[1]Ailin Li
[2]Rong Jing
[3]Qi Guo
[4,*]Bin Wei

Paper Recommendation Method based on Attention Mechanism and Graph Neural Network

Journal of Electrical Systems

Abstract: **-** At present, most recommendation technologies only consider text or citation information, which suffers from data sparseness and cold start problems. Therefore, an academic paper recommendation method based on attention mechanism and heterogeneous graph CAH is proposed. This method considers textual information and heterogeneous graph structure information to obtain a richer and more complete feature representation. Finally, cosine similarity is calculated to generate recommendations. The results show that compared with the content-based recommendation method, the accuracy rate, recall rate and f value of CAH method are increased by nearly 5.6%, 5.8% and 8.7%, respectively, which are significantly improved compared with the basic method. This method is expected to promote the in-depth application of recommendation systems in the field of artificial intelligence.

Keywords: Paper Recommendation, Heterogeneous Graph, Artificial Intelligence, Random Walk.

I. INTRODUCTION

With the rapid development of artificial intelligence technology, academic papers play an indispensable and crucial role in academic research. However, with the continuous growth in the number and diversity of academic papers, scholars are facing an increasingly challenging task of finding the literature they need within limited time. This situation has presented the academic community with an urgent issue: how to effectively manage and utilize this vast academic literature resource.

The goal of academic paper recommendation is to leverage recommendation algorithms, utilizing scholars' historical behaviors and personal interests, to provide tailored recommendations of academic papers for research users. This recommendation system operates by intelligently filtering and suggesting academic literature that aligns closely with the users' needs and interests based on their past search and reading behavior, as well as their personal preferences and field preferences. Through this process, academic paper recommendation systems aim to significantly enhance the retrieval efficiency and reading experience for research users, enabling them to access research papers related to their work quickly and accurately. The application of this technology holds the promise of advancing academic research and facilitating the dissemination and sharing of knowledge [1].

This paper proposes a hybrid recommendation method, which combines text information and heterogeneous map to recommend academic papers. Text information can well represent papers and scientific research users, but the representation is single and lacks structural information. Therefore, heterogeneous map is used to capture structural information between scientific research users and papers as a supplement, and text information can also alleviate the problems of sparse data and cold start.

At present, the research on recommendation of academic papers mainly focuses on three aspects.

The first aspect is to realize recommendation from the perspective of content characteristics. Wu [2] used Term Frequency-Inverse Document Frequency (TF-IDF) to extract key words from abstracts of academic papers and vectorize them, calculate the similarity between papers and authors, and integrate the similarity into the process of negative case extraction and probability matrix decomposition to generate recommendations. Chen [3] used TF-IDF and Word2vec to represent the features of research users and papers in the title, keywords and abstractions of papers, and integrated time weight and similarity to realize recommendation. Xiong [4] recommended the semantic types of keywords and the time value of academic papers. Bansal T [5] used GRU method to carry out vector mapping for paper texts, so as to realize academic paper recommendation.

The second aspect is to implement recommendations from the perspective of network characteristics. Xu [6] proposed a recommendation method for academic papers based on heterogeneous network embedding based on information construction of scientific research users and papers. Haruna K [7] made use of the potential correlation between target papers and citations to realize recommendation. Ma [8] constructs a heterogeneous graph of

[1] School of Artificial Intelligence, Gansu University of Political Science and Law, Lanzhou City, China
[2] Key Laboratory of Linguistic and Cultural Computing Ministry of Education, Northwest Minzu University, Lanzhou City, China
[3] Dalian Meteorological Bureau, Dalian Meteorological Information Center, Dalian, China
[4] Key Laboratory of Linguistic and Cultural Computing Ministry of Education, Northwest Minzu University, Lanzhou City, China
*Corresponding author: Bin Wei

scientific research users and papers, and uses Doc2vec and Metapath methods for node representation and similarity calculation, so as to realize paper recommendation. Yu [9] developed a tailored recommendation system that incorporated various entity connections within diverse information networks to handle implicit feedback data sets. Shi [10] introduced the idea of a weighted heterogeneous network, adapted the similarity assessment method relying on Metapath, and developed a Metapath-based collaborative filtering model. Then, the HIN embedding method using Metapath [11] guidance is proposed to recommend.

