Scholarly Paper Recommendation via Related Path Analysis in Knowledge Graph

Xiao Wang, Hanchuan Xu, Wenjie Tan, Zhongjie Wang, Xiaofei Xu
School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
{wxlxq, xhc, twj, rainy, xiaofei}@hit.edu.cn

***Abstract*—Recommending helpful and interesting scholarly papers for researchers from a large number of scholarly papers is the main way to improve research efficiency. Traditional collaborative filtering or content-based recommendation methods do not have a better-fused knowledge graph and have method bottlenecks such as cold start and poor interpretation. Based on the knowledge-aware path recurrent network (KPRN), this paper proposes a method for recommending scholarly papers that combines user preferences and knowledge graph path information. Firstly, a delayed extension bi-directional breadth-first search path algorithm is proposed to find the path between two nodes in the knowledge graph with low time complexity. Then, the user preference vector is generated by the user's historical paper operation. Finally, the LSTM cyclic neural network model is used to extract the information of multiple paths and combine it with user preferences to obtain the list of recommended papers. The experimental results show the validity and good interpretability of this method.**

***Index Terms*—Scholarly Paper Recommendation, Knowledge Graph, Related Path, KPRN**

I. Introduction

For researchers and students, the scholarly paper is an essential research reference. Faced with the increasingly large scale of scholarly papers, how to help the person in need of scholarly papers find the desired papers quickly and accurately is an essential means to improve the research efficiency. Moreover, it is also the main problem to be solved by the scholarly paper search and recommendation system.

There are two central challenges to scholarly paper recommendations. The first is the accurate modeling of researchers' research interests and preferences to achieve personalized and accurate recommendations. Researchers' research interests and content change over time, and the time-varying factors need to be taken into account when preferences are modeled. The other is to establish correlations between papers, between papers and users, and between users to enhance the interpretability and accuracy of the recommendation results. Existing methods include content-based recommendations [1], collaborative filtering-based recommendations [2], hybrid recommendation techniques [3], and graph-based scholarly paper recommendation methods [4] [5]. These methods have shortcomings in the modeling of researchers' preferences and paper correlations, resulting cold start and poor interpretability of recommendation results, and bottlenecks in the further improvement of recommendation effects.

Therefore, this paper proposes a method for recommending scholarly papers based on knowledge graph and related paths. The main ideas are: the establishment of inter-document correlations needs to be based on content such as title, author, working unit, abstract, keywords, and reference citation relationships; the knowledge graph technique can extend the related path through open domains and verticals to improve the richness, accuracy, and interpretability of the recommendation results, and the prior knowledge in the knowledge graph can alleviate the problem of cold start. In this paper, we first construct a knowledge graph of scholarly papers and then model user preferences based on the way users use scholarly paper. Then, combining the two information, long short-term memory neural network (LSTM) [6] is used to mine the related path information between nodes from the knowledge graph to complete the recommendation task based on the improved knowledge-aware path recurrent network (KPRN) recommendation algorithm [7].

This paper is organized as follows: The second part includes the system framework and the problem's mathematical definition. The third part gives the bi-directional BFS path generation method based on delay expansion. The fourth part introduces keyword-based research preference vector generation. The fifth part recommends scholarly papers using preference and related path information. Part six presents experimental validation, and the seventh part is related works, followed by a conclusion in part eight.

II. System Framework and Mathematical Definition

The problem model is as follows.

For research staff u, extract their history of preferences for papers over time $H_u = (p_1, v_1), (p_2, v_2), ..., (p_n, v_n)$ where $p_i | i \in [1, n]$ is the paper read historically by the researcher and $v_i | i \in [1, n]$ is the value of u's interest in the paper p_i. The list of existing candidate papers $Q = q_1, q_2, ..., q_m$. The goal of the paper recommendation service is to give a list of recommendations $M_u = (p_{l_1}, v_{l_1}), (p_{l_2}, v_{l_2}), ..., (p_{l_k}, v_{l_k})$ based on the historical information H_u of u, where $l_i \in [1, m]$.

