Hindawi
Mobile Information Systems
Volume 2022, Article ID 5198612, 9 pages
https://doi.org/10.1155/2022/5198612

Research Article

Personalized Recommendation Algorithm of Literary Works Based on Annotated Corpus

Guangping Lv

Zhejiang University of Science & Technology, Hangzhou 310023, China

Correspondence should be addressed to Guangping Lv; 118001@zust.edu.cn

Received 14 April 2022; Revised 16 May 2022; Accepted 18 May 2022; Published 29 June 2022

Academic Editor: Liping Zhang

Literary work personalized recommendation service is a service that focuses on users' needs, actively analyses users' interests and hobbies, and intelligently and efficiently discovers users' interesting information. Previous recommendation algorithms were unable to make effective and accurate recommendations in real time, resulting in poor recommendation outcomes. This study proposes a personalized recommendation algorithm for literary works based on the annotated corpus to address these issues. The dictionary is first used to mark the original text that was created by splitting words. The user's reading behavior is then analyzed using a combination of individual personality characteristics and a set of factors that differ from the individual background, and personality characteristics and situations. Finally, we count the frequency of each modifier and word in the modifier vector in the corpus for each word, create the feature vector, and perform cluster analysis. The results show that this method's MAE (mean absolute error) value is always lower than the traditional method, especially when the neighbor set size is 5, and that this method is clearly superior to the traditional method, with a maximum difference of 6.23%. *Conclusions*. The algorithm can produce satisfactory recommendation results and can be used to make personalized literary recommendations.

1. Introduction

Science and technology are rapidly evolving in today's information age, particularly the rapid development of computer information network technology and communication technology [1], which provides great convenience to our daily lives, and product and user data. Readers will spend more time selecting their favorite works when browsing literary works as the variety of literary works grow, which can easily bore them. In this context, recommendation technology has gotten a lot of attention because it can help with information overload. When users need to make a decision, they are given relevant information or decision-making suggestions based on the data, and artificial simulation is used to assist them. Users will not be perplexed when navigating the vast ocean of knowledge once they have completed the selection and decision-making process. Knowledge navigation can benefit from the use of personalized recommendation technology.

It is difficult to accurately and effectively get the information you need in the face of vast data, so you need to use information filtering technology. At present, information filtering technology is mainly divided into retrieval and search engine technology and recommendation system technology. Guzmán-Cabrera proposed a content-based weighted granularity sequence recommendation algorithm [2] by analyzing the attribute relationship of items. Chen proposed a model based on content and interest drift to be applied to the movie recommendation algorithm [3]; Yuan et al. used the characteristics of association rule mining to mine the association between user attributes and items, and proposed a classification random walk algorithm based on association rule mining [4]; With the wide application of CF(collaborative filtering) recommendation system in practical systems, some shortcomings are gradually exposed. For example, the history of users in the system is relatively poor, which is the so-called data sparseness problem. In the case of sparse data, it is difficult to accurately measure the similarity between users, resulting in the inability to form a reliable nearest-neighbor set, which seriously reduces the recommendation accuracy. Data sparseness has become a bottleneck restricting the development of collaborative recommendation technology [5, 6].

At present, research on Chinese vocabulary and word formation in the field of Chinese language and literature is limited to grammar, parts of speech, and the like, with no comprehensive study of semantics. The integration of research findings and information technology is still lacking. Personal preference errors in readers' ratings, such as failing to consider the position weight of feature words when processing the review data text, can have a negative impact on recommendation accuracy. These issues must be immediately addressed [7]. We try to strictly and rigorously establish semantic components on the basis of previous research results, combining with the actual situation reflected in the actual text; we do our best to describe each semantic component in a relatively complete manner. The goal of this research is to make advancements in both theory and practice. This study extracts and establishes the example database of modern Chinese special sentence patterns, and the model sentence system of special sentence patterns, using a large-scale tagging corpus. The thesis's research innovation is as follows:

(1) Based on the unsupervised Chinese part-of-speech tagging of a single corpus, this study applies unsupervised Chinese part-of-speech tagging to parallel corpora, constructs and designs related models, and verifies them through experiments

(2) In this study, a personalized recommendation algorithm for literary works based on tagged corpus is proposed, which makes full use of item category information and dynamically adjusts the user weights in neighbor sets according to different target items, which can more accurately depict the similarity between users. A modified overlap factor is proposed to make up for the deficiency of manual parameter adjustment in the existing methods, which enhances the practicability of the method.

2. Related Work

2.1. Annotated Corpus Research. Chinese word formation has been paid attention to by Chinese lexicology and grammar for a long time so that word formation-related research has become a hot topic in the Chinese language field. Huang et al. proposed a complete second-order hidden Markov model for Chinese part-of-speech tagging [8]; Yuan introduced the method of bidirectional Chinese part-of-speech tagging based on the traditional hidden Markov model [9]. Dalton et al. added the position information of words in sentences to the part-of-speech tagging, added this information as a feature to the algorithm, and constructed a novel maximum entropy Markov model. This information was more concretely expressed in the algorithm. After adding new features, the newspaper corpus was tested, and the accuracy rate exceeded 95% [10].

