Recommendation method for academic journal submission based on doc2vec and XGBoost

Huang ZhengWei[1] · Min JinTao[1] · Yang YanNi[1] · Huang Jin[1] · Tian Ye[1]

Abstract

With the continuous deepening of academic research in various disciplines and the continuous increase in the number of scientific researchers, exploring the mechanism of matching scientific research results and academic journal subjects is a key topic that can assist researchers in selecting suitable journals for submission. The classification and recommendation of academic journals based on a traditional text representation model cannot take advantage of the semantic relationship between words and cannot take into account the diversity of topics received by different journals, which affects the classification and recommendation effect. To solve these problems, this paper uses doc2vec to perform distributed representation of the bibliographic text so that the semantics between the text features are fully preserved. Then, the XGBoost algorithm is used to consider the impact of the different characteristics of the title, abstract, and keywords of the bibliography on the published journal. The academic journal submission recommendation model proposed in this paper can solve the problem that traditional methods cannot make full use of the contextual semantic information and improve the efficiency of scientific research personnel's academic achievement publications. Experiments on Common SCI English journals in the computer field show that when recommending three candidate journals, the accuracy rate reached 84.24%.

Keywords Doc2vec · XGBoost · Journal recommendation · Document bibliography · Machine learning

Introduction

With the continuous deepening of academic research in various disciplines, the number of researchers, publications and academic journals has grown rapidly. Selecting appropriate academic journals for submission is an inevitable and tedious task for researchers. For example, according to the contents of the DBLP dataset, there are more than 4152 foreign language journals in the field of computer science (Wang, He, et al., 2018; Wang, Liang, et al., 2018). As research horizons expand, researchers find it difficult to keep up with new

✉ Yang YanNi
  yangyanni614@foxmail.com

[1] China Three Gorges University, Yichang, China

discoveries even within their disciplines. Faced with so many journals, researchers usually obtain a small amount of journal information from colleagues and friends. According to the statistics of the top ten common reasons for rejection, including content in the proposed paper that does not match the subject accepted by the journal results in paper rejection. In addition, it will cause a delay in publication time and a decline in publication quality (Pradhan & Pal, 2020). Therefore, it is very important to explore the matching mechanism between research results and academic journal topics and to provide a reasonable academic journal submission recommendation model.

In existing academic fields, according to the direction of application, recommendation models can be divided into collaborator recommendations (Chaiwanarom & Lursinsap, 2015; Kong et al., 2016), paper recommendations (Wang & Ishuga 2018; Xia et al., 2014), citation recommendations (Huang et al., 2015; Liu et al., 2015), academic journal recommendations (Pradhan & Pal, 2020; Yu et al., 2012, 2018), etc. These studies objectively provide users with personalized information services, which are very helpful to researchers. Although there have been many different aspects of academic recommendation work, the research work recommended in academic journals is limited. In the research on academic journal recommendations, there are mainly the following methods. First, based on collaborative filtering (Liang et al. 2016; Yang & Davison, 2012), recommendations are made by analyzing a researcher's historical academic journal submissions. Although the method based on collaborative filtering can effectively utilize the historical record information of scholars' journals, this method has the problem of a cold start; that is, the recommendation results are often ineffective for researchers with less submission history. The second is based on academic social network recommendations (Chen et al., 2015; Silva et al., 2015), which are recommended based on a network graph of social relationships between authors and collaborators. Although this method can effectively discover related journals that have not published the author according to the journals that have published the collaborators, it has the problem that the recommendation effect is not sufficiently accurate for researchers. In addition, more importantly, when a researcher needs to become involved in a new research field because of a certain topic, that is, when there is an "interest drift" situation, the recommendation effect of this method is often greatly reduced, even for those with publishing experience. Abundant scientific researchers are also helpless. The third is the recommendation method based on the content of the paper (Schuemie & Kors, 2008; Wang, He, et al., 2018; Wang, Liang, et al., 2018), which extracts several topics or keywords from the content of published academic papers and recommends according to the same topics or keywords. Although this method can effectively solve the abovementioned cold start problem, it has the semantic problem that the context cannot be fully utilized in the current research situation. For example, when two papers have different keywords but discuss the same content, this method considers that the two papers are not related and will not recommend the journal that published the other paper.

Aiming at the above problems, this paper proposes a journal recommendation model and conducts experiments to verify the effectiveness of the model. The main contributions of our work include the following points.

(1) To the best of our knowledge, we propose a new journal recommendation model that uses doc2vec to represent bibliographic text and uses the XGBoost algorithm to classify bibliographic information after text representation. By entering the title, abstract, and keyword-related bibliographic information of a paper to be submitted, several foreign SCI journals suitable for the subject of the paper to be submitted are recom-

mended and sorted according to the degree of matching suitability, helping authors reduce the probability of the paper being rejected.