Currently, recommendations with HIN embedding guided by Metapath are a relatively novel approach. The third aspect is to combine content features and network features to recommend academic papers. Dai [12] combined the relationship between author preference and paper citation to realize scientific paper recommendation, and designed a prototype system. Pan [13] constructed the attribute graph on the basis of feature extraction of the paper, and used PPR and SVD++ graph algorithms to sort and recommend the paper. The above research demonstrates the potential of artificial intelligence technology in building academic paper recommendation systems.

II. PAPER RECOMMENDATION METHOD

The content-based paper recommendation method is predominantly reliant on the textual data found within the papers, but it encounters specific challenges, particularly related to the cold start problem. To tackle these challenges, this study introduces an innovative strategy. This approach involves the integration of user vectors that encompass both the paper and research domains and harnesses user vectors derived from the diverse graph structure. These enhancements serve to bolster the content-based paper recommendation system significantly.

Our recommended approach seamlessly fuses together the textual content and the intricate web of relationships within the graph structure of the papers. It takes into consideration the nuanced interplay between paper content and various features, ultimately yielding more precise and comprehensive recommendations for scholarly papers.

A. Model Frame Design

Since scientific research users only have name information and cannot effectively represent the characteristics, papers written by scientific research users are selected to represent the characteristics of scientific research users. In our work, TF-IDF [14] serves to extract and merge the title and abstract from every research paper written by a scientific user, thereby characterizing their research attributes. Select the title and abstract for feature representation. Based on the SciBERT [15] model, the obtained feature words are vectored. Finally, the attention mechanism applies varying weights to words within the text, facilitating the acquisition of vector representations for both scientific research users and papers. In the heterogeneous map part, HIN embedding based on Metapath-based feature representation method is employed to acquire obtain the feature vector representation of scientific research user node and paper node. Finally, the vectors obtained from the two parts are fused and the cosine similarity is calculated for recommendation. The model framework design is shown in Figure 1.

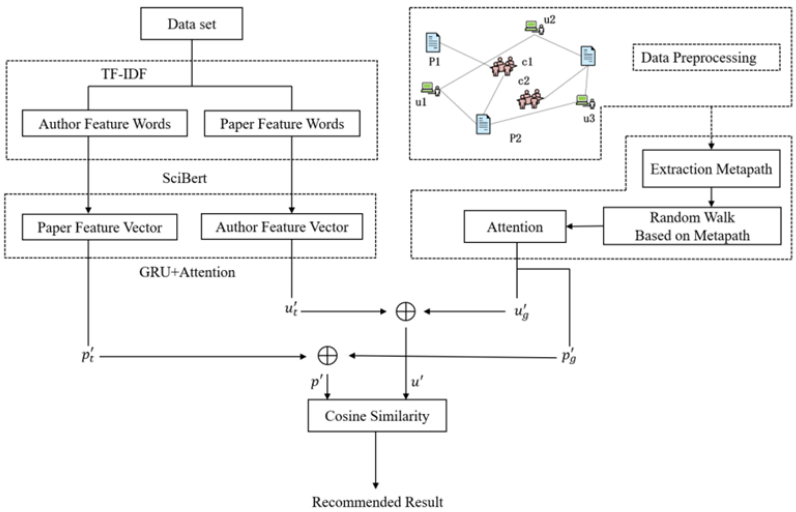

Figure 1: Recommended Model Framework

B. Feature Word Extraction

The academic papers written by scientific research users can directly and accurately represent the research interests of scientific research users, so the representation of the content of academic papers is particularly important. In the part of text feature extraction, metadata is first processed, and a series of pre-processing such as stopping words and word normalization are removed. After that, TF-IDF is used to select more important feature words, and then vector characterization is carried out.