Saif et al. put forward a method of part-of-speech tagging based on unsupervised multilingual learning. This method uses the hierarchical Bayesian model to predict the part-of-speech tagging sequence of two languages, and the results verify the effectiveness of multilingual learning [11]. Zheng et al. put forward a complete Bayesian method for unsupervised part-of-speech tagging, which integrates all possible parameter values, unlike only estimating a single set of parameters. Using the Bayesian method for part-of-speech tagging can achieve better performance than using maximum likelihood estimation [12]. Shi and Zhu explored the corresponding mechanism between semantic components and syntactic components of sentences. Through the preliminary investigation of the corresponding relationship between the two semantic components of agent and patient and the three syntactic positions of subject, adverbial, and object, it is clear that the semantic features of nouns, predicate verbs, and sentence patterns restrict the semantic components from appearing in syntactic positions [13].

2.2. Research on Personalized Recommendation Algorithm. As the core of the recommendation system, the personalized recommendation algorithm collects some previous information from users, analyzes users' preferences, and makes recommendations to users. Nowadays, personalized recommendation algorithm has been deeply researched and applied in the fields of e-commerce, education, and tourism services.

Warren Wang et al. considered expressing the knowledge system in the knowledge network, introduced the nearest-neighbor first candidate knowledge selection strategy, and proposed a personalized knowledge recommendation method based on constructivist learning theory—the constructive recommendation model [14]. Hu et al. proposed the research of the CF recommendation algorithm by integrating big data technology, social network analysis technology, and key user analysis technology [15]. It can be seen that the hybrid recommendation algorithm combining machine learning, data mining, and other knowledge is the main direction of future research and application. Dai et al. used data mining knowledge and the CF algorithm to propose a hybrid recommendation algorithm that combines user clustering and rating preferences [16]; Piao et al. made recommendations by using the similarity between resources and users' interests. When creating a user interest profile, you can comprehensively analyze the user's interest and behavior and abstract it into a vector expression.

Some previous studies have completed the commodity clustering operation based on fuzzy clustering [17, 18], which effectively improved the recommendation effect of the recommendation system. In Guo and Deng, in order to effectively solve the defects of the high-dimensional sparse matrix model, combined with the multilevel association rule algorithm, the final experimental results show that the recall rate and calculation time have been optimized, effectively solving the problems faced by the recommendation system in the case of sparse readers [19]. Hu et al. put forward CF based on a neural network, which selects the candidate nearest-neighbor set according to the intersection of the user's score vectors and uses the BP neural network to predict users' scores on items, thus reducing the sparsity of candidate nearest-neighbor datasets [20]. Ping solved the

problem of data sparseness to some extent by increasing user context information [21]. Under the social network environment, Xian et al. studied the book e-commerce recommendation system, integrated the trust rating into it, and thought that the trust mechanism played a very important role in the social network. Finally, they proposed a trust mechanism and recommendation method based on the social network recommendation system model [22].

3. Methodology

3.1. Semantic Analysis. There is usually no consensus on the scope or existence of subject-predicate sentences due to differing interpretations of Chinese grammatical features, subjects, and topics, especially because the semantic and pragmatic analyses of subject-predicate sentences are not thorough enough. We discovered that tagging the subject-predicate sentences is a difficult task during the corpus tagging process. The internal components of a sentence should not only be semantically compatible, but we have been thinking about how to classify subject-predicate sentences.

Let w_i be any word in the text. If its first two words $w_i - 2w_i - 1$ in the text are known, the conditional probability $P(w_i|w_i - 2w_i - 1)$ can be used to predict the probability of w_i appearing. This is the concept of the statistical language model.

It consists of n words in sequence, i.e., $W = w_1 w_2, \ldots, w_n$; then, the statistical language model is the probability $P(W)$ of the word sequence W appearing in the text. Using the product formula of probability, $P(W)$ is expanded as follows:

$$P(W) = P(w_1)P(w_2|w_1)P(w_3|w_1w_2)\cdots P(w_n|w_1w_2\cdots w_{n-1}). \quad (1)$$

It is not difficult to see that in order to predict the occurrence probability of the word w_n, it is necessary to know the occurrence probability of all the words before it.

Because metaphor is rarely generated through grammatical structure, only contextual semantic information can be used to determine whether a word or phrase is metaphorical. Some traditional methods cannot be identified at this time. The current deep learning model is used to learn contextual semantic information before performing metaphor recognition in this problem. As a result, the deep learning model will be used in this section to learn the semantic information between sentences, design a semantic-based algorithm framework based on the actual needs and feasibility of the algorithm, and make the final recognition of verb metaphors based on the dependency relationship between semantic information. Figure 1 shows the algorithm framework flowchart for the semantic framework flowchart in this document.

