A Hybrid News Recommendation Approach Based on Title-Content Matching

Shuhao Jiang  
Tianjin University of Commerce

Yizi Lu  
Tianjin University of Commerce

Haoran Song  
Tianjin University of Commerce

Zihong Lu  
Tianjin University of Commerce

Yong Zhang  
zhangyong@tjcu.edu.cn

Tianjin University of Commerce

Article

Keywords: News recommendation, user interest models, Siamese Neural Network, Neural Collaborative Filtering (NCF), Factorization Machines (FM).

Posted Date: April 3rd, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4123064/v1

License: ☒ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.
A Hybrid News Recommendation Approach Based on Title-Content Matching

Shuhao Jiang1, Yizi Lu2, Haoran Song3, Zihong Lu4 & Yong Zhang5

Personalized news recommendation can alleviate the information overload problem, and accurate modeling of user interests is the core of personalized news recommendation. Existing news recommendation methods integrate the titles and contents of news articles that users have historically browsed to construct user interest models. However, this method ignores the phenomenon of “title-content mismatching” in news articles, which leads to the lack of precision in user interest modeling. Therefore, a hybrid news recommendation method based on title-content matching is proposed in this paper.(1) An interactive attention network is employed to model the correlation between title and content contexts, thereby enhancing the feature representation of both; (2) The degree of title-content matching is computed using a Siamese Neural Network, constructing a user interest model based on title-content matching; and (3) Neural Collaborative Filtering (NCF) based on Factorization Machines (FM) is integrated, taking into account the perspective of the potential relationships between users for recommendation, leveraging the insensitivity of neural collaboration to news content to alleviate the impact of title-content mismatching on user feature modeling. The proposed model was evaluated on a real dataset, and the experimental results demonstrate that the proposed method effectively improved the performance of news recommendation.

Keywords News recommendation; user interest models; Siamese Neural Network, Neural Collaborative Filtering (NCF); Factorization Machines (FM).

In recent years, with the development of information technology, especially the popularization of mobile Internet, online reading has gradually become the mainstream choice for users to browse news. Various types of news are aggregated on online news platforms and presented to users, which are rich in content and diverse. However, users find it challenging to navigate through the massive amount of news to discover content that aligns with their interests. Personalized news recommendation methods have been developed to assist users in finding news of interest and effectively alleviate the problems caused by information overload1-3. Accurate user interest modeling is the core problem of personalized news recommendation. Existing news recommendation methods usually construct user interest models based on the content and titles of the news that users have historically clicked on the news platform. Many researchers have done a lot of work and achieved promising results. Okura et al.4 proposed to use Recurrent Neural Networks (RNN) to construct user preferences with browsing history as input sequences. Wang et al.5 proposed to learn the user representation from the news content and title of user’s historical clicks and evaluate the correlation between the clicked news and the candidate news through a knowledge-aware Convolutional Neural Network (CNN). Fan et al.6 proposed the News Recommendation Algorithm Based on Multiple Perspectives (BTEC), which utilizes the Bidirectional Encoder Representations from Transformers (Bert) model and the attention mechanism to vectorize the content, titles, and events in the news, respectively, and performs fusion processing for candidate news as well as news browsed by the user’s history based on the above view. As a matter of fact, some news platforms, in order to gain traffic or promote advertisement for profit, publish news with exaggerated titles, aiming to cause readers to click on the news, however, the actual news content is not exactly a reflection of the headline, leading to the issue of “title-content mismatching”. Such news content
news recommendation is an important task in the field of natural language processing, which has become the focus of more and more research institutions and personnel. The algorithms currently only take the title as features of the news. Furthermore, a Gate Recurrent Unit (GRU) network is used to both the title and content of the news as features of the news, and less than the threshold, deciding degree, and a threshold is reasonably determined. If the matching degree exceeds the threshold, historically browsed by users, and an interactive attention network is used to model the interaction between the titles and content of the news article, and these interactions serve as important clues for modeling the correlation between titles and content. The 10 pounds of the title in the third news is closely related to the weight loss in the text, while in fact the features extracted for the title and the content are independent of each other, and the extracted features are only the semantic vectors of their respective last. Finally, neural collaborative filtering makes recommendations based on the clicking behavior of users and their nearest neighbors, independent of the text content. Therefore, the integration of neural collaborative filtering for hybrid recommendation can play a role in improving the issue of title-content mismatching.

The work presented in this paper is motivated by the following observations: Firstly, detecting whether the titles and content of news articles match is crucial for modeling user interests. As illustrated in Figure 1, in the second news it can be found that the user has clicking behavior on the news because of the interest in entertainment news indicated by the title, but the content is not the entertainment news that has relevance to the title, and the content of the news is not representative of the user’s interest. Secondly, there exists a connection between the titles and content of the same news article, and these interactions serve as important clues for modeling the correlation between titles and content. The 10 pounds of the title in the third news is closely related to the weight loss in the text, while in fact the features extracted for the title and the content are independent of each other, and the extracted features are only the semantic vectors of their respective last. Finally, neural collaborative filtering makes recommendations based on the clicking behavior of users and their nearest neighbors, independent of the text content. Therefore, the integration of neural collaborative filtering for hybrid recommendation can play a role in improving the issue of title-content mismatching.

