Academic Paper Recommendation Method Combining Heterogeneous Network and Temporal Attributes

Weisheng Li, Chao Chang(✉), Chaobo He, Zhengyang Wu, Jiongsheng Guo,
and Bo Peng

South China Normal University, Guangzhou 510631, China
{weishengli,changchao,wuzhengyang,johnsenGuo,
bpeng}@m.scnu.edu.cn

Abstract. In the case of information overload of academic papers, the demand for academic paper recommendation is increasing. Most of the existing paper recommendation methods only utilize scholar friendship or paper content information, and ignore the influence of temporal weight on research interest, and hence they are hard to obtain good recommendation quality. Aiming at this problem, the method HNTA for academic paper recommendation based on the combination of heterogeneous network and temporal attributes is proposed. HNTA firstly constructs a heterogeneous network composed of different types of entities to calculate the similarity between two papers, and then the temporal attribute is introduced into scholars' research interests which are divided into instant interests and continuous interests to calculate the similarity between scholars and papers. Finally, by weighting the above two similarities, the purpose of recommending papers to scholars is achieved. Overall, HNTA can not only comprehensively utilize both relationships of scholars and the content information of papers, but also it considers the impact of the temporal weight of scholars' research interests. By conducting comparative experiments on the data set of the real academic social network: SCHOLAT, the results show that HNTA performs better than traditional paper recommendation methods.

Keywords: Academic paper recommendation · Heterogeneous network · Temporal attributes · Academic social networks

1 Introduction

In recent years, with the rapid development of the scientific research field, academic papers are increasing exponentially. Scholars need to quickly obtain papers related to their research interests among thousands of academic papers. In this case, the social recommendation system [1, 2] stands out to help users alleviate the problem of information overload and recommend papers to users with their relevant interests. Therefore, the academic paper recommendation system has become an indispensable tool for scholars to find papers.

Y. Sun et al. (Eds.): ChineseCSCW 2020, CCIS 1330, pp. 456–468, 2021.
https://doi.org/10.1007/978-981-16-2540-4_33

At present, the existing recommendation methods at home and abroad are mainly divided into three categories: recommendation based on content information [3–5]; recommendation based on Collaborative filtering [6–8]; recommendation based on hybrid approach [9–11]. At the same time, with the emergence of a large number of academic papers, a number of academic paper recommendation methods have been proposed.

In paper recommendations, Liu et al. [12] utilize keyword-driven and popularization consciousness, and then propose a paper recommendation algorithm based on undirected paper citation maps. Parvin et al. [13] propose a new collaborative filtering algorithm to predict the similarity rating of users. Manju et al. [14] construct a heterogeneous graph of papers and use the random walk method to alleviate the problem of cold start in the paper recommendation system. Pan et al. [15] combine the citation relationship between two papers and the contextual knowledge in the paper, and then propose a method based on heterogeneous graphs for academic papers recommendation. Meng et al. [16] use the coupling relationship to implement TOP-N recommendations of keywords. Catherine et al. [17] propose a recommendation method based on the knowledge graph. Guo et al. [18] calculate the similarity of Title-Abstract Attentive Semantics for recommending. Yue et al. [19] propose a listwise learning-to-rank recommendation method based on heterogeneous network analysis in social networks, and then construct a heterogeneous network and utilize the link of relationships on the meta path to perform list-level ranking for recommendation.

The above methods mainly considers the binary relationship between users and items and is more based on the scholar's basic information and the content of the paper itself. However, in the actual situation, the paper recommendation method is usually not only based on the content information associated with the scholars themselves, but it should also take into account the changes in scholars' research interests at different time periods. The research interest and research history of scholars represent the progress and progress of scholars. Research interest includes the decay and growth of interest, and the time factor will have a certain impact on the research interest of scholars [20]. At the same time, academic social networks usually contain rich and interrelated academic information features, which can be used as a heterogeneous network containing multiple entity types and relationship types [21], from which link relationships of scholars and papers can be more conveniently obtained. In this paper, we propose an academic paper recommendation method (HNTA) that combines heterogeneous network and temporal attributes.

