

## Research Article

# Design of Personalized News Recommendation System Based on an Improved User Collaborative Filtering Algorithm

Xuejiao Wang <sup>1</sup> and Chao Liu<sup>2</sup>

<sup>1</sup>College of Art and Design, Xi'an Mingde Institute of Technology, Xi'an 710124, Shaanxi, China

<sup>2</sup>School of International Economics and Trade, Universidad Complutense Madrid, Madrid 28040, Spain

Correspondence should be addressed to Xuejiao Wang; wangxj@mdit.edu.cn

Received 31 August 2022; Revised 27 September 2022; Accepted 1 October 2022; Published 22 February 2023

Academic Editor: R. Mo

Copyright © 2023 Xuejiao Wang and Chao Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To solve the problem of information overload in the field of news, this paper designs and implements a feasible news recommendation system, where the front-end web page is made by Django framework, whose performance is optimized by bootstrap and jquery, while in the back-end design, the original user similarity calculation method is improved by adding the time attenuation factor, and a news recommendation model based on user collaborative filtering (CF) algorithm is proposed. Experimental results show that the proposed algorithm achieves highest recall, accuracy, and *F1* score ratio compared with other algorithms, which indicate that the proposed algorithm has better performance.

## 1. Introduction

With the advent of the information age, the Internet provides a convenient way to obtain news. However, there are millions of news reports from different channels and fields on the Internet, which makes users lost in massive data, so that they need to spend a lot of time and energy to identify the news they want to browse, that is, “information overload” problem [1, 2]. Recommendation system is a kind of information filtering system, which can predict the user's preference for items according to the analysis of users' interest characteristics and historical behavior data, so as to help users make decision and analysis, which can improve the users' experience.

In the application of news field, the huge amount of news produced makes it extremely important to recommend news efficiently and quickly. At present, many websites generally adopt popular recommendation and click through rate recommendation to deal with massive news, but these methods often do not consider the interest difference between different users, and all the recommendation results are the same, which cannot really meet the actual needs of users [3, 4]. On the one hand, the current popular

recommendation system can reduce users' time to find news and improve their browsing experience. On the other hand, this system provides the output efficiency of news platform, so that more corresponding news can be timely browsed by users, which improves the utilization value of news and provides great commercial value for news platform companies.

Among them, news recommendation is a hot research field. By calculating the similarity of articles and considering topics, categories, etc., the content-based news recommendation system will generate a news list for users, which is similar to the news users as in [5]. In addition, users in the news field are more likely to be affected by popular items. At the same time, due to the timeliness of news recommendation, user interest is constantly evolving [6]. The collaborative filtering (CF) algorithm is useful for personalized news recommendation (PNC). For news recommendation based on CF, item CF is generally not adopted because the number of news is far greater than the number of users, and the update is fast, resulting in high computational complexity [7]. Therefore, a user-based CF algorithm is adopted in this paper, the basic principle of which is to find “neighbor” user groups similar to the current user's

preferences (ratings) according to all users' preferences (scores) for items or information.

The purpose of this paper is to develop a solution for the news recommendation and try to make this solution have strong compatibility and applicability. Through the research of hybrid recommendation technology, a feasible news recommendation system is designed and implemented. In the algorithm design, the existing similarity calculation method does not consider the time factor. Therefore, this paper considers the time factor and improves the original user similarity calculation formula; in addition, the time decay factor is considered.

## 2. Research Status of News Recommendation Algorithm

The principle of CF-based recommendation algorithm is to assume that users who like similar items have similar interests [8]. It first generates a set of users with similar interests for the target users and then recommends the items that the target users have not clicked on but the users in the collection have clicked on. Because the principle of recommendation algorithm based on CF is relatively simple, it has been popular in the field of recommendation. Wu et al. [9] proposed a cooperative noise reduction coder to recommend to users, where the automatic noise reduction encoder formulates the user project feedback data to generate a distributed structure of users and projects. Han [10] improved the two stages of data preprocessing and nearest neighbor selection in the CF algorithm, filled the user project evaluation matrix, introduced the tag factor and time factor, integrated CF with dichotomy K-means, and improved the similarity calculation formula. Saranya and Sadasivam [11] adopted a CF algorithm based on rough sets to score news categories, which improve the ranking of novel news. Jiang et al. [12] combined the knowledge graph with the CF algorithm, connected web API and mashup related information through the knowledge graph, and then calculated the distance between web API vectors to make recommendations. Xue et al. [13] used the matrix decomposition model to construct user news matrix by displaying rating and implicit feedback information of users and projected users and news into low-dimensional space through neural network.