The following will introduce the steps of the framework process and the related technologies involved:

(1) We acquire original text data through other technologies such as crawlers

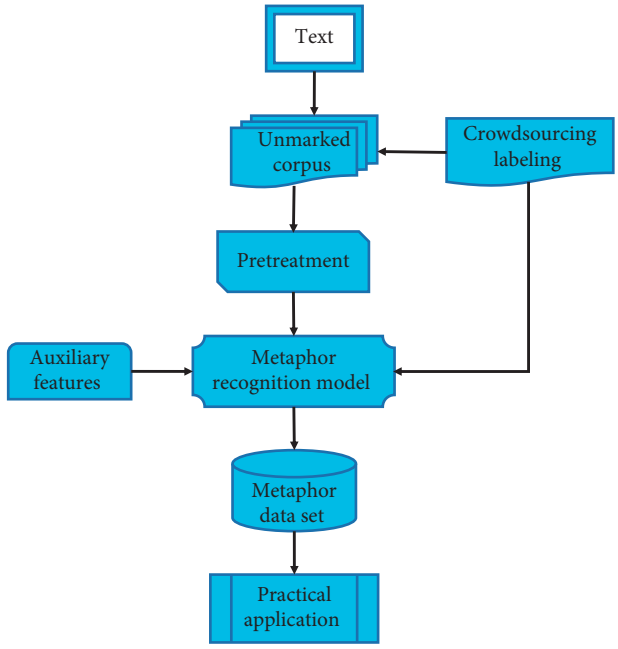

Figure 1: Flowchart of algorithm framework.

(2) The unlabeled corpus is marked by word segmentation, and then, a frequency dictionary is constructed to convert sentences into model input format;

(3) For the marked text data that meet the evaluation requirements, we construct the final metaphor corpus according to the requirements

(4) We use a crowdsourcing platform to manually label metaphors

(5) For the data that are not marked as required, we temporarily put it into the unlabeled corpus

First, we select features from the original data and then generate feature subsets. If TF-IDF (term frequency-inverse document frequency) feature weighting algorithm is not used, only numbers 0 and 1 can be used to measure whether there are feature words in the text.

The more times a word appears in a particular article, the greater the weight of the word in the document, as shown in the following formula:

$$w_{\mathrm{TF}}(t) = \mathrm{TF}(d,t). \quad (2)$$

The TF-IDF algorithm is practical at present, and it is often used in text processing, but there are still some shortcomings in some aspects. We need to improve TF-IDF by integrating the weights, and the improved algorithm is shown in the following formula:

$$\mathrm{TF-IDF-DW}(t_i) = \mathrm{TF}(d,t_i) \times \mathrm{IDF}(t_i) \times \mathrm{DW}(t_i) = \mathrm{TF}(d,t_i) \times \log\frac{N}{n} \times \overset{(c)}{\max}_{j=1} \mathrm{DD}(C_i,t)\mathrm{CD}(C_i,t). \quad (3)$$

Here, it is represented by DW weight value. If a feature item is evenly dispersed in different classified documents, even if the inverse document frequency value is large, then the value of interclass dispersion after analysis will be small, and the weight value obtained by it will not be correspondingly large.

3.2. Corpus Part-of-Speech Tagging. Readers in the same community often have a common tendency when choosing literary works. At the same time, through analysis, it is found that readers in different communities have different tendencies when choosing literary works. Therefore, we come to the conclusion that it is very important to divide readers' preferences in reader communities, which can improve the accuracy of recommendation. To accurately measure this phenomenon of community division in recommendation, the most important thing is to use readers' scores of literary works to divide communities.

Behavior is determined by personality traits and situations, that is, behavior is a function of personality traits and situations, as shown in formula (4).

$$B = f(\text{Per} \times \text{Situ}). \tag{4}$$

Therefore, by integrating individual personality characteristics, deducing from the set of individual background factors, and combining personality characteristics and context, this study analyzes users' literary reading behavior, so as to improve users' recommendation rate of literary works, improve the reading effect, and recommend literary works for users.

In order to prevent the recommended users from liking the recommended literary works, this study puts forward the concept of personality compatibility. Character compatibility reflects the user's tolerance for the character characteristics of each candidate literary work. It is calculated by the degree of compatibility between users who belong to this personality trait and each candidate literary work, and the grade reflects the user's preference for literary works, as shown in Figure 2.

In this study, users' preference for literary works type B_t is divided into two categories: preference and nonpreference, and literary works and literary works' types are divided into belonging and nonbelonging. The compatibility of end users with literary works depends on the number $|B^b|$ of literary works' types contained in literary work b.

After analyzing the user's compatibility ranking of literary works, formula (5) is adopted for specific calculation, which can fully reflect these eight categories.

$$C_{u,b} = \frac{\sum_{B_t}\left(P_{u,B_t} - \overline{P}_u\right)\left(f_{b,B_t} - \overline{f}_b\right)}{\sqrt{\sum_{B_t}\left(P_{u,B_t} - \overline{P}_u\right)^2 \times \sqrt{\sum_{B_t}\left(f_{b,B_t} - \overline{f}_b\right)^2}}}, \tag{5}$$

where $C_{u,b}$ represents the compatibility of user u with literary works b; P_{u,B_t} indicates the user's preference degree of literary works' type of user u to the t literary works' type B_t;

