REGULAR PAPER

Neural news recommendation with negative feedback

Chuhan Wu[1] [iD] **· Fangzhao Wu[2] · Yongfeng Huang[1] · Xing Xie[2]**

Abstract

News recommendation is important for online news services. Precise user interest modeling is critical for personalized news recommendation. Existing news recommendation methods usually rely on the implicit feedback of users like news clicks to model user interest. However, news click may not necessarily reflect user interests because users may click a news due to the attraction of its title but feel disappointed at its content. The dwell time of news reading is an important clue for user interest modeling, since short reading dwell time usually indicates low and even negative interest. Thus, incorporating the negative feedback inferred from the dwell time of news reading can improve the quality of user modeling. In this paper, we propose a neural news recommendation approach which can incorporate the implicit negative user feedback. We propose to distinguish positive and negative news clicks according to their reading dwell time, and respectively learn user representations from positive and negative news clicks via a combination of Transformer and additive attention network. In addition, we propose to compute a positive click score and a negative click score based on the relevance between candidate news representations and the user representations learned from the positive and negative news clicks. The final click score is a combination of positive and negative click scores. Besides, we propose an interactive news modeling method to consider the relatedness between title and body in news modeling. Extensive experiments on real-world dataset validate that our approach can achieve more accurate user interest modeling for news recommendation.

Keywords News recommendation · Dwell time · Negative feedback

1 Introduction

Online news services such as Google News[1] and Microsoft News[2] can collect news from various sources and display them to users in a unified view (Das et al. 2007; Wu et al. 2020). However, a large number of news articles are generated every day and it is overwhelming for users to find their interested news (Okura et al. 2017). Thus, personalized news recommendation is critical for these news services to target user interests and alleviate information overload (Wu et al. 2019c).

Accurate user interest modeling is a core problem in news recommendation. Existing news recommendation methods usually learn representations of users from their historical news click behaviors (Okura et al. 2017; Wang et al. 2018; Wu et al. 2019b, e). For example, Okura et al. (2017) proposed to learn user representations from the clicked news via a gated recurrent unit (GRU) network. Wang et al. (2018) proposed to learn user representations from clicked news with a candidate-aware attention network that evaluates the relevance between clicked news and candidate news. In fact, users usually click news due to their interests in the news titles. However, news titles are usually short and the information they condensed is very limited and even misleading. Users may be disappointed after reading the content of their clicked news. Thus, news click behaviors do not necessarily indicate user interest, and modeling users solely based on news clicks may not be accurate enough.

✉ Chuhan Wu
  wuchuhan15@gmail.com

  Fangzhao Wu
  wufangzhao@gmail.com

  Yongfeng Huang
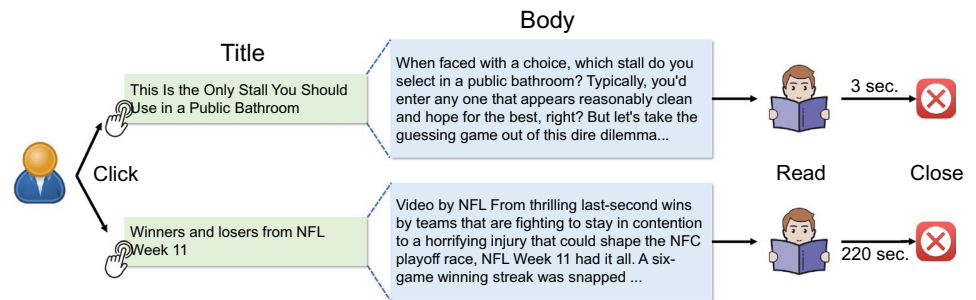 yfhuang@tsinghua.edu.cn

  Xing Xie
  xingx@microsoft.com

[1] Department of Electronic Engineering and BNRist, Tsinghua University, Beijing 100084, China

[2] Microsoft Research Asia, Beijing 100080, China

---

[1] https://news.google.com.

[2] https://www.msn.com.

Fig. 1 Two news clicked by a user. The time represent the news reading dwell time

Besides news clicks, users also provide other implicit feedback like the dwell time of news reading, which is an important indication of user interest. As shown in Fig. 1, the user clicks two news articles and respectively reads them for 3 and 220 s before closing the news webpages. We can infer that this user may not be interested in the first news because she reads the body of this news for only 3 s before her leaving. Thus, a short dwell time of news reading can be regarded as a kind of implicit negative user feedback, which can help model user interest more accurately. In addition, since users usually decide which news to click according to news titles and read news bodies for detailed information, incorporating both news title and body is useful for news and user interest modeling. Moreover, the title and body of the same news usually have some inherent relatedness in describing news content. For example, in Fig. 1 the title of the second news indicates that this news is about NFL games, and the body introduces the details of these events. Capturing the relations between news title and body may help better understand news content and facilitate subsequent user interest modeling.

In this paper, we propose a neural news recommendation approach with negative feedback (NRNF). We propose to distinguish positive and negative news click behaviors according to the reading dwell time, and learn separate user representations from the positive and negative news clicks with a Transformer to capture behavior relations and an additive attention network to select important news for user modeling. In addition, we propose to respectively compute a positive click score and a negative click score based on the relevance between candidate news representation and the user representations learned from positive and negative news clicks. Besides, we propose an interactive news modeling method to incorporate both news title and body by using an interactive attention network to model their relatedness. Massive experiments conducted on real-world dataset demonstrate that our approach can effectively improve the performance of user modeling by incorporating implicit negative user feedback.