Although the recommendation algorithm based on CF can obtain satisfactory recommendation results in some cases, it often has the problem of cold start. With the development of neural network, it has achieved good results in image processing, natural language processing, and other fields. Because the traditional recommendation model is a general algorithm model, it does not combine with specific business scenarios, so it is not effective in some specific application scenarios. In this case, the recommendation algorithm needs to make some changes according to different applications, so there is a recommendation algorithm model for a certain application scenario. Veličković et al. [14] proposed a novel attention mechanism network, which assigns different weights to different nodes in the nearest neighbor through the self-attention mechanism layer. These recommended algorithm models for

a certain field need to be combined with the actual business scenarios and need to design the models from the perspective of business and consider the actual application effect of the models. Neural network can find the hidden feature information in user behavior records and capture the interaction characteristics between the project and the user, which improves the accuracy of the recommendation algorithm.

## 3. Design of PNC System

*3.1. Overall Structure.* The architecture design of the system is shown in Figure 1. The whole system architecture design is divided into the following parts.

The data layer processes and stores user information, news information, and log file information, respectively. Pandas data processing framework is used in the process, and MySQL and Redis are used in the storage. There is user's historical information in the user's news platform. In addition, the historical browsing behavior of users is the ID of news, which needs to be mapped to the corresponding news content. In this system, the news refers to the news title. Log files record all operations in the system and are saved in the MySQL database.

The strategy layer includes recall strategy, sorting strategy, cold start strategy, and reordering strategy. This layer is to model the user according to the news content of user behavior sequence, so as to recommend the interested news content to the user according to the user interest of the news platform. The process steps of this layer are as follows:

- (1) The candidate news is recalled in the news dataset according to the news in the user behavior. In this step, several kinds of candidate news are recalled in the mass news dataset.
- (2) In the news dataset, the recommended range of news is reduced, and then the news ranking algorithm mentioned above is used to further sort the candidate news to generate the news recommendation list.
- (3) When there is no history browsing record in user behavior sequence, there is no sorting strategy. In this case, we need to use cold start strategy to recommend to users.
- (4) After sorting, the news needs to be reordered from different angles, and the final sorting results are returned to the user.

In the strategy layer, the model code is edited and trained through the Python framework.

The feedback layer includes feedback evaluation and recommendation list. This layer is to evaluate and adjust the results of the strategy layer. After the strategy returns the recommendation result, it evaluates the recommendation result according to the evaluation indicators such as accuracy, freshness, and popularity. When the strategy does not meet the requirements of the index, it needs to adjust the strategy again. In the evaluation process, SKlearn framework is used for evaluation.

The application layer provides interactive support for users and administrators, including user registration and