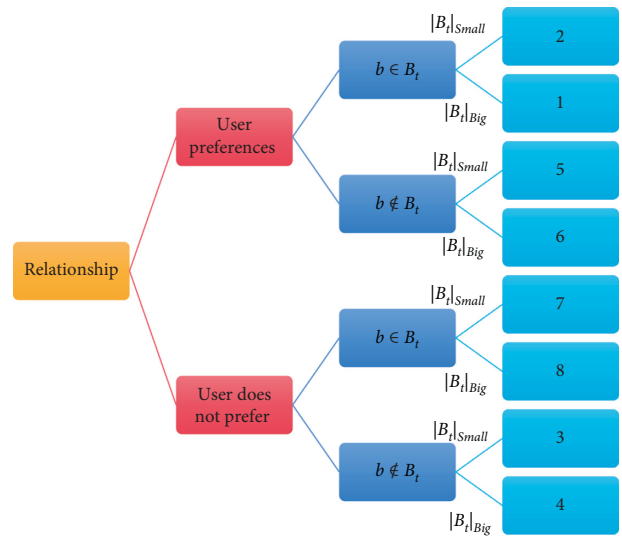

Figure 2: Relationship diagram between users and literary works.

and $\overline{P}_u$ represents the average score of user u preference for all types of literary works.

f_{b,B_t} indicates whether literary work b belongs to literary work type B_t; if so, its value is 1; otherwise, it is 0; and $\overline{f}_b$ represents the average score of specific values in the 0–1 vector of literary works' type set of literary work b.

The NB (naive Bayes) model is a common probability and statistics model with Bayes theorem as its mathematical foundation. The basic idea behind classification is to first learn the set of training samples. The learning process calculates the conditional probability of each category using statistical methods before sorting the test samples. The conditional probability and prior probability obtained from statistics are used in the classification process, and the category of test samples can be determined using the Bayesian formula.

In the NB classification model, it is precisely because it assumes that each characteristic attribute variable in the data sample is conditionally independent, so we can make the following derivation:

$$P(C_i|A) = \frac{P(A|C_i)P(C_i)}{P(A)}. \tag{6}$$

In the above formula, since the value of the denominator is a constant value, there is no need to calculate it, so we only need to calculate the value of the numerator and compare the largest one.

Among them, $P(a_1|C_j), P(a_2|C_j), \ldots, P(a_n|C_j)$, $j = 1, 2, \ldots, m$ can be obtained by statistical calculation of training samples, and the data of training samples are discrete values. We use the following formula:

$$P(a_k|C_i) = \frac{S_{ik}}{S_i}, \tag{7}$$

where S_{ik} is the number of samples with the attribute value of a_k and class C_i, where S is the total number of samples with class C.

3.3. Implementation of Personalized Recommendation Algorithm. For the CF algorithm itself, we want to improve the accuracy of recommendation, mainly considering two aspects: the first is the similarity calculation between readers and literary works; the second aspect is the prediction score of the project to be predicted; and first of all, we only use mathematical calculation methods to calculate the similarity. The second aspect is to consider the influence of neighboring readers on the prediction of literary works.

We also take into account the impact of English parts of speech on marks and include them as features in the feature template. Our quality templates, which include basic quality templates, part-of-speech quality templates, and English word parts of speech corresponding to Chinese words, have all been configured so far. The reading history data analysis found in this study can reflect readers' interests and reading habits, and the cluster analysis is used to analyze the reading data of literary works so that readers can have a clear understanding of the actual information needs. As a result, we use the ratio of average reader reading time to an average reading time of all books borrowed by readers to calculate the similarity of literary works that reflect readers' interest level in borrowing time.

$$\operatorname{sim}(x, y) = \frac{1/2\left(T_{1x} + T_{2y}\right)}{1\,(m+n)\left(\sum_{j=1}^{m} T_{1j} + \sum_{j=1}^{n} T_{2j}\right)}, \tag{8}$$

where x, y represent two literary works with identical classification index numbers borrowed by two readers; T_{1x}, T_{2y} indicate the borrowing time of literary works x, y by two readers s_1, s_2; m, n are the number of behaviors of borrowing literary works by two readers s_1, s_2, respectively; T_{1j}, T_{2j} indicate the borrowing time of literary works j by two readers s_1, s_2; and denominator means the average borrowing time of all literary works borrowed by two readers s_1, s_2.

We add the factor $r_{\min}$ to adjust so that the improved formula can accurately reflect the similarity of two readers, as shown in formula (9)as follows:

$$\operatorname{sim}(u, v) = \frac{r_{\min}}{2} \times \frac{\left|I_{uv}\right|}{\sum_{i \in I_{uv} \omega_p} \left|r_{u,v} - r_{v,i} + r_{\min}/2\right|}. \tag{9}$$

In the above formula, ω_p represents the influence weight of the user's scoring criteria, I_{uv} represents the literary works evaluated by two readers together, and $r_{\min}$ represents the value difference of the scoring range.

By calculating the books borrowed by readers by publicity formulas (8) or (9), the similarity of each book borrowed by two readers can be obtained, which can form a similarity matrix of literary works, in which readers who borrow a relatively small number of literary works are listed, and readers who borrow a relatively large number of literary works are listed.

