A Review of Academic Recommendation Systems Based on Intelligent Recommendation Algorithms

Huaiyuan Yang
Southwest Forestry University
College of Big Data and Intelligent Engineering
Kunming, China
prof_maintainer@swfu.edu.cn

Hua Zhou*
Southwest Forestry University
College of Big Data and Intelligent Engineering
Kunming, China
hzhou@swfu.edu.cn

Yucheng Li
Southwest Forestry University
College of Big Data and Intelligent Engineering
Kunming, China
lycan@swfu.edu.cn

2022 7th International Conference on Image, Vision and Computing (ICIVC) | 978-1-6654-6734-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIVC55077.2022.9886104

***Abstract*—With the continuous increase and rapid iteration of the number of academic achievements published, it is difficult for researchers to find academic materials and literature related to their research fields. Recommendation system is the main method to deal with and alleviate network information overload at present, and it has been widely studied and applied in academia. In particular, the research on the academic recommendation system based on the intelligent recommendation algorithm has achieved certain development in recent years. This paper gives a brief overview of intelligent recommendation algorithms, and summarizes the current academic recommendation systems based on traditional intelligent recommendation algorithms and deep learning recommendation algorithms, hoping to provide assistance for related research.**

Keywords*—*recommendation system, academic, intelligent recommendation algorithm, deep learning

I. Introduction

In recent years, with the explosive growth of various academic information on the Internet, people can easily obtain a large number of academic materials and documents, which brings great convenience for researchers to carry out academic work and research. The recommendation system [1] is essential to abstract the user's interest characteristics from a pile of seemingly disorganized raw data, and mine the user's preferences. In recent years, combining deep learning with recommender systems has become a new direction in the development of recommender systems. At present, the number of scientific papers in the field of academic research is increasing at a rate of 6% to 8% every year, which has led to researchers having to spend a lot of time and energy on paper retrieval in related fields. Although there are professional academic search engines and retrieval service providers in the field of dissertation retrieval, such as Google scholar search, Microsoft academic search, CNKI, etc., However, in the face of the returned popular search results, users are often dazzled and cannot find the information they need in the shortest time. Researchers hope that the paper retrieval system can actively recommend papers in related fields according to their actual needs, so as to save retrieval time and obtain high-quality retrieval results; At the same time, you need to quickly and efficiently promote your academic achievements to your peers, so as to improve your academic reputation, increase your influence in the academic world, and lay the foundation for better scientific research resources.

Regional science and Technology Information Center project for South and Southeast Asia (2018IA018).

II. A Brief Description of the Recommendation Algorithm

A. Traditional Recommendation Algorithms

Traditional recommendation algorithms [4] are mainly divided into three categories: content-based recommendation algorithms [5], collaborative filtering recommendation algorithms [6], [7], and hybrid recommendation algorithms [8].

1) Content-based recommendation algorithm: The content-based recommendation algorithm is the earliest recommendation algorithm proposed in the intelligent recommendation algorithm and applied in real life. The content-based recommendation algorithm mainly recommends items or items that have a high similarity to them based on the items or items that the user once liked. The method is to extract the characteristics of information such as user history browsing records and logs, etc. Use these features to establish user preference documents and compare the features of the items to be recommended. The items to be recommended with the greatest relevance are selected for recommendation [9]. Since users have different preferences for each item feature, the TF-IDF method is generally used to determine the weight for each item feature.

The basic idea of TF-IDF [10] is: the more times the keyword k appears in the document D, the greater the importance of k to the document D, the more the semantics of the document D can be expressed through k .In addition, the higher the frequency of the keyword k in other documents, the less the contribution of k to distinguishing documents. Let the number of documents contained in the document set be N, the number of documents containing the keyword K_i in the document set be n_i, and f_{ij} represents the number of times the keyword K_i appears in the document d_j,the term frequency TF_{ij} of K_i in document d_j is defined as(1):

$$TF_{ij} = \frac{f_{ij}}{\max_z f_{zj}} \quad (1)$$

where z represents the keyword appearing in the document d_j. The inverse frequency IDF_i that k_i appears in the document set is defined as (2):

$$IDF_i = \log \frac{N}{n_i} \quad (2)$$

K-dimensional vectors $\vec{d}_j = (w_{1j}, w_{2j}, \cdots, w_{kj})$ and $\vec{d}_c = (w_{1c}, w_{2c}, \cdots, w_{kc})$ are used to represent the project document and the configuration document of user c, and each component in each vector is calculated by the following formula (3):

$$w_{ij} = TF_{ij} \cdot IDF_i = \frac{f_{ij}}{\max_z f_{zj}} \cdot \log \frac{N}{n_i} \quad (3)$$