The main work of this paper include:

1. Constructing a heterogeneous network through different types of entities in the papers.
2. Adding temporal attributes to scholars' research interest which is divided into instant interest and continuous interest.
3. Calculating the similarity between papers and papers and the similarity between scholars and papers through the above heterogeneous network and temporal attributes. By using the data set of the real academic social platform: SCHOLAT[1], our experiment verifies that HNTA is practical and effective.

[1] http://www.scholat.com/.

2 Methodology

2.1 Constructing Heterogeneous Network

Firstly, the recommended words are extracted. The text information in the original paper data set of SCHOLAT is subjected to word segmentation, and a list of stop words is used to filter the text information to remove irrelevant stop words and symbols. The TF-IDF algorithm [22] and the Information Gain algorithm [23] are used to segment the word After processing the data, the recommended words are obtained. During this process, the keywords in the papers is added to a custom dictionary to improve the precision in the segmentation of texts.

$$TF - IDF = TF_{ij} \times IDF_j = \frac{n_{ij}}{\sum_k n_{k,j}} \times \log \frac{|D|}{1 + |\{j; t_i \in d_j\}|} \tag{1}$$

$$\begin{aligned} IG(T) = H(C) - H(C \mid T) = -\sum\nolimits_{i=1}^{m} p(c_i) \log p(c_i) + \\ p(t) \sum\nolimits_{i=1}^{m} p(c_i \mid t) \log p(c_i \mid t) + p(\bar{t}) \sum\nolimits_{i=1}^{m} p(c_i \mid \bar{t}) \log p(c_i \mid \bar{t}) \end{aligned} \tag{2}$$

Suppose there is a collection of n papers, $P = \{P_1, P_2, \dots, P_n\}$. Each paper in the collection is represented by m different types of features, $F = \{F_1, F_2, \dots, F_m\}$. Each feature Fi consists of a set of feature sets, that is, $F_i = \{f_{1i}, f_{2i}, \dots, f_{ni}\}$. The heterogeneous network is a huge network graph with semantic relations, which consists of the relationship between entities and entities that exist in the paper. After extracting the feature, the basic information of author, title, abstract, keywords and other information are obtained, and then different types of entities are represented as interrelated and different-shaped nodes to construct a heterogeneous network based on scholars and papers (Fig. 1).

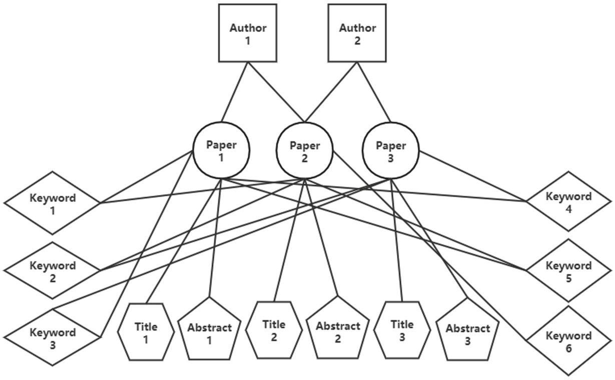


Fig. 1. Example of a heterogeneous network.

2.2 Similarity Calculation Based on Heterogeneous Network

Similarity of Pairwise Recommended Words. The extracted recommended words are constructed into a set of word vectors through the Word2Vec model [24], and then the word vectors are used to calculate the semantic similarity between word pairs: $W_{similaity}$. The calculating processes are shown in Eqs. 3 and 4,

$$W(X,Y) = \begin{cases} 0 & X,Y \notin \text{keywords} \\ W_{\text{similarity}} & X,Y \in \text{keywords} \end{cases} \tag{3}$$

$$W\left(\omega_i,\omega_j\right) = \alpha \times W_T\left(\omega_i,\omega_j\right) + \beta \times W_K\left(\omega_i,\omega_j\right) + \gamma \times W_A\left(\omega_i,\omega_j\right) \tag{4}$$

Where W_T, W_k, W_T respectively represent the similarity of the word pairs ω_i and ω_j in the thesis among the topics, keywords and abstracts, α, β, γ respectively represent the similarity of the topics, keywords and abstracts in the thesis.