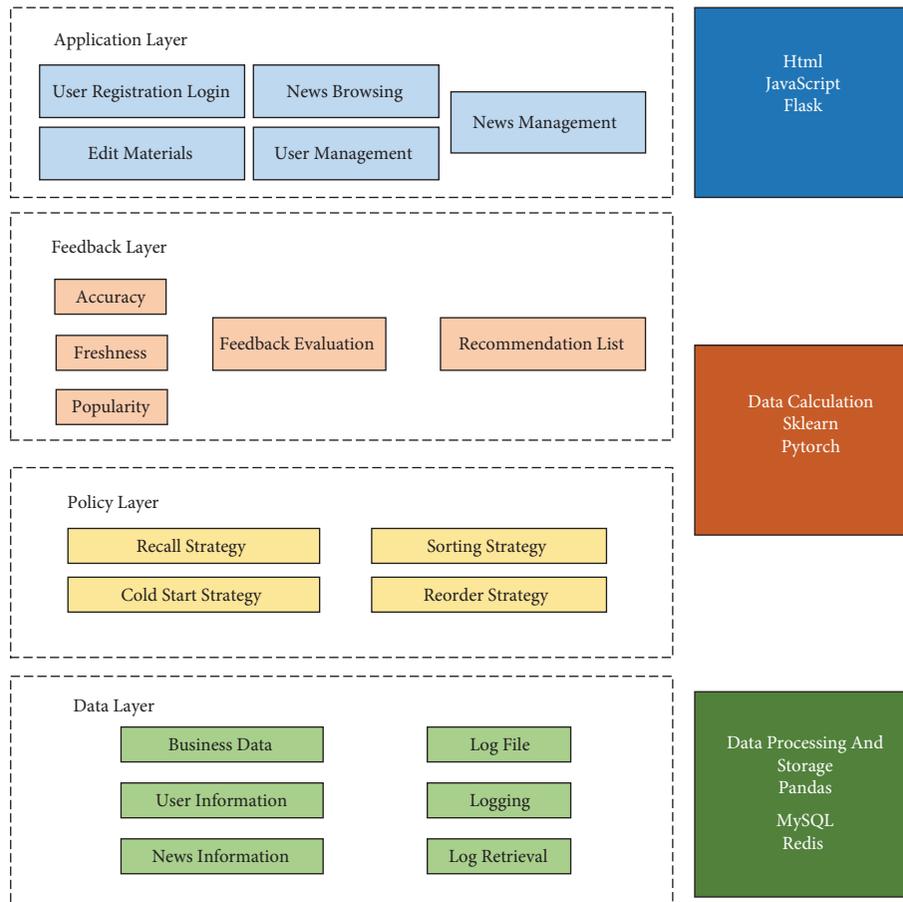


FIGURE 1: Overall system architecture.

login, news browsing, editing materials, user management, and news management. The whole framework uses Flask, HTML, and JavaScript to participate in the interaction process between the front end and back end.

**3.2. Function Module.** The system function module is divided into five parts, registration and login module, editing data module, news browsing module, user management module, and news management module. The system function module is shown in Figure 2. The registration and login module includes user registration and user login. The editing data module includes nickname modification, avatar modification, and password modification. News browsing module includes news browsing, category news browsing, historical news browsing, and recommended news browsing. The user management module includes user search and user deletion. News management module includes news release, news editing, and news deletion.

#### 4. Realization of PNC System

The recommendation system designed in this paper is mainly presented in the form of web pages. Since the back-end recommendation algorithm is designed by Python, in order to facilitate the interaction between the back end and the front end, the front-end web page is planned to use

Django framework, and bootstrap and jquery are used to optimize the web page effect. As a mainstream and popular web framework, Django uses a program structure similar to MVC, using MVT (Model, View, and Template) architecture. The full name of MVC is Model View Controller. M refers to Model, which mainly encapsulates the access to the database layer and adds, deletes, modifies, and checks the data in the database. V refers to View, which is used to encapsulate the results and generate the HTML content displayed on the page. C refers to Controller, which is used to receive requests, process business logic, interact with Model and View, and return results. MVT is the proper name of Django framework, and its function is the same as that of MVC. In MVT, Model is mainly used to define and interact with the database. View is similar to controller in JSP, which is used to control the behavior of the website in the background. Template is used to display the front-end effect of a web page.

**4.1. Front-End Functions.** The front-end design of website mainly involves the design and development of template. Different from JSP technology, under the Django framework, an HTML web page is used not only as a page but also as a template that can be reused many times.

The realization of template function: when a user logs in to a website of Django framework, the URL controller is used

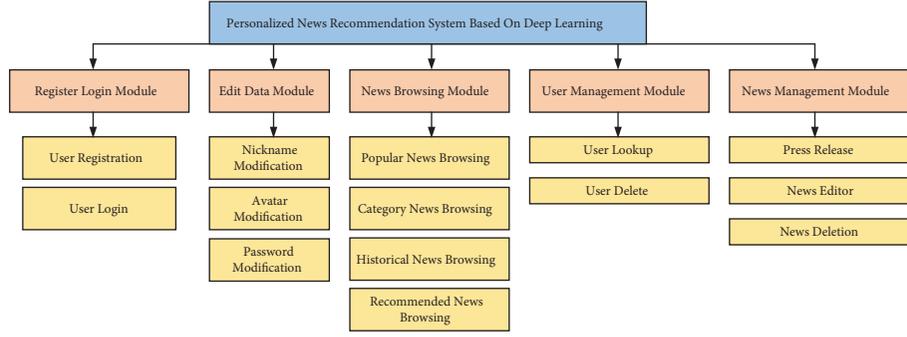


FIGURE 2: System function module.

for routing distribution to match the appropriate view function, and then the view function will read models or render a template directly and return to the user interface according to the process written by the programmer. Different view functions can render the same template, and the contents of the template will change according to the parameters passed by the view, which can reuse the pages with little change in style and reduce a lot of development costs. The running process of the whole framework is shown in Figure 3.

In addition to the design of templates, the front end also involves the display of back-end data. In this framework, the data are first processed by the controller and then transmitted to the template through the corresponding API. Taking the home page as an example, the parameters of the page include news title, news classification label, and user information. Among them, the news title and news tag are imported into the page as a list, and the user information only needs to pass in a user name, while in the background, the news list will be sorted according to the user's information and then output after processing.

The method of calling parameters in the template is similar to that in JSP technology. Each element in the list is displayed in the page by traversing through the parameters passed in. It is worth mentioning that all the pages of this project use bootstrap to beautify the appearance of web pages. Because bootstrap is a responsive framework, using bootstrap can avoid the problem of display dislocation in different size devices.