2 Related work

News recommendation is an important task in the data mining field, and has gained increasing attention in recent years (Zheng et al. 2018). Learning accurate representations of users is critical for news recommendation. Many existing news recommendation methods rely on manual feature engineering to build user representations (Liu et al. 2010; Capelle et al. 2012; Son et al. 2013; Karkali et al. 2013; Garcin et al. 2013; Bansal et al. 2015; Ren et al. 2015; Chen et al. 2017; Zihayat et al. 2019). For example, Li et al. (2010) proposed to represent users using their demographics, geographic information and behavioral categories which summarize the consumption history on Yahoo!. Garcin et al. (2012) proposed to model users by aggregating the LDA features of all clicked news into a user vector by averaging. Lian et al. (2018) proposed to build user representations from various handcrafted features like user ID, demographics, locations and the feature collection of clicked news. However, these methods rely on manual feature engineering, which needs massive domain knowledge to craft. In addition, handcrafted features are usually not optimal for representing user interest.

In recent years, several news recommendation methods based on deep learning techniques are proposed (Okura et al. 2017; Khattar et al. 2018; Wang et al. 2018; Wu et al. 2019c; An et al. 2019; Wu et al. 2019a, d, e; Ge et al. 2020; Hu et al. 2020). For example, Okura et al. (2017) proposed to learn representations of users from the representations of their browsed news using a GRU network. Wang et al. (2018) proposed to learn user representations from clicked news based on their relevance to the candidate news. Zhu et al. (2019) proposed to generate user representations using a combination of a CNN network and an attention-based LSTM network. Wu et al. (2019e) proposed to use multi-head self-attention network to learn user representations by capturing the interactions between clicked news. These methods only rely on news click behaviors to learn
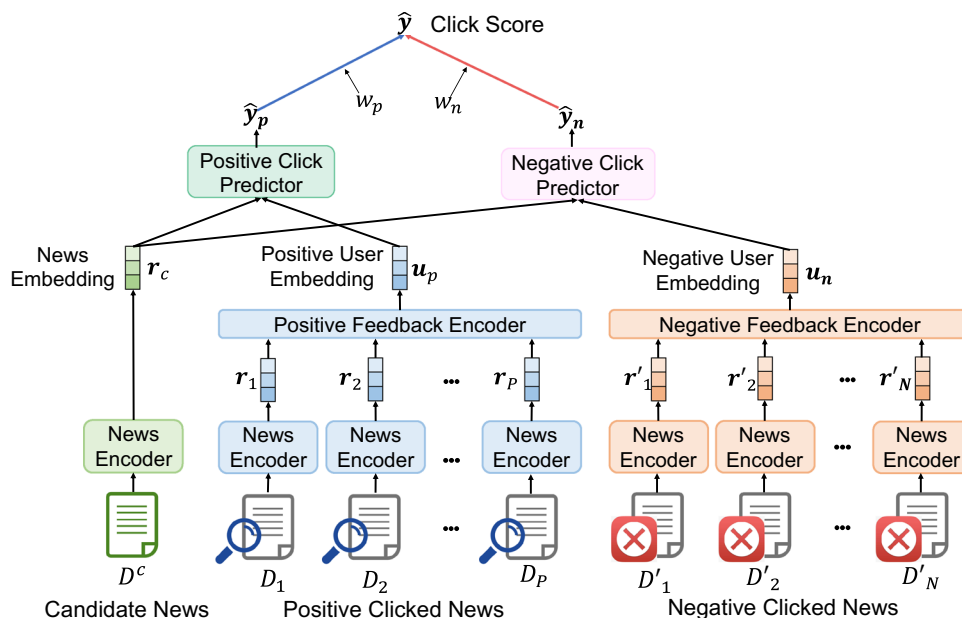
user representations, which may be insufficient because news clicks may not exactly reflect user preferences. Different from these methods, our approach can learn representations of users by incorporating the implicit negative feedback of users inferred from their reading dwell times, which can provide useful clues to calibrate the user interest inferred from news click behaviors.

3 Our approach

In this section, we introduce the details of our neural news recommendation approach with negative feedback (NRNF). The architecture of *NRNF* is illustrated in Fig. 2. We will introduce each core technique in our method, including user modeling, click prediction, news modeling, and model training, in the following sections.

3.1 User modeling with negative feedback

In this section, we introduce how to model user interests from their clicked news by considering negative feedback. Usually, the interests of users on the news they read carefully are very different from those they closed quickly after having a glance at the body. For example, as shown in Fig. 1, we can infer that the user may be indeed interested in the second news since she reads this news for about 220 s. However, she is probably not interested in or unsatisfied with the first news since she reads this news only for 3 s after click. Thus, incorporating short reading dwell time as the implicit negative feedback of users has the potential to enhance the modeling of user interest. To characterize the differences of news

in representing the real user interests, we use their reading dwell times t to distinguish positive news clicks from the negative ones. We regard the clicked news with $t < T$ as negative news clicks and those with $t \geq T$ as positive news clicks, where T is a time threshold. Since simply aggregating all news together will ignore the differences of positive and negative news, in the *user encoder* module we propose to use a multi-view learning framework to incorporate positive and negative news clicks as different views of users. Their details are introduced as follows.

