Veracity-aware and Event-driven Personalized News Recommendation for Fake News Mitigation

Shoujin Wang
shoujinwang@foxmail.com
RMIT University, Macquarie University
Australia

Xiaofei Xu
s3833028@student.rmit.edu.au
RMIT University
Australia

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

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

Wenzhuo Song
wzsong17@gmail.com
Northeast Normal University
China

ABSTRACT

Despite the tremendous efforts by social media platforms and fact-check services for fake news detection, fake news and misinformation still spread wildly on social media platforms (e.g., Twitter). Consequently, fake news mitigation strategies are urgently needed. Most of the existing work on fake news mitigation focuses on the overall mitigation on a whole social network while ignoring developing concrete mitigation strategies to deter individual users from sharing fake news. In this paper, we propose a novel veracity-aware and event-driven recommendation model to recommend personalised corrective true news to individual users for effectively debunking fake news. Our proposed model Rec4Mit (Recommendation for Mitigation) not only effectively captures a user's current reading preference with a focus on which event, e.g., US election, from her/his recent reading history containing true and/or fake news, but also accurately predicts the veracity (true or fake) of candidate news. As a result, Rec4Mit can recommend the most suitable true news to best match the user's preference as well as to mitigate fake news. In particular, for those users who have read fake news of a certain event, Rec4Mit is able to recommend the corresponding true news of the same event. Extensive experiments on real-world datasets show Rec4Mit significantly outperforms the state-of-the-art news recommendation methods in terms of the capability to recommend personalized true news for fake news mitigation.

CCS CONCEPTS

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

KEYWORDS

Fake news mitigation, News recommendation, Recommender systems, Fake news detection

*Corresponding author.

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

ACM ISBN 978-1-4503-9096-5/22/04…$15.00
https://doi.org/10.1145/3485447.3512263

ACM Reference Format:
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 (WWW '22), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3485447.3512263

1 INTRODUCTION

Fake news has been recognised as a huge threat to societies. For example, fake news and misinformation such as political propaganda [1] and financial propaganda [4] have been a long-standing issue. Therefore, combating fake news has attracted increasingly more attention in recent years.

Social media platforms, government agencies and fact-checking services all have paid attention to detect fake news. Automated fake news detection methods [24, 27, 32, 33, 53, 60] have also been reported in the research literature. Despite these tremendous efforts, fake news still spreads on the social media. Therefore, fake news mitigation has become an urgent task.

Most of the existing studies on fake news mitigation focus on strategies to introduce true news to counteract the spread of fake news on social networks [29]. They generally model the spread and influence of news on social networks based on information diffusion models such as Independent Cascade (IC) and Linear Threshold (LT) models [17], point process models [59], and reinforcement learning models [36]. These methods often formulate the mitigation task as an optimization problem to maximize the propagation of true news on networks. On one hand, such mitigation strategies only focus on the mitigation on the whole network and thus can hardly work on individual users. On the other hand, it is hard to deploy them on real-world social network sites whose information diffusion structure is dynamic and user behaviours vary across news topics.

In this paper, we focus on the concrete task of how to design practical intervention strategies to deter *individual users* from sharing fake news. We propose a novel model to recommend personalised, corrective true news to individual users for fake news mitigation.

In recent years, more and more users read news recommended by social media platforms (e.g., Twitter). News recommender systems have been playing an increasingly important role in influencing and even changing users' reading behaviours [9]. However, conventional news recommender systems can not address the unique challenges imposed by fake news. Towards fake news mitigation, news recommendation are based on particular *data* with unique characteristics and have a different *goal*. Specifically, from the data perspective, each piece of news is often associated with an *event*, e.g., US election. For each event, there are usually multiple pieces of news with different veracity levels (true or fake). In addition, a user may read the true and/or fake news for the same event, and may often be interested in the news pieces for the same or related events in a certain period. From the goal perspective, news recommendation for fake news mitigation not only needs to recommend personalized news preferred by users to best match their personal preferences, but also needs to recommend true news only. In particular, if a user has read some fake news for an event, she should be recommended the corresponding true news for this event to match her preference so as to mitigate the fake news.

These unique characteristics and goal essentially trigger the following unique challenges (CHs) for fake news mitigation oriented news recommendation when given a user's reading history consisting of true and/or fake news. **CH1**: How to model the complex transitions over the potential latent events behind news for predicting the next piece of relevant news preferred by a user? **CH2**: How to only recommend true news to a user given a set of candidate news whose veracity is unknown? **CH3**: How to model the transition over latent events while avoiding the interference from news veracity related information (e.g., news content style)? (Note that true news and fake news for the same event often have different content styles [18], and thus are easily regarded as news for different events [28], misleading the modelling of event transition.)

Existing conventional news recommender systems, including collaborative filtering methods [9, 21] and content-based methods [20, 39], are oblivious to the veracity of news [16]. Their goal is to find the relevant news to match a user's personal preference only [42, 44]. In addition, the event information behind news is less studied. Obviously, these conventional news recommender systems cannot address the aforementioned three challenges.

To address the unique challenges of fake news mitigation, we propose a novel news recommendation model Rec4Mit (Recommendation for Mitigation) for fake news mitigation task. Specifically, Rec4Mit first effectively disentangles the event-specific information and veracity-specific information from each news embedding to enable the modelling of event and veracity separately to reduce the interference between them. Afterwards, taking the event-specific information as input, a novel event detection and transition module is devised to first detect the latent event(s) behind each piece of news, and then model the transitions between events over a sequence of historical news pieces read by a user. Simultaneously, a news veracity predictor is built to predict the veracity label of each candidate news piece by taking the veracity-specific information as input. Finally, a novel next-news predictor is designed to predict the next piece of true news that may interest the user by considering both the event transition information and the label of the candidate news. The event transition information ensures that the recommended next news is highly related to the event(s) that the user is focusing on while veracity information guarantees that only the predicted true candidate news is recommended.

The main contributions of this work are summarized below:

- We propose news recommendation to mitigate fake news. For the first time, we analyze the data characteristics and challenges for this novel problem, formalize it and propose a novel solution.
- We propose a novel model Rec4Mit to recommend predicted true news to individual users tailored for mitigating fake news.
- In Rec4Mit, a novel event-label disentangler is designed to disentangle the event- and veracity-specific information from a given news piece, and an event detection and transition module together with a veracity classifier is carefully devised to model event transitions and predict news veracity respectively.

Extensive experimental results on two real-world datasets demonstrate that our proposed method consistently and significantly outperforms the representative and state-of-the-art methods.

2 RELATED WORK

There has been extensive research in the literature on fake news detection [8, 26, 28, 29, 47]. In this paper, we focus on the fake news mitigation task via news recommendation.

2.1 Fake News Mitigation

Approaches for fake news mitigation can be roughly categorized into three classes: (1) independent cascade (IC) and linear threshold (LT) model based approaches, (2) point process model based approaches, and (3) reinforcement learning (RL) based approaches.

Both IC and LT are well-known information diffusion models. They model the information diffusion process on a network by taking each user as an node while each edge between a pair of users indicates the influence between them. The news spreads from one user to another according to the influence [17]. Based on IC and LT models, some decontamination methods were proposed to select the best set of seed users to start the diffusion of true news on the network so that the users exposed to fake news can be decontaminated [22]. However, these methods can be only suggested as a corrective measure *after* the spread of fake news [29]. To alleviate this drawback, another mitigation strategy based on competing cascades was proposed to introduce a true news cascade to compete with the fake news cascade when the fake news begin to propagate [3, 11]. But this method heavily relies on the competition between two cascades and has no external moderation.