Figure 1. News clicked by users

In this paper, a hybrid news recommendation method based on title-content matching is proposed to address the problem of insufficiently accurate modeling of user interest preferences due to text-title mismatch. Initially, the Bert model is used to extract the title and content features of news articles historically browsed by users, and an interactive attention network is used to model the interaction between the title and the content. Subsequently, a Siamese neural network is applied to calculate the title-content matching degree, and a threshold is reasonably determined. If the match degree exceeds the threshold, both the title and content of the news are taken as the features of the news, and less than the threshold, only the title is taken as the features of the news. Furthermore, a Gate Recurrent Unit (GRU) network is employed to build users’ time sequential features based on the news they have browsed, processing all historically viewed news to obtain user features based on title-content matching. For candidate news, Bert is used to extract features and CNN is used to extract local features to enhance the feature representation of the candidate news, and the user features based on title-content matching are fitted with the candidate news features. Additionally, an improved neural collaborative filtering approach using FM is adopted to capture low- and high-order interactive relationships between users and news, mitigating data sparsity while improving the issue of title-content mismatching in content-based recommendation. Finally, the content recommendation based on title-content matching detection is mixed with the FM-based neural collaborative filtering recommendation to predict the ultimate click probability.

The contributions and innovations of this paper are as follows:

- Using an interactive attention network to model the relevance of title and content. On the basis of extracting features for the title and content separately, the interactive features of the title and content are considered to enhance the feature representation of the title and content, at the same time facilitate the accurate assessment of the relevance between the title and content.
- Modeling user interest features based on the detection of title-content matching. Detecting title-content matching for news articles that users click on, determining user preferences for each news article based on a threshold, and selecting content for modeling. Extracting time sequence features from the user's history of browsed news to construct a user interest preference model based on title-content matching.
- Adopting an improved neural collaborative filtering approach to alleviate the title-content mismatching problem. Optimizing traditional neural collaborative filtering with FM and combining it with Multi-Layer Perceptron (MLP) to learn low- and high-order feature interactions between users and news, comprehensively capturing interactions to improve data sparsity. Only the clicking behaviors of users and neighboring users are considered, alleviate the problem of title-content mismatching.
- Experiments are conducted on benchmark datasets, and the results demonstrate that this approach can effectively alleviate the title-content mismatching problem and significantly improve the performance of news recommendations.

Related work

News Recommendation Algorithm

Personalized news recommendation is an important task in the field of natural language processing, which has become the focus of more and more research institutions and personnel. The algorithms currently
The architecture of our approach is illustrated in Figure 2: a candidate news. The architecture mainly consists of four modules: a user interest modeling module based on the reading time, which integrates content recommendation and neural collaborative filtering recommendation to high-order interaction features between the user and the clicked news; and a hybrid recommendation module, which integrates content recommendation and neural collaborative filtering recommendation to realize the predictive ranking of candidate news and predict the probability of the user finally clicking on the candidate news. The architecture of our approach is illustrated in Figure 2:

Differing from the above work, this paper proposes to use convolutional neural networks to obtain local representations of words from the titles, body text through Bi-GRU and measures the global and local similarity between the titles and the content. Liu et al.21. proposed a convolutional neural network injected with an attention mechanism for news text feature extraction, which extracts the user's interest features by adding time sequential predictions to the news that the user has already browsed and injecting a multi-head self-attention mechanism. Hybrid recommendation methods mainly combine two or more recommendation methods, aiming to improve the recommendation effectiveness.

**Detection of Title-Content Matching Degree**

The content of an article is the most important implicit information in news, and its degree of association with the title and matching with user interests will directly affect the user’s recommendation experience. Therefore, title-content matching detection is an important task for online platforms. In recent years, many scholars have explored the use of deep learning technology for the detection of title-content matching. Peter Bourgonje et al.23. developed an algorithm to detect the relevance between titles and the main body of articles, based on Term Frequency, Inverse Document Frequency, N-gram models, and logistic regression. Dong et al.24. proposed a deep similarity-aware attention model that learns the representation of titles and body text through Bi-GRU and measures the global and local similarity between the titles and the content. Yin et al.25. utilized the Chinese Bert model to vectorize the title and content representations of news, applying fusion attention to fuse the features of title and content, and finally use Bi-GRU to detect the title-content matching. The above approach detects title-content matching degree by extracting features of titles and content separately, which only considers the phenotypic features of titles and content, ignoring the interactive features between titles and content, which is crucial for assessing title and content relevance.26. Differing from the above work, this paper adopts the Bert model to deeply mine the text semantics and uses an interactive attention network to model the interaction between the titles and content, which enhances the representation of the title and content features and improves the accuracy of the calculation of the title-content matching degree at the same time.

Furthermore, most current detections of title-content mismatching are commonly used to identify 'clickbait' issues. In fact, applying the detection of title-content mismatching in the field of news recommendation can identify whether titles and content match, discover the content users are genuinely interested in, and accurately and effectively model user interests.