TF-IDF is a text representation method, which combines the frequency of each word in an academic paper and the importance of the word in the whole corpus. Its calculation formula is as follows:

$$TF - IDF(w,d,D) = TF(w,d) * IDF(w) \tag{1}$$

In the recommendation task of the paper, if a word's frequency in a paper is considered as TF and the number of papers containing the word is set as DF, then IDF=log(N/DF), N is the total quantity of papers. Finally, multiply TF, IDF to get the TF-IDF value of the word in this article.

C. Feature Word Vectorization

After select the research interest feature words of scientific research users, vectorization should be carried out on the selected feature words. In this paper, the vector representation of experimental data is fine-tuned using the SciBERT model pre-trained in scientific literature. Take the vector representation of scientific users as an example.

$$\{u_1', u_2', ..., u_n'\} = SciBert(u) \tag{2}$$

This paper used the attention method to assign different weights to the words in the final vector representation.

First, a layer of Bi-GRU [16] neural network is used to encode the vector representation of the words in the title to capture the semantic relationship between the words.

$$h_{fi} = GRU_f(u_i'), i \in [1, n] \tag{3}$$

$$h_{bi} = GRU_f(u_i'), i \in [n, 1] \tag{4}$$

The word vector obtained by concatenation operation represents the hidden state u_i^' corresponding to h_i.

$$h_i = [h_{fi}; h_{bi}] \tag{5}$$

Then, the importance weight α_i of each word in the sequence is obtained by the self-attention mechanism.

$$score_i = W_2 \tanh(W_1 u_i' + b_1) + b_2 \tag{6}$$

$$\alpha_i = \frac{\exp score_i}{\sum_{i=1}^{n} score_i} \tag{7}$$

Where, W_1, W_2, b_1, b_2 are the parameters of model autonomous learning. Finally, weighted to get the research user vector representing u_t^' in the text section.

$$u_i' = \alpha_i h_i \tag{8}$$

$$u_t' = \{\alpha_1 h_1, \alpha_2 h_2, ..., \alpha_n h_n\} \tag{9}$$

Similarly, the paper vector in the text section represents p_t'. The process graph of research user vector and paper vector is obtained by using Bi-GRU and Attention mechanism for text feature words. Text vectorization is shown in Figure 2.

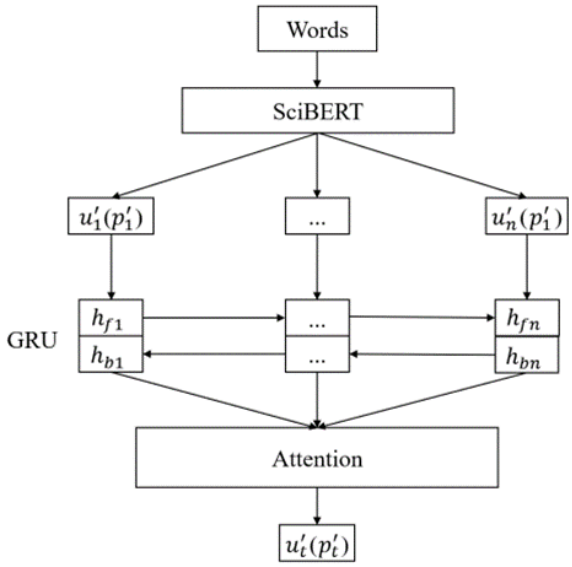

Figure 2: Text Vectorization

D. Random Walk Based on Metapath

In this experiment, the heterogeneous graph G (V, E), including author, paper, conference node, paper - author and paper - conference edge, is constructed, and the random walk strategy based on Metapath is used for graph embedding representation [17].

Given a heterogeneous information network G(V, E) and a specific Metapath, a random walk approach is utilized to generate a node sequence based on the provided Metapath, following the rules below:

$$P(n_{t+1}=x \mid n_t=v,\rho)=f(x)=\begin{cases}\frac{1}{|N^{S_{t+1}}|}, & (v,x)\in E, \varphi(x)=S_{t+1}\\ 0, & others\end{cases} \tag{10}$$

Where, n_t represents the t node of the node sequence generated in a random walk;S_t represents the type of node v; $N^{S_{t+1}}$ is the set of first-order neighbours of node v, whose type is S_{t+1}. The random walk repeats the pattern of generating nodes by following the Metapath until the length of the node sequence reaches a predefined length.