**4.2. Back-End Functions.** Assume that the user named  $U$  needs personalized recommendation. First, find the users with similar interests. Those users are called the nearest neighbor users. Then, recommend the news that the nearest neighbor users like but the user  $U$  has not seen. This is the CF algorithm based on users. The advantage of this recommendation system is that the recommended news may be completely irrelevant in content, so it can discover the potential interests of users and generate personalized recommendation results for each user.

**4.2.1. Algorithm Design.** The CF algorithm mainly uses the similarity of behavior to calculate the similarity of interest. Given user  $u$  and user  $v$ , formula (1) is used to calculate the similarity of interest of  $u$  and  $v$ , where  $N(u)$  represents the

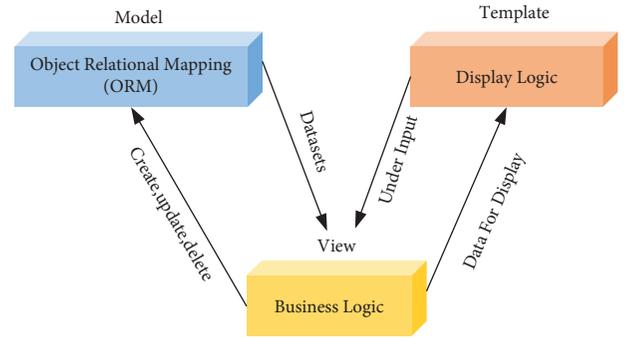


FIGURE 3: Operation process of MTV structure.

news set that user  $u$  has read, and  $N(v)$  represents the news set that user  $v$  has browsed.

$$W_{uv} = \frac{|N(u) \cap N(v)|}{\sqrt{|N(u)||N(v)|}}. \quad (1)$$

User  $u$  has read news  $\{a, b, d\}$ , and user  $v$  has read news  $\{a, c\}$ . Formula (1) is used to calculate the similarity between user  $u$  and user  $v$ , and the calculation of  $W_{uv}$  can be rewritten by the news sets  $\{a, b, d\}$  and  $\{a, c\}$ , which are shown in the following formula:

$$W_{uv} = \frac{|\{a, b, d\} \cap \{a, c\}|}{\sqrt{|\{a, b, d\}||\{a, c\}|}} = \frac{1}{\sqrt{6}}. \quad (2)$$

**4.2.2. Improved Algorithm.** Formula (1) only considers whether the user has read a certain news, but not the reading time. Due to the strong timeliness of news recommendation, in the user's reading history, the farther away from the present, the more limited the effect of user's behavior on prediction. According to the characteristics of news recommendation, we adjust the similarity formula. At the same time, the influence of hot news and news reading time on similarity is considered. In order to punish the influence of the popular news in the common interest list of user  $u$  and user  $v$  on their similarity, the popular news is downgraded to avoid the Harry Potter effect. The calculation formula considers the time attenuation factor to prevent the similarity between user  $u$  and user  $v$  who like news  $i$  from being too small within a certain time range. Formula (1) is

improved by considering the characteristics of news recommendation, as shown in the following formula:

$$\frac{\sum_{i \in N(u) \cap N(v)} m^* (Ta - To/Tc - To) + (1 - m)1/1 + N(i)}{\sqrt{|N(u)||N(v)|}}, \quad (3)$$

where  $N(u)$  represents the reading set of user  $u$ ,  $N(v)$  represents the reading set of user  $v$ ,  $N(i)$  represents the occurrence times of the news item  $i$ ,  $Ta$  represents the execution time,  $Tc$  represents the current time,  $To$  represents the initial time, and  $m$  represents the adjusting factor, which ranges from 0 to 1. The larger  $m$  is, the more factors are considered in the change of interest, while the smaller  $m$  is, the more punishment is imposed on popular news.

For the convenience of description, this improved CF algorithm is renamed as UserCF algorithm.

**4.2.3. Algorithm Flow.** After obtaining the user's similarity, UserCF algorithm recommends  $k$  news that users are most interested in but not read by the recommended user. Formula (3) measures user  $U$ 's interest in news  $i$  in UserCF algorithm:

$$P(u, i) = \sum_{v \in S(u, k) \cap N(i)} W_{uv} R_{vi}, \quad (4)$$

where  $S(u, k)$  represents the  $k$  nearest neighbor sets that are most similar to user  $u$ 's interest,  $N(i)$  represents the user set that has read news  $i$ ,  $W_{uv}$  represents the interest similarity between user  $u$  and user  $v$ , which can be calculated by formula (3),  $R_{vi}$  represents user  $v$ 's preference degree or score for news, user  $v$  has read news  $i$ , and  $R_{vi}$  is 1. The recommended news is the one that user  $u$  has not read before. The algorithm of user  $u$ 's interest in news  $i$  is described as follows. for  $((v = 1; v \leq \text{all users}; v + 1))$

Formula (3) is used to calculate  $W_{uv}$ .