**Method**

In this paper, we propose a hybrid news recommendation method based on title-content matching, which mainly consists of four modules: a user interest modeling module based on the reading time, which integrates content recommendation and neural collaborative filtering recommendation to realize the predictive ranking of candidate news and predict the probability of the user finally clicking on the candidate news. The architecture of our approach is illustrated in Figure 2:
User modeling based on title-content matching degree detection

In order to solve the problem of insufficiently accurate user interest feature modeling caused by title-content mismatching, user interest feature modeling based on title-content matching degree detection is proposed to construct a title-content matching user interest model. Learning user representations from the titles and content of news browsed historically by users, this builds user interest preferences, which mainly includes three parts: feature extraction of titles and content based on phenotypic and interactive characteristics, title-content matching degree detection using Siamese neural networks, and user interest modeling based on title-content matching. The architecture of the user modeling method based on title-content matching degree detection is illustrated in Figure 3:

**Figure 3.** User interest modeling framework based on the detection of the consistency between title and content

Extraction of Phenotypic and Interactive Title and Content Features

Firstly, the features of the news titles and body text browsed historically by the user need to be extracted. Traditional neural networks, when extracting features, are unable to fully capture contextual information...
and are insufficient in precision for extracting deep semantic information. Therefore, this paper utilizes the large-scale pre-trained language model, Bert, replacing the traditional Word2Vec model in the embedding layer. The feature extraction of titles and body text includes two steps: generating word embedding matrices and performing average pooling.

Based on the encoded inputs of the news titles and body text $\text{code}_t$, $\text{code}_b$, Bert word embedding matrices $\text{text}_t$, $\text{text}_b$ are generated, as represented in equations (3-1) and (3-2):

$$\text{text}_t = \text{Bert} (\text{code}_t) = \begin{pmatrix} \text{word}_1 \\
\vdots \\
\text{word}_s \end{pmatrix}_{s \times m}$$  \hspace{1cm} (3-1)

$$\text{text}_b = \text{Bert} (\text{code}_b) = \begin{pmatrix} \text{word}_1 \\
\vdots \\
\text{word}_s \end{pmatrix}_{s \times m}$$  \hspace{1cm} (3-2)

Where $\text{word}_j$ is the $j$th word embedding, $s$ is the length of the title or content, and $m$ represents the word embedding dimension determined in the pre-training phase. The Bert word embedding matrix $\text{text}_t$ and $\text{text}_b$ is average pooled to obtain the semantic vectors of title and content $c_t$, $c_b$, as shown in equations (3-3) and (3-4):

$$c_t = \text{meanpool}(\text{text}_t)$$  \hspace{1cm} (3-3)

$$c_b = \text{meanpool}(\text{text}_b)$$  \hspace{1cm} (3-4)

Average pooling involves summing all the word embeddings and then dividing each element by $s$, thereby smoothly aggregating the semantic information of the word embeddings to generate semantic vectors that represent the title or content. This method can provide a more holistic representation of the title or content compared to [CLS] word embeddings and max pooling.

Titles and content are interconnected, yet during feature extraction, they are completely separated and independent, unable to perceive each other’s information. The extraction yields only their respective semantic vectors, lacking interactive features between the title and content, which makes it difficult to learn the latent relationship and relevance between the two. An interactive attention network is employed to simulate the interplay of context between titles and content, capturing the association between the article content and its title. Specifically, for the Bert output of the title $c_t$, it is treated as the query, with the content output $c_b$ serving as both key and value, which contains the context in the body and its interaction with each word in the title, resulting in a title-aware body text representation $tb-rep$; similarly, the Bert output of the content $c_b$ is used as the query, with the title output $c_t$ serving as both key and value, conveying the context within the title and its interaction with every word in the body text 19, resulting in a body text-aware title representation $bt-rep$:

$$tb-rep = \text{MultiHead}(c_t, c_b, c_b)$$  \hspace{1cm} (3-5)

$$bt-rep = \text{MultiHead}(c_b, c_t, c_t)$$  \hspace{1cm} (3-6)

The final output is the features of the title and content that are based on phenotypic and interactive.

Title-Content Matching Degree Detection Based on Siamese Neural Network

The Siamese neural network, also referred to as a conjoined network with symmetry, which is naturally suitable for the calculation of title-content matching degree. By sharing the weights between the two sub-networks of the Siamese neural network, it reduces the number of parameters to be trained, reduces the complexity of the model, and at the same time maps vectors with different spatial dimensions to the same dimension, ensuring a consistent data distribution between titles and body text. When calculating the degree of compatibility between the title and the content, the meaning of each word in the title and the content will be affected by a number of words in the front and a number of words in the back. Hence, a Bidirectional Long Short-Term Memory Network (Bi-LSTM) is employed at the encoding layer in this paper. The structure is shown in Figure 4:

![Figure 4. Structure of Bi-LSTM neural network model](image-url)
are the state information of the current moment and the previous one, it means that the content of $\text{title}$ and $\text{content}$ is the candidate hidden state, $\text{title}$ and $\text{content}$ are the final feature vectors for title and content, $\text{title}$ and $\text{content}$ are the feature vectors of each dimension in the two vectors, respectively.

$\text{GRU}$ includes two gates: an update gate $\text{gate}$ and a reset gate $\text{gate}$, respectively signify the forward propagated state at time $t$ and backward propagated state at time $t+1$; $W_1$, $W_2$, $W_3$ and $W_4$ represent the weight matrices corresponding to the different components, respectively.