In this paper, two meta paths of UPU and UPVPU are designed to construct heterogeneous information networks, in which U represents scientific research users, P represents papers, and V represents conferences. Specific meta paths are shown in Figure 3.

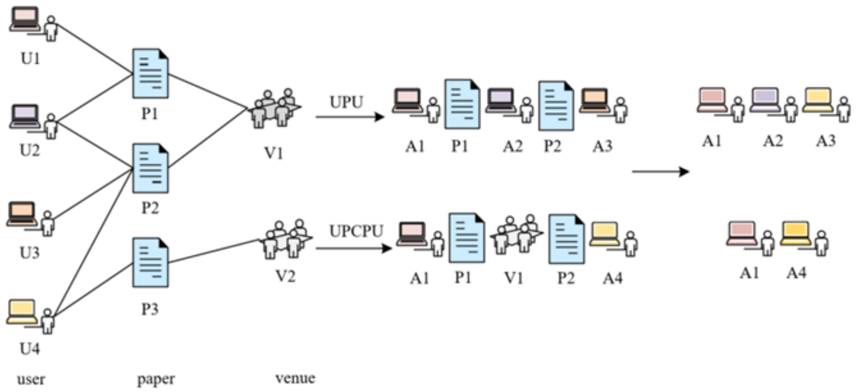

Figure 3: Basic principle of random walk strategy based on Metapath in academic paper network

In the recommendation task, the goal is to improve the recommendation performance, so it is necessary to obtain the effective representation of the research user node and the paper node. Therefore, this paper only focuses on the node sequence starting or ending with the research user node or the paper node.

E. Node Modeling

In the academic network, every node has its own neighbourhood, and the influence of the nodes in the neighbourhood is different. Given a meta-path of a specific type, this paper uses a basic attention mechanism to learn node weights under the path. The construction of the attention mechanism is depicted in Figure 4. Firstly, for each sample, there is a query vector of d_q dimension, which constitutes the query vector matrix Q of $m \times d_q$ dimension, that is, the query vector is considered to be the characteristic of the sample. In addition, for every piece of information, there is a key vector d_q dimension and a value vector dimension, forming a key-value pair.

$$H = Attention(Q,K,V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{11}$$

Where, $1/\sqrt{d_k}$ is the scale factor, and dividing by $\sqrt{d_k}$ an ensure that the gradient value remains stable during model training. The node weight information in a specific path is calculated as follows:

$$w_{ij} = Softmax(Attention(x_i,x_j)) = \frac{\exp(\rho(\alpha_\pi \cdot [x_i \parallel x_j]))}{\sum_{k\in N_i^\pi}\exp(\rho(\alpha_\pi \cdot [x_i \parallel x_k]))} \tag{12}$$

Where, x_i and x_j represent the low-dimensional node characteristics of nodes i and j, π represents a certain Metapath type, N_i^πrepresents the neighbourhood of node i, and α_π represents the attention coefficient under a path of type π. Finally, weighted fusion of nodes is performed to obtain:

$$U_i^\pi = \rho(\sum_{j\in N_i^\pi} w_{ij}^\pi \cdot x_i) \tag{13}$$

U_i^{π} is the representation of node i under path π, and is the weighted fusion of its neighbour nodes. Thus, the representation of any node under a specific path can be obtained, and the representation of nodes under different types of metaphaths can be obtained by analogies, which are then input into the path importance modelling layer to integrate multiple metaphaths learning and finally embed.