- (i) *Step 1.* Sort all  $W_{uv}$  to get  $k$  nearest neighbor sets  $S(u, k)$  with the most similar interests of user  $u$ .
- (ii) *Step 2.* Find the users  $S(u, k) \cap N(i)$  who have read news  $i$  from  $k$  nearest neighbor users.
- (iii) *Step 3.* The similarity degree  $W_{uv}$  of each user  $v \in S(u, k) \cap N(i)$  in user  $u$  and set  $S(u, k) \cap N(i)$  is multiplied by the user's interest degree  $R_{vi}$  and the sum is the interest degree of user  $u$  in news  $i$ .
- (iv) *Step 4.* According to  $P(u, i) = P(u, i) + W_{uv} * R_{ui}$ , sort  $P(u, i)$  to find the news that  $k$  neighbors are most interested in and recommend it to user  $u$ .

When a new user registers, the user will select several categories of interest. These categories are displayed dynamically through the word segmentation results of news in the database, which are not fixed. After users choose, the back end will give priority to the categories they are interested in during the process of visiting the website. This recommendation method is mainly to deal with the problem of cold start. Before the recommendation data are formed, tag recommendation is used to deal with the situation that

the recommendation system does not have sufficient data for recommendation analysis. Similarly, SQL statements are used to organize the required news objects into a linked list and then render the output. In the page, the news headlines are read in an iterative loop. The recommendation process is shown in Figure 4.

## 5. Experiment and Analysis

**5.1. Dataset.** The dataset used in this experiment is from the user data of a domestic news website, including 5000 users and nearly 120000 reading records. Each dataset contains six parts, and the details are shown in Table 1. In this experiment, the dataset is divided into two parts, in which the ratio of training set and test set is 7:3. The data information is shown in Table 1:

**5.2. Evaluation Index.** In this experiment,  $F1$  score, precision, and recall were used to evaluate the algorithm proposed in this chapter.  $F1$  score is the harmonic mean of accuracy and recall rate. Accuracy refers to the ratio of recommended correct news items to the recommended list, while recall ratio refers to the ratio of recommended correct news items to the list clicked by users. The calculation formulas are as follows:

$$F1 = 2 * \frac{(\text{precision} * \text{recall})}{(\text{precision} + \text{recall})},$$

$$\text{precision} = \frac{\sum_u |R(s) \cap T(s)|}{\sum_u |R(u)|}, \quad (5)$$

$$\text{recall} = \frac{\sum_u |R(s) \cap T(s)|}{\sum_u |T(s)|},$$

where  $R(s)$  represents the recommended list of news titles recommended to user  $s$  and  $T(s)$  represents the news headlines clicked by user  $s$  in the test set.

### 5.3. Results and Discussion

**5.3.1. Parameter Optimization.** First of all, in the fusion similarity calculation, through the weight influence factor  $\alpha$ , traverse the value, observe the change of precision value, and measure the weight of similarity calculation method. The experimental results are shown in Figure 5.

The change of weight influence factor  $\alpha$  can affect the prediction accuracy of news recommendation algorithm. When  $\alpha = 0.5$ , the news recommendation algorithm that integrates user reading records and news title similarity proposed in the model has the best performance, and when  $\alpha = 0.4$ , the accuracy is the highest. Therefore,  $\alpha$  is selected as 0.4.

**5.3.2. Comparison of Different Algorithms.** In order to verify the effectiveness of the back-end design of the system, the experimental results are compared and analyzed with other algorithms. In this paper, three algorithms are selected for

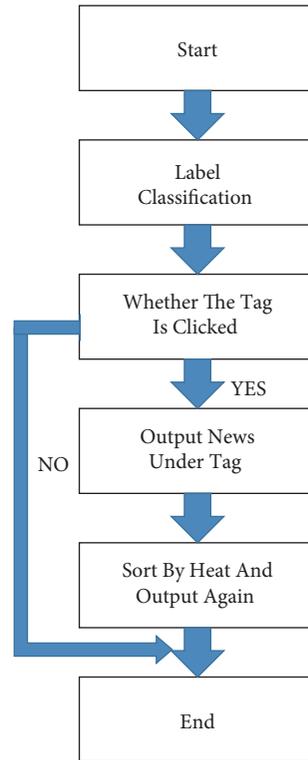


FIGURE 4: News recommendation process.