Finally, the forward and backward hidden vectors are spliced to get the final output semantic vector that contains all the information in both forward and backward directions. The Siamese neural network is used to make the distance between similar titles and content as small as possible. The vectors processed by the coding layer need to be fused to calculate the gap between the semantics expressed by the title-content vectors. When it is necessary to consider whether the texts are semantically similar or not, the cosine similarity can be utilized to measure in the direction of the text’s semantics. As shown in equation (3-10):

$$\text{similarity} = \cos(\theta) = \frac{X \cdot Y}{\|X\| \|Y\|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}}$$

Where $X$ and $Y$ are the final feature vectors for title and content, $x_i$ and $y_i$ are the feature representations of each dimension in the two vectors, respectively.

Users usually decide whether to click on a news article based on its title and read the content of the news to get detailed information, hence, the computed title-content matching is used to determine user interest preferences. If the title-content matching degree is greater than a threshold $\lambda$, it means that the content of the article meets the user's expectations, and the title and content of the article are spliced as the characterization of the news; if the title-content matching degree is less than a threshold $\lambda$, it suggests that the news article exhibits a title-content mismatching, and only the title is selected as the characterization of the news.

User interest modeling based on title-content matching

$\text{GRU}$ is employed to learn the user's interest preference representation from the user's browsing history. The $\text{GRU}$ can model the information in the user's past behavior sequence, where each time step corresponds to a browsing behavior, and $\text{GRU}$ updates the internal state at each time step to capture dynamic changes in the user's interest.

A matrix of users' historical browsing news that matches the title of the content $E_{\text{news}}$ is fed into the $\text{GRU}$ network for computation, and the hidden state of the last time step is output. This allows for the consideration of the recency of browsing time$^{28}$. The final user interest feature is computed as shown in the following equation:

$$E_{\text{news}} = \{e_1, e_2, \ldots, e_k\}$$

For each time step $t$, the $\text{GRU}$ includes two gates: an update gate $i_t$ and a reset gate $r_t$, $e_t$ is the input of the moment $t$, $h_t$ and $h_{t-1}$ are the state information of the current moment and the previous moment respectively, $\sigma$ is the sigmoid function, $\tilde{h}_t$ represents the candidate hidden state, $\odot$ is the term-
by-term product, $W_u$, $W_c$, and $W_h$ are the weight matrices of the update gate, the reset gate, and the candidate hidden state, respectively.

The farthest to the nearest $l$ news items are considered as $l$ moments of user browsing events, with the final representation of user interest being the features of the news that the user is predicted to click on in the next moment by the GRU network, denoted as $U_i = h_i$, resulting in the user interest features based on title-content matching.

**Candidate news feature modeling based on Bert-CNN**

To efficiently extract the contextual relationships, local features and global features of candidate news text, the Bert model is first used to deeply extract the textual semantic information and consider the global features of candidate news. On this basis, the CNN model is combined to obtain short-range local features of words from candidate news, to comprehensively enhance the text feature representation of candidate news.

The content of the candidate news is represented as $[H_1, H_2, \ldots, H_n]$, which is transformed into a sequence of word vectors $[h_1, h_2, \ldots, h_n]$ by the Bert model, where $n$ is the length of the text content. The contextual semantic information of the word is very important for characterizing the news, and utilizing CNN allows for obtaining the contextual representation $b_i$ of the $i$-th word, calculated as follows:

$$e_i = ReLU\left(F \times b_{i-K\mid i+K} + d\right)$$

(3-17)

where $b_{i-K\mid i+K}$ is the concatenation of word embeddings from position $i-K$ to $i+K$, $F$ and $d$ are the kernel and bias parameters of the CNN filter, $ReLU$ is the activation function, and the output is the Bert-CNN based sequence of contextual representations of the candidate news $[e_1, e_2, \ldots, e_n]$, and ultimately the output of the feature representation of the candidate news $V_{news}$.

**Neural Collaborative Filtering Based on FM**

Traditional neural collaborative filtering is constrained by the relatively simple structure of Generalized Matrix Factorization (GMF), primarily focusing on the direct interactions between users and news. GMF models user-item interactions mainly by dot product, which is somewhat deficient in modeling nonlinear relationships and limited in the case of sparse data. Unlike GMF, FM can learn the first-order and second-order feature interactions between users and news, and introduces a factorization mechanism to better capture the potential relationship between users and news.

Therefore, this paper adopts FM instead of the traditional GMF, which can flexibly deal with the complex relationship between higher-order nonlinear features by learning the implied feature vectors of the interactions between users and news, which is more advantageous in the face of data sparsity. The neural collaborative filtering model based on FM mainly consists of two parts: FM and MLP, which can capture low-order feature interactions as well as learn high-order feature interactions, improve the model's ability to model the complex relationship between users and news, and improve the overall recommendation performance.