To allow external interventions, a multi-stage intervention strategy [6] based on multivariate point process [59] was proposed. Specifically, a social influence model based on multivariate point processes is utilized to model the news propagation. However, these methods rely on an strict assumption that fake news has been already identified and its propagation is tracked.

In recent years, RL-based methods have been proposed to learn a misinformation prevention strategy to seek the seed nodes of true news cascade (protector) against those of fake news (attacker) when given the examples of attacker-protector pairs [36]. However, such method faces the same issue as point process based methods.

In summary, all of these information diffusion model based methods often only focus on the mitigation on the whole network while ignoring the concrete mitigation on individual users. In addition, they are hard to be deployed on real-world social network sites whose information diffusion structure is dynamic and volatile and user behaviours vary across news topics. This actually drives the need of more practical fake news mitigation strategies to effectively deter individual users from sharing fake news.

2.2 News Recommendation

A variety of methods have been proposed for news recommendation, and recent years have witnessed the success of deep learning based methods in extracting news' semantic features and mining users' preferences [50, 54]. Specifically, various models including recurrent neural networks (RNNs) [2, 23], attention mechanism [49, 61], dilated convolution [38], graph neural networks [13], knowledge distillation [39] and reinforcement learning [58] were explored for news recommendation. For example, self-attention and additive attention were employed to represent words within each news and a sequence of news read by a user for news recommendation [51]. Dilated convolutions were utilized to capture fine-grained interest matching signals for news recommendation [38]. However, all of these methods are for conventional news recommendation and cannot be employed to fake news mitigation task which has unique data and goal (cf. Paragraph 5 in Introduction).

Only a few researchers proposed the utilization of recommender systems for mitigating fake news, such as fact-check URL recommendation to mitigate fake news [37, 55]. However, these methods only recommend limited fact-checked URLs to a small group of users, i.e., fact checkers. In real-world cases, it is more common to recommend news (rather than URLs) which may be fact-checked or unchecked to all potential users of a social network. To the best of our knowledge, there is no existing technical work that directly recommends news to all users to mitigate fake news.

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 which are sequentially interacted by a given user u. Each news is indexed by its interaction timestamp where $\mathcal{V} = \mathcal{V}_{true} \cup \mathcal{V}_{fake}$ denotes the whole news set which consists of both true news and fake news. In addition, a news information table $\mathcal{N}$ records the meta information (i.e., news title and news description) of each news piece occurred in $\mathcal{D}$.

For each user u, given the $(t-1)$ historical true and/or fake news pieces interacted by u, denoted as context $C_u = \{v_1, \cdots, v_{t-1}\}$ together with the news label and meta information, we build a recommendation model $\mathcal{M}$ (i.e., Rec4Mit). $\mathcal{M}$ aims to learn the user's dynamic reading preference from C_u and predict the label (i.e., true or fake) of each piece of candidate news based on its meta information. As a result, $\mathcal{M}$ generates a recommendation list of predicted true news which can best satisfy the user's reading preference at the moment. In other words, for the sake of mitigating fake news, the recommended news should not only be closely relevant to the user's reading history but also be true.

4 THE REC4MIT MODEL

As shown in Figure 1, the Rec4Mit model contains three main modules: (1) *event-veracity disentangler*, (2) *event detection and transition module*, and (3) *next-news prediction module*. First, given a context consisting of a sequence of $(t-1)$ interacted news pieces, the event-label disentangler disentangles the embedding of each news piece into two parts: event embedding which represents the underlying event(s) behind this news and veracity embedding which encodes the veracity of the news. Then, taking the event embedding as the input, the event detection and transition module detects which event(s) is associated with each contextual news piece and then comprehensively model the complex transitions of events over the sequence of contextual news. Consequently, the user's current preference towards events is well captured. Finally, the next-news prediction module, on the one hand, takes the captured preference as the input to predict the next news which can best match the user's preference. On the other hand, it takes each candidate news' label embedding as the input to predict its veracity label. As a result, a list of predicted true news which can best match the user's reading preference is selected as the recommendation result. Next, we introduce the three modules successively.

4.1 Event-veracity Disentangler

To avoid the interference between event information and label information of a piece of news (cf. CH3 in Introduction), we devise a novel event-label disentangler to effectively disentangle the information for these two parts from a news piece. Event-label disentangler is built on the basis of an adversarial auto-encoder framework [7] which consists of *an encoder*, *an event decoder*, *a label decoder* and a specially designed *loss function module*.

4.1.1 *Encoder.*

An encoder is a three-layer network architecture which takes the news embedding as the input. Given a piece of news v_i, we learn an informative news embedding $\mathbf{v}_i \in \mathbb{R}^{dim(v)}$ by combining the embedding $\mathbf{v}_i^{id}$ of news ID and the embedding $\mathbf{v}_i^{m}$ of news meta information, i.e., news title and description texts. The former encodes the news co-occurrence information while the latter encodes the news content information. Formally,

$$\mathbf{v}_i = FC([\mathbf{v}_i^{id}; \mathbf{v}_i^{m}]), \tag{1}$$

where $[\mathbf{a}; \mathbf{b}]$ indicates the concatenation of vectors $\mathbf{a}$ and $\mathbf{b}$ and FC indicates a fully connection layer. $\mathbf{v}_i^{m}$ is a concatenation of both news title embedding and news description embedding. Following the common practice in text embedding [14, 56], both types of embedding are first obtained with the commonly used pre-trained BERT model and then are fine-tinned on our task. The dimension of $\mathbf{v}_i^{id}$ and $\mathbf{v}_i$ are empirically set to 128 and 256 respectively using the grid search method. The dimension of title embedding and description embedding is set to 768 by default in BERT model. Once the news embedding $\mathbf{v}_i$ is ready, it is fed into encoder to obtain the hidden state $\mathbf{h}_i$. The operations in the encoder are specified as:

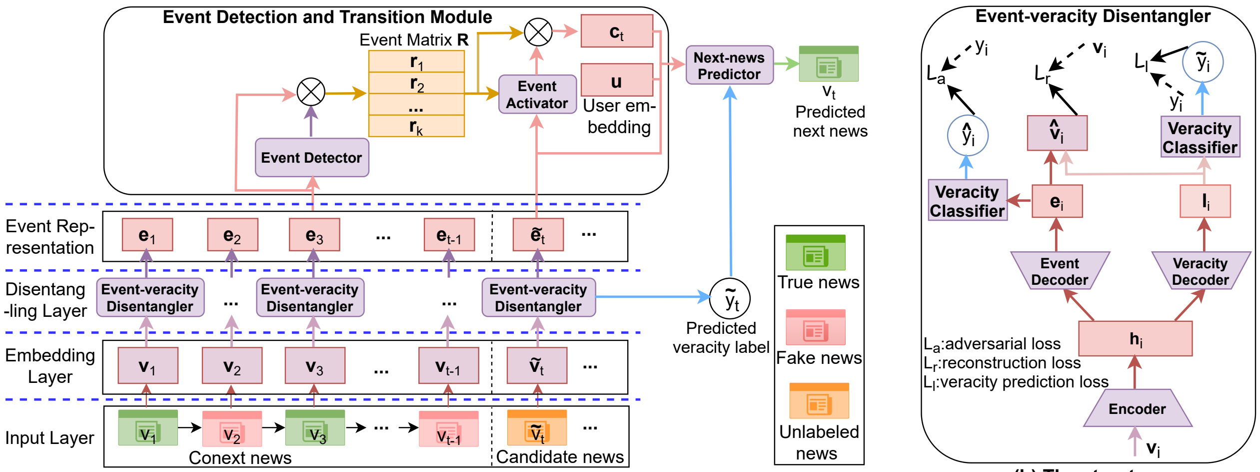
$$\mathbf{z}_i^1 = LeakyReLU(Dense(\mathbf{v}_i)), \tag{2}$$

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

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

where the parameter α in LeakyReLU is empirically set to 0.1. We utilize the residual connections for efficient feature extraction. *Dense* indicates a dense layer.