The positive news click view is used to learn a positive user embedding from positive clicked news. We denote the P positive clicked news of a user as $[D_1, D_2, \ldots, D_P]$. We first use a news encoder to obtain their representations, which are denoted as $\mathbf{R}^p = [\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_P]$. Then, we use a positive feedback encoder to learn a unified user embedding from these news representations. It uses a similar architecture with the user encoder used by Wu et al. (2019e). Since different clicked news usually have some interactions with each other, we use a news-level Transformer (Vaswani et al. 2017) to capture the interactions among them,[3] and the output sequence is computed as:

$$\mathbf{D}^p = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_P] = Transformer(\mathbf{R}^p, \mathbf{R}^p, \mathbf{R}^p), \quad (1)$$

where the three parameters in *Transformer*(·) respectively denote the input query, key and value. In addition, since different clicked news have different importance in modeling user interests, we use an additive attention (Yang et al. 2016)

---

[3] We add position embeddings to the news embeddings to capture behavior orders.

network to select important news to learn more informative user representations. The final positive user embedding $\mathbf{u}_p$ learned from this view is computed as:

$$\mathbf{u}_p = Attention(\mathbf{q}^p, \mathbf{D}^p, \mathbf{D}^p), \tag{2}$$

where $\mathbf{q}^p$ is a query parameter vector, $\mathbf{D}^p$ is regarded as the key and value. It summarizes user interests inferred from the positive user feedback.

The negative news click view is used to learn representations of users from their negative clicked news. Similar to positive news, we use a combination of Transformer and additive attention network to learn the negative user embedding from the representations of the negative clicked news. We denote the negative user embedding learned from this view as $\mathbf{u}_n$, which condenses the information of negative user interest inferred from the negative clicked news.

3.2 Click prediction

In this section, we introduce the details of click prediction in our approach. Usually, if a candidate news is similar to the news that a user reads carefully, it may have a high probability to be clicked by this user. However, if it is very similar to the news that are previously closed by a user very quickly, it will probably not be clicked by this user. Motivated by the observations above, we propose to calculate a positive click score $\hat{y}_p$ and a negative click score $\hat{y}_n$ respectively based on the relevance of candidate news to the positive and negative user interest. We denote the embedding of a candidate news D^c as $\mathbf{r}_c$. Following Okura et al. (2017), $\hat{y}_p$ and $\hat{y}_n$ are respectively computed by the inner product of the candidate news embedding and the user embeddings learned from the positive and negative views, which are formulated as:

$$\begin{aligned} \hat{y}_p &= \mathbf{r}_c^T \mathbf{u}_p, \\ \hat{y}_n &= \mathbf{r}_c^T \mathbf{u}_n. \end{aligned} \tag{3}$$

The final click probability score $\hat{y}$ is a weighted summation of the positive score $\hat{y}_p$ and the negative score $\hat{y}_n$, which is formulated as:

$$\hat{y} = w_p \hat{y}_p + w_n \hat{y}_n, \tag{4}$$

where w_p and w_n are learnable parameters. A positive parameter means that the relevance between candidate news and the corresponding user embedding has a positive correlation with the final click score, and this correlation is negative if the parameter is negative.

3.3 News modeling

In this section, we introduce the architectures of the news encoder in our approach for news modeling. It is used to
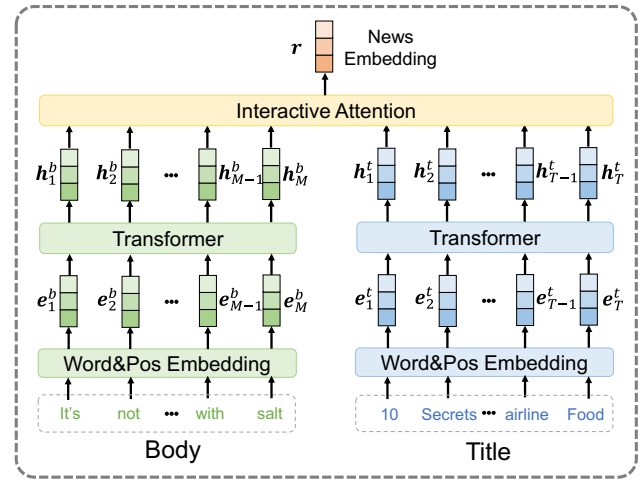

Fig. 3 The architecture of the news encoder

learn representations of news from their titles and bodies. Since news titles and bodies usually have very different characteristics, simply concatenating title and body together may not be optimal. Thus, in this module we use a multi-view learning framework to incorporate the information of title and body as different views of news. In addition, since there is usually inherent relatedness between news title and body, we propose to use an interactive attention network to capture their relations and form the unified news embedding. The details of each module in the news encoder are introduced as follows (Fig. 3).