Obtaining the similarity between the project document and the user profile by the angle cosine similarity [11] is the most common method, the formula is as follows (4):

$$sim(c, \ d_j) = \cos(\vec{d}_c, \vec{d}_j) = \frac{\sum_{i=1}^{k} w_{ic} w_{ij}}{\sqrt{\sum_{i=1}^{k} w^2{}_{ic}} \sqrt{\sum_{i=1}^{k} w^2{}_{ij}}} \quad (4)$$

2) Collaborative filtering recommendation algorithm: The concept of collaborative filtering recommendation was first proposed by Goldberg, Nicols, Oki, and Terry in 1992 and applied to the Tapestry system [12], which is only suitable for a small user group. At present, there are two main types of collaborative filtering recommendation algorithms: Memory-based collaborative filtering recommendation algorithm [13] and model-based collaborative filtering recommendation algorithm [14]. Memory-based recommendation algorithms are further divided into user-based recommendation and item-based recommendation.

Memory-based recommendation mainly collects user rating data for items, establishes a user-item rating matrix, and then uses the cosine similarity method to calculate the similarity between users, and recommends users according to the size of the similarity [15]. Let I_{ij} denote the item set jointly rated by user i and user j, and I_i and I_j denote the item set scored by user i and user j respectively, then the similarity between user i and user j is (5):

$$sim = \frac{\sum_{c \in I_{ij}} \left(R_{i,c} - \overline{R}_i\right)\left(R_{j,c} - \overline{R}_j\right)}{\sqrt{\sum_{c \in I_i} \left(R_{i,c} - \overline{R}_i\right)^2} \cdot \sqrt{\sum_{c \in Ij} \left(R_{j,c} - \overline{R}_j\right)^2}} \quad (5)$$

where $R_{i,c}$ represents the rating of item c by user i, and R_i and R_j represent the average rating of item c by users i and j, respectively. However, with the development of the Internet and the increase of network resources and network registered users, memory-based recommendation algorithms face serious problems such as cold start, data sparsity, and scalability. For example, a new user in the system has not rated or interacted with many items yet, a user has not rated a new item in the system, both users and items are new objects in the system. To solve the cold start problem, content-based nearest neighbor search techniques are commonly used [16].

The model-based recommendation algorithm is generally used when the amount of data is large, to establish a model to describe the user's scoring behavior of items, and to determine the model from the training data set through statistics, machine learning, and data mining techniques, and then estimate the model according to the model. Then estimate the unknown item score according to the model [17], [18].

3) Hybrid recommendation algorithm: The hybrid recommendation algorithm is a mixture of at least two or more recommendation algorithms, which is proposed to solve the drawbacks of a single recommendation algorithm. For example, the goal of the collaborative filtering recommendation algorithm is to convert the relationship between users and items into a scoring prediction problem, establish a user scoring matrix based on the user's scoring information on the item, and then analyze and recommend it, but there are problems such as cold start. The content-based recommendation can well solve the cold start problem of the recommender system [19], and reduce the impact of rating sparsity on the recommendation effect [20]. The hybrid recommendation algorithm draws on the advantages of each recommendation algorithm, learns from each other's strengths and complements the weaknesses, can achieve better recommendation effect, and improve the effectiveness of recommended content to users. Hybrid recommendation algorithms include weighted, switched, hybrid, cascading, and cascaded forms.

B. Recommendation Algorithm Based on Deep Learning

In recent years, with the development of deep learning technology research. Deep learning has achieved good results in image recognition, natural language processing, speech recognition, and other fields [21]. Now more and more recommendation research projects hope to apply and integrate deep learning technology into intelligent recommendation systems. Deep learning has excellent feature learning and the ability to fit higher-order functions, and can effectively learn corresponding features from a large amount of data [22]. Many models integrate textual and structural information through deep learning techniques to improve the performance of text-based recommender systems.