Once the hidden state $\mathbf{h}_i$ of news v_i is ready from encoder, it will be taken as the input of both event decoder and label decoder.



(a) The overall structure of Rec4Mit model

(b) The structure of Event-veracity Disentangler

Figure 1: (a) Rec4Mit is built on three main components: Event-veracity Disentangler, Event Detection and Transition Module, and Next-news Predictor; (b) The Event-veracity Disentangler is built on the Encoder, Event Decoder, Veracity Decoder and Veracity Classifier.

From the representation learning perspective, $\mathbf{h}_i$ can be regarded as the high-level abstract representation of news v_i.

4.1.2 Event Decoder. An event decoder extracts the event-specific information of a piece of news from the news abstract representation $\mathbf{h}_i$. Similar to encoder, the structure of an event decoder is also a three-layer network with residual connections. Specifically,

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

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

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

where the α in LeakyReLU is set to 0.1. The output $\mathbf{e}_i$ of the event decoder is the latent representation of event(s) behind news v_i.

4.1.3 Veracity Decoder. The label decoder contains a normal decoder which has the same structure as that of an event decoder and a news veracity classifier. The classifier is designed to predict the label of the news v_i by taking the output of the normal decoder as the input. By minimizing the difference between the predicted veracity label and the ground-truth label in training set, we can enforce the label decoder to extract the veracity-specific information. In the test stage, this veracity classifier is used to predict the label of each candidate news so that only true news would be recommended. To be specific, taking the news abstract representation $\mathbf{h}_i$ as the input, the normal decoder performs the operations similar to those described in Eqs. (5) to (7) to output the label representation $\mathbf{l}_i$ of news v_i. Then, the classifier built on a sigmoid layer takes $\mathbf{l}_i$ as the input to predict the label $\tilde{y}_i$ of v_i,

$$\tilde{y}_i = sigmoid(Dense(\mathbf{l}_i)). \tag{8}$$

We use sigmoid layer here since it is commonly used for binary classifications and we leave the exploration of more powerful classifier as the future work.

4.1.4 Loss Function Module. To well train event-label disentangler, we specially design three losses to optimize it: *reconstruction loss*, *label prediction loss*, *adversarial loss*. During the training stage, these loss will be as part of total loss. Reconstruction loss is natural for any auto-encoder structure. Here, we concatenate the disentangled event representation vector $\mathbf{e}_i$ and label representation vector $\mathbf{l}_i$ to reconstruct the news embedding $\mathbf{v}_i$ to ensure there is no information loss during the disentangling process. Formally,

$$\mathcal{L}_r = \frac{1}{2}([\mathbf{e}_i; \mathbf{l}_i] - \mathbf{v}_i)^2. \tag{9}$$

Label prediction loss is to ensure that the disentangled label representation $\mathbf{v}_i$ can correctly indicate the label (true or fake) of the news. Here we use the commonly used binary cross entropy loss [57] to quantify such loss from one piece of news v_i:

$$\mathcal{L}_l = -y_i log(\hat{y}_i), \tag{10}$$

where y_i and $\hat{y}_i$ are the ground-truth label and predicted label respectively.

Adversarial loss is to ensure the disentangled event representation is pure and does not contain label-related information. To achieve this goal, we use the disentangled event representation $\mathbf{e}_i$ of news v_i to predict the label v_i and then maximize the prediction error (minimize the reciprocal of the error),

$$\mathcal{L}_a = -\frac{1}{y_i log(\widetilde{y}_i)}, \tag{11}$$

where $\widetilde{y}_i = sigmoid(Dense(\mathbf{e}_i))$.

Finally, the overall loss of the event-label disentangler is the sum of all the aforementioned three losses:

$$\mathcal{L}_d = \mathcal{L}_r + \mathcal{L}_l + \mathcal{L}_a. \tag{12}$$

4.2 Event Detection and Transition Module

The event detection and transition module consists of two novel components: (1) *an event detector* to detect the possible latent event(s) associated with each news (no explicit event information is available), and (2) *an event transition net* to model the complex transition relations within and between events over a sequence of context news read by a user. These two particularly designed components equip the module with the capability to identify which event(s) the user is focusing on and how her/his reading preference regarding event is changing along with her/his reading history.

4.2.1 Event Detector. Given the event representation $\mathbf{e}_i$ of news v_i, we employ a softmax [12] layer to obtain an event distribution vector $\beta_i \in \mathbb{R}^k$ to indicate the possibility of v_i belonging to each of the k latent events:

$$\beta_i = softmax(\mathbf{e}_i \mathbf{W}_1), \tag{13}$$

where $\mathbf{W}_1 \in \mathbb{R}^{dim(e)*k}$ is a learnable weight matrix. The event number k is a hyper-parameter to be tuned according to the diversity of contents in the news meta data and more diverse contents indicate a larger k. In this work, k is empirically set to 10 and 20 on Politi dataset and Gossip dataset respectively via the validation set.

Once β_i is ready, we split the event representation $\mathbf{e}_i$ into k event-specific representation vectors where each corresponds to one latent event. For instance, the j^{th} event-specific representation for news v_i is calculated as:

$$\mathbf{e}_{i,j} = \beta_{i,j} \mathbf{e}_i, \tag{14}$$

where $\beta_{i,j}$ denotes the value on the j^{th} dimension of the vector β_i.

4.2.2 Event Transition Net. Event transition net contains *an event memory module* to store the detected k event-specific representations for each piece of news in the context and *an event activator* to activate those events in the memory matrix which are closely related to the next target news.

Once each piece of news in a given context C has been processed by the event detector, we obtain an event-specific representation tensor $\mathcal{T} \in \mathbb{R}^{|C|*k*dim(\mathbf{e}_{i,j})}$ where $|C|$ is the length of the context, i.e., the number of news pieces in the context. To obtain a cohesive representation for each event, we attentively aggregate the event-specific representations from all of the context news in C. Specifically, in order to model the sequential dependencies between news in C, we concatenate the learnable position embedding $\mathbf{p}_i$ of news v_i and its event representation $\mathbf{e}_i$ as the input of the attention model, i.e., $\mathbf{f}_i = [\mathbf{e}_i; \mathbf{p}_i]$ Accordingly, the attentive aggregation weight vector $\gamma \in \mathbb{R}^{|C|}$ is calculated as:

$$\gamma = softmax(tanh(\mathbf{F}\mathbf{W}_2)\mathbf{W}_3), \tag{15}$$

where the i^{th} row of matrix $\mathbf{F}$ equals to $\mathbf{f}_i$. $\mathbf{W}_2$ is a learnable projection matrix to project $\mathbf{F}$ to a low dimensional space and $\mathbf{W}_3$ is another learnable matrix for the attention model.

Then, the cohesive representation $\mathbf{r}_j$ of the j^{th} event in C is calculated by aggregating the j^{th} event-specific representation of each context news in C with the attention vector γ:

$$\mathbf{r}_j = \sum_{i=1}^{|C|} \gamma_i \mathbf{e}_{i,j}, (j \in \{1, ..., k\}). \tag{16}$$

Once the cohesive representation of each of the k latent events is ready for a given context C, C can be represented by an event-aware matrix $\mathbf{R} \in \mathbb{R}^{k*dim(\mathrm{r})}$ whose j^{th} row equals to $\mathbf{r}_j$.