(2) The factors that affect the recommendation effect are studied, including the title, abstract and keyword factors in the bibliographic information, the factors of the dimension of the bibliographic text representation, and the factors of the number of recommended journals.

(3) According to measurement indicators such as precision rate and recall rate of different journals, the differences and similarities of accepted topics among different journals are studied.

Additionally, in response to the increasing number of journal papers, a dynamic update part of the training model is added to enable the trained model to continuously learn new paper data so that it has good stability and real-time performance.

The rest of this article is organized as follows. The "Related work" section examines existing work related to journal recommendation. In the "Proposed method" section, we detail our proposed model. Then, the factors that influence the recommendation effect and the similarities and differences in the accepted topics of different journals are analyzed in the "Results" section. Finally, the main conclusions of this paper and future work are described in the "Conclusion" section.

Related work

From the perspective of the specific information utilized, a recommendation model can be divided into academic journal recommendations based on the social network of paper collaborators, academic journal recommendations based on paper content, and academic journal recommendations based on collaborative filtering. In this section, the related work of the academic journal recommendation method based on the social network of the paper collaborators and the academic journal recommendation method based on the paper content will be expounded.

The academic journal recommendation method based on the social network of the paper collaborators uses the information of the paper collaborators to build a social network among the paper collaborators and recommends journals that published researchers who are closely related to the authors of the proposed paper. According to whether to introduce information other than collaborator information, it can be divided into two categories: traditional recommendation based on the social network of the paper collaborators and improved recommendation based on the social network of the paper collaborators. The traditional recommendation method based on the social network of paper collaborators only considers the collaborator information of the paper. For example, Luong et al. (2012) proposed a social network-based method to recommend academic journals by analyzing the information of coauthors in related fields of published papers. Huynh and Hoang (2012) proposed a collaborative knowledge model based on a combination of graph theory and probability theory. The improved recommendation method based on the social network of the paper collaborators not only considers the collaborator information of the paper but also considers the author and publication information of the network relationship within the company. In addition, Yu et al. (2018) proposed a novel personalized academic journal recommendation model, PAVE, which not only includes the coauthorship relationship and the relationship between authors and journals

but also takes into account the frequency of copublishing and the weight of authorship relationships and the academic level of researchers.

The academic journal recommendation method based on the content of the paper extracts relevant features by analyzing the text information of the paper bibliography or the content of the paper to classify and recommend according to the similarity of the features. According to the different technical methods used, they can be divided into two categories: traditional technical methods and deep learning techniques. The traditional method of academic journal recommendation based on the content of the paper uses manual extraction of features from the content of the paper. For example, Wang, Liang, et al. (2018), Wang, He, et al. (2018)) proposed a recommendation model based on keyword similarity to recommend appropriate journals in the field of computer science. The system extracted keywords from the abstract and title information of each paper through the chi-square feature selection method, calculated the keyword weights, and finally used softmax regression as a classifier to rank and recommend journals and conference categories. Yang and Davison (2012) proposed a memory-based recommendation model, believing that the writing styles of papers accepted by different academic journals were different. They researched the relationship between academic journals and writing style and recommended appropriate academic journals using the stylistic features of vocabulary, grammar and structure. The academic journal recommendation method based on the content of the paper using deep learning mainly uses deep learning technology to automatically extract features from the content of a paper and does not require human judgment. For example, Pradhan et al. (2020) proposed an academic journal recommendation system based on a bidirectional LSTM and an integrated attention mechanism, which extracted features from the title and abstract information of a paper to be submitted to identify the category of academic journals to submit to. In addition, they introduced a new academic journal recommendation model, CNAVER (Pradhan & Pal, 2020). The system model adopts the fusion of two models based on a paper-paper network and venue-venue network. Abstracts can provide academic journal recommendations, which largely solve the problem of a cold start. Feng et al. (2019) introduced a journal recommendation system, Pubmender, based on deep learning convolutional neural networks. The system combines pretraining methods to represent paper abstracts and uses a fully connected softmax model to recommend the best academic journals in the field of biomedicine.

Reviewing related research, we found that previous research did not consider that different papers have different keywords but discuss the same topic and did not consider the relationship between the paper to be submitted and all papers in a journal. Papers may be similar to a few papers in a journal but not similar to most papers in the journal, and the above problems lead to inaccurate recommendations for academic journal submissions. Furthermore, none of the above studies take into account that the subject scope of the content of papers accepted by journals can change dynamically over time. Therefore, in view of the above problems, the recommendation model for academic journal submission proposed in this paper uses doc2vec to represent the bibliographic text, which solves the problem that the traditional methods cannot fully utilize the semantic information of the context. Based on the bibliographic information of a paper to be submitted and a large amount of training data, the classification model is trained based on the bibliographic information to solve the problem that traditional methods do not take into account the similarity with all papers in a journal. Finally, the training model is updated regularly and dynamically to ensure the stability and real-time performance of the model.