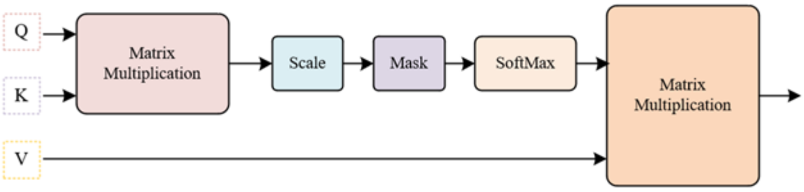

Figure 4: Structure of Attention Mechanism

F. Metapath Fusion

Node representations obtained under different Metapath have different semantics. In order to learn more comprehensive node embedding representations, node embedding representations under different Metapath need to be integrated. In this module, we refer to the Metapath fusion strategy proposed by HAN model [18]. The specific equation can be expressed as follows.

$$w_{\pi_i} = \frac{1}{|v|}\sum_{i \in v} q^T \cdot \tanh(W \cdot z_i^{\pi} + b) \tag{14}$$

Where W is the weight matrix, b is the bias vector, q is the semantic-level attention vector, and all Metapath and semantic-specific embedment share all of the above parameters. After the importance of each Metapath is obtained, they are normalized through the SoftMax function. The weight of the Metapath π_i, normalized by SoftMax function for the importance of all the metaphaths above, can obtain α_{π_i}.

$$\alpha_{\pi_i} = \frac{\exp(w_{\pi_i})}{\sum_{i=1}^{N}\exp(w_{\pi_i})} \tag{15}$$

This can be explained by the importance of different Metapath π_i to a particular task. The higher α_{π_i} is, the more important the Metapath π_i is. Next, the learned weight is taken as the coefficient, and the nodes obtained under different metagroup are weighted and fused to obtain the final embedded U, as shown below:

$$u_g{}' = \sum_{i=1}^{N} \alpha_{\pi_i} \cdot U_{\pi_i} \tag{16}$$

According to the above methods, the vector u_g' of scientific research user and the vector p_g' of paper under the heterogeneous graph are finally obtained. Finally, the obtained research user vector u_t' and u_g', paper vector p_t' and p_t' are spliced respectively to get the final research user vector $u'$and paper vector p', The specific process is as follows:

$$u' = sigmoid(u_t{}', u_g{}') \tag{17}$$

$$p' = sigmoid(p_t{}', p_g{}') \tag{18}$$

G. Similarity Calculation and Recommendation Result Generation

After obtaining the final feature vectors of scientific research users and papers, the cosine similarity between vectors is calculated by using the calculation method of cosine similarity, and the similarity is sorted from high to low. The 15 academic papers with the highest similarity value are selected to recommend to scientific research users, and the recommendation effect is compared on the utilization accuracy, recall rate and F value of the results.

Cosine similarity is equal to the dot product of two vectors divided by the lengths of two vectors.

$$Similarity(u, p) = \frac{u \cdot p}{\|u\| \times |p|} = \frac{\sum_{i=1}^{n}(u_i \times p_i)}{(\sum_{i=1}^{n} u_i^2)^{1/2} \times (\sum_{i=1}^{n} p_i^2)^{1/2}} \tag{19}$$

III. EASE OF USE

A. Data Set

This paper selected part of the data in the public data set DataBase systems and Logic Programming (DBLP) for processing, and extracted 21044 pieces of metadata including the title of the paper, the abstract of the paper, the users of the research paper and the conference information published in the paper as the initial corpus. Table 1 shows the quantity relationship of each field.

Table 1: Corresponds to the Number of Each Field

Paper	Author	Conference
21044	28646	18

B. Evaluation Method

This paper focuses on the recommendation performance of the model, so an evaluation method conforming to this method is designed to measure the effectiveness of the model.