The main idea of the method proposed in this paper is to extract path information from the knowledge graph of scholarly papers and recommend papers that may be of interest to users, taking into account their paper preferences. The system framework of the proposed method is shown in Figure 1 and consists of three phases.

978-1-7281-8531-6/20/$31.00 ©2020 IEEE
DOI 10.1109/ICSS50103.2020.00014

- Delayed extension of the bidirectional BFS path algorithm. The algorithm finds the paths between two nodes in the knowledge graph with low time complexity.
- Generation of user preference vectors. The vector of the researcher's scientific preferences is generated from the vector of keywords in the paper for optimizing recommendations.
- Complete the list of paper recommendations based on preference and related path information. Based on the path information obtained, the cosine of the user preference vector and the paper node vector are used as the user's preference for nodes on the path. The LSTM [6] extracts the path information and researchers' preference information and fuses the information from multiple paths to give a ranking of the user's rating for the target scholarly paper preference.

In the knowledge graph of scientific and technical services, nodes are connected to certain patterns of critical paths based on researchers' browsing habits.

TABLE I
EXAMPLES OF THE SCHOLARLY PAPER CRITICAL PATH

Meta-path	Meta-path assumptions
$P \to A \leftarrow P$	Other scholarly paper by the same author may be considered relevant to the current paper.
$P \to A \leftrightarrow A \leftarrow P$	Two papers with authors who have co-authored can be considered relevant paper.
$P \to V \leftarrow P$	Two publications in the same journal with a high degree of keyword similarity can be considered related papers.
$P \to C \to P \leftarrow K$	The paper cited and the keyword similarity between the two documents is considered relevant.
$P \to A \leftarrow P \to C \to P$	Another publication by the same author, where the paper cited in that publication can be considered relevant

Table 1 shows some examples of critical paths, where P is the paper node in the science and technology services knowledge graph, A is the author node, V is the journal node, K is the keyword node, and C represents the references between the paper. As shown in Table 1, the paths between the scholarly paper nodes reflect the reasons for the paper preferences of the research staff.

Among the critical path-based recommendation methods, KPRN [7] is a path-inference-based recommendation method. It believes that the related paths between users and items reflect the reasons for users' preferences for items and make recommendations for new users by mining the combined information of the related paths between general users and items. KPRN extracts the information of different paths through LSTM and then digs up the user's preferences for these paths to make recommendations for new users. The basic assumption of KPRN is that the reasons for preferences between different users are presented as paths in the knowledge graph. The combinations of paths are independently and equally distributed. However, different scientific researchers may have different preferences for different path combinations due to their different scientific interests and preferences. In this paper, we improve the KPRN recommendation method based on the above considerations and propose a recommendation method for scholarly papers based on preferences and related path information.

III. Bi-directional BFS Path Generation Method based on Delay Expansion

When there are some nodes in the knowledge graph with large outliers, if the BFS algorithm is directly used to extend, the number of nodes in the expansion queue will increase rapidly. As shown in Figure 2, when expanding outward from node:ijcnn, if all the nodes associated with it are directly pressed into the queue, it will make the path search time complex highly.

To solve this problem, we propose a bi-directional BFS path generation method based on delay expansion. The method can quickly generate a valid path between two nodes in the knowledge graph. The method adopts a bi-directional BFS strategy to expand the search from the starting node and the target node to the intermediate node simultaneously to reduce the amount of node scaling by half. For the two bi-directional scaling queues, two other corresponding delayed scaling queues are set up. For the node with a large outage, it is not scaled directly but is placed in the delayed expansion queue, and is determined at each step of expansion whether the node meets the opposite expansion node. For path R, if it contains only one node n_i with a large outlier or contains only two nodes n_i and n_j with higher outliers and the two nodes are directly connected. Our methods can significantly reduce the number of node extensions and thus generate paths quickly. The algorithm flows as Algorithm 1. The path between the two documents in Figure 2 is re-extended as shown in Figure 3.