Given a candidate news piece v_t, only those relevant events from context C are activated to build a candidate news-specific context embedding for better calculating the relevance between C and v_t. Specifically, the disentangled event representation $\mathbf{e}_t$ of v_t is used as the activation signal. Formally, the activation degrees of all the k events behind v_t is calculated as a k-dimensional activation vector:

$$\delta = softmax(\mathbf{R}\mathbf{e}_t), \tag{17}$$

where the value on the j^{th} dimension of δ indicates the activation degree of the j^{th} event for news v_t.

Now, we can calculate the candidate news-specific context embedding $\mathbf{c}_t$ by carefully aggregating the representations of those events in C which are relevant to the given candidate news v_t:

$$\mathbf{c}_t = \sum_{j=1}^{k} \delta_j \mathbf{r}_j. \tag{18}$$

In addition to the news context C, user ID information is also taken as an input to benefit the personalization of the recommendation since it records the relatively stable user's preference embedded in the whole dataset. Accordingly, the user-aware and candidate news-specific context embedding is calculated as:

$$\mathbf{c}_{tu} = tanh([\mathbf{c}_t; \mathbf{u}]\mathbf{W}_4), \tag{19}$$

where $\mathbf{W}_4$ is a learnable matrix for the non-linear transformation.

4.3 Next-news Prediction Module

Once $\mathbf{c}_{tu}$ is ready, it is input into the next-news prediction module to predict the next news for user u. Specifically, first, the event-label disentangler is used to disentangle the event representation $\mathbf{e}_t$ and label representation $\mathbf{l}_t$ for each candidate news v_t. Then, the commonly used inner product operation is performed on $\mathbf{c}_{tu}$ and $\mathbf{e}_t$ to calculate a score to quantify the relevance between the given context C_u and each candidate news v_t from the event perspective. Finally, sigmoid function transfers the score to a probability value which indicates the possibility of news v_t to be liked by user u:

$$p_t = sigmoid(\mathbf{c}_{tu} \cdot \mathbf{e}_t). \tag{20}$$

At the same time, to only recommend true news to users, the label $\hat{y}_t$ of each candidate news v_t is predicted using the veracity classifier (cf. Eq (8)) by taking $\mathbf{l}_t$ as the input. Consequently, out of the predicted true candidate news, those top-K news with the highest possibilities are selected to form the recommendation list.

4.4 Model Optimization and Training

4.4.1 The Loss Function. Since multiple tasks are jointly optimized, the total loss consists of three parts: (1) the loss generated during the next-item prediction task, denoted as $\mathcal{L}_p$, (2) the loss $\mathcal{L}_d$ (cf. Eq (12)) generated during the disentangling operation on context news, denoted as $\mathcal{L}_d^{con}$, and (3) the loss $\mathcal{L}_d$ generated during the disentangling operation on candidate news, denoted as $\mathcal{L}_d^{can}$. As a result, the total loss $\mathcal{L}$ for our model is defined as:

$$\mathcal{L} = \mathcal{L}_p + \mathcal{L}_d^{con} + \mathcal{L}_d^{can}. \tag{21}$$

The calculation of the loss $\mathcal{L}_d$ has been introduced in Section 4.1.4, and thus we only introduce how to calculate $\mathcal{L}_p$. We use the commonly used cross-entropy loss to specify the loss $\mathcal{L}_p$ for the next-item prediction task. Specifically, given a news context C, a contrastive pair $\langle s^+, S^- \rangle$ is built by taking the ground truth next news v_s as the positive sample s^+ and randomly sampling n pieces of

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

news from the news set $\mathcal{V} \setminus v_s$ as the negative sample set S^-. The loss from the positive sample and each negative sample are $log(p_s^+)$ and $log(1-p_{s^-})$ respectively where p_s^+ is the predicted probability (cf. Eq (20)). Hence, the loss from one contrastive pair is calculated:

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

4.4.2 Model Training. Our model is implemented using Tensorflow 1.13 running under Python 3.6 environment. Model parameters are learned by minimizing the total loss $\mathcal{L}$ based on a mini-batch learning procedure. The Adam optimizer [19] is used for gradient learning. The initial learning rate is determined by grid search in the range of [0.0006, 0.0014] with a step=0.002 and is empirically set to 0.001. The batch size is empirically set to 64. The number of negative samples for training is set to 4. All parameters are tuned on the validation set. Our experiments are run on a cluster where each node is with a setting of 32Core, 2.0Ghz CPU and 128G RAM.

5 EXPERIMENTS AND EVALUATIONS

5.1 Data Preparation

Public datasets for conventional news recommendation task lack of news veracity label, while datasets for fake news detection/mitigation task usually lack of the reading history of each individual user. Hence, there is no existing dataset ready for our experiments. Fortunately, we found the publicly available and commonly used FakeNewsNet dataset[1] [31] not only has label for each piece of news, but also hides users' reading history information in it. To the best of our knowledge, this is the only public dataset which can be used for our experiments. To this end, we processed FakeNewsNet dataset to extract the sequence of news (contain true and/or fake news) read by each user on Twitter. FakeNewsNet contains two comprehensive data sets (i.e., PolitiFact and GossipCop) with diverse features in news content, social context, and spatiotemporal information. In each dataset, for each user, we extract all the news read by her/him, and then order these pieces of news by the reading timestamp to form a news sequence $\mathcal{S}$. All the sequences together form the user-news interaction dataset $\mathcal{D}$. At the same time, the meta information including the title and one-paragaph description of each news is extracted to form the news information table $\mathcal{N}$ (cf. Section 3).

Following common practice in processing sequence data [43, 45], given a user u's news sequence $\mathcal{S}$, we build sequence instance(s) for training and test. Each instance is of length t and is in the form of context-target news pair $\langle C, v_t \rangle (C = \{v_1, \cdots, v_{t-1}\})$ where the context C, user ID and the news meta information are used as the input to predict the target next news v_t for user u. Note that, to mitigate fake news, only really true news is expected to be recommended to users. Hence, only those instances with a ground-truth true target news piece are kept in our experiments. The commonly used sliding window technique [35] is used to split $\mathcal{S}$ into multiple instances when $|\mathcal{S}| > t$, while padding and masking technique [5, 34] is used when $|\mathcal{S}| < t$. To feed the data into the model, following the common practice [41, 45], t is fixed to 5 in this paper. Finally, we randomly select 10%, 20% and 30% of the built sequence instances to form three test sets. Similarly, we select 10% as the corresponding validation set while the remainder for the corresponding training set respectively. For each split ratio, we perform 10-fold cross validation and report the average result. Our method consistently outperforms all the baselines on all the three splits, and only the results w.r.t. the 10% split are reported for saving space. The characteristics of experimental datasets are shown in Table 1.

[1] https://github.com/KaiDMML/FakeNewsNet

5.2 Experiment Settings

5.2.1 Baseline Methods. Our task to recommend the next news piece is actually a sequential/session-based news recommendation task [30, 43, 46]. Hence, we select the representative and/or state-of-the-art approaches for both general (item) sequential recommendation and sequential news recommendation as baselines approaches. This results in the following nine approaches built on various models including nearest neighbour model, memory network, recurrent neural network (RNN), convolutional neural network (CNN), graph neural network (GNN), and attention model. Note that the approaches for combating fake news via URL recommendation [37, 55] take the recommendation as a normal user-URL interaction matrix completion task without modeling the sequential behaviours of each user. Therefore, they have different task setting and input data from ours and thus can not be used as baselines.