Proposed approach

Transforming a journal recommendation problem into a classification problem is a challenging task. In the context of journal recommendations, it is crucial to accurately mine the relationship between papers and the topics accepted by journals. The recommendation framework for academic journal submissions proposed in this paper includes three steps. The first step is the embedding of the bibliographic information of the papers to obtain a text representation of the bibliographic information of each paper. The second is to train the matching classification model corresponding to the paper and journal according to the different topics accepted by different journals. The third is to input the bibliographic information of a paper to be submitted into the trained paper classification matching model for matching and obtain several academic journals that match the theme of the paper to be submitted.

Model building

The academic journal recommendation model proposed in this paper is composed of the following parts. First, the bibliographic text representation part is based on doc2vec. The bibliographic information of a certain number of published papers in each journal is obtained. The bibliographic information includes field information such as keywords, abstracts, and titles of the papers. The information of the keyword fields in the bibliographic data is extracted to make a special dictionary in the field of scientific research, and the title and abstract of the bibliography are word-segmented using the dictionary. Meaningless words are filtered out from the result of the word segmentation, and a text feature vector of each bibliographic information is generated using doc2vec training. The second is the bibliographic text classification part based on XGBoost, which is generated according to the bibliographic text representation part. The text features of the bibliography are used to train the XGBoost literature and journal topic classifier and obtain a paper classification model that learns the topic range of papers accepted by different journals. The third is the recommendation of academic journals part. The abstract, keywords, and title information of a paper to be submitted are input and doc2vec is used to represent the textual feature information of the paper. The result is compared with the trained paper classification model and suitable publications for the paper to be submitted are recommended. The fourth is the dynamic update part of the training model, which is responsible for the model continuously learning new data generated in the paper database to ensure that the model has certain stability and real-time performance. The structure of the submission recommendation model is shown in Fig. 1.
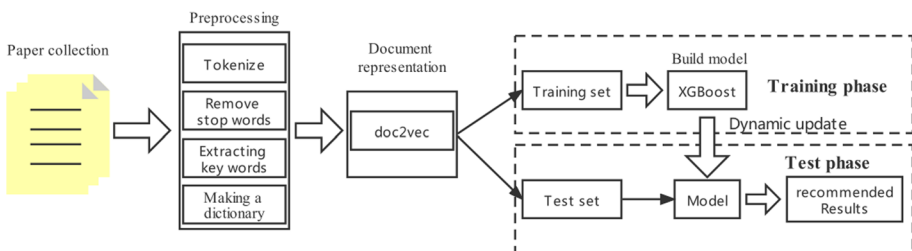

Fig. 1 The structure of the recommendation model for academic journal submissions

Bibliography text feature representation based on doc2vec

Text feature representation processes words into vectors or matrices. Currently, commonly used text representation models include bag-of-words models, topic models, and word embedding models. Because word embedding models have good text representation ability, this paper adopts a word embedding model based on doc2vec to represent the text information of the bibliography and solve the problem that traditional methods cannot make full use of the context information of the bibliography. At the same time, doc2vec can reduce the difference between different bibliographies and the effect of the large difference in text length. The doc2vec algorithm is a deep learning algorithm for text vectorization proposed by Mikolov et al. on the basis of word2vec, which extends the calculation of vector features from the word level to the sentence (or paragraph) level. The difference from word2vec is that in the input layer of the neural network, doc2vec adds a sentence vector. During each training process, long text is introduced into the corpus as a special paragraph ID so that the word input of each document may be different. However, the output is a fixed length vector. During the training process, the algorithm combines the context, word order and paragraph features to train the probability distribution of word vector occurrences. Practice shows that doc2vec shows good results in terms of text similarity and text classification.

The framework proposed in this paper uses the doc2vec model to convert bibliographic information into text embedding vectors. The specific implementation is as follows. First, the Jieba word segmentation tool is used to extract all words from the entire bibliographic information corpus to construct a vocabulary set of size. The word vector for each word contained in this vocabulary has a fixed dimension and is randomly initialized. Each bibliographic information consists of a subset of words belonging to

$$v = \{w_1, w_2, w_3, \ldots, w_n\} \tag{1}$$

Then, word and document representations are learned using the PV-DM structure of doc2vec. Considering the large difference in text length between different bibliographic information, the paragraph vector in doc2vec is used, and it is shared during several training processes of the same sentence so that the input information in each training contains the paragraph vector. As a sentence slides and takes several words for training, the shared paragraph vector, which is part of the input layer of each training, will become more accurate. After training is completed, all word vectors in the training samples and the document vector of each bibliographic information are obtained.

The parameters we used for training and representation on the document bibliography corpus were as follows: the document embedding vector dimension was 300, the batch size was 256, the window size was 5, and the AdaGrad optimization method was used with the learning rate set to 0.2.