TABLE 1: Dataset information.

| Attribute   | Details                    |
|-------------|----------------------------|
| User_id     | ID of the user             |
| Time        | When the user clicks       |
| Click       | Does the user click        |
| Click_count | Number of user clicks      |
| Category_id | Category of news headlines |
| Title       | News headline text         |

comparison: PNR, FALS, and NR\_LDA. The results are shown in Figures 6–8.

PNR is an improved PNC algorithm based on consumer click behavior [15]. Firstly, association rules between adjacent news are established in the news browsing sequence of consumers, and then the restriction of browsing time difference is added to the construction of association rules to recommend news to users.

Fused ALS (FALS) uses the weighted hybrid recommendation strategy to fuse ALS model recommendation, content-based recommendation, and improved cyclic neural network algorithm and obtains a hybrid algorithm model through logical regression training weight [16].

NR\_LDA combines the user’s interest and news timeliness [17] and adds the influence of news release time to the topic extraction by using the LDA model to improve the effect of the recommendation algorithm.

As can be seen from the above figures, with the increase of recommendation list  $N$ , the accuracy first increases and then decreases. When  $N=10$ , the accuracy reaches the

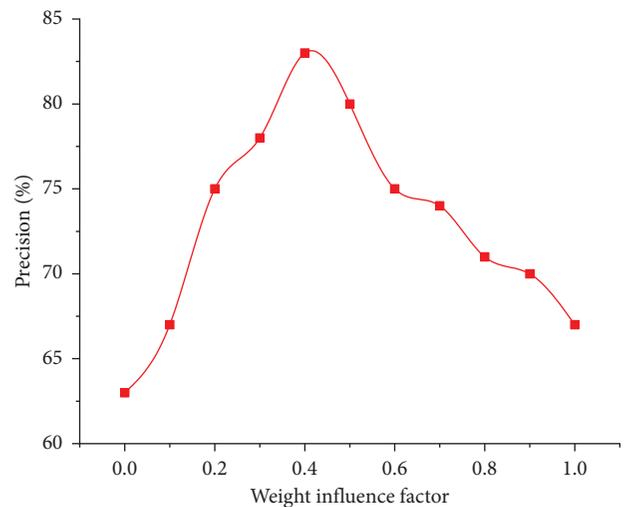


FIGURE 5: Influence of weight influence factors on algorithm precision.

highest. Under the same  $N$  value, compared with the four algorithms, the proposed algorithm improves by 2.7% compared with PNR, 3.9% compared with NR\_LDA, and 12.6% compared with FALS, with the highest accuracy among the four algorithms. With the increase of recommendation list  $N$ , the recall rate increases gradually. The recall rate of the proposed algorithm is higher than that of the other three algorithms in the same recommendation list  $N$ , indicating that the proposed recommendation algorithm has better performance.

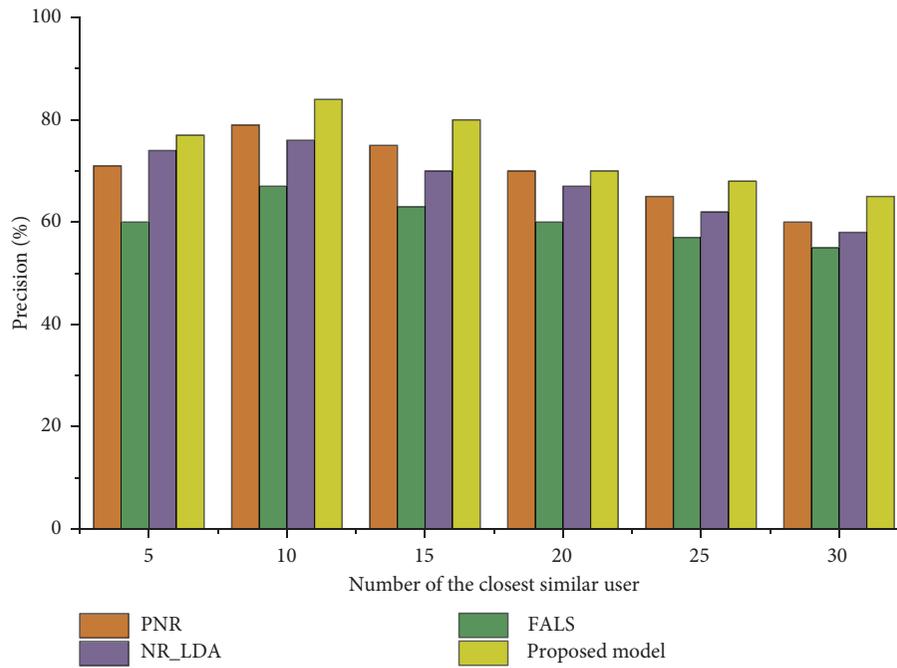


FIGURE 6: Precision comparison of different models.