Similarity of Pairwise Papers. The similarity between two papers can be measured according to the similarity between recommended words. At the same time, there are other similar papers in the papers and scholars' entities to jointly construct a similarity model of the papers. Then, the similarity between the recommended words is obtained by training the word vector: $\mathrm{S}_w(p_1,p_2)$. The calculating processes are shown in Eqs. 5,

$$\mathrm{S}_w(p_1,p_2) = \alpha \times W_T(p_1,p_2) + \beta \times W_K(p_1,p_2) + \gamma \times W_A(p_1,p_2) \tag{5}$$

Where $W_T(p_1,p_2)$ represents the similarity between the recommended phrases of the topics between the two papers, $W_K(p_1,p_2)$ represents the similarity between the recommended phrases of the keywords between the two papers, and $W_A(p_1,p_2)$ represents the similarity of the recommended words in the abstracts of the two papers. α, β, and γ respectively represent the proportion of attributes in similarities, at the same time, different types of papers have different parameter proportions.

2.3 Research Interest Similarity Based on Temporal Attributes

Interest in Continuous Research. Scholars' continuous research interest mainly refers to the relevant research interests that scholars continuously have research achievement in the process of research for a long time. In the time distribution of the data set of SCHOLAT, the time span of the paper is large, so a time decay function is used with a small time decay in the initial process: exponential time decay function. The calculating processes are shown in Eqs. 6,

$$W\left(t_y,t\right) = \frac{1}{1+e^{\mu\times\left(t-t_y\right)}} \tag{6}$$

The sum of word vectors for sexual research interests is:

$$CV_{user} = \sum\nolimits_{i=0}^{m} W(MAX(L_i(\text{year}))) \times V(L_i) \tag{7}$$

Where m is the total number of persistent keywords, L_i is the word vector of persistent keywords, W is the time decay function, V is the word vector of recommended words after training. Then, the similarity between scholars and scholars in continuing research interests is:

$$Csimilarity = \cos(CV_{\text{user1}}, CV_{\text{user2}}) = \frac{\Sigma_{i=1}^{n}\text{CVuser}_1 \times CV_{\text{user2}}}{\sqrt{\sum_{i=1}^{n}(CV_{\text{user1}})^2} \times \sqrt{\sum_{i=1}^{n} CV_{\text{user2}}}} \tag{8}$$

Interest in Instant Research. The scholar's instant research interest represents the research interest of scholars within a certain year. The key words of instant research are the concentrated expression of scholars' research interest in a certain year, which mainly represents the change of scholars' research interest.

$$R_T = R - R_c \tag{9}$$

$$IV_{user} = \Sigma_{i=0}^{k} W(R_T(i, year)) \times V(R_T(i)) \tag{10}$$

$$Isimilarity = cos(TV_{\text{user1}}, TV_{\text{user2}}) = \frac{\Sigma_{i=1}^{n} IV_{\text{user1}} \times IV_{\text{user2}}}{\sqrt{\sum_{i=1}^{n}(IV_{user1})^2} \times \sqrt{\sum_{i=1}^{n} IV_{user2}}} \tag{11}$$

Where R represents the original recommended words extracted by the instant research interest, R_c represents the Continuous research keywords of scholars, R_T represents the instant research interests after removing the persistent keywords, k represents It is the length of R_T. IV_{user} is the sum of word vectors of scholars' instant research interest.*Isimilarity* is the similarity of the instant research interests between scholars and scholars. Finally, the similarity of research interests between scholars and scholars is:

$$Ssimilarity = \sigma \times Csimilarity + \tau \times Isimilarity \tag{12}$$

Where $\sigma + \tau = 1$.

2.4 HNTA Recommendation Model

In the method of calculating the similarity proposed above, the similarity is measured through two aspects as the final recommendation algorithm. Figure 2 shows the overall process diagram of the HNTA for Top-N recommendation.