Bibliography text classification based on XGBoost

The bibliographic text classification part based on XGBoost uses the XGBoost classifier to train the journal classification model on the basis of a large number of bibliographic information data sets based on the features generated by doc2vec and the corresponding category label information, thereby indirectly learning about the subject categories of papers accepted by different journals.

The specific details of this part are as follows. Doc2vec is used to represent each bibliographic information, a feature vector with a fixed dimension length is obtained, and the

feature vector is used as a feature of the classification model. The published journals in each bibliographic information field are used, sklearn is used for one-hot category encoding, and the encoded discrete results are the target of the classification model. Finally, the feature and the corresponding target are input into the XGBoost classifier for training and learning, and a trained classification model is obtained.

According to the characteristics of the bibliographic information data set, the important parameters related to XGBoost used in the training process are as follows: a tree-based model is used for boosting calculation, and the parameter is set to gbtree; the maximum number of iterations in the spanning tree process is set to 95; the step size of each iteration in the process is set to 0.05; the random sampling ratio of each tree is 0.6; the maximum depth parameter of the tree is set to 7; and the loss function used is merror, that is, the multiclassification error rate.

Academic journal recommendation implementation

The academic journal recommendation implementation part recommends several academic journals suitable for publishing the topic based on the title, abstract, and keyword information of the paper to be submitted by the user and the trained classification model.

The specific implementation details of this part are as follows. After the academic journal recommendation system receives the title, abstract, and keyword information of a paper to be submitted by a scientific researcher, it calls the text representation part of the bibliography based on doc2vec and generates the paper to be submitted based on the information input by the scientific researcher as a bibliographic information feature vector. According to the model trained by the bibliographic text classification module based on XGBoost, the overall distance between the feature vector of the bibliographic information of the paper to be submitted and the feature vector space range of the bibliography received by different journals is compared. The closer the distance is, the more suitable the journal is in line with the subject of the proposed paper. According to the distance, several suitable journals for publication are recommended for the paper to be submitted to, and the recommended journals are sorted by matching degree.

Training model dynamic update

In practical situations, the data are usually of great value, given the ever-increasing number of published papers. By adding a dynamic update part to the training model, the trained model can continuously learn new paper data from the paper database, enabling the trained model to have good stability and real-time performance.

Taking into account the advantages and disadvantages of a full update and incremental updates, we adopt a method combining full and incremental updates. That is, the new bibliographic data added by relevant journals in the dissertation database are obtained every month, and a new XGBoost spanning tree structure is added on the basis of the previously trained model by means of incremental updates so that the model is incrementally updated and trained. The time period is short and has high flexibility and real-time performance. At the same time, every three months, new bibliographic data added by relevant journals in the dissertation database added within the past three months will be summarized, the new bibliographic data will be merged with the previously collected bibliographic data, and the full amount of bibliographic data will be used. The update method retrains using the XGBoost algorithm to achieve the update coverage of the original training model so

that the trained model has high accuracy and stability. The dynamic update process of the training model is shown in Fig. 2.

Results

### Experimental data collection and preprocessing

This paper selects 45 common foreign SCI core journals in the category of computer technology. The bibliographic data of 450 articles recently published by each journal as of December 25, 2021 were collected, of which each journal contained 400 pieces of training sample data and 50 pieces of test sample data, for a total of 20,250 pieces of bibliographic information data. Each bibliographic data point includes the title, abstract, keywords, and publishing journal information.

The preprocessing of the collected bibliographic data includes the following steps. First, all the keyword information in the bibliographic information is extracted, each keyword is separated with a newline character, and a dictionary in the field of computer scientific research is developed. The purpose is to not use the default dictionary that comes with the Jieba Word segmentation tool not being sufficiently accurate for some professional nouns in a paper. This preprocessing step is also a highlight of our experiment. Second, word segmentation is performed on the abstract and title information of the bibliography. Third, words that are meaningless to the results, high-frequency words and low-frequency words, such as "data", "research", etc. are removed.

Experimental evaluation index

This paper selects three recommended numbers of candidate journals, Top@1, Top@3, and Top@5. Top@1 recommends the journal that is most suitable for publication, and the evaluation and recommendation results are more stringent. Top@3 and Top@5, respectively provide three and five journals that are suitable for the publication of papers to be submitted, and their sorting order descends according to the size of the score.