- **SKNN**: a sequence and time aware neighborhood model for next-item recommendation [10].
- **CSRM**: a neighborhood session enhanced framework based on memory network and RNN for next-item recommendation [40].
- **SR-GNN**: a representative GNN based model for next-item recommendation [52].
- **SASRec**: a representative self-attention based model for next-item recommendation [15].
- **DAN**: a deep neural network which combines RNN, CNN and attention model for next-news recommendation [61].
- **NRMS**: a multi-head self-attention based model for news recommendation according to each user's reading history [51].
- **LSTUR**: an RNN and attention based model for next-news recommendation. It learns users' long and short term preferences [2].
- **FedNewsRec**: a decentralized next-news recommendation model based on RNN, CNN and attention mechanism [25].
- **FIM**: a 3D CNN-based next-news recommendation model which preforms fine-grained interest matching [38]. It is claimed to be the state-of-the-art approach in the literature [14, 56]

5.2.2 Evaluation Metrics. Given the unique *goal* (i.e., to mitigate fake news) of our approach, we evaluate the performance of all the approaches from two perspectives: (1) the prediction accuracy to measure whether an recommendation approach can accurately recommend the right news to well match a user's reading preference, and (2) the ratio of true news included in the recommendation list

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

	PolitiFact						GossipCop					
	REC@5	**REC@20**	**MRR@5**	**MRR@20**	**NDCG@5**	**NDCG@20**	**REC@5**	**REC@20**	**MRR@5**	**MRR@20**	**NDCG@5**	**NDCG@20**
SKNN	0.2176	0.6088	0.1171	0.1553	0.1414	0.2524	0.1697	0.6252	0.0394	0.0911	0.0703	0.2074
CSRM	0.3752	0.6773	0.2629	0.2923	0.2906	0.3763	0.4764	0.6387	0.3496	0.3661	0.3813	0.4281
SR-GNN	0.3678	0.6741	0.2562	0.2865	0.2837	0.3711	0.4920	0.6239	0.3933	0.4067	0.4180	0.4560
SASRec	0.2962	0.6608	0.1582	0.1933	0.1924	0.2954	0.2419	0.4655	0.1009	0.1244	0.1358	0.2010
DAN	0.1874	0.7405	0.0784	0.1338	0.1049	0.2637	0.3174	0.4541	0.3157	0.3257	0.3161	0.3512
NRMS	0.4752	<u>0.8260</u>	0.3103	0.3449	0.3511	0.4511	0.6354	0.8239	0.4505	0.4702	0.4966	0.5516
LSTUR	<u>0.4827</u>	0.8111	<u>0.3166</u>	<u>0.3491</u>	<u>0.3577</u>	<u>0.4515</u>	<u>0.6950</u>	<u>0.8817</u>	<u>0.4955</u>	<u>0.5156</u>	<u>0.5454</u>	<u>0.6005</u>
FedNewsRec	0.3584	0.7949	0.1940	0.2377	0.2344	0.3596	0.2267	0.4892	0.1248	0.1498	0.1499	0.2237
FIM	0.3711	0.7042	0.1930	0.2250	0.2371	0.3311	0.3521	0.5911	0.2340	0.2570	0.2631	0.3312
Rec4Mit	**0.5561***	**0.8868***	**0.3462***	**0.3808***	**0.3979***	**0.4944***	**0.7552***	**0.9543***	**0.4984***	**0.5205***	**0.5625***	**0.6220***
Improvement[2]	15.21%	7.36%	9.35%	9.08%	11.24%	9.50%	8.66%	8.23%	0.59%	0.95%	3.14%	3.58%

[2]The improvement over the best-performing baseline methods whose performance is underlined.

to measure whether an approach is able to recommend sufficient true news to counteract fake news. On the one hand, following the common practice [2], the prediction accuracy is assessed by how well the ground-truth next news is ranked in the recommendation list, the higher the ranking, the better the performance. We adopt three widely used ranking-based metrics: recall, mean reciprocal rank (MRR) and normalized discounted cumulative gain (NDCG) [48], which are commonly used to evaluate the next-news recommendation performance [25, 38]. We calculate each of these metrics based on the top K ranked news (denoted as REC@K, MRR@K and NDCG@K respectively). On the other hand, the *ratio of true news* (denoted as RT) in each recommendation list is a novel metric proposed by us. It is tailored for fake news mitigation and is defined:

$$RT@K = \frac{\#True\ news}{K} * 100\%. \tag{23}$$

For all these four metrics, we report them at K = 5 and K = 20 respectively in our experiments. Following [40], a paired t-test with p<0.05 is used for significance test.

5.2.3 Parameter Settings. To ensure fair comparison, all the model parameters including hyper-parameters of both baseline methods and our method are well tuned in the same way on validation set. Specifically, for the parameters in each baseline, we first initialize them with the values reported in the original paper and then carefully tune them on our datasets for best performance. The embedding dimension for user ID and news ID are set to 32 and 128 respectively using grid search with a step of 4 and 16 respectively in all methods. Other important parameters for each baseline will be specified during the experimental result analysis in Section 5.3. In our model, the dimension of position embedding is set to 32 using grid search with a step of 4. The number of latent events k is empirically set to 10 on PolitiFact and 20 on GossipCop dataset via the validation set. The number of training epochs for both datasets is set to 15. We set the number of negative samples in training to 4 (2 true news pieces plus 2 fake news pieces).

5.3 Performance Comparison with Baselines

5.3.1 Comparisons w.r.t. Prediction Accuracy. The prediction accuracy of our proposed Rec4Mit and those of the nine baselines are reported in Table 2. For the best performance, we carefully tuned the parameters of each baseline on validation set. In SKNN, the number of neighborhoods is set to 500. In CSRM, the number of nearest neighbors in the OME component is set to 256. In SR-GNN, the decay ratio of learning rate is set to 0.1 and the L2-penalty is set to 10^{-5}. In SASRec, the dropout rate is set to 0.25. DAN requires entity as the additional input. We directly treat each word as an entity to make the model run since there is no entity information in our datasets. In NRMS and FedNewsRec, the number of heads for self-attention is set to 14. In LSTUR, the number of filters in CNN is 300, and the window size of the filters is set to 3.

Overall, the representative general sequential/session-based recommendation approaches including SKNN, CSRM, SR-GNN and SASRec performs worse compared with the representative sequential news recommendation approaches such as NRMS and LSTUR. This demonstrates the superiority of sequential news recommender systems in handling the news data with unique characteristics, such as the existence of news meta information (e.g., news title and news description) and the strong real-time nature of news events. Out of the last five baseline approaches designed for news recommendation, DAN relies on the extra entity information which does not exist in our dataset and thus it performs not so well. FedNewsRec is originally designed for privacy-preserving purpose and should be trained in a special decentralized way. Due to the limitation of available devices, we can only train FedNewsRec in a general way, which may limit its performance [25]. Interestingly, FIM does not perform well on our datasets, this may be caused by the lack of capability to explicitly model the sequential dependency which is very important in our datasets. owing to the well modelling of both users' sequential behaviour information and the rich semantic information embedded in news meta data, NRMS and LSTUR achieve good performance. Although promising, on one hand, these sequential news recommendation approaches still fail to model the events and event transitions behind news sequence, which are one of the most important driving force of users' reading behaviours. On the other hand, they lack the capability to judge whether a piece of news is true or fake before recommending it to users and thus cannot be employed for mitigating fake news.

In contrast, our proposed Rec4Mit not only carefully takes both sequential behaviour information and news meta information into account, but also carefully models each user's reading preference

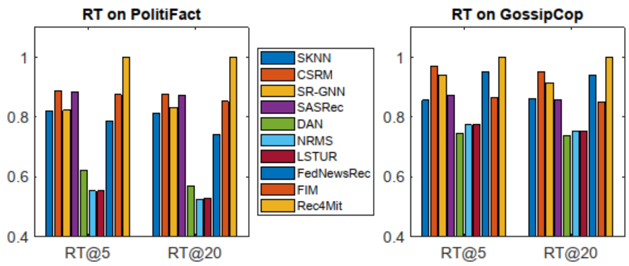


Figure 2: The ratio of true news (RT) in recommendation lists.

transitions over the news events she/he is focusing on along with the time. As a result, Rec4Mit can achieve the best performance on both datasets. Specifically, Rec4Mit consistently and significantly outperforms the best-performing baseline(s) from 0.59% to 15.21% with an average of 7.24% w.r.t. accuracy (cf. the last row in Table 2).