The user and news IDs are expanded into one-hot feature vectors $o_u$ and $o_i$, respectively, which, after passing through the embedding layer, are transformed into the embedding vectors $p_u^F$ and $q_i^F$ for FM:

$$p_u^F = embed_u(o_u) = W_{u}^T o_u$$

(3-18)

$$q_i^F = embed_i(o_i) = W_{i}^T o_i$$

(3-19)

FM learns a hidden vector for each second-order crossover feature, and the weights of the crossover features can be expressed as the inner product of the corresponding hidden vectors of the features:

$$\hat{y} = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle v_i, v_j \rangle x_i x_j$$

(3-20)

The second-order part is simplified as follows:

$$\sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle v_i, v_j \rangle x_i x_j = \frac{1}{2} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} \langle v_i, v_j \rangle x_i x_j - \sum_{i=1}^{n} \langle v_i, v_i \rangle x_i^2 \right)$$

(3-21)
Where $v_i$ is the latent vector for the $i$-th feature, and $v_{ij}$ is the $f$-th element of the $i$-th feature. The aforementioned steps achieve low-order feature crossing. Splice the embedding vectors of user and news:

$$x = \text{concat}(p_u^T, q_i^T) = \begin{bmatrix} p_u^T \\ q_i^T \end{bmatrix}$$

(3-22)

Input the spliced vectors into the FM model, and finally get the output of FM $\hat{y}^{FM}$:

$$\hat{y}^{FM} = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} <v_i, v_j > x_i x_j$$

$$= w_0 + \sum_{i=1}^{n} w_i x_i + \frac{1}{2} \sum_{j=1}^{n} \left( \sum_{i=1}^{n} v_i^2 x_i^2 - \sum_{i=1}^{n} v_i^2 x_i^2 \right)$$

(3-23)

The MLP introduces nonlinear elements through the activation function, and it can learn higher-order feature interactions through the multilayer neural network structure, which makes the model adapt to the complex user-news interaction relations. Similar to FM, the one-hot vectors of users and news are first input into the embedding layer, which outputs the embedding vectors for MLP $p_u^M$ and $q_i^M$:

$$p_u^M = \text{embed}_u(o_u) = W_u^T o_u$$

(3-24)

$$q_i^M = \text{embed}_i(o_i) = W_i^T o_i$$

(3-25)

Then the embedding vectors of MLP are spliced together and inputted into the MLP model. Since the ReLU activation function has higher computational efficiency and better interpretability compared to sigmoid and tanh functions, MLP introduces nonlinear activation function ReLU, which makes the neural network to learn the higher order nonlinear relationship between users and news, and finally get the output of MLP $\hat{y}^{MLP}$:

$$\hat{y}^{MLP} = W_y^T \left( W_y^T a \left( W_y^T \left[ p_u^M \quad q_i^M \right] + b_1 \right) + b_2 \right) + b_3$$

(3-26)

$$\text{ReLU}(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases}$$

(3-27)

**Hybrid Recommendation**

The neural collaborative filtering model relies on deep neural networks with strong learning ability to achieve interactive feature learning between users and news, considering the similarity between users and their nearest neighbors, finding similar users based on their clicking behavior, and news preferred by similar users is recommended. Meanwhile, Content-based news recommendation models user interest features based on the consideration of title-content matching. A hybrid approach combining content recommendation based on title-content matching and neural collaborative filtering recommendation based on FM to predicts the final click probability.

User interest preference features based on title-content matching $U_s$ are fitted with candidate news features $V_{new}$ constructed based on Bert-CNN, yielding an output of $\hat{y}^{C}$, respectively. In neural collaborative filtering, the outputs of FM and MLP are $\hat{y}^{FM}$ and $\hat{y}^{MLP}$, and these three outputs are mixed. Due to the weaker learning capability of linear methods and the propensity for overfitting when directly modeling all orders of cross-features with only three features, resulting in reduced generalization ability, a single hidden layer MLP is utilized as the hybrid recommendation module.

$$\phi = a \left( W_y^T \left[ \hat{y}^{FM} \quad \hat{y}^{MLP} \right] + b_1 \right)$$

(3-28)

$$\hat{y} = \sigma \left( W_y^T \phi + b_2 \right)$$

(3-29)

$$\text{Leaky ReLU}(x) = \begin{cases} x, x \geq 0 \\ \alpha x, x < 0 \end{cases}$$

(3-30)

where $\sigma$ is a sigmoid function that maps values to between 0 and 1 to get the final predicted click probability $\hat{y}$. $a$ is the Leaky ReLU function, The Leaky ReLU function as shown in equation (3-30), the
value is negative when the input is less than 0, allowing for a small gradient. Where $\alpha$ is a small value, often set to 0.01.

During the training process, Binary Cross Entropy (BCE) is used to compute the loss, as shown in equation (3-31):

$$BCELoss \left( \hat{y}, y \right) = -\frac{1}{n} \sum_{i=1}^{n} \left( y_i \log_{e} \hat{y}_i + (1 - y_i) \log_{e} (1 - \hat{y}_i) \right)$$

(3-31)

After calculating the loss, gradients are backpropagated, and the model iteratively updates its parameters until convergence.

**Experimental results and analysis**

**Dataset and Parameter Settings**

**Dataset and Data Processing**

This study selected the publicly available dataset released by the Datacastle competition, which was collected from the well-known financial news website - Caixin. The dataset statistics are as shown in Table 1. This dataset mainly includes six items: user ID, news ID, browsing timestamp, news title, news content, and news publication time. This paper primarily utilizes the first five items.