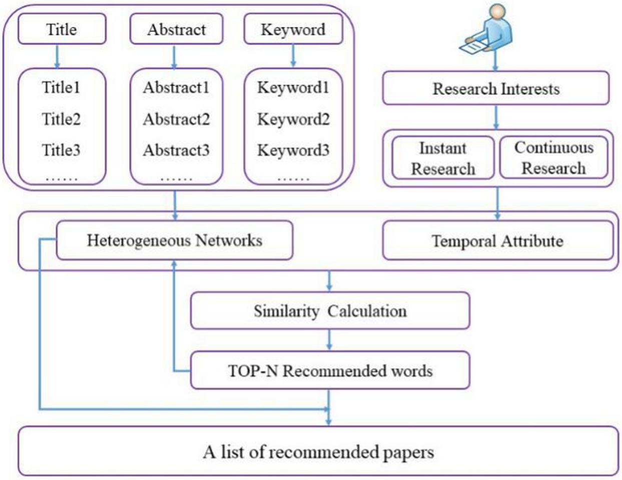


Fig. 2. The framework of HNTA.

Similarity Between Search Terms and Recommended Terms. By calculating the same keywords with the highest similarity to the current text as recommended words, where the similarity of keywords between different attributes is measured according to the similarity between different attributes. The calculation of semantic similarity between search words and candidate words is shown in Eqs. 13,

$$S(key1, key2) = \frac{\sum_{i=1}^{m} S_i(\mathrm{key}1, key2)}{m} \tag{13}$$

Where m is the number of attributes to be compared in this article, which mainly represent the abstract of the paper, keywords and topics.

Similarity between Scholars and Recommended Words. Calculate the similarity between recommended words and scholars by combining scholars' instant research interest and continuation research interest.

$$S_{sw}(SCHOLAR, CW) = \psi \times SCV_{SCHOLAR} + \omega \times SIV_{SCHOLAR} \tag{14}$$

Where $SCV_{SCHOLAR}$ represents the similarity of scholars' continuous research interest, and $STV_{SCHOLAR}$ represents the scholar's instant research interest. ψ and ω respectively represent the proportion of each module. Finally, the recommended words between user-search words are constructed, and the comprehensive standard of the above two measures is used as the final TOP-N recommended standard:

$$S_{TOP-N}(SCHOLAR, KEY) = \alpha \times S(KEY) + \beta \times S(SCHOLAR) \tag{15}$$

Finally, finding the corresponding papers of TOP-N recommended words in the heterogeneous network, so as to obtain a list of recommended papers for scholars. Table 1 shows the overall process of the HNTA.

Table 1. The overall process.

HNTA Process:
Input: a scholar enters the search term;
Output: a list of Recommended papers;
Step1. Constructing a heterogeneous network;
Step2. Constructing the scholar's research interest model based on temporal attributes;
Step3. Obtaining recommended words and find candidate papers through the heterogeneous network;
Step4. Calculating the paper similarities: $S(key1, key2)$;
Step5: Calculating the scholar similarities: $S_{sw}(SCHOLAR, CW)$;
Step6: Weighting the two similarities: $\alpha \times S(KEY) + \beta \times S(SCHOLAR)$, and getting a list of top-n recommended paper: $S_{TOP-N}(SCHOLAR, KEY)$

3 Experiments and Results

3.1 Datasets

The algorithm proposed in this paper is recommended for academic papers in academic social networks, so we choose the real paper data set of the online academic information service platform-SCHOLAT for experiments. The data set mainly included scholars' basic information, academic paper titles, abstracts, and keywords from 34,518 papers after data preprocessing. Figure 3 shows a relationship diagram of a user's paper in the SCHOLAT data set.