5.3.2 Comparisons w.r.t. the Ratio of True News. The ratio of true news (RT) in the recommendation list generated by each of the compared approaches is reported in Figure 2. It is clear that our proposed Rec4Mit outperforms all compared baseline methods with a clear margin w.r.t. RT@5 and RT@20 on both datasets. This is natural since all the baseline approaches are not capable of predicting the veracity of candidate news and thus they recommend both true and/or fake news to a user as long as the user likes it. In contrast, in our proposed Rec4Mit, a special news veracity classifier is designed to predict weather a candidate news piece is true or fake and only those predicted true news are recommended to users. Hence, owing to the high accuracy of the classifier (F1 > 0.98 on both datasets), Rec4Mit is able to recommend nearly only really true news to each user (RT@5 and RT@20 equal to 100% on PolitiFact, and 99.96% and 99.91% respectively on GossipCop). This would significantly benefit the mitigation of fake news.

It should be noted that the absolute value of RT is also high for some baselines including CSRM, SASRec, FIM, etc. The reason is that the true news is much more frequent than fake news in PolitiFact dataset while there is much more true news than fake news in GossipCop dataset. Specifically, the average frequency of true news and fake news are 483.83 and 276.81 respectively in PolitiFact dataset, and the number of true news and fake news are 6,792 and 2,737 respectively in GossipCop dataset. Such imbalance leads the models to recommend true news with much larger possibilities.

5.4 Ablation Study

5.4.1 Settings. To analyze the rationality and the effectiveness of the designed components in our model, we conduct an ablation study. Specifically, we compare the next-news prediction performance of the original model Rec4Mit with that of its three variants:

- **i. Rec4Mit-Disen**: The variant which removes the event-label disentangler (cf. Section 4.1) by taking the input of the disentangler (i.e., news embedding $\mathbf{v}_i$) as its outputs (i.e., event representation $\mathbf{e}_i$ and label representation $\mathbf{l}_i$ of the news v_i).
- **ii. Rec4Mit-Event**: The variant which removes event detector (cf. Section 4.2.1). This variant does not model the different events behind news pieces and thus the complex transitions over multiple events are simplified to the transitions within one event.
- **iii. Rec4Mit-Label**: The variant which removes the news veracity classifier insider the label decoder (cf. Section 4.1.3). Consequently, no task for news label prediction will be performed.

5.4.2 Findings. The results are reported in Table 3 (see the appendix) and Rec4Mit significantly outperforms each of its three variants with $p < 0.05$. From this table, we have the following findings:

- *Finding 1: Event-label disentangler can obviously boost the next-news prediction accuracy* by disentangling the event-specific and label-specific information from news to avoid the interference between them. Specifically, by comparing the values in the first two rows in Table 3, we can see that the disentangler leads to an improvement of from 0.51% to 4.27% with an average of 2.08% on PolitiFact dataset and from 3.18% to 12.81% with an average of 9.32% on GossipCop dataset.
- *Finding 2: The event detector can greatly improve the performance* by modelling the possible multiple latent events. By comparing the values in the first row and the third row in Table 3, we can see that the event detector leads to an improvement w.r.t. all the metrics ranging from 0.97% to 4.47% with an average of 3.01% on PolitiFact dataset and from 2.08% to 10.19% with an average of 7.54% on GossipCop dataset.
- *Finding 3: The veracity classifier and the label prediction task clearly benefit the next-news prediction task.* This can be verified by the improvement of all the metrics ranging from 0.35% to 1.05% with an average of 0.76% on PolitiFact dataset and from 11.67% to 16.23% with an average of 14.39% on GossipCop dataset when the first row and the fourth row in Table 3 are compared.

6 CONCLUSIONS

In this paper, we focused on a novel and significant research problem: how to effectively and concretely mitigate fake news from the individual user's perspective? To address this problem, we have proposed a novel veracity-aware and event-driven news recommender system called Rec4Mit to provide personalized and corrective true news recommendation for mitigating fake news. owing to the special design, Rec4Mit is able to model the event transitions behind a sequence of news and also classify the veracity of each piece of candidate news. As a result, given a user's reading history consisting of a sequence of true and/or fake news, Rec4Mit is able to recommend the most suitable predicted true news to the user to best match her/his reading preference as well as to mitigate fake news. Extensive experiments on real-world datasets demonstrated the superiority of Rec4Mit over the state-of-the-art next-news recommendation methods and verified the rationality and effectiveness of the specific design in Rec4Mit model. Future work includes the exploration of more accurate approaches to better detect the events behind news and model their transitions over the news sequence.

ACKNOWLEDGMENTS

This work was supported by ARC Discovery Project DP200101441.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31, 2 (2017), 211–36.