It can be seen that from the starting paper node:A Robust SVM Des... to its corresponding published journal: ausai, the node will be placed in the delayed expansion queue and will not be expanded yet due to its large output. From the target node:Nonlinear dimensi... forward, expand in the same way. Place the journal node:ijcnn in the delayed expansion queue. At this point, node:ausai is found to have a similar relationship with node:ijcnn, which generates a path from the starting paper node to the target paper node. It can be seen that the number of node extensions in the queue is much less than the original BFS extensions in this process.

TABLE II
TIME FOR DIFFERENT LENGTH PATHS IN BI-DIRECTIONAL BFS SCALING BASED ON LATENCY SCALING

path length	4	6	8
Time(s)	0.065	0.369	2.697
Number of paths	9	3	23

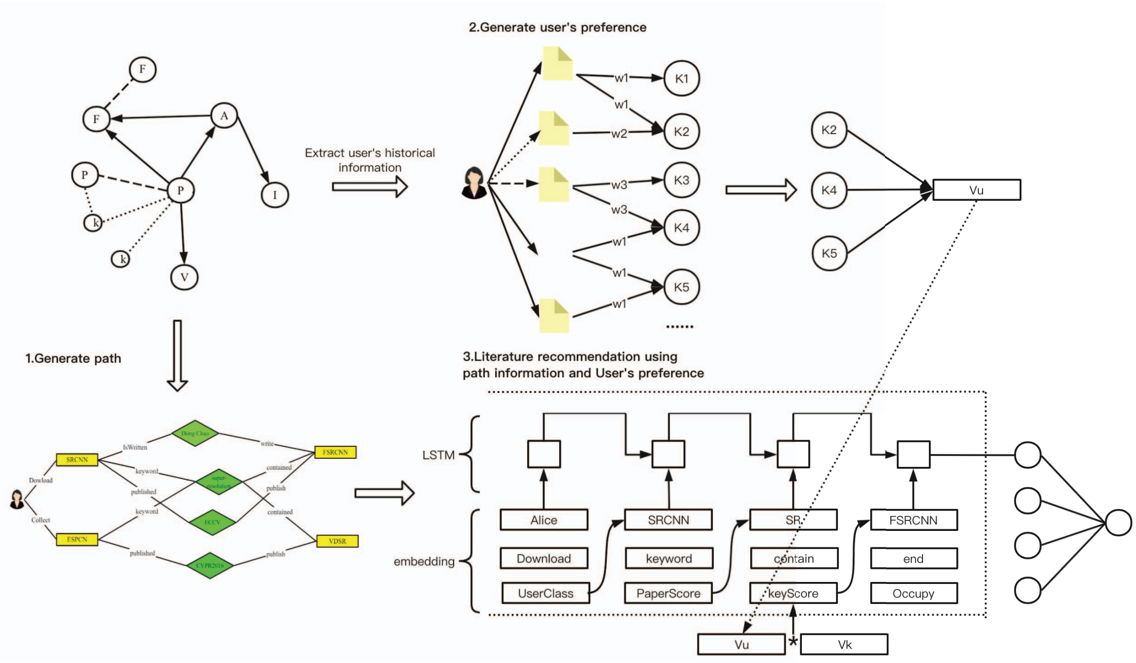

Fig. 1. A Methodological Framework for Recommending Scholarly Papers Based on Knowledge Graph and Related Path Information