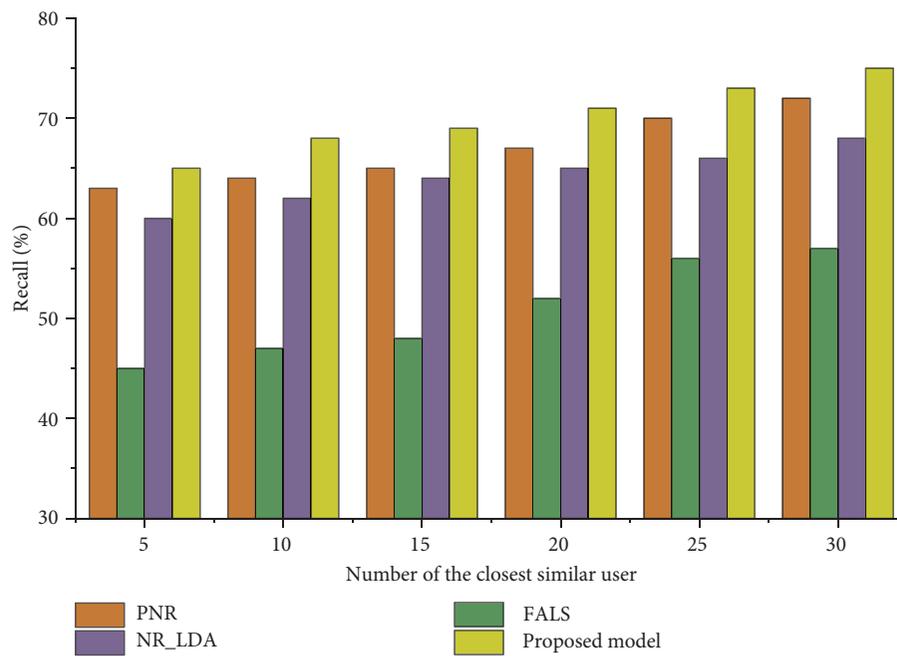


FIGURE 7: Recall comparison of different models.

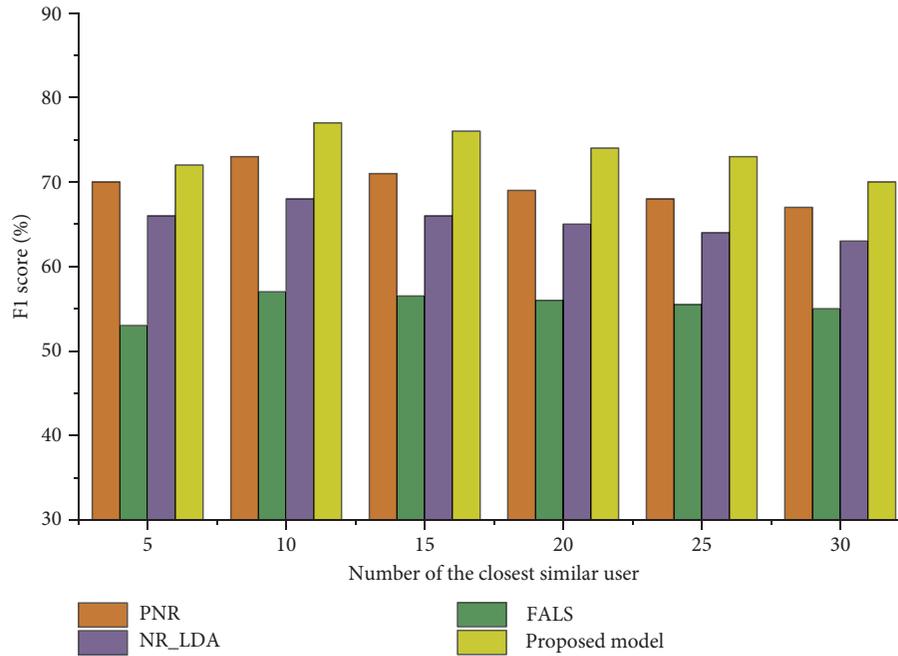


FIGURE 8: F1 score comparison of different models.