[2] 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.
[3] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. 2011. Limiting the spread of misinformation in social networks. In *Proceedings of the 20th International Conference on World Wide Web*. 665–674.
[4] Jonathan Clarke, Hailiang Chen, Ding Du, and Yu Jeffrey Hu. 2020. Fake news, investor attention, and market reaction. *Information Systems Research* 32, 1 (2020), 35–52.
[5] John E Collins and et al. 2012. Incorporating RNA-seq data into the zebrafish Ensembl genebuild. *Genome Research* 22, 10 (2012), 2067–2078.
[6] Mehrdad Farajtabar, Jiachen Yang, Xiaojing Ye, Huan Xu, Rakshit Trivedi, Elias Khalil, Shuang Li, Le Song, and Hongyuan Zha. 2017. Fake news mitigation via point process based intervention. In *Proceedings of the 34th International Conference on Machine Learning*. PMLR, 1097–1106.
[7] 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 & Knowledge Management*. 474–483.
[8] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.
[9] 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*. IEEE, 437–441.
[10] 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.
[11] Xinran He, Guojie Song, Wei Chen, and Qingye Jiang. 2012. Influence blocking maximization in social networks under the competitive linear threshold model. In *Proceedings of the 2012 Siam International Conference on Data Mining*. SIAM, 463–474.
[12] Geoffrey E Hinton and Russ R Salakhutdinov. 2009. Replicated softmax: an undirected topic model. *Advances in Neural Information Processing Systems* 22 (2009), 1607–1614.
[13] 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.
[14] 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.
[15] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *Proceedings of the 2018 IEEE International Conference on Data Mining*. IEEE, 197–206.
[16] Mozhgan Karimi, Dietmar Jannach, and Michael Jugovac. 2018. News recommender systems–Survey and roads ahead. *Information Processing & Management* 54, 6 (2018), 1203–1227.
[17] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the 9th SIGKDD Conference on Knowledge Discovery and Data Mining*. 137–146.
[18] Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. 2019. Mvae: Multimodal variational autoencoder for fake news detection. In *Proceedings of 28th International World Wide Web Conference*. 2915–2921.
[19] 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.
[20] Michal Kompan and Mária Bieliková. 2010. Content-based news recommendation. In *Proceedings of the 11th International Conference on Electronic Commerce and Web Technologies*. Springer, 61–72.
[21] Zhongqi Lu, Zhicheng Dou, Jianxun Lian, Xing Xie, and Qiang Yang. 2015. Content-based collaborative filtering for news topic recommendation. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*. 217–233.
[22] Nam P Nguyen, Guanhua Yan, My T Thai, and Stephan Eidenbenz. 2012. Containment of misinformation spread in online social networks. In *Proceedings of the 4th Annual ACM Web Science Conference*. 213–222.
[23] 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.
[24] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. Automatic detection of fake news. In *Proceedings of the 27th International Conference on Computational Linguistics*. 3391–3401.
[25] Tao Qi, Fangzhao Wu, Chuhan Wu, Yongfeng Huang, and Xing Xie. 2020. Privacy-preserving news recommendation model learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 1423–1432.
[26] Feng Qian, Chengyue Gong, Karishma Sharma, and Yan Liu. 2018. Neural user response generator: fake news detection with collective user intelligence. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. 3834–3840.
[27] Julio CS Reis, André Correia, Fabrício Murai, Adriano Veloso, and Fabrício Benevenuto. 2019. Supervised learning for fake news detection. *IEEE Intelligent Systems* 34, 2 (2019), 76–81.
[28] Karishma Sharma, Xinran He, Sungyong Seo, and Yan Liu. 2021. Network inference from a mixture of diffusion models for fake news mitigation. In *Proceedings of the 15th International AAAI Conference on Web and Social Media*. 668–679.
[29] Karishma Sharma, Feng Qian, He Jiang, Natali Ruchansky, Ming Zhang, and Yan Liu. 2019. Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology* 10, 3 (2019), 1–42.
[30] Heng-Shiou Sheu, Zhixuan Chu, Daiqing Qi, and Sheng Li. 2021. Knowledge-guided article embedding refinement for session-based news recommendation. *IEEE Transactions on Neural Networks and Learning Systems* (2021), 1–7.
[31] 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.
[32] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter* 19, 1 (2017), 22–36.
[33] Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond news contents: The role of social context for fake news detection. In *Proceedings of the 12th ACM International Conference on Web Search and Data Mining*. 312–320.
[34] Wenzhuo Song, Shoujin Wang, Yan Wang, and Shengsheng Wang. 2021. Next-item recommendations in short sessions. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 282–291.
[35] 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.
[36] Guangmo Tong. 2020. StratLearner: learning a strategy for misinformation prevention in social networks. *Advances in Neural Information Processing Systems* 33 (2020), 1–10.
[37] Nguyen Vo and Kyumin Lee. 2018. The rise of guardians: Fact-checking url recommendation to combat fake news. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 275–284.
[38] 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.
[39] 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.
[40] 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.
[41] Shoujin Wang, Longbing Cao, Liang Hu, Shlomo Berkovsky, Xiaoshui Huang, Lin Xiao, and Wenpeng Lu. 2021. Hierarchical attentive transaction embedding with intra-and inter-transaction dependencies for next-item recommendation. *IEEE Intelligent Systems* 36, 04 (2021), 56–64.
[42] Shoujin Wang, Longbing Cao, Yan Wang, Quan Z Sheng, Mehmet A Orgun, and Defu Lian. 2021. A survey on session-based recommender systems. *ACM Computing Surveys (CSUR)* 54, 7 (2021), 1–38.
[43] Shoujin Wang, Liang Hu, Yan Wang, Longbing Cao, Quan Z. Sheng, and Mehmet Orgun. 2019. Sequential recommender systems: challenges, progress and prospects. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. AAAI Press, 6332–6338.
[44] Shoujin Wang, Liang Hu, Yan Wang, Xiangnan He, Quan Z Sheng, Mehmet A Orgun, Longbing Cao, Francesco Ricci, and Philip S Yu. 2021. Graph learning based recommender systems: A review. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*. AAAI Press, 4644–4652.
[45] 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.
[46] Shoujin Wang, Liang Hu, Yan Wang, Quan Z Sheng, Mehmet Orgun, and Longbing Cao. 2020. Intention2Basket: a neural intention-driven approach for dynamic next-basket planning. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence*. IJCAI, 2333–2339.
[47] Zhaoxu Wang, Wenxiang Dong, Wenyi Zhang, and Chee Wei Tan. 2014. Rumor source detection with multiple observations: Fundamental limits and algorithms. *ACM SIGMETRICS Performance Evaluation Review* 42, 1 (2014), 1–13.

[48] 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.

[49] 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.

[50] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. Npa: Neural news recommendation with personalized attention. In *Proceedings of the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2576–2584.

[51] Chuhan Wu, Fangzhao Wu, Suyu Ge, Tao Qi, Yongfeng Huang, and Xing Xie. 2019. Neural news recommendation with multi-head self-attention. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 6389–6394.

[52] 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.

[53] Xiaofei Xu, Ke Deng, and Xiuzhen Zhang. 2022. Identifying cost-effective debunkers for multi-stage fake news mitigation campaigns. In *Proceedings of the 15th ACM Conference on Web Search and Data Mining*. 1–9.

[54] Jiahao Xun, Shengyu Zhang, Zhou Zhao, Jieming Zhu, Qi Zhang, Jingjie Li, Xiuqiang He, Xiaofei He, Tat-Seng Chua, and Fei Wu. 2021. Why do we click: visual impression-aware news recommendation. In *Proceedings of the 29th ACM International Conference on Multimedia*. 3881–3890.

[55] Di You, Nguyen Vo, Kyumin Lee, and Qiang Liu. 2019. Attributed multi-relational attention network for fact-checking url recommendation. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1471–1480.

[56] 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.

[57] 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.

[58] 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.

[59] Ke Zhou, Hongyuan Zha, and Le Song. 2013. Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics*. PMLR, 641–649.

[60] Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: fundamental theories, detection methods, and opportunities. *Comput. Surveys* 53, 5 (2020), 1–40.

[61] 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.

7 APPENDIX

7.1 Ablation Study Result

The ablation study result is reported in Table 3.

7.2 Parameter Test

We test the sensitivity of the next-news prediction performance of our proposed model Rec4Mit on two key parameters: the number of latent events k (cf. Section (4.2.1)) and the dimension of news embedding $\mathbf{v}_i$ (cf. Eq (1)).

7.2.1 Performance w.r.t. the number of latent events. We vary the number of latent events from 1 to 19 on PolitiFact dataset with a step of 3 and from 1 to 25 on GossipCop dataset with a step of 5. Figure 3 shows the results on both datasets. The best performance is achieved when the number of latent events is set to 10 and 20 on PolitiFact and GossipCop respectively. Either a smaller or a larger value leads to worse performance. On the one hand, a small value is not adequate to model the possible multiple events with different transition patterns. On the other hand, the number of events is usually limited, and larger values may not match the real-world cases but bring more unnecessary model parameters, thereby hurting the performance.

7.2.2 Performance w.r.t. the dimension of news embedding. We vary the news embedding dimension from 208 to 304 on both datasets with a step of 16. Figure 4 shows the results on PolitiFact and GossipCop datasets. We have found that an appropriate dimension value, i.e., 256 on both datasets, can achieve the best performance by providing an informative representation for each piece of news. A small embedding dimension would fail to embed the rich information from both news occurrence patterns and news meta information, while a larger dimension may bring unnecessary model parameters and thus hurt the performance.

7.3 Case Study

To demonstrate how our proposed model Rec4Mit can benefit fake news mitigation in a more straightforward way, we conduct a case study on one of the real-world datasets (i.e., GossipCop dataset) used in our experiments. To be specific, we sampled 5 users from GossipCop dataset, for each user, we show the given reading history (i.e., context news), and the corresponding recommendation list (i.e., recommended news) generated by our proposed model Rec4Mit.