<table>
<thead>
<tr>
<th>#Users</th>
<th>9457</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.#Words Per News Title</td>
<td>14.0</td>
</tr>
<tr>
<td>#News</td>
<td>100197</td>
</tr>
<tr>
<td>Avg.#Words Per News Content</td>
<td>584.0</td>
</tr>
</tbody>
</table>

**Table 1. Detailed dataset statistics**

Preprocess the dataset: remove data in the dataset where any field is empty, remove data with title and content length less than 10, and remove duplicate news IDs and user IDs to get 9,395 users and 5,853 news items. For duplicate browsing records, only the records with the latest timestamp are kept. Due to the timeliness of news, the preprocessed news browsed by the user's history was divided in chronological order, with a training set to test set ratio of 5:1. The training set's positive to negative sample ratio was 1:4, where negative samples were drawn from news not browsed by users, selected randomly. The dataset contains 474850 data samples, of which the test set contains 82545 samples, and the training set contains 392305 samples. There are 78461 positive samples and 313844 negative samples in the training set.

**Parameter Settings**

In this paper, the model is implemented based on PyTorch framework, the specific model parameters are set as shown in Table 2 below, and the optimizer uses Adam.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bert model</td>
<td>alibaba-pai/pai-bert-tiny-zh</td>
</tr>
<tr>
<td>Batch size</td>
<td>64</td>
</tr>
<tr>
<td>Learning rate</td>
<td>2.00E-03</td>
</tr>
<tr>
<td>Num epochs</td>
<td>8</td>
</tr>
<tr>
<td>Embedding dim</td>
<td>256</td>
</tr>
<tr>
<td>Click num</td>
<td>35</td>
</tr>
<tr>
<td>Title length</td>
<td>60</td>
</tr>
<tr>
<td>Content length</td>
<td>300</td>
</tr>
<tr>
<td>Candidate news word length</td>
<td>300</td>
</tr>
<tr>
<td>Num attention heads</td>
<td>16</td>
</tr>
<tr>
<td>Num GRU layers</td>
<td>2</td>
</tr>
</tbody>
</table>

**Table 2. Hyperparameter value**

To evaluate the performance of the model presented in this paper, the following five widely used metrics are adopted: the Area Under Curve (AUC), which measures the ability of the recommendation system to distinguish between news of interest to the user and news not of interest, as shown in equation (4-1); the Mean Reciprocal Rank (MRR), focusing on the ranking of the first positive sample in the recommendation list, emphasizing the accuracy of the list's ranking, as shown in equation (4-2); the Normalized Discounted Cumulative Gain (nDCG), an indicator for assessing the quality of ranking tasks, as shown in equation (4-3); the F1 Score, which is the harmonic mean of precision and recall, aiming to consider both the accuracy and coverage of the classification model; and the Gini Coefficient, used to measure the inequality of the recommendation results and to assess whether the distribution of user interest across different news is balanced. The final performance comparison is conducted on the test set. To ensure the accuracy of the experimental results, each experiment is repeated three times, and the average value of the evaluation metrics is taken as the final result.

$$AUC = \frac{1}{|S|} \sum_{i=1}^{|S|} \sum_{v \in v^c} r_{uv} \times \frac{|v^c|^2 \times (|v^c| + 1)}{2}$$

(4-1)
\[ \text{MRR} = \frac{1}{|S|} \sum_{i=1}^{n} \frac{1}{\text{rank}_i} \]  

\[ \text{nDCG} @ K = \frac{\text{DCG}_K @ K}{\text{IDCG}_K} \]  

\[ \text{DCG}_K @ K = \sum_{i=1}^{K} \frac{2^{\text{rel}_i} - 1}{\log_2(i+1)} \]  

\[ \text{IDCG}_K = \max \left\{ \sum_{i=1}^{m} \frac{2^{\text{rel}_i} - 1}{\log_2(i+1)} \right\} \]

Where \( u \) represents the user, \( S \) is the number of users, \( v_u^+ \) and \( v_u^- \) denote positive and negative samples, respectively, \( r_{ui} \) is the ranking of positive samples sorted by probability scores from lowest to highest. \( \text{rank}_i \) is the rank of the \( i \)-th sample. \( K \) is the size of the recommendation list, \( \text{relu}_i \) is the recommendation result at position \( i \), where \( \text{relu}_i \) equals 0 when the user does not click on the \( i \)-th news; otherwise, it equals 1. \( \text{DCG}_K \) is the summation of values in the recommendation list with the predicted scores sorted from large to small, while \( \text{IDCG}_K \) is the normalized \( \text{DCG}_K \) in the ideal condition with the ground-truth labels sorted from 1 to 0.

**Performance Evaluation**

To evaluate the performance of the method presented in this paper, it is compared with recent advanced news recommendation algorithms, including:

- **NRMS**\(^{16}\), utilize word-level attention and multi-head self-attention to learn news representations from news titles and employ news-level multi-head self-attention to capture the relationships between historically browsed news for learning user representations.

- **NAML**\(^{18}\), propose a multi-view news learning method that considers titles, content, and categories as different views. View-level attention and word-level attention are used to learn news representations, and user representations are learned based on the news browsed by users.

- **LSTUR**\(^{30}\), propose a recommendation method incorporating long and short-term user interests, learning news representations from titles and topic categories, and employing a GRU network to learn users’ short-term representations while shaping long-term user interests from the entire click history.