Fig. 2 Dynamic update diagram of the training model
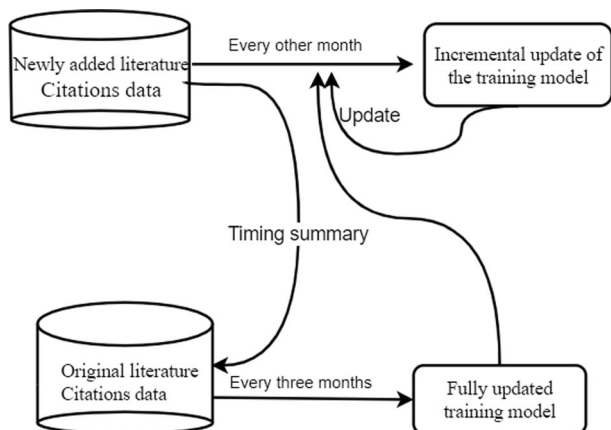
The indicators used to evaluate the text classification recommendation model are the accuracy (Acc), precision (P), recall (R) and F1. The Acc indicates the proportion of correctly predicted samples in the total samples:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

The P refers to the ratio of the number of correctly classified texts to the number of all classified texts:

$$P = \frac{TP}{TP + FP}$$

R refers to the ratio of the number of correctly classified texts to the actual number of texts in that class:

$$R = \frac{TP}{TP + FN}$$

where TP specifies the number of positive samples predicted to be positive; FP specifies the number negative samples predicted to be positive; FN specifies the number positive samples predicted to be negative; and TN specifies the number of negative samples predicted to be negative. Both P and R take values between 0 and 1. The closer the value is to 1, the higher the precision and recall.

The F1 indicator comprehensively considers the precision rate and the recall rate. The F1 value is more inclined to the indicator with a smaller value of precision and recall. The larger the F1 value is, the higher the quality of the model.

$$F1 = \frac{2 * P * R}{P + R}$$

Analysis of the results

Considering that most academic journals in the computer field have more than one publication topic, and the publication topics in different journals are repeated, if only one journal is recommended at a time, the requirements for the recommendation model are too strict, and researchers cannot be provided sufficient choices. Therefore, we studied the accuracy of different recommended candidate journal numbers and analyzed the accuracy of the recommended results when the number of recommended candidate sets was from 1 to 20, as shown in Fig. 3. As the number of recommended candidate journals increases, the accuracy rate increases from 46.17% for Top@1 to 84.24% for Top@3 to 89.43% for Top@5. When the number of recommended candidate journals is less than four, the increase in the accuracy of the recommended results increases with the increase in the number of recommended candidate journals, but when the number of candidate journals is greater than four, the increase in the accuracy of recommended results decreases significantly.

Furthermore, we investigate the effect of the feature vector dimension on the results when the bibliography is represented based on doc2vec. As shown in Fig. 4, with the increase in the number of selected features, the accuracy of the Top@N recommended journal candidate set has a certain degree of improvement, but when the feature selection dimension increases to 250 dimensions, the increase in the accuracy rate significantly decreases. At the same time, as the number of recommended candidate journals increases

Fig. 3 Accuracy of Top@N recommendation results
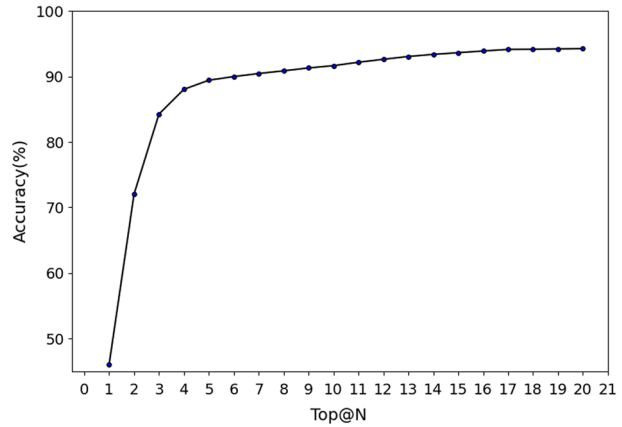
Fig. 4 Influence of the bibliographic feature vector dimension on accuracy
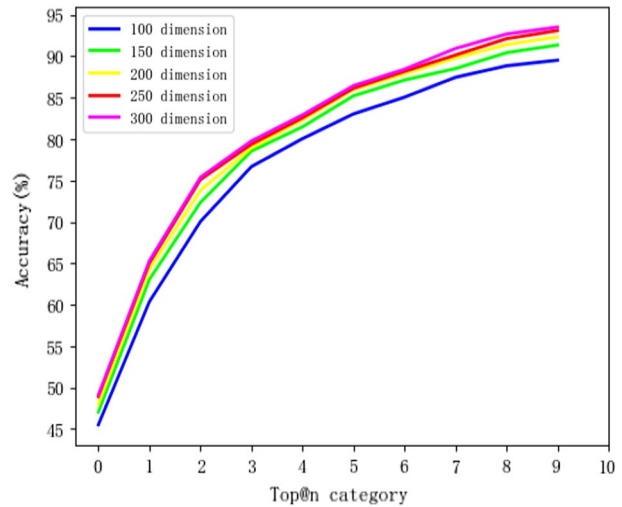
to more than seven, the dimensions of the word vector and bibliographic feature vector have a more obvious impact on the results. Therefore, considering the factors affecting the accuracy and the time cost of program operation, the default number of feature dimensions is selected to be 250 in this paper.