Then, the maximum similarity in each column is compared with the similarity matrix to find the similarity of literary works that may reflect the same interests and hobbies of two readers. That is, the number of columns in the matrix. Then, we calculate the distance of interest between two readers according to formula (10).

$$D\left(S_1, S_2\right) = \frac{1}{1 + \sqrt{\sum_{i=1}^{k} \operatorname{sim}_i^2}}. \tag{10}$$

Total algorithm flow is as follows:

(1) We preprocess the reader rating dataset to obtain reader characteristics

(2) The cluster analysis method is applied, and then, the number of communities is used to divide the reader community.

(3) We calculate the similarity between readers and each community representative point, and then, we select the community with the highest similarity for recommendation within the community.

(4) By calculating the target readers' predicted scores of unrated books by the neighboring readers, the predicted scores here are calculated by incorporating the user scoring criteria factors

(5) We form a recommendation list of books with the highest prediction score and recommend them to readers

By using the above methods, this chapter designs and implements a personalized recommendation service system model of literary works based on the annotated corpus, as shown in Figure 3.

The personalized recommendation service system of literary works based on tagged corpus mainly realizes two functions, one is the mining function. Mining data association rules look for readers' potential borrowing patterns when borrowing literary works. The second is the personalized recommendation function, which applies the mined association rules to the personalized recommendation service of literary works.

4. Experiment and Results

In this section, we analyze the above training and results based on the semantic algorithm framework. As the metaphor generated based on semantics is not limited by language form, and the deep learning model needs a large number of datasets, in this algorithm experiment, the VUA corpus is mainly selected for training and verification, and the other two corpora (TriFi and MOH) are also used. The experimental results will be analyzed.

In this study, due to the attention mechanism, the internal weight distribution of the algorithm will be analyzed, and the learning mechanism is more suitable for metaphorical verb recognition. Figure 4 shows the experimental results in various corpora.

Different corpora have different precision ratios and F1 values, according to the statistics in Figure 4, which could be related to the size and quality of the corpora. The larger the corpus, the wider the distribution, the more semantic information the model learns, and the better the model's generalization ability. The model has multiple outputs at the same time, which helps to prevent overfitting and improve the model's generalization ability. The difference between

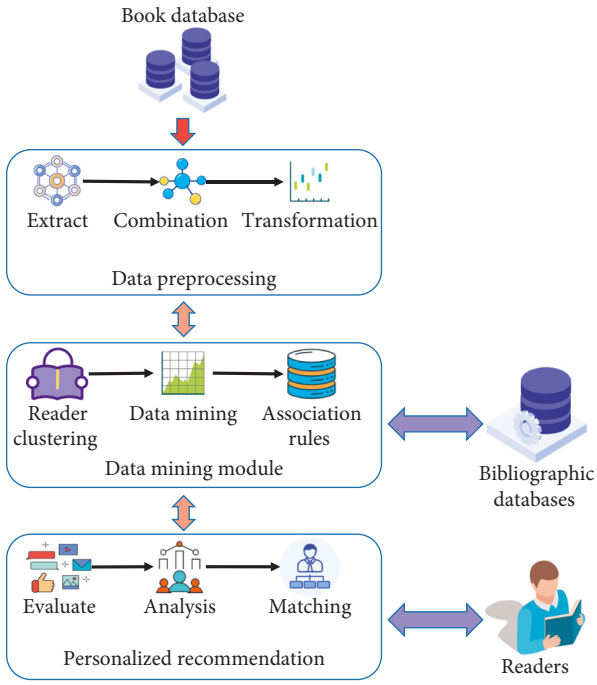

Figure 3: Personalized information recommendation service system model of literary works.

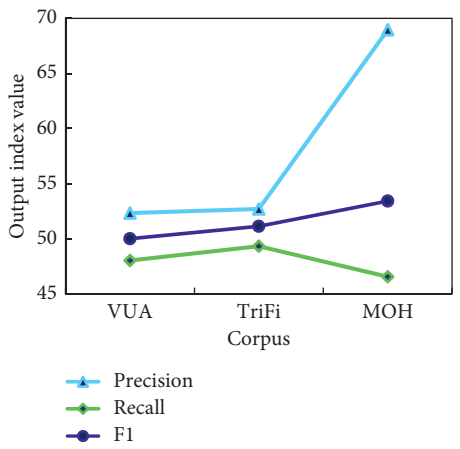

Figure 4: Experimental results under different corpora.

Chinese characters and English paragraphs is that they are only presented in the form of words when they are formed, whereas Chinese texts are not. That one word is linked to another. Because combining these words in the analysis will reduce analysis efficiency, it is necessary to delete them when processing the text.

After applying the feature weighting algorithm, we can use a more accurate number to represent the contribution of this feature element to document ranking. If feature elements play a small role in text classification, their proportion is small, and vice versa. If we use the feature weighting algorithm, the results obtained in text recognition will be more accurate. As shown in Figures 5–7, the enhanced TF-IDF algorithm is superior to the basic TF-IDF algorithm in F value of accuracy and recovery rate.