At present, the recommendation algorithms based on deep learning include deep learning recommendation algorithms using auxiliary information, model-based deep learning recommendation algorithms, and dynamic deep learning recommendation algorithms. The deep learning recommendation algorithm using auxiliary information is one of the most commonly used deep learning recommendation algorithms. The main method is to extract auxiliary information about users or items, such as content features of items, habit features of users, etc [23]. Deep learning recommendation algorithms using auxiliary information can fuse similar features or different types of features. Automatically extracting higher-level implicit feature representations of users and items, the recommendation effect is better than the general recommendation algorithm that only uses auxiliary information [24]. The model-based deep learning recommendation algorithm [25] obtains the user's latent representation by directly learning the user's item rating matrix to predict the future rating to improve the recommendation effect. The restricted boltzmann machine (RBM) [26] is the earliest neural network model applied in recommender systems. The dynamic deep learning recommendation algorithm [27] mainly deals with some dynamic features that will change with time, such as user interests and item content. We generally consider adding deep recurrent neural networks and convolutional neural networks. Deep recurrent neural networks and convolutional neural networks have great advantages in extracting time-series features and contextual features of semantic content [28].

III. Academic Recommendation System Based on Traditional Recommendation Algorithm

At present, because most academic materials are stored on the Internet in the form of literature, the development trend of academic recommendation systems with content-based recommendation algorithms, collaborative filtering-based recommendation algorithms, and hybrid recommendation algorithms as the core is still increasing day by day. Generally speaking, it is very difficult to develop an academic recommendation system that recommends accurate and efficient content, because a good academic recommendation system needs to consider various information from the literature, including the abstract of the literature, when recommending. Keywords, titles, information given by the author himself, and some abstract questions. However, this information still belongs to the literature content, as well as the analysis of users and project items. According to a large number of research literature, it is shown that academic recommendation systems using traditional recommendation algorithms can achieve better recommendation results.

Now there are some good academic recommendation systems on the Internet, such as Academia. edu [29], ResearchGate [30], Semantic Scholar [31], PubMed [32], These academic recommendation websites use traditional content-based recommendation algorithms to recommend content that users may need after comparing the literature keywords and abstracts in the database with the words and sentences that users search for. At present, a lot of effective progress has been made in the research of academic recommendation systems using traditional recommendation algorithms at the academic research level. For example, there are Kazunari Sugiyama [33] and others recommend academic papers for professional researchers. They recommend scientific papers based on the most recent articles published by users, and also classify users into junior researchers and advanced researchers according to the number of published articles. and give different parameters in the model. In the experiment, Kazunari Sugiyama et al. asked each junior and senior researcher to tag papers that were relevant to their recent research interests. These tags are stored in a database to form a user-rating matrix. The experimental database provides nearly 600 scientific articles selected from the ACL Anthology Reference Corpus (ACL ARC) [34]. Use information such as references in these papers to construct feature vectors for articles. Due to the continuous deepening of academic research, the difficulty of research topics for researchers themselves is also increasing, and the literature that they can search and consult on the Internet faces serious data sparsity. Reference [35] uses a collaborative filtering recommendation algorithm to model target papers such as abstracts, keywords, and introductions of cited papers and cited papers in the literature that are helpful to users. They used NDCG and MRR to evaluate the information retrieval quality of their system and found that this method of applying a collaborative filtering recommendation algorithm to find potential documents that users need is the best recommendation accuracy in the academic recommendation system at that time. This approach greatly alleviates the data sparsity problem that senior researchers face when researching projects. Joonseok Lee [36] et al. used a collaborative filtering recommendation algorithm-based approach to recommending their potentially preferred articles to researchers. The academic paper recommendation system studied by Joonseok Lee et al. uses the K-Nearest Neighbor (KNN) [37] algorithm to automatically monitor the research directions that the user may be researching or are interested in based on the similarity of the documents searched by the user to recommend them to the user. Braja Gopal Patra [38] et al. were the first to develop and propose an academic recommendation system based on document datasets for a recommendation. This recommendation system is based on document similarity, and the recommendation algorithm used adds additional title similarity to title similarity. weights to achieve a better recommendation effect.