Accuracy. Design idea: Generate 15 paper recommendation results for each scientific research user according to cosine similarity. If there are papers written by scientific research users among the 15 recommended papers, the recommendation will be regarded as successful. The accuracy rate is the ratio of the number of successful recommendations to the total number of scientific research users.

$$\mathrm{Pr}ecision = \frac{\mathrm{Re}commended\ number\ of\ \text{successes}}{\text{Total number of users}} \tag{20}$$

Recall. The calculation method of recall rate is designed as the ratio of the number of academic papers written by scientific research users to the total number of academic papers written by scientific research users in the recommended 15 academic papers.

$$\mathrm{Re}call = \frac{papers@15}{All\_papers} \tag{21}$$

F. F value is the harmonic average of accuracy and recall rate considering the accuracy and recall rate.

$$F = \frac{2\times \mathrm{Pr}ecision \times \mathrm{Re}call}{\mathrm{Pr}ecision + \mathrm{Re}call} \tag{22}$$

C. Contrast Experiment

In the experimental section of this paper, three sets of comparative experiments have been established with the aim of delving deeply into the research question and drawing conclusions based on reliable experimental results.

He present work centers on the model’s recommendation efficiency, thus an evaluation approach compatible with this is designed to assess the model’s effectiveness:

a. Content: The most basic content-based recommendation is made by calculating cosine similarity between text vector and scientific user vector.

b. Content + Attention (CA): In the representation process of text vector, the attention mechanism is added to calculate the importance of each word to get the text vector.

c. Content + Attention + Heterogeneous Graph (CAH)

The above experiments verify the effectiveness of attention mechanism and heterogeneous graph structure to feature vector complement.

D. Experimental Result

The experimental environment of our work is Ubuntu(20.04.6LTS) operating system, GPU is TITAN RTX, video memory capacity is 24G, processor is Intel(R) Xeon(R) Silver 4114 CPU @2.20GHz, Programming using python3.6 and the pytorch framework.

The outcomes are presented in Table 2, the academic paper recommendation method designed by us based on text and heterogeneous graph can obtain better recommendation results than the method based solely on content and adding attention mechanism.

Table 2: Accuracy, Recall and F

Method	Accuracy	Recall	F
Content	0.862	0.126	0.220
CA	0.896	0.165	0.279
CAH	0.918	0.184	0.307

In content-based paper recommendation, the vector representation of users has the problem of sparse data. The importance of different words was not considered when word vector was used to represent user and paper vector. This problem was alleviated after the addition of attention mechanism. However, focusing solely on text to represent user and paper vectors continues to struggle with issues related to sparse data and initialization challenges. Therefore, this paper considers to obtain structural information from the heterogeneous graph to represent the text vector, so as to obtain more accurate and abundant user vector and paper vector for scientific research. It can be seen from the obtained results that the recommendation effect is greatly improved after the addition of heterogeneous map information, and the accuracy rate is increased by nearly 5.6%, the recall rate by 5.8% and the F value by 8.7% compared with the content-based method. Heterogeneous maps can well alleviate the problems of data sparsity and cold start.

E. Usefulness Evaluation of The Model

On the premise that the recommendation effect is good, the usefulness of the recommendation model is evaluated. If most of the 15 academic papers recommended are written by scientific research users themselves, the usefulness of the recommendation model is extremely low. Therefore, the calculation method of usefulness is the proportion of the recommended 15 academic papers excluding those written by scientific research users. The usefulness value of the recommendation model was obtained by summing and averaging the usefulness indexes calculated by each scientific research user, and the usefulness value was finally calculated to be 0.826, indicating that the recommendation model proposed in this paper is able to recommend academic papers that meet the research interests of scientific research users.

IV. CONCLUSION

The research in this paper has certain limitations. This paper only collected recommendations based on the author of the paper as a scientific research user, and focused on the vector representation of content and simple cosine similarity in the design to produce recommendation results, without considering the more complex deep learning technology. In practical application, if you want to recommend scientific research users who have not written papers, consider other means of user representation, such as scientific social networking of all kinds of useful information.

Future studies, we will utilize more complex graph neural networks and employ deep learning methods, such as large-scale pre-trained language models, to design the model in order to achieve more accurate and personalized recommendation results. It's worth noting that the CAH method presented in this paper has relatively low requirements for the production environment, making it easier to deploy in real-world project applications while minimizing the model's implementation costs, all while ensuring accuracy.