The first module in the news encoder is a title transformer, which is used to capture the contexts of words in a news title. We first use a word and position embedding layer to convert the sequence of words in a news title into a sequence of low-dimensional word embeddings with their position information. Denote a news title with T words as $[w_1^t, w_2^t, \ldots, w_T^t]$. It is converted into an embedding sequence $\mathbf{E}^t = [\mathbf{e}_1^t, \mathbf{e}_2^t, \ldots, \mathbf{e}_T^t]$. Then, we use a Transformer to learn the contextual representation of each word, which is computed as follows:

$$\mathbf{H}^t = [\mathbf{h}_1^t, \mathbf{h}_2^t, \ldots, \mathbf{h}_T^t] = Transformer(\mathbf{E}^t, \mathbf{E}^t, \mathbf{E}^t). \tag{5}$$

The second module in the news encoder is a body transformer, which aims to capture the contexts of words in a news body. Similar to news title, we also use a word and position embedding layer to obtain the embeddings of the M words in a news body, which are denoted as $\mathbf{E}^b = [\mathbf{e}_1^b, \mathbf{e}_2^b, \ldots, \mathbf{e}_M^b]$. Then, we use a Transformer to capture the contexts of words, and its output $\mathbf{H}^b$ is computed as

$$\mathbf{H}^b = [\mathbf{h}_1^b, \mathbf{h}_2^b, \ldots, \mathbf{h}_M^b] = Transformer(\mathbf{E}^b, \mathbf{E}^b, \mathbf{E}^b). \tag{6}$$

The third module is an interactive attention network, which is used to capture the relations between news title and body

Table 1 Statistics of the dataset

# Users	337,699	Avg. # words per title	10.99
# News	249,038	Avg. # words per body	719.34
# Sessions	500,000	Avg. dwell time	294.91 s
# Samples	40,400,206	Avg. # negative news per user	5.00
# Clicked samples	990,624	# Unclicked samples	39,409,582

when forming the unified news embedding. We first use two additive attention network to respectively attend to the important words in news title and body to form the title and body representations (denoted as $\mathbf{r}^t$ and $\mathbf{r}^b$), which are formulated as

$$\mathbf{r}^t = Attention(\mathbf{q}^t, \mathbf{H}^t, \mathbf{H}^t), \tag{7}$$

$$\mathbf{r}^b = Attention(\mathbf{q}^b, \mathbf{H}^b, \mathbf{H}^b), \tag{8}$$

where $\mathbf{q}^t$ and $\mathbf{q}^b$ are learnable query vectors. Then, we use the title representation as the query of the body attention network to select words in news body according to their relevance to the content of news title, and the output title-aware body representation $\mathbf{c}^b$ is computed as follows:

$$\mathbf{c}^b = Attention(\mathbf{r}^t, \mathbf{H}^b, \mathbf{H}^b). \tag{9}$$

We learn a body-aware title representation $\mathbf{c}^t$ in a similar way by using the body representation as the title attention query, which is formulated as:

$$\mathbf{c}^t = Attention(\mathbf{r}^t, \mathbf{H}^b, \mathbf{H}^b). \tag{10}$$

We finally aggregate all title and body representations into a unified news representation $\mathbf{r}$, i.e., $\mathbf{r} = \mathbf{r}^t + \mathbf{r}^b + \mathbf{c}^t + \mathbf{c}^b$.

3.4 Model training

Following (Huang et al. 2013; Wu et al. 2019b), we use negative sampling techniques to construct labeled data from the raw news impression logs for model training. For each news clicked by a user, we randomly sample Q news that are displayed in the same impression but not clicked by this user. We denote the click probability scores of the clicked news sample as $\hat{y}^+$ and the Q unclicked news samples as $[\hat{y}_1^-, \hat{y}_2^-, \ldots, \hat{y}_Q^-]$. we re-formulate the news click prediction problem as a pseudo $Q + 1$-way classification task to jointly predict these scores. We normalize these scores via the softmax function to compute the posterior click probability of the clicked news, which is formulated as follows:

$$p_i = \frac{\exp(\hat{y}_i^+)}{\exp(\hat{y}_i^+) + \sum_{j=1}^{Q} \exp(\hat{y}_{i,j}^-)}. \tag{11}$$

The final loss function we used is the negative log-likelihood of all clicked news samples, which is formulated as:

$$\mathcal{L} = -\sum_{i \in \mathcal{S}} \log(p_i), \tag{12}$$

where $\mathcal{S}$ is the set of clicked news samples for training.

4 Experiments

4.1 Datasets and experimental settings

Our experiments were conducted on a real-world news recommendation dataset collected from MSN News[4] logs from Oct. 31, 2018 to Jan. 29, 2019. The detailed statistics are shown in Table 1. We sorted all sessions by time, and we used the first 80% of sessions for training, the next 10% for validation and the rest 10% for test.

In our experiments, the word embeddings were 300-dimensional and initialized by Glove (Pennington et al. 2014). The Transformers had 8 attention heads, whose output dimension was 32. The maximum lengths of title and body were respectively set to 30 and 200, respectively. The numbers of positive and negative clicked news were respectively limited to 50 and 20. Following (Huang et al. 2013), the negative sampling ratio Q was set to 4. The threshold T was set to 10 s. Adam (Kingma and Ba 2014) was used as the optimizer, and the batch size was 30. To mitigate overfitting, we applied dropout (Srivastava et al. 2014) techniques to the word embeddings and Transformer networks, and the dropout ratio was set to 0.2. These hyperparameters were tuned on the validation set. We independently repeated each experiment 10 times and reported the average results in terms of AUC, MRR, nDCG@5 and nDCG@10.

4.2 Performance evaluation

We evaluate the performance of our approach by comparing it with many baseline methods, including:

- *LibFM* (Rendle 2012), a factorization machine based method for recommendation. We extracted the TF-IDF

---

[4] https://www.msn.com/en-us/news.