To balance recall rate and accuracy, F1 score was used as the evaluation standard of this experiment. As shown in Figure 8, with the increase of  $N$ , F1 score first increases, then decreases, and finally tends to be stable, which reaches its maximum when  $N=10$ . In the case of a small number of similar users recently, the number of news recommended by users may not be comprehensive enough. However, if too many recent similar users are selected, some users with low similarity degree will be included, and the news with high popularity clicked by these users that has no interest of target users will be recommended to target users, resulting in reduced accuracy. Therefore, in the news recommendation algorithm, it is necessary to reasonably select similar users as recent similar users to obtain the ideal CF recommendation effect.

## 6. Conclusion

This paper designs a personalized news recommendation system. The functional requirements of this system need to meet the acquisition and storage of news data, user behavior recording, news content display, etc. Then, the architecture design of PNC system is introduced, including multilayer structure, where the front-end web uses Django framework to reduce development costs and greatly improve development efficiency, while the back-end design adopts the news recommendation model based on the UserCF algorithm. The test results show that the recall, accuracy, and F1 score of the proposed algorithm are higher than those of the other three algorithms in the same recommendation list  $N$ .

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request.

## Conflicts of Interest

This study was conducted in the absence of any commercial or financial relationship that could be interpreted as potential conflicts of interest.

## References

- [1] Y. Wu, X. Wang, and Y. X. Ding, "Adaptive on-line web topic detection method for web news recommendation system," *Acta Electronica Sinica*, vol. 38, no. 11, pp. 2620–2624, 2010.
- [2] M. Zihayat, A. Ayanso, X. Zhao, H. Davoudi, and A. An, "A utility-based news recommendation system," *Decision Support Systems*, vol. 117, pp. 14–27, 2019.
- [3] Y. Peng, *Research and Implementation of PNC Algorithm Based on Improved LDA Topic Model*, Fuyang Normal University, Fuyang, China, 2021, in Chinese.
- [4] Y. Wu, M. Qi, and R. Yang, "A news recommendation system based on improved CF algorithm," *Computer engineering and science*, vol. 39, no. 06, pp. 1179–1185, 2017, in Chinese.
- [5] Y. Xu, "News recommendation analysis and exploration based on the integration of content-based recommendation and CF," *Guide to news research*, vol. 9, no. 13, p. 111, 2018, in Chinese.
- [6] F. Garcin, K. Zhou, and B. Faltings, "PNC based on CF," in *Proceedings of the 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, vol. 1, pp. 437–441, Macau, China, December 2012.
- [7] S. Jiang and W. Hong, "A vertical news recommendation system: CCNS—an example from Chinese campus news reading system," in *Proceedings of the International Conference on Computer Science & Education*, pp. 1105–1114, British Columbia, Canada, August 2014.
- [8] Z. Cui, X. Xu, F. Xue et al., "Personalized recommendation system based on collaborative filtering for IoT scenarios," *IEEE Transactions on Services Computing*, vol. 13, no. 4, pp. 685–695, 2020.

- [9] Y. Wu, C. DuBois, A. X. Zheng, and M. Ester, "Collaborative denoising auto-encoders for top-N recommender systems," in *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pp. 153–162, San Francisco, CA, USA, February 2016.
- [10] K. Han, "Personalized news recommendation and simulation based on improved collaborative filtering algorithm," *Complexity*, vol. 2020, Article ID 8834908, 12 pages, 2020.
- [11] K. G. Saranya and G. S. Sadasivam, "Personalized news article recommendation with novelty using CF based rough set theory," *Mobile Networks and Applications*, vol. 22, no. 4, pp. 1–11, 2017.
- [12] B. Jiang, J. Yang, Y. Qin, T. Wang, M. Wang, and W. Pan, "A service recommendation algorithm based on knowledge graph and collaborative filtering," *IEEE Access*, vol. 9, pp. 50880–50892, 2021.
- [13] H. J. Xue, X. Dai, and J. Zhang, "Deep matrix factorization models for recommender systems," *IJCAI*, vol. 17, pp. 3203–3209, 2017.
- [14] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017, <https://arxiv.org/abs/1710.10903>.
- [15] Y. Wang and W. Shang, "PNC based on consumers' click behavior," in *Proceedings of the International Conference on Fuzzy Systems & Knowledge Discovery*, pp. 634–638, IEEE, Zhangjiajie, China, August 2016.
- [16] Hukaida, *Design and Implementation of PNC System Based on Hybrid Recommendation Algorithm*, Beijing University of Posts and telecommunications, Beijing, China, 2021, in Chinese.
- [17] Z. Ding, Y. Peng, and G. Sun, "PNC algorithm based on improved LDA topic model," *Journal of Chifeng University (Natural Science Edition)*, vol. 37, no. 06, pp. 28–32, 2021, in Chinese.