- **TANR**\(^{31}\), propose a neural news recommendation approach with topic-aware news representations, utilize CNN to extract features from news titles and append news topic categorization. learn the representations of users from their browsed news and use attention networks to select informative news for user representation learning.

- **DFFA**\(^{21}\), extract the feature matrix of news text through the CNN injecting attention mechanism. By adding time series prediction to the news that users had browsed and injecting multi-head self-attention mechanism, the interest characteristics of users were extracted.

- **base**, extract features of candidate news using the Bert model, combine semantic vectors of news with sequences of user historical behavior, employ GRU for sequential modeling of users, and integrate neural collaborative filtering based on FM.

<table>
<thead>
<tr>
<th>Methods</th>
<th>MRR</th>
<th>nDCG@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAML</td>
<td>75.59</td>
<td>78.25</td>
</tr>
<tr>
<td>NRMS</td>
<td>76.40</td>
<td>79.51</td>
</tr>
<tr>
<td>TANR</td>
<td>75.72</td>
<td>80.23</td>
</tr>
<tr>
<td>LSTUR</td>
<td>78.88</td>
<td>82.38</td>
</tr>
<tr>
<td>DFFA</td>
<td>78.22</td>
<td>82.14</td>
</tr>
<tr>
<td>My Model-base</td>
<td>81.36</td>
<td>82.48</td>
</tr>
<tr>
<td>My Model</td>
<td>81.88</td>
<td>83.03</td>
</tr>
</tbody>
</table>

**Table 3.** The performance of different methods on Caixin dataset

<table>
<thead>
<tr>
<th>Methods</th>
<th>MRR</th>
<th>AUC</th>
<th>nDCG@5</th>
<th>nDCG@10</th>
<th>F1 Score</th>
<th>Gini</th>
</tr>
</thead>
<tbody>
<tr>
<td>My Model-base</td>
<td>81.36</td>
<td>84.77</td>
<td>82.48</td>
<td>83.56</td>
<td>34.85</td>
<td>37.43</td>
</tr>
<tr>
<td>My Model</td>
<td>81.88</td>
<td>85.22</td>
<td>83.03</td>
<td>84.16</td>
<td>35.10</td>
<td>37.15</td>
</tr>
</tbody>
</table>

**Table 4.** The performance of basic model

The performance of this paper’s model and all the baseline models on the Caixin dataset is evaluated and the results are shown in Table 3. All the above methods are deep learning-based news recommendation methods, and the modules used are similar to the models in this paper, which makes the comparison experiments meaningful. There are several observations from the summary in Table 3.
Models considering user sequence features, such as LSTUR, DFFA, and My Model-base, outperform NAML, NRMS, and TANR. This indicates that the time series features of users’ historically browsed news can better reflect user characteristics, and accurate user interest modeling is crucial for news recommendation.

(2) My Model-base and DFFA outperform other models, suggesting that the Bert model can delve deeper into news content semantics and consider news contextual features more effectively than other models.

(3) My Model-base outperforms NAML, NRMS, TANR, LSTUR, and DFFA indicating that fusing FM-based neural collaborative filtering can alleviate data sparsity while improving the quality of recommendations.

(4) The method presented in this paper consistently outperforms compared baseline models, further indicating that addressing the issue of title-content mismatching is vital for accurately modeling user interest preferences and improving the quality of news recommendations. Unlike baseline methods, our approach models the context of titles and content with interactive attention network on top of deeply extracting news semantics with Bert, enhancing news understanding to help more accurately measure titles and content relevance, and judging the degree of title-content matching to help more accurately model user interest preferences. Furthermore, integrating improved neural collaborative filtering and considering the potential relationships between users for recommendations, alleviates the issue of title-content mismatching without considering text content.

As shown in Table 4, My Model metrics are better than My Model-base, because modeling users based on title-content matching can more accurately capture user interest preferences. The Gini coefficient is used to measure the degree of diversity of user interests in a recommender system, and the diversity of the two methods varies less, because both methods use hybrid recommendation, which can effectively improve the performance of recommendation.

Ablation Study

Hyperparameter Analysis

(1) User Historical Sequence Length

The impact of the length of user historical sequences on recommendation effectiveness is considered. The variation of recommendation performance with different numbers of historical browsing actions is illustrated in Figure 5, where the horizontal axis represents the evaluation metrics, and the vertical axis represents the values of these metrics. F1 Score and Gini Coefficient are shown in graph (b), with values ranging from 0.34 to 0.38. The results indicate that recommendation effectiveness improves with the increase in the length of user historical sequences. This improvement occurs because, when the user historical behavior sequence is too short, only a recent period of browsing records is used, lacking sufficient behavioral data to accurately understand user interest preferences; When the user historical sequence is too long, performance begins to decline. This decline is due to user interests changing over time, where some older news has less reference value, and not all news will influence user interest preferences. Therefore, the length of user historical sequences affects the accuracy of user interest preference modeling, thereby influencing recommendation performance. Setting the user historical sequence to a moderate value, such as 35, is more suitable for our method. When user historical reading records exceed this range, the earliest part of the reading time is taken.