Table 2 The performance of different methods

Methods	AUC	MRR	nDCG@5	nDCG@10
LibFM (Rendle 2012)	0.6296±0.0049	0.2233±0.0041	0.2467±0.0045	0.3044±0.0039
DSSM (Huang et al. 2013)	0.6514±0.0031	0.2300±0.0028	0.2539±0.0032	0.3099±0.0030
Wide&Deep (Cheng et al. 2016)	0.6367±0.0029	0.2257±0.0026	0.2496±0.0024	0.3050±0.0022
DeepFM (Guo et al. 2017)	0.6410±0.0041	0.2275±0.0035	0.2502±0.0037	0.3069±0.0032
DFM (Lian et al. 2018)	0.6444±0.0037	0.2288±0.0033	0.2511±0.0038	0.3082±0.0035
EBNR (Okura et al. 2017)	0.6554±0.0022	0.2324±0.0019	0.2556±0.0024	0.3125±0.0021
DKN (Wang et al. 2018)	0.6531±0.0036	0.2311±0.0028	0.2544±0.0030	0.3112±0.0026
DAN (Zhu et al. 2019)	0.6621±0.0016	0.2354±0.0012	0.2588±0.0014	0.3153±0.0012
NAML (Wu et al. 2019a)	0.6739±0.0015	0.2412±0.0014	0.2634±0.0013	0.3215±0.0016
NRMS (Wu et al. 2019e)	0.6789±0.0019	0.2440±0.0017	0.2671±0.0017	0.3272±0.0018
NRNF-basic	<u>0.6818</u>±0.0016	<u>0.2456</u>±0.0016	<u>0.2688</u>±0.0017	<u>0.3290</u>±0.0019
NRNF*	**0.6896**±0.0015	**0.2506**±0.0014	**0.2780**±0.0015	**0.3323**±0.0016

The best results are in bold and the second-best ones are underlined

*The improvement over other methods is significant at $p < 0.01$

features from the titles and bodies of news clicked by users to build their representations.

- *DSSM* (Huang et al. 2013), deep structured semantic model. We regarded the concatenation of the news clicked by a user as the query and candidate news as documents, and we used the same feature with *LibFM* for news and user representation.
- *Wide&Deep* (Cheng et al. 2016), a famous neural recommendation method which consists of a wide linear channel and a deep neural network channel. We used the same features with *LibFM* as the input of both channels.
- *DeepFM* (Guo et al. 2017), another widely used neural recommendation method which uses a combination of factorization machines and deep neural networks. We used the same TF-IDF features to feed for both components.
- *DFM* (Lian et al. 2018), a deep fusion model for news recommendation, which combines fully connected layers with different depths. We used TF-IDF features and word embeddings as the input.
- *EBNR* (Okura et al. 2017), an embedding based neural news recommendation method which uses denoising autoencoders to learn news representations and a GRU network to learn user representations.
- *DKN* (Wang et al. 2018), a neural news recommendation method which learns news representations via knowledge-aware CNNs and learns user representations via an attention network based on the relevance between clicked news and candidate news.
- *DAN* (Zhu et al. 2019), a neural news recommendation method which learns news representations via two parallel CNNs and learns user representations using a combination of attentional LSTM and CNN;

- *NAML* (Wu et al. 2019a), a neural news recommendation approach with attentive multi-view learning.
- *NRMS* (Wu et al. 2019e), a neural news recommendation method that uses multi-head self-attention for news and user modeling.
- *NRNF-basic*, a variant of our *NRNF* approach that does not consider negative feedback.
- *NRNF*, our news recommendation approach with negative feedback.

For fair comparison, in *GRU*, *DKN*, *DAN* and *NRMS* we used the concatenation of news title and news body to learn news representations. The results of these methods are summarized in Table 2.

From Table 2, we have several observations. First, compared with baseline methods that use handcrafted features to represent news and users (e.g., *LibFM*, *DSSM* and *DeepFM*), the models using neural networks for learning news and user representations (e.g., *EBNR*, *NAML* and *NRNF*) achieve better recommendation performance. This is probably because neural networks can learn more informative news and user representations than hand-crafted features. Second, compared with other neural news recommender like *DAN*, *NAML* and *NRMS*, *NRNF* and its variant *NRNF-basic* achieve better performance. This is because the latter methods can capture the relatedness between news title and body to enhance news modeling, while other methods cannot. Third, *NRNF* outperforms other baseline methods, and further hypothesis tests show that the improvement is significant. This is probably because incorporating the implicit negative user feedback can help learn more accurate user interest representations.
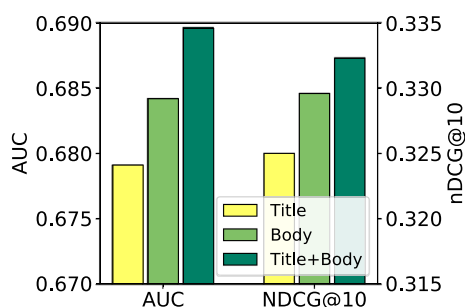
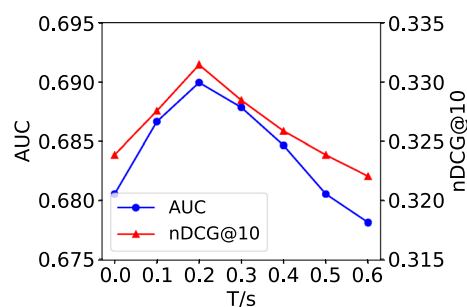**Fig. 4** Effectiveness of different news information