IV. Academic Recommendation System Based on Deep Learning Recommendation Algorithm

Compared with traditional recommendation models (such as content-based recommendation, collaborative filtering recommendation, and hybrid recommendation), deep learning technology can more effectively and deeply mine users' potential needs, item characteristics, and user-item historical interaction records [39], [40].On the one hand, through Self-help representation learning, deep user-item features can be learned and represented, On the other hand, through automatic feature learning from multi-source heterogeneous data, there is no need to map the data to the same latent space. Inspired by this, in recent years, more and more researchers have carried out research on deep learning-based recommender systems.

Hassan [41] et al. proposed an academic paper recommendation model using Recurrent Neural Networks (RNNs) [42]. They crawl the titles and abstracts of papers from the PubMed [43] database, and use RNNs to learn these keywords to discover their deep latent semantic features to improve the quality of academic paper recommendation. Some scholars also use the method of recommending citations to make academic recommendations. In order to alleviate the data sparsity problem of top literature materials in different fields.

Wu Junchao [44] et al proposed an academic paper recommendation algorithm (WN-APR) that integrates convolutional text and heterogeneous information networks. In the experiment, in order to fully explore the impact of different features on the performance of the recommender system and enhance the recommendation diversity, they used a three-dimensional convolutional neural network instead of a two-dimensional convolutional neural network and improved the Bayesian personalized ranking loss function. The recommendation results are improved compared to the traditional baseline model in terms of accuracy and diversity. Gao Yan [45] et al. proposed an implicit feedback-based convolutional depth semantic matching paper recommendation model. The model can learn the distributed semantic expressions of users and papers, map users and papers to the same low-dimensional semantic space to calculate similarity, and achieve semantic level matching. The model uses an end-to-end supervised algorithm targeting ranking. Finally, Gao Yan et al. developed the intelligent academic assistant Acad. Bot based on this model to recommend the required academic literature for users. Ali Zafar [46] et al. proposed a network embedding model based on Generative Adversarial Network Global Citation Recommendation employing Generative Adversarial Network (GCR-GAN), which uses Heterogeneous Bibliographic Network (HBN) [47] to recommend that users may need 's citations. The recommendation results generated by this citation recommendation model on the DBLP and ACM datasets achieve over 10% gains in both the Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (nDCG) metrics. There are also teams that carry out some systematic research on improving learning methods in order to improve the effectiveness of academic research. Tomohiro Saito[48] and others designed a learning path recommendation system based on neural network. They made the scholars' current academic performance and future goals into a learning ability chart, and then applied the recurrent neural network to the learning ability chart to recommend the target users. Recommend applicable learning paths.

In general, deep learning provides a new solution for the research and development of recommendation algorithms with its powerful representation learning mode. It can effectively mine multi-source heterogeneous data such as images, and has good data fitting and generalization ability in practical applications. It can promote the application and development of joint feature learning, multi-task learning and cross-domain learning.

V. CONCLUSION

This paper gives an overview of the recommendation algorithm and the academic recommendation system and introduces the traditional recommendation algorithm and the application of the deep learning-based recommendation algorithm in the academic recommendation system.

Although the academic recommendation system based on deep learning also has the problem of overfitting, compared with the traditional recommendation system, the cold start problem and the large workload of artificial construction of features, data sparsity, and single recommendation content are well alleviated. which has played a role in promoting the development of academic recommendation systems. At present, most of the research on recommendation systems based on deep learning is aimed at e-commerce and multimedia entertainment projects (movies, music, short videos). The research on deep learning academic recommendation algorithms is still in its infancy. Scholars and teams in this direction also less. However, since the recommendation algorithm based on deep learning plays a very good role in the academic recommendation system, it is believed that the use of deep learning technology to solve the problem of recommendation algorithms is just around the corner.

REFERENCES

[1] Sharma L, Jera A. Review of Recommendation Systems: research challenges [J]. International Journal of Engineering Trends and Technology, 2013, 4(5): 1989-1992.