3.2 Baseline Methods and Evaluation Metrics

In order to verify the recommended effect of HNTA method, the following experimental methods are used for comparative analysis. The three methods are:

(1) PWFC [25], which uses a co-authored network constructed by academic achievements among scholars and the theme of published papers, builds a three-layer paper recommendation model, constructs keyword vectors and classifies academic achievements, and then researches on scholars. The co-authored network has added a random walk model to extract relevant features for TOP-N recommendation.

(2) UPR [26], which mainly proposes to construct the recommendation model of scholars' papers according to the recent research results of users and the relationship between citations and citations between two papers. Among them, it mainly constructs the recent research interest model of scholars and the feature vector of candidate papers.

(3) CB [17], which mainly adopts the classical content-based recommendation model. Through the similarity of the research results published by scholars in the content to make recommendations related to TOP-N.

宋峰华
王小明
谭文崇
李明
张伟
傅云
方家轩
张文升
胡珊
印鉴
朱定局
蔡国民
杨志康
梁路
黄佳怡
申宇铭
吴正洋
陈启买
陈玲
林美华
杨佐伟
毛超丹
陈莉莉
韩坚华
何庆
汤娜
贺超波
叶小平
姚良超
黄玲
左亚尧
王津凌
张锋
袁成哲
朱佳
赵美华
张华
陈万德
李建国
郑凯
林海
王常吉
汤庸
毛承洁
李春英
王勇
刘海
王萍
洪少文
赵淦森
汤非易
蒋运承
钟一捷
李佳玲
潘炎
陈健
刘冬宁
黄昌勤
陈明
舒忠梅
郑进嘉
刘礼
陈国华
黄泳
王鹏
彭泽武
蒋达峰
郭欣
李鹤
黄世平
刘双印
郭昆
张智博
郑泽林
王静
石玉强
周武

Fig. 3. The relationship diagram of the paper.

In this paper, Recall, Precission and F1-score value are used to evaluate the effect of TOP-N recommendation.

$$Recall = \frac{1}{N}\sum_{i=1}^{n}\frac{L_i}{L_n} \tag{16}$$

$$Precission = \frac{1}{N}\sum_{i=1}^{n}\frac{L_i}{R_i} \tag{17}$$

$$F1 - Score = \frac{2 \times Precission \times Recall}{Precission + Recall} \tag{18}$$

Where L_n represents all the research keywords that all users like among the recommended research keywords. N represents the number of samples recommended by test research interest, L_i represents the number of research keywords recommended in the sample, and R_i represents the total number of data recommended by TOP-N in the fourth sample.

3.3 Parameter Tuning

In the time attenuation function, we could find that scholars will have a great interest in a certain research in a period of time, and grow rapidly in a certain period of time.

Temporal weight parameter plays an important role in the calculation of similarity of research interest. In the data set, the time attenuation function have a large time span. Compared with exponential time attenuation function, linear time attenuation function, Logistic time attenuation function and Ebbinghaus time attenuation function, the exponential attenuation function with smaller attenuation is selected finally. After analyzing the temporal weight, it is found that when the attenuation factor is $\gamma = 0.3$, it will have the best effect. Figure 4 shows the temporal attribute weights. Besides, we set $\alpha = 0.4$, $\beta = 0.4$, $\gamma = 0.2$ in formula (4) and $\alpha = 0.3$, $\beta = 0.5$, $\gamma = 0.2$ in formula (5).

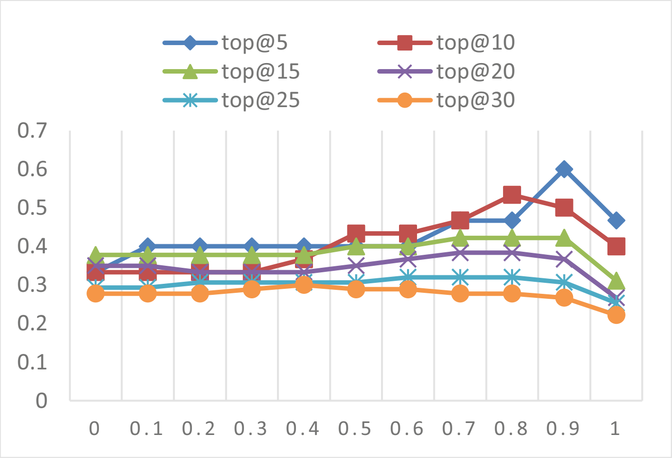


Fig. 4. According to the proportion of research interest, 0.1 is used as the step size, and the range is [0, 1]. The recommended results of different proportion are compared, and the time attenuation is exponential attenuation. In the selection of parameters, choose the one with higher recommendation Precision, that is, scholars' research interest accounts for a higher proportion: 0.9.