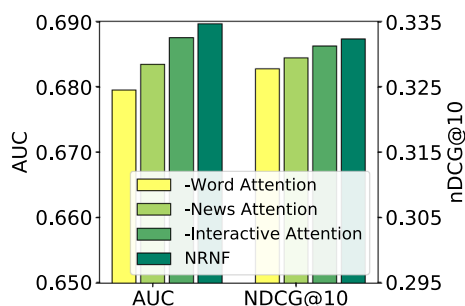**Fig. 6** Influence of the threshold T on our approach

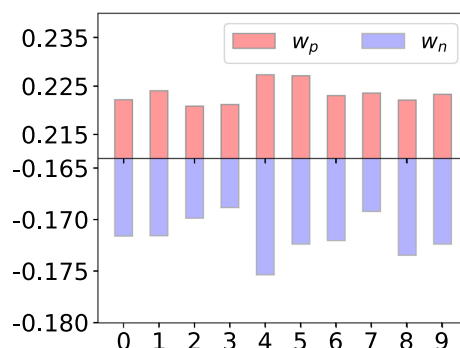**Fig. 5** Effectiveness of different attention networks

Fig. 7 Values of w_p and w_n in 10 independent experiments

4.3 Effectiveness of different news information

Next, we explore the influence of incorporating different news information for news modeling. We compare the performance of our approach with its variants with different combinations of news information, and the results are shown in Fig. 4. From Fig. 4, we have several observations. First, both news title and body are useful for news recommendation. This is because users usually decide which news to click based on its title, and read news body for detailed information. Thus, both news title and body can provide important information for news recommendation. Second, news body is more useful than news title in representing news. This is intuitive because news body usually contains the details of news, which can provide richer information than news title. Besides, combining both title and body of news can further improve the performance of our approach. These results validate the effectiveness of the multi-view learning framework in our approach.

4.4 Effectiveness of attention network

Then, we explore the influence of different attention networks on the performance of our approach. The results are shown in Fig. 5. From Fig. 5, we find that word-level attention networks are very useful for our approach. This is because different words in news titles and bodies usually
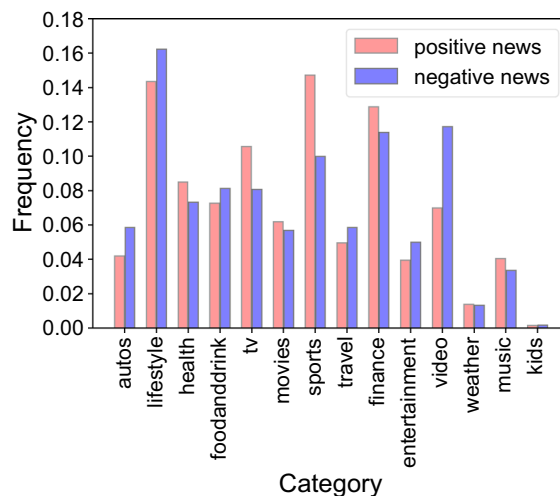

Fig. 8 Topic distributions of positive and negative clicked news

have different importance in representing news. Thus, selecting important words can help learn more informative news representations. In addition, news-level attention network is also important. This is because different news usually have different informativeness in representing the preferences of users, and selecting informative news can benefit

Table 3 Top News with the highest negative feedback ratios

Clicked news	NF ratio (%)
The 30 most heartwarming random acts of kindness from 2018	75.8
Pink lettuce is the newest pink produce	73.8
10 secrets of airline food	73.4
Get ready to load up your plates! 20 restaurants where kids eat free	73.2
Christmas coat drive for Paterson students	73.0
The best all-you-can-eat deal in every state	72.0
The most ingenious foods featured on shark tank	71.7
The five bikes sabrina is most excited to see in 2019	70.3
The greatest sports siblings in history	67.1
2018 holiday gift guide: the 25 best gift ideas under $500	66.7

user representation learning. Besides, the interactive attention network can also boost the performance. This may be because it can help capture the relations between news title and body to enhance news modeling. Moreover, combining all these kinds of attention networks can further improve the performance of our approach. These results verify the effectiveness of each kind of attention mechanism in our approach.

4.5 Influence of hyperparameters

In this section, we study the influence of an important hyperparameter on our approach, i.e., the dwell time threshold T. We compare the performance of our approach w.r.t different values of T. The results are illustrated in Fig. 6. We find the performance of our approach improves when the threshold T increases. This is probably because when T is too small, many negative news clicks cannot be distinguished from the positive ones, and the modeling of user interest is not accurate enough. In addition, the performance of our approach declines when T goes too large. This is also intuitive because when T is too large, many positive news clicks will be mistakenly regarded as negative ones and the errors will be encoded into the negative user embedding, which leads to the sub-optimal performance. Therefore, a moderate selection of T, e.g., 10 s may be more appropriate for our approach, which is also consistent with our news browsing experiences.