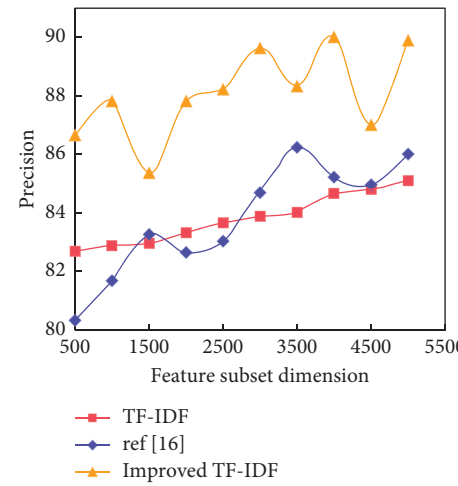

Figure 5: Accuracy comparison of algorithms.

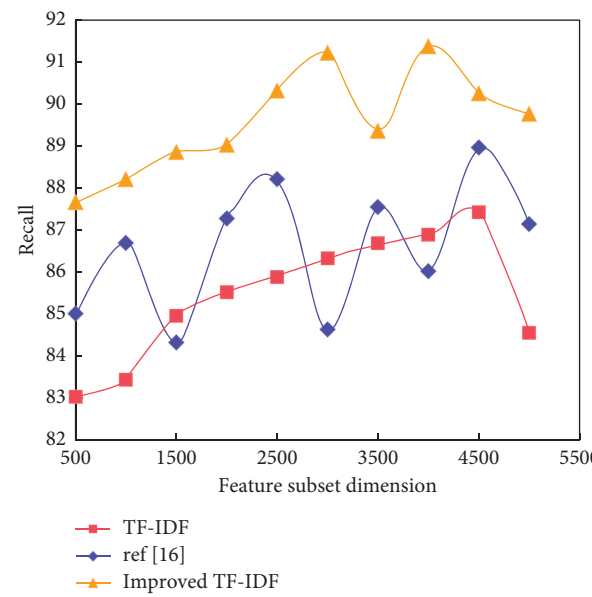

Figure 6: Comparison of algorithm recall rate.

When the dimension of the feature subset is 500 dimensions, the improved TF-IDF is superior to the original algorithm in all indexes, and the F value of accuracy and recall tends to be stable. The accuracy of the TF-IDF algorithm is 87.821%, the recall rate is 89.036%, and the F value is 91.368%, which is still higher than that of the original TF-IDF algorithm and ref [16]. Therefore, the improved TF-IDF algorithm proposed in this study effectively improves the performance of the traditional TF-IDF algorithm.

There are linguistic differences in the ambiguity models involved in part-of-speech tags. For example, ambiguous words in one language may not be ambiguous in another language. We take the simple word "development" as an example. It can be either a verb or a noun. In order to avoid some inconsistencies in the corpus, it is necessary to preprocess the corpus. It should be noted that as a modifier of verbs, it can be one of many parts of speech such as verbs, adverbs, and nouns or a combination of many parts of

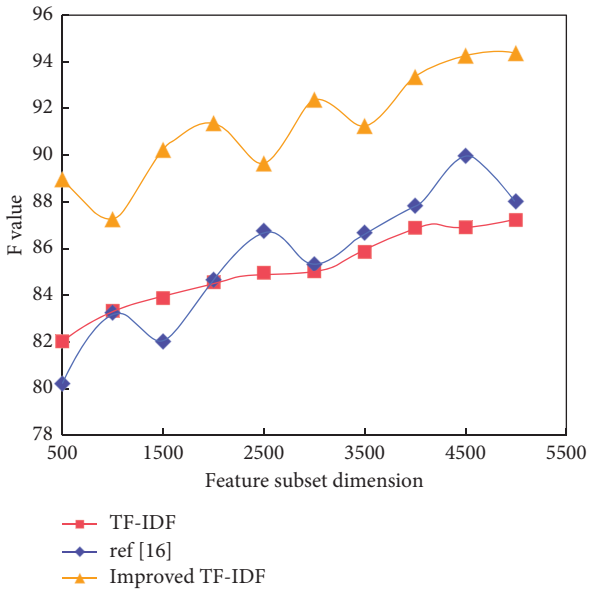

Figure 7: Comparison of algorithm F values.

speech, but here only adverbs are extracted as modifiers. According to the verbs that have already appeared, we must calculate the probability that a verb belongs to a specific category, and we must obtain the modifier vector for that category. Obviously, if this modifier vector modifies a verb and the probability is relatively high, the likelihood that it belongs to this category is relatively high. The issue then becomes determining the co-occurrence probability of adverbs and verbs in modifier vectors for known verbs, which has to do with the subject's characteristics. Because if there is a receiver of an action in a specific scene, there must also be a sender of the action, and the receiver and the agent must rely on one another and appear in pairs. In general, determining whether a noun in the subject-object position is the agent or the recipient is influenced not only by the noun's semantic features but also by the noun's features and sentence patterns.

In this experiment, we use the same Chinese corpus and the same evaluation methods mentioned above. One-to-one situations can be directly handled with the corresponding English part-of-speech, but one-to-many situations cannot use the corresponding English part-of-speech. Only the pronunciation of the first word part and the corresponding word sequence is selected. The part-of-speech tagging results of cluster analysis are given in Table 1.