Considering the diversity of content topics received by different journals, this paper studies the performance of precision, recall, and F1 value of different journals when only three candidate journals are recommended. As shown in Fig. 5, it can be found that journals with higher accuracy are journals with a single topic, such as "ACM Transactions on Computer Systems ", "Journal of chemical Information and Modeling", "ACM Transactions on Graphics" and other journals. Journals with lower accuracy are mostly comprehensive journals with diverse content topics, such as "IEEE Transactions on Multimedia" and "Mobile Networks and Applications". For example, "ACM Transactions on Graphics" publishes papers on the subject of computer vision, while "IEEE Transactions on Multimedia" publishes papers on natural language processing, image recognition, embedded, software systems, and more. Therefore, the accuracy rate of "ACM Transactions on Graphics"
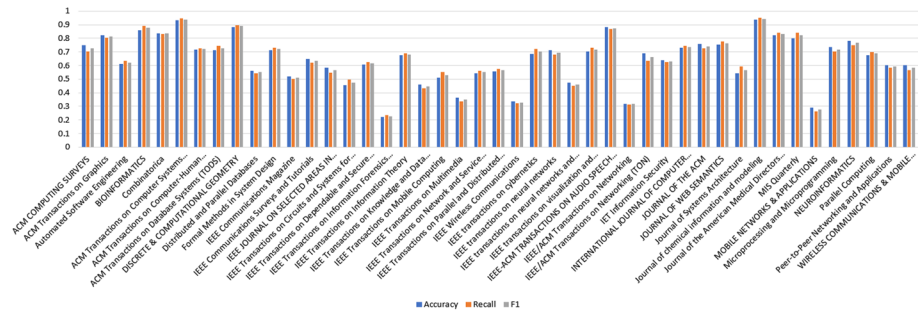
Fig. 5 The indicators of different journals when three candidate journals are recommended

is higher than that of "IEEE Transactions on Multimedia". In addition, we found that most of the journals that published multiple topics and similar topics had lower recall than precision, and journals that published articles on a single topic had higher recall than precision. For example, the publication topics of "Journal of Systems Architecture" and "IEEE Transactions on Circuits and Systems for Video Technology" are independent, and the publication topics of these two journals rarely appear in other journals, so the recall rate is relatively high. However, the publication topics of "IEEE Journal on Selected Areas in Communications", "IEEE Communications Surveys and Tutorials" and "Wireless Communications and Mobile Communicating" have more repetitive topics, so the recall rate is low.

We select the bibliographic data of all papers in the training set of "ACM Transactions on Graphics", "IEEE Transactions on Multimedia", "IEEE Transactions on Circuits and Systems for Video Technology", and "IEEE Communications Surveys and Tutorials", and the high-dimensional feature vector of the bibliographic information of each paper is reduced to three-dimensional space using the PCA dimensionality reduction method and displayed, as shown in Fig. 6. According to the spatial distribution map of the bibliographic information features, it is not difficult to see that the aggregation density of different journals is not the same. "ACM Transactions on Graphics" accepts papers with relatively single topics, and the distribution in the three-dimensional feature space is relatively concentrated; "IEEE Transactions on Multimedia" accepts papers with richer topics, which are more divergent in spatial distribution. At the same time, it can be found that different journals have different degrees of repetition in the scope of accepted topics. For example, "IEEE Transactions on Multimedia" and "IEEE Transactions on Circuits and Systems for Video Technology" have more overlapping topics, and "IEEE Transactions on Multimedia" and "IEEE Communications Surveys and Tutorials" have fewer repetitive topics.

Considering the influence of different structural information in the bibliographic information on the accuracy rate, this paper studies the accuracy of a case without title structure information, keyword structure information, and abstract structure information compared with a case with complete structure information. As shown in Table 1, in the absence of title, keyword, and abstract structure information, the accuracy rates drop by 2.38%, 3.22%, and 4.37%, respectively. The experimental results show that the textual information contained in the abstract structure information is more abundant than that in the keyword and title structure information.