This paper mainly studies the recommendation methods of academic papers. In addition to the previous content-based methods, heterogeneous maps are combined as feature supplements, and different attention mechanisms are used to capture importance in the research process. The final recommendation model CAH is significantly better than the basic content-based recommendation method in the recommendation effect. In CAH method, TF-IDF algorithm and pre-trained SciBERT were used for word vector representation of feature words. The importance weight of each word was obtained by using attention mechanism, and the basic vector representation of scientific research users and papers was obtained by weighting. Then, the heterogeneous graph is used to obtain the supplementary vector of scientific research users and papers through the metaphaths random walk method and different attention mechanisms. Finally, the basic vector and the supplementary vector are spliced together to get the final vector, and the cosine similarity is used to get the recommendation result.

ACKNOWLEDGMENT

This work was supported by Gansu Province Higher Education Institution Innovation Capacity Enhancement Project (2019A-090) and Gansu University of Political Science and Law University-Level Key Research Project (GZF2019XZDLW22).

REFERENCES

[1] Ma HF, Hu DL and Liu YH 2021 A Method to recommend scientific Papers based on author and literature Influence. Journal of Northwest Normal University (Natural Science Edition) 57(02) 55-62.

[2] Wu L, Yue F and Wang HR 2020 A Method to recommend academic papers based on tags of researchers Journal of Computer Science (02) 51-57.

[3] Chen CH, Li XT and Zou XZ 2019 An algorithm for recommending academic Papers Based on Word2vec and Time Factors Library Forum 39(05) 110-117.

[4] Xiong HX, Meng X and Ye JX 2021 Academic paper recommendation based on keyword semantic type and literature aging Modern Information 41(01): 13-23.

[5] Bansal T, Belanger D and McCallum A 2016 Ask the gru: Multi-task learning for deep text recommendations proceedings of the 10th ACM Conference on Recommender Systems 107-114.

[6] Xu K, Liu RX and Lin HF 2020 Paper recommendation method based on heterogeneous network embedding Journal of Shandong University (Science Edition) 55(11): 35-45.

[7] Haruna K, Ismail M A and Bichi A B 2018 A citation-based recommender system for scholarly paper recommendation Computational Science and Its Applications–ICCSA 2018: 18th International Conference, Melbourne, VIC, Australia, July 2-5, 2018, Proceedings, Part I 18. Springer International Publishing 514-525.

[8] Ma X and Wang R 2019 Personalized scientific paper recommendation based on heterogeneous graph representation IEEE Access 7: 79887-79894.

[9] Yu X, Ren X and Sun Y 2014. Personalized entity recommendation: A heterogeneous information network approach Proceedings of the 7th ACM international conference on Web search and data mining 283-292.

[10] Shi C, Zhang Z and Luo P 2015 Semantic path based personalized recommendation on weighted heterogeneous information networks Proceedings of the 24th ACM International on Conference on Information and Knowledge Management 453-462.

[11] Shi C, Hu B and Zhao WX 2018 Heterogeneous information network embedding for recommendation IEEE Transactions on Knowledge and Data Engineering 31(2) 357-370.

[12] Dai DW 2018 Research on recommendation algorithm of scientific Papers based on user preference and citation relationship Chongqing University.

[13] Pan F, Huai LB and Cui RY 2019 A paper recommendation Algorithm Based on Distributed Graph Computing Application Research of Computers 36(06) 1629-1632.

[14] Qaiser S and Ali R 2018 Text mining: use of TF-IDF to examine the relevance of words to documents International Journal of Computer Applications 181(1): 25-29.

[15] Beltagy I, Lo K and Cohan A 2019 SciBERT: A pretrained language model for scientific text arXiv preprint arXiv:1903.10676.

[16] Tang Q, Chen J and Lu H 2019 Full attention-based Bi-GRU neural network for news text classification 2019 IEEE 5th International Conference on Computer and Communications (ICCC) 1970-1974.

[17] La ZY, QIAN YR and Leng HY 2022 A review of graph Embedding based on Random Walk Computer Engineering and Applications 58(13) 1-13.

[18] Wang X, Ji H and Shi C 2019 Heterogeneous graph attention network The world wide web conference 2022-2032.