4.6 Parameter analysis

In this section, we conducted several studies on the results of other two key parameters learned by our model, i.e., w_p and w_n. The results in 10 independent experiments are illustrated in Fig. 7. According to these results, we find w_p is consistently positive. It is intuitive because the news read by users carefully can usually represent the preferences of users. Thus, the similarities between the candidate news and the positive clicked news should have a positive impact on

the click probability. In addition, we find w_n is consistently negative. This may be because if a user reads a news article very quickly, we can infer that he/she is probably not interested in this news. Thus, the click probability should have a negative correlation with the similarities between candidate news and negative clicked news.

4.7 Qualitative analysis of negative news

In this section, we will conduct several qualitative analysis on the characteristics of the negative clicked news. First, we want to compare the topic distributions of positive and negative news, and the results are shown in Fig. 8. According to Fig. 8, we find the ratios of negative news in different topic categories have some differences. For example, it is interesting that the ratio of negative news is the highest in the "video" category. This may be because many users are attracted by the titles of the news in this category, but they are not interested in the content of these news. Thus, incorporating the information of negative feedback may also implicitly encode topical information into user representations, which is useful for more accurate news recommendation.

In addition, we aim to discover some common patterns of negative news, e.g., the contents and writing styles of their titles. We calculate the ratio of negative feedback (NF) for each news,[5] and the titles of the top 10 news with the highest NF ratios are listed in Table 3. From Table 3, we find many news with top NF ratios are clickbaits, and they have very similar writing styles, i.e., describing something in a sensationalized way but no details are provided. Thus, many users may be attracted by the news titles and click these news. However, after taking a glance at the news body, many users may be disappointed and close these news quickly. Thus, it is harmful for news platforms to improve user experiences if

---

[5] We ignore the news with less than 10 clicks to filter some possible noisy news.

Table 4 Several clicked and candidate news of a randomly selected user

Clicked news	Dwell time		
23 Celebs who still live in their hometown	75s (+)		
Ranking the Oscar's best picture winners from every year	1s (−)		
Oscars 2019: 'the favourite,' 'Roma' lead with 10 nominations	2s (−)		
Gladys Knight to perform the national anthem at super bowl LIII	44s (+)		
Five players who could be traded during super bowl week	187s (+)		
Candidate news		Score (w/o NF)	Score (w/ NF)
Ranking the possible Super Bowl LIII matchups		0.669	**0.893**
Bradley Cooper snubbed by oscars for best director and more nomination shocker		**0.912**	0.075
Olympic swimmer Nathan Adrian reveals he has testicular cancer		0.212	0.158

The maximum predicted score in each column and the candidate news clicked by this user are in bold. NF represents negative feedback

clickbaits are frequently recommended to users. Fortunately, our approach may have the potential to avoid recommending too many clickbaits to user by incorporating the negative feedback of users, which can improve the reading experiences of users.

4.8 Case study

In this section, we conducted several case studies to show the effectiveness of our approach. First, we want to visually explore the effectiveness of incorporating the negative feedback of users. The clicked news and candidate news of a randomly selected user are listed in Table 4. We calculate the click probability score respectively using our approach and its variant without negative feedback. According to Table 4, the second candidate news is assigned the highest click score when negative feedback is not considered. This is probably because it is very similar to the second and third clicked news, since all of them are about Oscars. However, we can infer that this user is probably not very interested in this topic since she closes the second and third clicked news very quickly. Thus, it may be ineffective to recommend Oscars related news to this user. Fortunately, our approach can assign the second candidate news a low click score. This is because our approach can take the implicit negative user feedback into consideration, which is beneficial for modeling user preferences more accurately.

5 Conclusion

In this paper, we propose a neural news recommendation approach that can consider the implicit negative feedback of users. We propose to distinguish positive and negative

news clicks according to users' reading dwell time, and we propose to respectively learn two different embeddings of a user from her positive and negative news clicks. In addition, we propose to compute a unified news click score based on a combination of the relevance scores between candidate news and these two user embeddings. Besides, we propose an interactive news modeling method that can learn unified news representations from title and body with full consideration of their relatedness. Extensive experiments on real-world dataset show that incorporating negative user feedback can effectively improve the performance of user modeling for news recommendation.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

An, M., Wu, F., Wu, C., Zhang, K., Liu, Z., Xie, X.: Neural news recommendation with long-and short-term user representations. In: ACL, pp. 336–345 (2019)

Bansal, T., Das, M., Bhattacharyya, C.: Content driven user profiling for comment-worthy recommendations of news and blog articles. In: RecSys., pp. 195–202. ACM (2015)

Capelle, M., Frasincar, F., Moerland, M., Hogenboom, F.: Semantics-based news recommendation. In: WIMS, p. 27. ACM (2012)

Chen, C., Meng, X., Xu, Z., Lukasiewicz, T.: Location-aware personalized news recommendation with deep semantic analysis. IEEE Access **5**, 1624–1638 (2017)

Cheng, H.T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., et al.: Wide and deep learning for recommender systems. In: DLRS, pp. 7–10. ACM (2016)

Das, A.S., Datar, M., Garg, A., Rajaram, S.: Google news personalization: scalable online collaborative filtering. In: WWW, pp. 271–280. ACM (2007)

Garcin, F., Dimitrakakis, C., Faltings, B.: Personalized news recommendation with context trees. In: Proceedings of the 7th ACM Conference on Recommender Systems, pp. 105–112. ACM (2013)

Garcin, F., Zhou, K., Faltings, B., Schickel, V.: Personalized news recommendation based on collaborative filtering. In: WI-IAT, vol. 1, pp. 437–441. IEEE (2012)