3.4 Comparisons and Analysis

From the similarity among abstracts, titles and keywords of the previously trained papers, the first m are selected for comparison. In this experiment, the experimental results are compared in terms of Recall, Precission and F1-Score, in which the TOP-N is 5, 10, 15, 20, 25, 30 respectively. Table 2 resports the relevant recommended words of a scholar on "friend recommendation". Table 3 is a list of papers recommended by the scholar after searching for the keyword "friend recommendation".

The following figures are comparative analysis of HNTA and other algorithms: PWFC, UPR, CB about Precission, Recall and F1-score.

As shown in Fig. 5 and Fig. 6, they respectively show the verification of the Precision, Reacll and F1-score of the relevant results of the SCHOLAT data set using the three algorithms. It could be found that, to a certain extent, the HNTA has good results in various indicators.

Table 2. The result of the recommended word of “friends recommendation”.

Number	Recommended words	Comprehensive similarity score
1	Personalized recommendation	0.6772
2	Recommendation algorithm	0.6681
3	User similarity	0.6511
4	Friend relationship	0.6428
5	Link prediction	0.6412
6	Social network	0.6322
7	User interest	0.6333

Table 3. Scholar Tang Yong’s recommended list of papers.

Num	Recommendation_Paper_title
1	Friend Recommendation in Social Network Using Nonnegative Matrix Factorization
2	Explicit and Implicit Feedback Based Collaborative Filtering Algorithm
3	Design of Learning Resource Recommendation Model Based on Interest Community
4	A Novel Hybrid Friends Recommendation Framework for Twitter
5	Multiple Criteria Recommendation Algorithm Based on Matrix Factorization and Random Forest

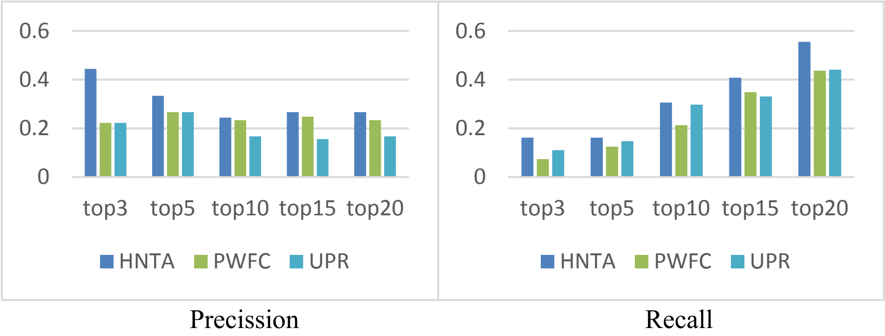
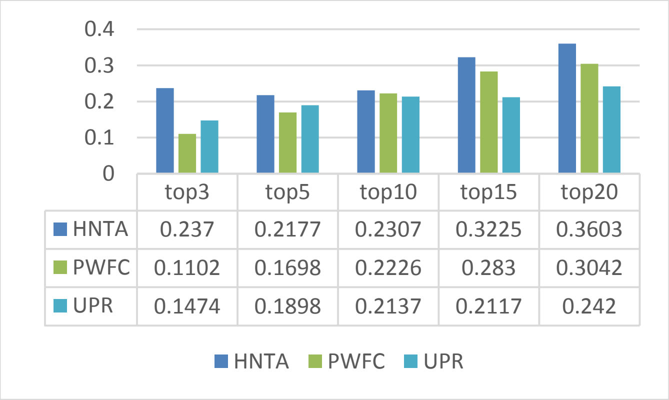


Fig. 5. Precision and recall results of HNTA in different recommended lengths.