In addition, it is worth noting that considering that there are many professional terms in academic research papers, this paper studies the accuracy of word segmentation using keyword fields in the bibliographic information to create a custom dictionary in the process of data preprocessing. The custom dictionary is built by extracting keyword information in the
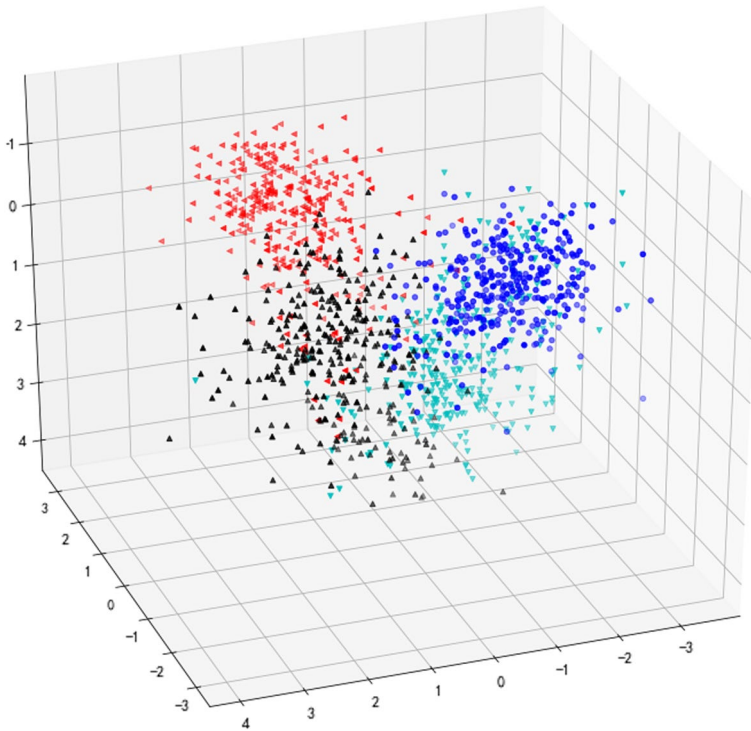
Fig. 6 Spatial distribution of bibliographic information features

Table 1 The influence of the bibliographic information structure of the paper on the accuracy

Text structure	Accuracy (%)	Degree of decrease
All structures	84.24	
No title, with keywords, abstract	81.86	2.38%
No keywords, title, abstract	81.02	3.22%
No abstract, with title and keywords	79.87	4.37%

bibliographic information of a paper according to the format of the Jieba word segmentation tool default dictionary, and the custom dictionary has a richer professional terminology in the computer field. The research results show that using a custom dictionary for word segmentation in the data preprocessing process has a 2.21% improvement in accuracy compared to the default dictionary.

Conclusion

Based on the methods of the recommendation of academic collaborators, recommendation of academic papers, recommendation of citations, and recommendation of reviewers, recommendations for submission to research academic journals is relatively lacking, and a good recommendation mechanism for academic journals is undoubtedly beneficial for researchers to find suitable academic publication journals for their academic achievements. To improve the efficiency of academic publications and help researchers quickly find candidate journals, the proposed academic journal recommendation model only needs the abstract, title, and keyword information of a paper to recommend several journals that are suitable for the paper to be submitted to in an academic field. The method proposed in this paper is applicable to the recommendation of academic journals in various disciplines, only needing to replace the corresponding data set, which provides strong practical application significance.

The academic journal submission recommendation model proposed in this study uses doc2vec to represent bibliographic information to solve the problem that traditional methods cannot make full use of the semantic information of the context by comparing whether they have the same topics and keywords for recommendation. The XGBoost algorithm is used to learn the rules of the topics received by different academic journals and then the bibliographic data is classified to solve the problem that existing methods do not take into account the relationship with all papers in a journal. Lastly, a dynamic update of the training model is provided for to ensure the training model has good stability and real-time performance.

Although the accuracy of the recommendation results for academic journal submissions in this study reached 84.24% in the case of recommending three candidate journals, in future research, with the continuous increase in bibliographic data, it is possible to try to target words in different disciplines. The generation of the feature vector of the bibliographic information is optimized by combining the dynamic word vector and the static word vector to better represent the bibliographic information and high-dimensional bibliographic features are considered to improve the generalization ability and accuracy of the model.

References

Chaiwanarom, P., & Lursinsap, C. (2015). Collaborator recommendation in interdisci plinary computer science using degrees of collaborative forces, temporal evolution of research interest, and comparative seniority status. *Knowledge-Based Systems, 7*(5), 161–172.

Chen, Z., Xia, F., Jiang, H., Liu, H., & Zhang, J. (2015). AVER: random walk based academic venue recommendation. In *Proceedings of the 24th International CONFERENCE on World Wide WEB (WWW'15)* (pp. 579–584). https://doi.org/10.1145/2740908.2741738

Chen, X., Zhao, H. J., Zhao, S., Chen, J., & Zhang, Y. P. (2019). Citation recommendation based on citation tendency. *Scientometrics, 121*(2), 937–956. https://doi.org/10.1007/s11192-019-03225-6

Dawen, L., Laurent, C, James, M., & David M. (2016). Modeling user exposure in recommendation. In *Proceedings of the 24th International Conference on World Wide Web (WWW'16)* (pp. 951–961). https://doi.org/10.1145/2872427.2883090.