In order to intuitively compare the performance of the experiment, we give the comparison column curve of the total part-of-speech tagging accuracy as shown in Figure 8.

From the experimental results in Table 1 and Figure 8, it can be seen that after adding the English part-of-speech function as the initial result on the basis of the sixth tagging result, the tagging accuracy of all parts of speech is improved. The corpus of this experiment is used for cyclic iteration. By comparing the experimental results of the two initial predictions, it can be seen that the accuracy of the second case is higher than that of the first case. The similarity calculation

Table 1: Part-of-speech tagging results.

Iterations	Corpus 1	Corpus 2	Corpus 3	Corpus 4
0	87.363	88.369	85.663	88.012
1	87.021	89.014	88.714	87.362
2	88.963	91.325	89.612	89.012
3	89.214	90.223	89.016	90.225
4	88.021	87.569	90.332	91.374
5	87.011	89.326	87.416	87.869
6	89.033	85.639	89.067	88.358

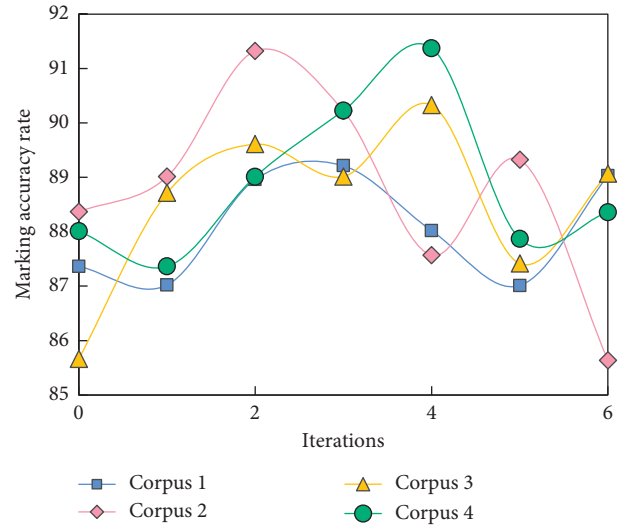

Figure 8: Comparison of the accuracy of total part-of-speech tagging.

method proposed in this study can describe the similarity between users in more detail through enhanced overlap factor and local similarity. The improved overlap factor can not only correct the possible deviation of users when there are fewer scoring items but also consider the user's scoring behavior when calculating the variance of the user's scoring, which reflects whether the user's rating is "credible."

Using information from different attributes as the target of feature investigation for cluster analysis, several representative clusters can be obtained. This study starts with the categories of readers who borrow literary works and makes a cluster analysis of readers. We divide them into different pools and then conduct targeted association mining. Figure 9 and Table 2 show the comparison of MAE (average absolute error) value and accuracy between the proposed algorithm and the traditional CF algorithm.

It can be seen from Figure 9 that the MAE value of this method is always lower than that of the CF method, especially when the number of users in the neighbor set is 5; this method is obviously better than the CF method; and the maximum difference can be 6.23%. This study compensates for the shortcomings of the CF method by introducing the modified overlap factor, incorporating the variance analysis of the user's score, and calculating the local similarity. The difference between them tends to decrease as the number of neighbor set users increases, but the method in this study is still better than the CF method in the case of 25 neighbor set

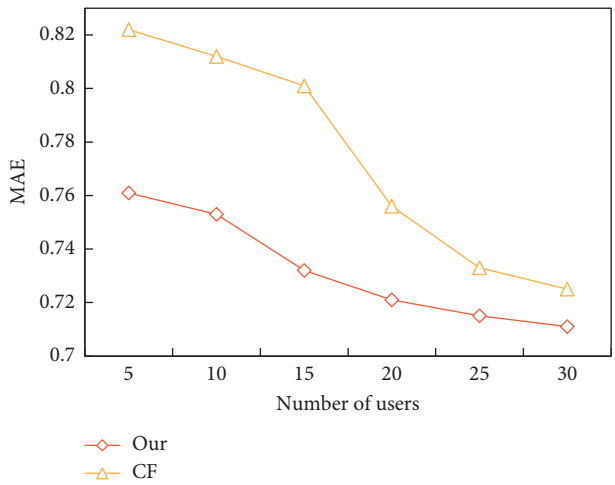

Figure 9: MAE value comparison.

Table 2: Accuracy comparison.

Number of users	Algorithm in this study	Traditional CF
5	0.401	0.363
10	0.423	0.372
15	0.411	0.355
20	0.436	0.378
25	0.433	0.381
30	0.429	0.379

users. Table 2 also provides that the method used in this study significantly improves the accuracy of the recommendation results. To summarize, the similarity calculation method presented in this study can more accurately describe user similarity and improve the method's practicability.

