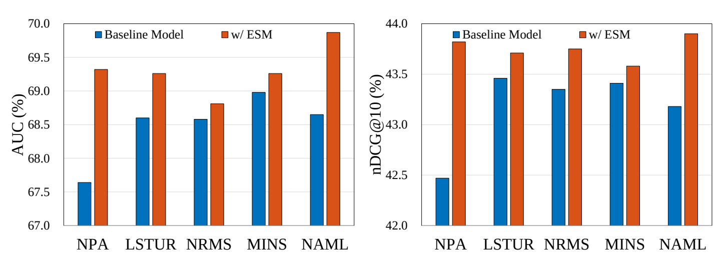
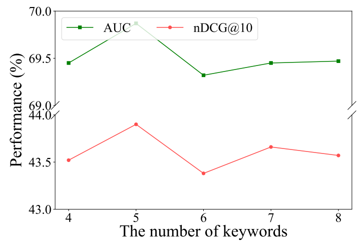
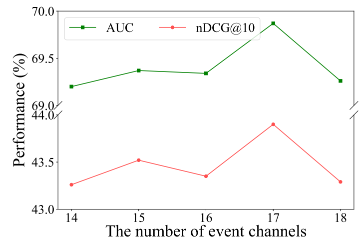


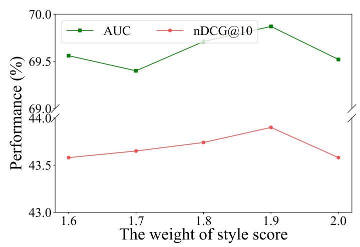
Fig. 4. Impact of news encoder in terms of AUC and nDCG@10 on MIND-500K.



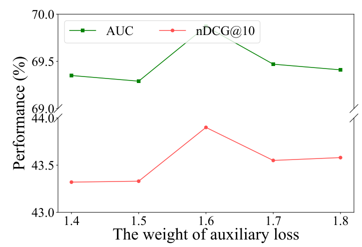
(a) Impact of keywords.



(b) Impact of event channels.



(c) Impact of style matching score.



(d) Impact of auxiliary loss.

Fig. 5. The performance of hyperparameter analysis on MIND-500K dataset.

and the weight of the auxiliary loss function. ESM performs best on the MIND-500K dataset when the number of keywords and the number of event channels are set to 5 and 17 respectively. More keywords benefits accurately calculating the event distribution, but too many can introduce noise and reduce accuracy. similarly, more channels can help to aggregate event information accurately, but too many can lead to overfitting and limited performance. The best results are achieved when the weight of style matching score is set to 1.9. Although style information can improve performance, a too-large weight will weaken the influence of event information. Lastly, the weight of the auxiliary loss function plays a crucial role in disentangling event and style information and ESM performs optimally when the weight is set to 1.6.

5 Conclusion

In this paper, we propose a novel news recommendation method via jointly modeling event matching and style matching (ESM). ESM considers news events and styles as two essential factors in recommending news to users. To achieve this, an event-style disentangler is devised to separate event and style information from news articles. Besides, an event matching module and a style matching module are designed to match news articles with users' preferences from the event and style perspectives, respectively. Extensive experiments have demonstrated the superiority of ESM over the representative and state-of-the-art methods. Future work will focus on explicitly modeling news style features to enhance performance.

6 Ethics Consideration

In this paper, we conducted experiment on two datasets, i.e., MIND-large and MIND-500K. MIND-large is a publicly available news recommendation dataset, which is constructed by Wu et al.[38]. In MIND-large, the privacy of each user was protected by mapping user into an anonymized ID. Since MIND-500K was extracted from MIND-large, it also protects user privacy. Therefore, we ensure that this work will not lead to any privacy issues.

References

1. An, M., Wu, F., Wu, C., et al.: Neural news recommendation with long- and short-term user representations. In: ACL. pp. 336–345 (2019)
2. Ge, S., Wu, C., Wu, F., Qi, T., Huang, Y.: Graph enhanced representation learning for news recommendation. In: WWW. pp. 2863–2869 (2020)
3. Han, S., Huang, H., Liu, J.: Neural news recommendation with event extraction. arXiv preprint arXiv:2111.05068 (2021)
4. Hu, L., Xu, S., Li, C., et al.: Graph neural news recommendation with unsupervised preference disentanglement. In: ACL. pp. 4255–4264 (2020)
5. Huang, P.S., He, X., Gao, J., et al.: Learning deep structured semantic models for web search using clickthrough data. In: CIKM. pp. 2333–2338 (2013)
6. Khattar, D., Kumar, V., et al.: Weave&rec: A word embedding based 3-D convolutional network for news recommendation. In: CIKM. pp. 1855–1858 (2018)
7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ILCR. pp. 1–15 (2015)
8. Li, J., Zhu, J., Bi, Q., et al.: Miner: Multi-interest matching network for news recommendation. In: ACL Findings. pp. 343–352 (2022)
9. Liu, D., Lian, J., Liu, Z., et al.: Reinforced anchor knowledge graph generation for news recommendation reasoning. In: KDD. pp. 1055–1065 (2021)
10. Liu, D., Lian, J., Wang, S., et al.: Kred: Knowledge-aware document representation for news recommendations. In: RecSys. pp. 200–209 (2020)
11. Lu, W., Wang, R., Wang, S., et al.: Aspect-driven user preference and news representation learning for news recommendation. TITS **23**(12), 25297–25307 (2022)
12. Mao, Z., Li, J., Wang, H., Zeng, X., Wong, K.: Digat: Modeling news recommendation with dual graph interaction. In: EMNLP Findings. pp. 6595–6607 (2022)
13. Mao, Z., Zeng, X., Wong, K.F.: Neural news recommendation with collaborative news encoding and structural user encoding. In: EMNLP Findings. pp. 46–55 (2021)
14. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: EMNLP. pp. 1532–1543 (2014)
15. Przybyla, P.: Capturing the style of fake news. In: AAAI. pp. 490–497 (2020)
16. Qi, T., Wu, F., Wu, C., et al.: Hierec: Hierarchical user interest modeling for personalized news recommendation. In: ACL. pp. 5446–5456 (2021)
17. Qi, T., Wu, F., Wu, C., Huang, Y.: Personalized news recommendation with knowledge-aware interactive matching. In: SIGIR. pp. 61–70 (2021)
18. Qi, T., Wu, F., Wu, C., Huang, Y.: News recommendation with candidate-aware user modeling. In: SIGIR. pp. 1917–1921 (2022)
19. Qiu, Z., Hu, Y., Wu, X.: Graph neural news recommendation with user existing and potential interest modeling. TKDD **16**(5), 1–17 (2022)