Ge, S., Wu, C., Wu, F., Qi, T., Huang, Y.: Graph enhanced representation learning for news recommendation. In: WWW, pp. 2863–2869 (2020)

Guo, H., Tang, R., Ye, Y., Li, Z., He, X.: Deepfm: a factorization-machine based neural network for ctr prediction. In: AAAI, pp. 1725–1731. AAAI Press (2017)

Hu, L., Li, C., Shi, C., Yang, C., Shao, C.: Graph neural news recommendation with long-term and short-term interest modeling. Inf. Process. Manag. **57**(2), 102142 (2020)

Huang, P.S., He, X., Gao, J., Deng, L., Acero, A., Heck, L.: Learning deep structured semantic models for web search using click-through data. In: CIKM, pp. 2333–2338. ACM (2013)

Karkali, M., Pontikis, D., Vazirgiannis, M.: Match the news: a firefox extension for real-time news recommendation. In: SIGIR, pp. 1117–1118. ACM (2013)

Khattar, D., Kumar, V., Varma, V., Gupta, M.: Weave & rec: A word embedding based 3-d convolutional network for news recommendation. In: CIKM, pp. 1855–1858. ACM (2018)

Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

Li, L., Chu, W., Langford, J., Schapire, R.E.: A contextual-bandit approach to personalized news article recommendation. In: WWW, pp. 661–670. ACM (2010)

Lian, J., Zhang, F., Xie, X., Sun, G.: Towards better representation learning for personalized news recommendation: a multi-channel deep fusion approach. In: IJCAI, pp. 3805–3811 (2018)

Liu, J., Dolan, P., Pedersen, E.R.: Personalized news recommendation based on click behavior. In: IUI, pp. 31–40. ACM (2010)

Okura, S., Tagami, Y., Ono, S., Tajima, A.: Embedding-based news recommendation for millions of users. In: KDD, pp. 1933–1942. ACM (2017)

Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: EMNLP, pp. 1532–1543 (2014)

Ren, R., Zhang, L., Cui, L., Deng, B., Shi, Y.: Personalized financial news recommendation algorithm based on ontology. Proc. Comput. Sci. **55**, 843–851 (2015)

Rendle, S.: Factorization machines with libfm. TIST **3**(3), 57 (2012)

Son, J.W., Kim, A., Park, S.B., et al.: A location-based news article recommendation with explicit localized semantic analysis. In: SIGIR, pp. 293–302. ACM (2013)

Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. JMLR **15**(1), 1929–1958 (2014)

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: NIPS, pp. 5998–6008 (2017)

Wang, H., Zhang, F., Xie, X., Guo, M.: Dkn: Deep knowledge-aware network for news recommendation. In: WWW, pp. 1835–1844 (2018)

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. AAAI Press (2019)

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)

Wu, C., Wu, F., An, M., Huang, Y., Xie, X.: Neural news recommendation with topic-aware news representation. In: ACL, pp. 1154–1159 (2019)

Wu, C., Wu, F., An, M., Qi, T., Huang, J., Huang, Y., Xie, X.: Neural news recommendation with heterogeneous user behavior. In: EMNLP-IJCNLP, pp. 4876–4885 (2019)

Wu, C., Wu, F., Ge, S., Qi, T., Huang, Y., Xie, X.: Neural news recommendation with multi-head self-attention. In: EMNLP-IJCNLP, pp. 6390–6395 (2019)

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)

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E.: Hierarchical attention networks for document classification. In: NAACL-HLT, pp. 1480–1489 (2016)

Zheng, G., Zhang, F., Zheng, Z., Xiang, Y., Yuan, N.J., Xie, X., Li, Z.: Drn: A deep reinforcement learning framework for news recommendation. In: WWW, pp. 167–176 (2018)

Zhu, Q., Zhou, X., Song, Z., Tan, J., Guo, L.: Dan: Deep attention neural network for news recommendation. AAAI **33**, 5973–5980 (2019)

Zihayat, M., Ayanso, A., Zhao, X., Davoudi, H., An, A.: A utility-based news recommendation system. Decis. Support Syst. **117**, 14–27 (2019)

Chuhan Wu is a Ph.D. student of Electronic Engineering at Tsinghua University under the direction of his advisor Prof. Yongfeng Huang. His research interests are in recommender systems, user modeling, natural language processing and machine learning.

Fangzhao Wu is a senior researcher at Microsoft Research Asia. His research mainly focuses on natural language processing, user profiling and social media mining.

Yongfeng Huang is currently a professor in the Department of Electronic Engineering, Tsinghua University, China. He received the PhD degree in computer science and engineering from Huazhong University of Science and Technology in 2000. His research interests include next-generation Internet and Web data mining.
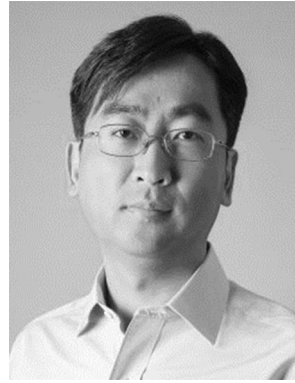
Xing Xie is currently a senior principal research manager at Microsoft Research Asia, and a guest Ph.D. advisor at the University of Science and Technology of China. He joined Microsoft Research Asia in July 2001, working on data mining, social computing and ubiquitous computing.