[2] Batmaz Z, Yurekli A, Bilge A. A review on deep learning for recommender systems: challenges and remedies[J]. Artificial Intelligence Review, 2019, 52(1): 1-37.

[3] Mu R. A survey of recommender systems based on deep learning[J]. IEEE Access, 2018, 6: 69009-69022.

[4] Bo Y, Pengfei Z. Review of Recommendation Algorithms [J]. Journal of Shanxi University (Natural Science Edition), 2011, 34(03): 337-350.

[5] Yanran W, Mei C, Hanhu W, Xin Z. A content filtering-based recommendation algorithm for scientific and technical literature[J]. Computer Technology and Development,2011,21(02):66-69.

[6] Su X, Khoshgoftaar T M. A survey of collaborative filtering techniques[J]. Advances in artificial intelligence, 2009, 2009.

[7] Xijun Y, Peisen Y, Xiaoqing G, Zhihui Y, Jing H. Collaborative filtering recommendation algorithm based on user interest and project cycle[J]. Journal of the Nanjing University of Technology,2018,42(04):392-400.

[8] Renckes S, Polat H, Oysal Y. A new hybrid recommendation algorithm with privacy[J]. Expert Systems, 2012, 29(1): 39-55.

[9] Jiangbo S, Xiaoxuan S, Hai L, Baolin Y, Zhaoli Z. A content-based recommendation algorithm for learning resources[J]. Multimedia Systems,2018,24(2).

[10] Ramos J. Using tf-idf to determine word relevance in document queries[C]//Proceedings of the first instructional conference on machine learning. 2003, 242(1): 29-48.

[11] Rahutomo F, Kitasuka T, Aritsugi M. Semantic cosine similarity[C]//The 7th International Student Conference on Advanced Science and Technology ICAST. 2012, 4(1): 1.

[12] Hongwei Ma, Guangwei Zhang, Peng LI. Review of collaborative filtering recommendation algorithm [J]. Journal of small microcomputer systems,2009,30(07):1282-1288.

[13] Yang Fuping, Wang Hongguo, Dong Shuxia, Zhao Xuechen. Collaborative filtering recommendation algorithm based on memory effect [J]. Computer Engineering, 2012, 38(23): 63-66.

[14] Liu Qingwen. Research on recommendation algorithm based on collaborative filtering [D]. University of Science and Technology of China, 2013.

[15] Liu H, Hu Z, Mian A. A new user similarity model to improve the accuracy of collaborative filtering[J]. Knowledge-Based Systems, 2014, 56(jan.):156-166.

[16] Dhanabal S, Chandramathi S. A review of various k-nearest neighbor query processing techniques[J]. International Journal of Computer Applications, 2011, 31(7): 14-22.

[17] Sandvig J J, Mobasher B, Burke R D. A survey of collaborative recommendation and the robustness of model-based algorithms[J]. IEEE Data Eng. Bull., 2008, 31(2): 3-13.

[18] Hrnjica B, Music D, Softic S. Model-based recommender systems[J]. Trends in Cloud-based IoT, 2020: 125-146.

[19] Guo Y H, Deng G S. Hybrid Recommendation Algorithm of Item Cold-start in Collaborative Filtering System[J]. Computer Engineering, 2008, 34(23):11-13.

[20] Yang F, Zheng Y, Zhang C. Hybrid recommendation algorithm based on probability matrix factorization[J]. Journal of Computer Applications, 2018.

[21] Mu R. A survey of recommender systems based on deep learning[J]. Ieee Access, 2018, 6: 69009-69022.

[22] Batmaz Z, Yurekli A, Bilge A. A review on deep learning for recommender systems: challenges and remedies[J]. Artificial Intelligence Review, 2019, 52(1): 1-37.

[23] Liu Y, Wang S, Khan M S. A novel deep hybrid recommender system based on auto-encoder with neural collaborative filtering[J]. Big Data Mining and Analytics, 2018, 1(3): 211-221.

[24] Liao B, Zhang J, Wu C. Deep sequence learning with auxiliary information for traffic prediction[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018: 537-546.