As shown in Fig. 7, through the comparison between the HNTA algorithm and the CB algorithm, it is further proved that the HNTA algorithm will have better paper recommendation results when there is no citation relationship or the citation relationship network between scholars in the social network is not complete. The above experiments prove the feasibility and practicability of HNTA.



	top3	top5	top10	top15	top20
HNTA	0.237	0.2177	0.2307	0.3225	0.3603
PWFC	0.1102	0.1698	0.2226	0.283	0.3042
UPR	0.1474	0.1898	0.2137	0.2117	0.242

Fig. 6. F1-score comparions with different recommended length.

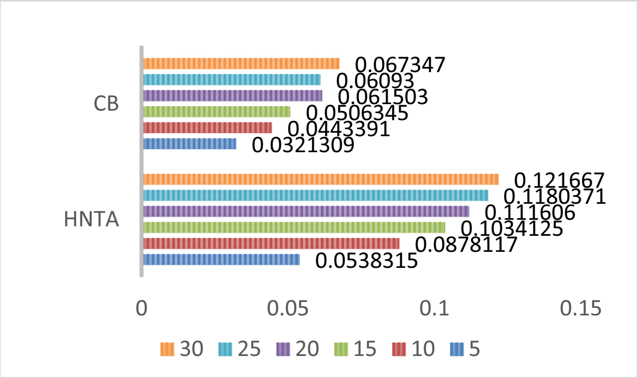


Fig. 7. The result of F1-score of HNTA and CB in different recommended length.

4 Conclusion

In this paper, on the research of paper recommendation methods, we conduct research from the scholar's research interests and the content of the paper, and then an academic paper recommendation algorithm combining heterogeneous network and temporal attributes (HNTA) is proposed. At the beginning, this paper constructs a heterogeneous network composed of different entities to calculate the similarity between papers. Then, the temporal attributes are introduced into the extraction of scholars' research interest in order to calculate the similarity between scholars and papers. Finally, by weighting the above two similarities, the purpose of recommending the papers to scholars is achieved. The data is extracted from the real paper data set in the academic platform: SCHOLAT, and the recommendation algorithm proposed in this paper was used for experimental verification. Precision, recall, and F1-score in this paper recommendation experiment were respectively carried out for the recommendation results. Experiment verifies that the HNTA method has a good recommendation effect. Since the paper contains a large amount of information that could be mined, our next step will be to consider adding trust between scholars and citation relationships between papers to further improve the accuracy of similarity in order to obtain better recommendation results.