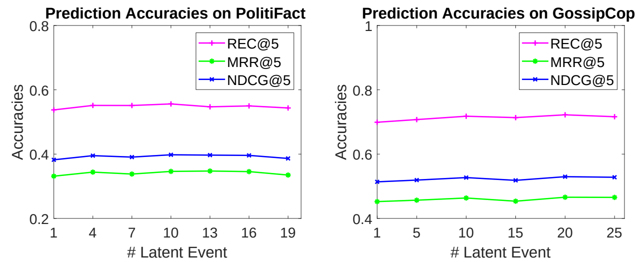


Figure 3: The impact of number of latent events on the prediction performance.

The results are shown in Table 4, where each user has two rows: the first row contains the context news read sequentially by the user and the second row contains the top-5 news pieces recommended to the user according to the given context news. For each piece of news, its veracity is indicated by green (true news) and red (fake news) color. Also, the ground truth news which has been really read by each user is pointed out in the recommend news list. For each user, we use underline to highlight the same/related events/topics from the context news and recommended news. From this table, we have the following findings:

- *Finding 1: Users may read true and/or fake news for the same/related events or for different events.* Table 4 shows that the reading history (i.e., context news) of user_1, user_2 and user_3 contains both true news and fake news, while that of user user_4 and user_5 contains fake news only. This directly verified the unique data characteristics in fake news mitigation which have been introduced in the Introduction part (cf. Paragraph 5 of Introduction on Page 1).
- *Finding 2: For the users who have read fake news for certain events, our model is able to recommend the corresponding true news for the same/related events to counteract the fake news.* For example, user_1 has read a piece of fake news on the event "chris pratt divorce", i.e., CN_4, our model recommended a piece of true news on the same event, i.e., RN_3, and hit the ground truth. Similar cases can be easily found for all the other four users listed in the table. This directly verified the effectiveness of our proposed model in mitigating fake news by recommending corrective true news.

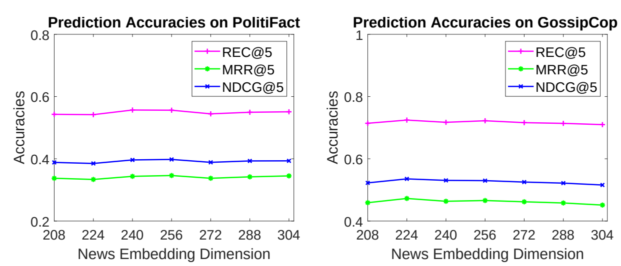


Figure 4: The impact of news embedding dimension on the prediction performance.

Table 3: Comparison of Rec4Mit with its Variants on two real-world datasets, *the improvement is significant at $p < 0.05$.

	PolitiFact						GossipCop					
	REC@5	REC@20	MRR@5	MRR@20	NDCG@5	NDCG@20	REC@5	REC@20	MRR@5	MRR@20	NDCG@5	NDCG@20
Rec4Mit	**0.5561***	**0.8868***	**0.3462***	**0.3808***	**0.3979***	**0.4944***	**0.7552***	**0.9543***	**0.4984***	**0.5205***	**0.5625***	**0.6220***
Rec4Mit-Disen	0.5527	0.8721	0.3377	0.3702	0.3816	0.4919	0.6932	0.9249	0.4418	0.4678	0.5053	0.5737
Rec4Mit-Event	0.5375	0.8783	0.3314	0.3687	0.3823	0.4855	0.6991	0.9349	0.4523	0.4778	0.5138	0.5838
Rec4Mit-Label	0.5502	0.8837	0.3427	0.3786	0.3938	0.4920	0.6673	0.8546	0.4288	0.4495	0.4883	0.5442

Table 4: Recommendation Lists for 5 Users Sampled from GossipCop Dataset.

User$_1$	**Context news (CN)**	CN$_1$: jennifer lawrence says u mother u led to darren split	CN$_2$: where is travis scott why kylie jenner s boyfriend avoids the spotlight	CN$_3$:jennifer lawrence says u mother u led to darren split	CN$_4$:chris pratt files for divorce from anna faris	
	Recomm-ended news (RN)	RN$_1$:tori kelly is engaged to basketball player boyfriend u e	RN$_2$:steven innovative co creator of u nypd blue u u hill street blues u dies at	RN$_3$**(ground truth)**: chris pratt and anna faris finalize divorce one year after separating reports	RN$_4$:rita ora kisses cardi b in the new video for controversial track u girls u	RN$_5$:harvey weinstein timeline how the scandal unfolded
User$_2$	**Context news (CN)**	CN$_1$: selena gomez brings a and a bikini to australia u but not justin bieber	CN$_2$:justin bieber selena gomez their time apart is driving him crazy	CN$_3$:justin bieber and selena gomez may have broken up for good this time	CN$_4$:justin bieber s ex baskin champion wows in a bikini amid his engagement to hailey baldwin	
	Recomm-ended news (RN)	RN$_1$**(ground truth)**:selena gomez u s mom responds to justin bieber relationship rumors	RN$_2$:taylor swift s stalker sentenced to year probation and gps monitoring	RN$_3$:celebrities with tattooed eyebrows including helen mirren rooney michelle	RN$_4$:prince harry and harry styles reunite	RN$_5$:kristen bell hosts sag awards in series of gowns see the stunning looks
User$_3$	**Context news (CN)**	CN$_1$: brad pitt he had a blast playing with kids during secret cambodian family reunion	CN$_2$:kim kardashian responds to claims she was attacked in los angeles such weird rumors	CN$_3$:pepsi pulls controversial kendall jenner ad after outcry	CN$_4$:girls cast spoofs golden girls on jimmy kimmel live	
	Recomm-ended news (RN)	RN$_1$**(ground truth)**:brad pitt u s red carpet surprise at u lost city of z u premiere	RN$_2$:the fast food guide	RN$_3$:kesha s mother drops against dr luke	RN$_4$:jason aldean and wife brittany kerr revealed the gender of their baby in the cutest way	RN$_5$:video justin timberlake announces opening act for man of the woods tour u z
User$_4$	**Context news (CN)**	CN$_1$:selena gomez demi lovato bond over boys possible duet more during epic reunion	CN$_2$:real reason behind justin bieber and selena gomez u s breakup has finally been revealed	CN$_3$:poor joe jonas is trying desperately to look like ex gigi hadid s new boyfriend zayn malik	CN$_4$:katie holmes pushing jamie foxx to go more public with their relationship u why he u s u hesitant u	
	Recomm-ended news (RN)	RN$_1$:first look at ryan murphy s new fox series	RN$_2$:video justin timberlake announces opening act for man of the woods tour u z	RN$_3$:jennifer aniston	RN$_4$:best royal wedding gowns of all time	RN$_5$**(ground truth)**: justin s wife, his character may have relationship issues
User$_5$	**Context news (CN)**	CN$_1$:kim kardashian kanye west snubbed by kate middleton and prince william report debunked	CN$_2$:selena gomez and justin bieber baby news	CN$_3$:kate carrying twins	CN$_4$: ellen degeneres and de rossi rock coordinating outfits to justin timberlake s concert	
	Recomm-ended news (RN)	RN$_1$**(ground truth)**:jessica simpson insists she s not pregnant on ellen	RN$_2$:how e is remaking the people s choice awards for	RN$_3$:louis c k is accused by women of sexual misconduct	RN$_4$:harry styles kisses james corden in christmas u carpool karaoke u	RN$_5$:james corden pays tribute to manchester in heartfelt monologue

Green color indicates true news.
Red color indicates fake news.
_____ is used to highlight the identical/related events/topics from the context news and the recommended news of each user, e.g., "chris prat divorce" appears both in the fourth context news and the third recommended news of User$_1$.
"**ground truth**" means the corresponding recommended news has been really read by the user in the test set.