[25] Zhang L, Luo T, Zhang F. A recommendation model based on deep neural network[J]. IEEE Access, 2018, 6: 9454-9463.

[26] Zhang N, Ding S, Zhang J. An overview on restricted Boltzmann machines[J]. Neurocomputing, 2018, 275: 1186-1199.

[27] Gan M, Cui H. Exploring user movie interest space: A deep learning based dynamic recommendation model[J]. Expert Systems with Applications, 2021, 173: 114695.

[28] Wang R, Li Z, Cao J. Convolutional recurrent neural networks for text classification[C]//2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019: 1-6.

[29] Niyazov Y, Vogel C, Price R. Open access meets discoverability: Citations to articles posted to Academia. edu[J]. PloS one, 2016, 11(2): e0148257.

[30] Thelwall M, Kousha K. ResearchGate articles: Age, discipline, audience size, and impact[J]. Journal of the Association for information Science and technology, 2017, 68(2): 468-479.

[31] Fricke S. Semantic scholar[J]. Journal of the Medical Library Association: JMLA, 2018, 106(1): 145.

[32] White J. PubMed 2.0[J]. Medical Reference Services Quarterly, 2020, 39(4): 382-387.

[33] Sugiyama K, Kan M Y. Scholarly paper recommendation via user's recent research interests[C]//Proceedings of the 10th annual joint conference on Digital libraries. 2010: 29-38.

[34] Bird S, Dale R, Dorr B J. The ACL anthology reference corpus: A reference dataset for bibliographic research in computational linguistics[J]. 2008.

[35] Sugiyama K, Kan M Y. A comprehensive evaluation of scholarly paper recommendation using potential citation papers[J]. International Journal on Digital Libraries, 2015, 16(2): 91-109.

[36] Joonseok Lee,Kisung Lee,Jennifer G. Kim. Personalized Academic Research Paper Recommendation System[J]. CoRR,2013,abs/1304.5457.

[37] Peterson L E. K-nearest neighbor[J]. Scholarpedia, 2009, 4(2): 1883.

[38] Braja Gopal Patra,Vahed Maroufy,Babak Soltanalizadeh,Nan Deng,W. Jim Zheng,Kirk Roberts,Hulin Wu. A content-based literature recommendation system for datasets to improve data reusability - A case study on Gene Expression Omnibus (GEO) datasets[J]. Journal of Biomedical Informatics,2020,104(prepublish).

[39] Da'u A, Salim N. Recommendation system based on deep learning methods: a systematic review and new directions[J]. Artificial Intelligence Review, 2020, 53(4): 2709-2748.

[40] Zhang L, Luo T, Zhang F. A recommendation model based on deep neural network[J]. IEEE Access, 2018, 6: 9454-9463.

[41] Hassan H A M. Personalized research paper recommendation using deep learning[C]//Proceedings of the 25th conference on user modeling, adaptation and personalization. 2017: 327-330.

[42] Medsker L R, Jain L C. Recurrent neural networks[J]. Design and Applications, 2001, 5: 64-67.

[43] Canese K, Weis S. PubMed: the bibliographic database[J]. The NCBI handbook, 2013, 2(1).

[44] Junchao Wu, Baisong Liu, Xiaofeng Shen, Xueyuan Zhang. Academic Paper Recommendation Algorithm for Convolution Fusion Text and Heterogeneous Information Network [J/OL]. Computer Application Research: 1-7 [2022-04-10].

[45] Yan G. The paper recommendation model and application based on deep learning [D]. University of Chinese Academy of Sciences, 2017.

[46] Ali Zafar,Qi Guilin,Muhammad Khan,Kefalas Pavlos,Khusro Shah. Global citation recommendation employing generative adversarial network[J]. Expert Systems with Applications,2021,180.

[47] Sebastian Y, Siew E G, Orimaye S O. Learning the Heterogeneous Bibliographic Information Network for Literature-Based Discovery[J]. Knowledge-Based Systems, 2017, 115(JAN.1):66-79.

[48] Tomohiro Saito,Yutaka Watanobe. Learning Path Recommendation System for Programming Education Based on Neural Networks[J]. International Journal of Distance Education Technologies (IJDET),2020,18(1).