20. Song, W., Wang, S., Wang, Y., Wang, S.: Next-item recommendations in short sessions. In: RecSys. pp. 282–291 (2021)
21. Wang, H., Wu, F., Liu, Z., Xie, X.: Fine-grained interest matching for neural news recommendation. In: ACL. pp. 836–845 (2020)
22. Wang, H., Zhang, F., Xie, X., Guo, M.: Dkn: Deep knowledge-aware network for news recommendation. In: WWW. pp. 1835–1844 (2018)
23. Wang, J., Chen, Y., Wang, Z., Zhao, W.: Popularity-enhanced news recommendation with multi-view interest representation. In: CIKM. pp. 1949–1958 (2021)
24. Wang, J., Jiang, Y., Li, H., Zhao, W.: Improving news recommendation with channel-wise dynamic representations and contrastive user modeling. In: WSDM. pp. 562–570 (2023)
25. Wang, N., Wang, S., Wang, Y., et al.: Exploiting intra- and inter-session dependencies for session-based recommendations. World Wide Web **25**(1), 425–443 (2022)
26. Wang, R., Wang, S., Lu, W., Peng, X.: News recommendation via multi-interest news sequence modelling. In: ICASSP. pp. 7942–7946 (2022)
27. Wang, R., Wang, S., Lu, W., Peng, X., Zhang, W., Zheng, C., Qiao, X.: Intention-aware user modeling for personalized news recommendation. In: DASFAA. pp. 179–194 (2023)
28. Wang, S., Guo, S., Wang, L., et al.: Multi-interest extraction joint with contrastive learning for news recommendation. In: ECML-PKDD. pp. 606–621 (2022)
29. Wang, S., Cao, L., Wang, Y., Sheng, Q.Z., Orgun, M.A., Lian, D.: A survey on session-based recommender systems. CSUR **54**(7), 1–38 (2021)
30. Wang, S., Hu, L., Wang, Y., Sheng, Q.Z., Orgun, M., Cao, L.: Modeling multi-purpose sessions for next-item recommendations via mixture-channel purpose routing networks. In: IJCAI. pp. 3771–3777 (2019)
31. Wang, S., Pasi, G., Hu, L., Cao, L.: The era of intelligent recommendation: editorial on intelligent recommendation with advanced ai and learning. IEEE Intelligent Systems **35**(5), 3–6 (2020)
32. Wang, S., Xu, X., Zhang, X., et al.: Veracity-aware and event-driven personalized news recommendation for fake news mitigation. In: WWW. pp. 3673–3684 (2022)
33. Wang, S., Zhang, X., Wang, Y., Liu, H., Ricci, F.: Trustworthy recommender systems. arXiv preprint arXiv:2208.06265 (2022)
34. Wu, C., Wu, F., An, M., Huang, J., Huang, Y., Xie, X.: Neural news recommendation with attentive multi-view learning. In: IJCAI. pp. 3863–3869 (2019)
35. Wu, C., Wu, F., An, M., Huang, J., Huang, Y., Xie, X.: Npa: Neural news recommendation with personalized attention. In: KDD. pp. 2576–2584 (2019)
36. Wu, C., Wu, F., Ge, S., Qi, T., Huang, Y., Xie, X.: Neural news recommendation with multi-head self-attention. In: EMNLP. pp. 6389–6394 (2019)
37. Wu, C., Wu, F., Qi, T., Huang, Y.: Empowering news recommendation with pre-trained language models. In: SIGIR. pp. 1652–1656 (2021)
38. Wu, F., Qiao, Y., Chen, J.H., Wu, C., Qi, T., Lian, J., Liu, D., Xie, X., Gao, J., Wu, W., et al.: Mind: A large-scale dataset for news recommendation. In: ACL. pp. 3597–3606 (2020)
39. Zhou, X., Li, J., Li, Q., Zafarani, R.: Linguistic-style-aware neural networks for fake news detection. arXiv preprint arXiv:2301.02792 (2023)
40. Zhou, Z., Ma, L., Liu, H.: Trade the event: Corporate events detection for news-based event-driven trading. In: ACL Findings. pp. 2114–2124 (2021)