Feng, X., Zhang, H., Ren, Y., Shang, P., Zhu, Y., Liang, Y., & Xu, D. (2019). The deep learning-based recommender system "Pubmender" for choosing a biomedical publication venue: Development and validation study. *Journal of Medical Internet Research, 21*(5), 27–35. https://doi.org/10.2196/12957

Hiep, L., Tin, H., Gauch, S., Loc, D., & Kiem, H. (2012). Publication venue recommendation using author network's publication history. *Intelligent Information and Database Systems, 7*(1), 426–435.

Huang, W., Wu, Z., Liang, C., Mitra, P., & Giles, C. L. (2015). A neural probabilistic model for context based citation recommendation. In *Proceedings of the 29 AAAI Conference on Artificial Intelligence* (pp. 2404–2410).

Huynh, T., & Hoang, K. (2012). Modeling collaborative knowledge of publishing activities for research recommendation. *Computational Collective Intelligence: Technologies and Applications, 7*(6), 41–50.

Kong, X. J., Jiang, H. Z., Yang, Z., Xu, Z. Z., Xia, F., & Tolba, A. (2016). Exploiting publication contents and collaboration networks for collaborator recommendation. *PLoS ONE, 11*(2), 246–261.

Liang, D., Charlin, L., McInerney, J., & Blei, D. M. (2016). Modeling user exposure in recommendation. In *Proceedings of the 25th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee* (pp. 951–961).

Liu, H. F., Kong, X. J., Bai, X. M., Wang, W., Bekele, T. M., & Xia, F. (2015). Context-Based Collaborative Filtering for Citation Recommendation. *Ieee Access, 32*(3), 1695–1703. https://doi.org/10.1109/access.2015.2481320

Luong, H., Huynh, T., Gauch, S., et al. (2012). Publication venue recommendation using author network's publication history. *In Proceedings of the 4th Asian conference on Intelligent Information and Database Systems*(*ACIIDS'12*) (pp. 426–435). https://doi.org/10.1007/978-3-642-28493-9_45

Paccanaro, A., & Hinton, G. E. (2001). Learning distributed representations of concepts using linear relational embedding. *Ieee Transactions on Knowledge and Data Engineering, 13*(2), 232–244. https://doi.org/10.1109/69.917563

Pradhan, T., Gupta, A., & Pal, S. (2020). HASVRec: A modularized Hierarchical Attention-based Scholarly Venue Recommender system. *Knowledge-Based Systems, 11*(6), 204–226. https://doi.org/10.1016/j.knosys.2020.106181

Pradhan, T., & Pal, S. (2020). CNAVER: A content and network-based academic VEnue Recommender system. *Knowledge-Based Systems, 12*(24), 189–208. https://doi.org/10.1016/j.knosys.2019.105092

Schuemie, M. J., & Kors, J. A. (2008). Jane: Suggesting journals, finding experts. *Bioinformatics, 24*(5), 727–728. https://doi.org/10.1093/bioinformatics/btn006

Silva, T., Ma, J., Yang, C., & Liang, H. (2015). A profile-boosted research analytics framework to recommend journals for manuscripts. *Journal of the Association for Information Science and Technology, 66*(1), 180–200. https://doi.org/10.1002/asi.23150

Wang, D. H., Liang, Y. C., Xu, D., Feng, X. Y., & Guan, R. C. (2018). A content-based recommender system for computer science publications. *Knowledge-Based Systems, 157*, 1–9. https://doi.org/10.1016/j.knosys.2018.05.001

Wang, G., He, X., & Ishuga, C. I. (2018). HAR-SI: A novel hybrid article recommendation approach integrating with social information in scientific social network. *Knowledge-Based Systems, 14*(8), 85–99. https://doi.org/10.1016/j.knosys.2018.02.024

Xia, F., Chen, Z., Wang, W., Li, J., & Yang, L. T. (2014). MVCWalker: Random walk-based most valuable collaborators recommendation exploiting academic factors. *Ieee Transactions on Emerging Topics in Computing, 2*(3), 364–375. https://doi.org/10.1109/tetc.2014.2356505

Yang, Z., & Davison, B. D. (2012). Venue Recommendation: Submitting your Paper with Style. *International conference on machine learning and applications (ICMLA 2012)* (pp. 681–686). https://doi.org/10.1109/ICMLA.2012.127

Yu, J., Xie, K., Zhao, H. H, & Liu, F. F. (2012). Prediction of user interest based on collaborative filtering for personalized academic recommendation. *Internation conference on computer science and network technology (ICCSNT 2012)* (pp. 584–588). https://doi.org/10.1109/ICCSNT.2012.6526005

Yu, S., Liu, J. Y., Yang, Z., Chen, Z., Jiang, H. Z., Tolba, A., & Xia, F. (2018). PAVE: Personalized Academic Venue recommendation Exploiting co-publication networks. *Journal of Network and Computer Applications, 10*(4), 38–47. https://doi.org/10.1016/j.jnca.2017.12.004