5. Conclusions

With the rapid development of the internet and the rapid rise of e-literature reading websites, the amount of data such as books and users is increasing, which makes the data shortage and startup problems in the literature recommendation system have a greater and greater impact on the cold storage, which makes the recommendation system more efficient, and the quality of recommendation is declining. In this study, a personalized recommendation algorithm for literary works based on tagged corpus is proposed, and the prediction score is calculated. For unregistered words in the tagging process, some rules are defined, and tagged corpus is obtained according to these rules, and readers are grouped by the cluster analysis algorithm to form different interest groups. It is found that the MAE value of this method is always lower than that of the CF method, especially when the number of users in the neighbor set is 5; this method is obviously superior to the CF method; and the maximum value can differ by 6.23%. The final experimental results show that the unsupervised part-of-speech tagging method proposed in this study further improves the performance of Chinese part-of-speech tagging and makes users more satisfied with the recommendation results.

Data Availability

The data supporting the results of this study are not available to protect participant privacy.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Q. Liu, L. Cheng, A. L. Jia, and C. Liu, "Deep reinforcement learning for communication flow control in wireless mesh networks," *IEEE Network*, vol. 35, no. 2, pp. 112–119, 2021.

[2] R. Guzmán-Cabrera, "Authorship attribution of Spanish poems using n-grams and the web as corpus," *Journal of Intelligent and Fuzzy Systems*, vol. 4, pp. 1–6, 2020.

[3] Y. Chen, Z. Xiao, X. Zhang, and Z. Tao, "DSTL: solution to limitation of small corpus in speech emotion recognition," *Journal of Artificial Intelligence Research*, vol. 66, pp. 381–410, 2019.

[4] Z. Yuan, W. Zheng, Z. Tong, and X. Yao, "Cross-corpus speech emotion recognition based on domain-adaptive least-squares regression," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 584–588, 2016.

[5] R. Ñanculef, I. Flaounas, and N. Cristianini, "Efficient classification of multi-labeled text streams by clashing," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5431–5450, 2014.

[6] Z. Huang, Y. Liu, C. Zhan, and J. Liu, "A novel group recommendation model with two-stage deep learning," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021.

[7] N. Bassiou and C. Kotropoulos, "Long distance bigram models applied to word clustering," *Pattern Recognition*, vol. 44, no. 1, pp. 145–158, 2011.

[8] J. Huang, J. Liu, and X. Yao, "A multi-agent evolutionary algorithm for software module clustering problems," *Soft Computing*, vol. 21, no. 12, pp. 3415–3428, 2016.

[9] L. C. Yuan, "A word clustering method based on mutual information," *Systems Engineering*, vol. 26, no. 5, pp. 120–122, 2008.

[10] L. A. Dalton, M. E. Benalcazar, M. Brun, and E. R. Dougherty, "Analytic representation of Bayes labeling and Bayes clustering operators for random labeled point processes," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1605–1620, 2015.

[11] A. Saif, N. Omar, U. Z. Zainodin, and M. J. Ab Aziz, "Building sense tagged corpus using wikipedia for supervised word sense disambiguation," *Procedia Computer Science*, vol. 123, pp. 403–412, 2018.

[12] C. T. Zheng, C. Liu, and H. S. Wong, "Corpus-based topic diffusion for short text clustering," *Neurocomputing*, vol. 275, pp. 2444–2458, 2018.

[13] Y. Shi and Y. Zhu, "Research on Fast Recommendation Algorithm of Library Personalized Information Based on Density Clustering," *Security and Communication Networks*, vol. 2022, Article ID 1169115, 9 pages, 2022.

[14] W. Wang, Z. Wang, and M. Zhang, "Personalized recommendation algorithm on microblogs," *Computer science and exploration*, vol. 6, no. 10, pp. 895–902, 2012.

[15] L. Hu, G. Song, Z. Xie, and K. Zhao, "Personalized recommendation algorithm based on preference features," *Tsinghua Science and Technology*, vol. 19, no. 3, pp. 293–299, 2014.

[16] Y. Dai, H. W. Ye, and S. J. Gong, "Personalized Recommendation Algorithm Using User Demography

Information," in *Proceedings of the Second International Workshop on Knowledge Discovery & Data Mining*, pp. 100–103, IEEE Computer Society, 2009.

[17] J. Chen, C. Du, Y. Zhang, P. Han, and W. Wei, "A clustering-based coverage path planning method for autonomous heterogeneous UAVs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, pp. 1–11, 2021.

[18] J. Chen, Y. Zhang, L. Wu, T. You, and X. Ning, "An adaptive clustering-based algorithm for automatic path planning of heterogeneous UAVs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, pp. 1–12, 2021.

[19] Y. H. Guo and G. S. Deng, "Improved personalized recommendation algorithm in collaborative filtering," *Application Research of Computers*, vol. 25, no. 1, pp. 39–38, 2008.

[20] L. Hu, W. Wang, F. Wang, X. Zhang, and K. Zhao, "The design and implementation of composite collaborative filtering algorithm for personalized recommendation," *Journal of Software*, vol. 7, no. 9, pp. 2040–2045, 2012.

[21] H. Ping, "The research on personalized recommendation algorithm of library based on big data and association rules," *The Open Cybernetics & Systemics Journal*, vol. 9, no. 1, pp. 2554–2558, 2015.

[22] J. Xian, Z. Qin, and L. Sun, "An optimization of collaborative filtering personalized recommendation algorithm based on time context information," *IFIP Advances in Information and Communication Technology*, vol. 449, pp. 146–155, 2015.