References

1. Chun-Hua, T., Peter, B.: Exploring social recommendations with visual diversity-promoting interfaces. ACM Trans. Interact. Intell. Syst. **10**(1), 5:1–5:34 (2020)
2. Venugopal, K., Srikantaiah, K., Nimbhorkar, S.: Web Recommendations Systems, 1st edn. Springer, Heidelberg (2020). https://doi.org/10.1007/978-981-15-2513-1
3. Yiu-Kai, N.: CBRec: a book recommendation system for children using the matrix factorisation and content-based filtering approaches. IJBIDM **16**(2), 129–149 (2020)
4. Dimosthenis, B., Christos, T.: Promoting diversity in content based recommendation using feature weighting and LSH. J. Artif. Intell. Appl. Innov. **16**(1), 452–461 (2020)
5. Braja, G., Vahed, M., Babak, S., Nan, D., et al.: A content-based literature recommendation system for datasets to improve data reusability - A case study on Gene Expression Omnibus (GEO) datasets. J. Biomed. Informat. **104**, 103399 (2020)
6. Kadyanan, I., Dwidasmara, I., Mahendra, I.: A hybrid collaborative filtering recommendation algorithm: integrating content information and matrix factorisation. IJGUC **11**(3), 367–377 (2020)
7. Shunmei, M., Qianmu, L., Jing, Z., et al.: Temporal-aware and sparsity-tolerant hybrid collaborative recommendation method with privacy preservation. Concurr. Comput. Pract. Exp. **32**(2), 5447 (2020)
8. Depeng, D., Chuangxia, C., Wenhui, Y., et al.: A semantic-aware collaborative filtering recommendation method for emergency plans in response to meteorological hazards. Intell. Data Anal. **24**(3), 705–721 (2020)
9. Antonio, G., Maxim, N., Dheevatsa, M., et al.: Mixed dimension embeddings with application to memory-efficient recommendation systems. CoRR abs/1909.11810 (2019)
10. Kaya, B.: Hotel recommendation system by bipartite networks and link prediction. Inf. Sci. **46**(1), 53–63 (2019)
11. Liulan, Z., Jing, L., Weike, P., et al.: Sequence-aware factored mixed similarity model for next-item recommendation. In: BigComp, pp. 181–188 (2020)
12. Hanwen, L., Huaizhen, K., Chao, Y., et al.: Keywords-driven and popularity-aware paper recommendation based on undirected paper citation graph. Complexity **2020**, 2085638:1–2085638:15 (2020)
13. Hashem, P., Parham, M., Shahrokh, E.: TCFACO: trust-aware collaborative filtering method based on ant colony optimization. Expert Syst. Appl. **18**, 152–168 (2019)
14. Manju, G., Abhinaya, P., Hemalatha, M.R., et al.: Cold start problem alleviation in a research paper recommendation system using the random walk approach on a heterogeneous user-paper graph. IJIIT **16**(2), 24–48 (2020)
15. Pan, L., Dai, X., Huang, S., Chen, J.: Academic paper recommendation based on heterogeneous graph. In: Sun, M., Liu, Z., Zhang, M., Liu, Y. (eds.) Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. LNCS (LNAI), vol. 9427, pp. 381–392. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25816-4_31
16. Xiangfu, M., Longbing, C., Xiaoyan, Z., et al.: Top-k coupled keyword recommendation for relational keyword queries. Knowl. Inf. Syst. **50**(3), 883–916 (2017). https://doi.org/10.1007/s10115-016-0959-3

17. Rose, C., Kathryn, M., Maxine, E., et al.: Explainable entity-based recommendations with knowledge graphs. CoRR. abs/1707.05254 (2017)
18. Guibing, G., Bowei, C., Xiaoyan, Z., et al.: Leveraging title-abstract attentive semantics for paper recommendation. In: AAAI, pp. 67–74 (2020)
19. Feng, Y., Hangru, W., Xinyue,Z., et al.: Study of listwise learning-to-rank recommendation method based on heterogeneous network analysis in scientific social networks. Comput. Appl. Res. 1–7 (2020)
20. Yongbin, Q., Yujie, S., Xiao, W.: Interest mining method of weibo users based on text clustering and interest attenuation. Comput. Appl. Res. **36**(05), 1469–1473 (2019)
21. Guojia, W., Bo, D., Shirui, Pa., et al.: Reinforcement learning based meta-path discovery in large-scale heterogeneous information networks. In: AAAI, pp. 6094–6101 (2020)
22. Wang, J., Xu, W., Yan, W., et al.: Text similarity calculation method based on hybrid model of LDA and TF-IDF. In: Computer Science and Artificial Intelligence, pp. 1–8 (2019)
23. Gonen, S., Roee, A., Irad, B.: A weighted information-gain measure for ordinal classification trees. Expert Syst. Appl. **152**, 113375 (2020)
24. Martin, G.: word2vec, node2vec, graph2vec, X2vec: towards a theory of vector embeddings of structured data. In: PODS, pp. 1–16 (2020)
25. Lantian, G., Xiaoyan, C., Fei, H., et al.: Exploiting fine-grained co-authorship for personalized citation recommendation. J. IEEE Access **5**, 12714–12725 (2017)
26. Kazunari, S., Min-Yen, K.: Scholarly paper recommendation via user's recent research interests. J. ACM. 29–38 (2010)