Figure 5. Variation of indicators with length of historical series

(2) Impact of the Title-Content Matching Threshold $\lambda$

The effect of the threshold $\lambda$ of title-content matching on the performance of the method in this paper is investigated. The recommendation performance with different $\lambda$ values is shown in Figure 6. The results indicate that the performance of our method gradually increases with the growth of $\lambda$. This is because the value of $\lambda$ is key to constructing user interest features, and a too small $\lambda$ would result in underutilization of title-content matching detection, with most news being judged as title-content matching, leading to inaccurate user interest modeling. When $\lambda$ is too large, performance begins to decline. This is because a too large $\lambda$ results in most news being deemed title-content mismatching, causing user interest features to overly derive from clicked news titles without adequately utilizing news content, thus losing important information and not sufficiently mining user interests. In graph (b), the Gini coefficient represents the diversity of recommendation. The better the performance of the recommendation, the lower the diversity of this method. Therefore, setting the value of $\lambda$ to 0.75 is more suitable for the dataset used in this paper.
In this section, experiments are conducted to verify the effectiveness of the title-content matching degree detection module used in this paper. To assess the effectiveness of the title-content matching degree detection module, two variant models are set up:

- **Title + Content:** This variant removes the title-content matching degree detection module and directly concatenates the title and content processed through Bert and interactive attention network, serving as the feature representation of the news.
- **Cosine similarity:** This approach directly uses cosine similarity to calculate title-content matching degree.
- **Siamese Network:** This method employs a Siamese Network to calculate title-content matching degree.

In the experiments, the optimal hyperparameters validated earlier were used, and the results are shown in Figure 7. The results indicate that directly concatenating the title and content yields the worst effect, demonstrating the crucial role of title-content matching degree detection in the methodology of this paper. This suggests that mismatching between titles and content can lead to imprecise modeling of user features, and conducting title-content matching degree checks before constructing user interest preferences helps to correct user interest preferences, which is vital for modeling user interests accurately. Moreover, when calculating title-content matching degree, using a Siamese Neural Network outperforms the direct use of cosine similarity. This indicates that the Siamese neural network, employing Bi-LSTM at the encoding layer, can help learn more accurate representations of titles and content for the computation of title-content matching degree, further demonstrating the effectiveness of the title-content matching degree detection module based on the Siamese Neural Network adopted in this paper.

---

**Figure 6.** The Variation of Each Metric with the Value of $\lambda$

The Effectiveness of the Title-Content Matching Degree Detection Module

The Effectiveness of interactive attention network

In this section, the impact of interactive attention on the performance of our method is explored, as shown in Figure 8. Comparing the model without interaction attention network with the method in this paper, from the results, the performance of our method is significantly better than that of the model without interaction attention network module, which shows the effectiveness of interaction attention network. This is because interactive attention network can help capture the relationship between news titles and content, combining the advantages of phenotypic and interactive types to enhance the representations of titles and content. This facilitates the assessment of the relevance between the title and the body, aiding in the calculation of title-content matching degree.
In this section, the effectiveness of the Neural Collaborative Filtering based on FM in our method is intuitively explored from three perspectives:

- **w/o NCF**: A method without the integration of Neural Collaborative Filtering, solely utilizing content-based deep learning for recommendations.
- **NCF**: A method that integrates deep learning approaches based on title-content matching degree detection with traditional Neural Collaborative Filtering.
- **FM-NCF**: A method that integrates deep learning methods based on title-content matching degree detection with Neural Collaborative Filtering based on FM.

The experiments use the optimal hyperparameters verified above, and the results are shown in Figure 9. The results indicate that recommendations integrating Neural Collaborative Filtering significantly outperform those based only on deep learning. This demonstrates that Neural Collaborative Filtering can alleviate the issue of title-content mismatching and enhance recommendation performance. This is because Neural Collaborative Filtering does not rely on any additional information, meaning it does not consider text content, thus avoiding the issue of title-content mismatching altogether and recommending based solely on the user’s nearest neighbors' other clicked news. As shown in graph (b), the diversity metric significantly decreases after integrating Neural Collaborative Filtering, which is due to NCF often emphasizing popular news in user-news interaction data, increasing inequality in recommendation results. Moreover, Neural Collaborative Filtering, by solving the title-content mismatching issue, makes user interest modeling too precise, overlooking users' diverse needs. Furthermore, Neural Collaborative Filtering based on FM significantly outperforms traditional Neural Collaborative Filtering, indicating that FM-based Neural Collaborative Filtering can effectively improve recommendation performance. This is because FM can learn the implicit feature vectors of interactions between users and news, flexibly handling the complex relationships between high-order non-linear features. Both ablation experiment results validate the effectiveness of integrating Neural Collaborative Filtering based on FM in this paper.

**Conclusion**

In this paper, a hybrid recommendation method that considers the degree of title-content matching is proposed. An interactive attention mechanism is employed in the deep learning recommendation module to learn the relevance between titles and content, and title-content matching is assessed before modeling user interest preferences. In the Neural Collaborative Filtering module, FM is utilized instead of the original GMF to model both low-order and high-order feature interactions between users and news, alleviating data sparsity and considering the latent relationships among users. The final recommendation result is a blend of deep learning recommendation based on title-content matching degree detection and Neural Collaborative Filtering.
Data availability

All data generated or analysed during this study are included in this published article [and its supplementary information files].

References


29 Rendle, S. Factorization machines. in 2010 IEEE International conference on data mining. 995-1000 (IEEE).


Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- tran.csv