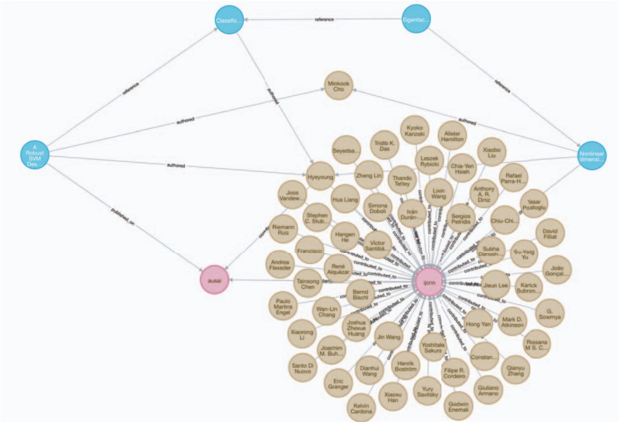

Fig. 2. An extended example of BFS

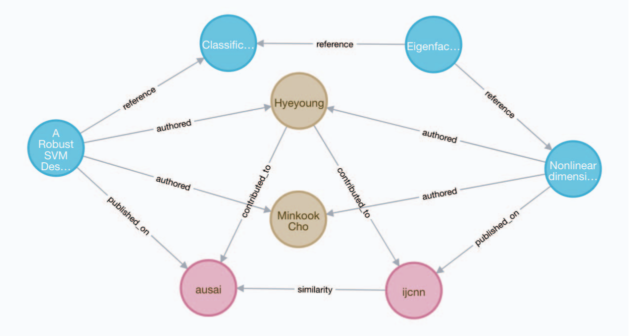

Fig. 3. Example of bi-directional BFS scaling based on latency scaling

Table 2 shows the time taken by the delayed expansion bi-directional BFS algorithm to generate different length paths between two paper nodes in the knowledge graph. It can be seen that the time consumed by the algorithm does not increase rapidly with the growth of path length. The time complexity of our method is still exponentially complex, but the floor is much smaller than the original BFS. In the case of path lengths less than 8, the path generation method in this paper is sufficient for subsequent recommendations. In the scholarly paper recommendations, the general path length is around 6, so this paper only considers paths with the length between nodes less than or equal to 6.

IV. Keyword-Based Vector Generation of Research Preferences

The preferences of the researchers can be represented using the vector $\vec{u} = (u_1, u_2, u_3, ..., u_n)$. Also, the vector $\vec{p} = (p_1, p_2, p_3, ..., p_n)$ is used to represent the paper node in the knowledge graph of scientific and technical services. So, the distance between the vector $\vec{u}$ and the vector $\vec{p}$ can be used to indicate the researcher's paper preference. The user's interest in research is usually reflected in the paper that he or she has historically browsed. The set of these paper keywords is $K = \{(k_1, \vec{v_1}), (k_2, \vec{v_2}), (k_3, \vec{v_3}), ..., (k_l, \vec{v_l})\}$, where k_i is the i-th keyword, $\vec{v_i}$ is the word vector corresponding to k_i. It is assumed that the researcher's research preference vector $\vec{u}$ and the paper representation vector $\vec{v}$ can be expressed as a function of the keyword vector $f(\vec{v_1}, \vec{v_2}, \vec{v_3}, ..., \vec{v_l})$. To explore in what way the keyword vectors can be combined to better express preferences and paper themes, this section first identifies the best vector combination solution through paper

Algorithm 1 Bi-directional BFS path generation algorithm based on delay expansion

```
Input: Start node, destination node, maximum path length, delay expansion out degree value.
Output: The path from the starting node to the target node.
Initialize que1, que2, waitQue1, waitQue2, history1, history2
while len(que1)!=0 and len(que2)!=0 and i<maxLength//2 do
    nextQue1=set()
    for startnow in que1 do
        if startnow out degree is greater than maxDegree then
            Add startnow to the waitQue1
        else
            Add nodes associated with startnow to nextQue1 and history1
        end if
    end for
    que1=nextQue1
    for endnow in que2 do
        if endnow out degree is greater than maxDegree then
            Add endnow to the waitQue2
        else
            Add nodes associated with endnow to nextQue2 and history2
        end if
    end for
    que2=nextQue2
end while
roads=[The path formed by que1 meeting que2,
The path formed by que1 meeting waitQue2,
The path formed by que2 meeting waitQue1,
The path formed by waitQue1 meeting waitQue2]
return roads
```

classification experiments. Then the method of generating the researchers' preference vector is designed according to this scheme.

A. Selection of the optimal vector combination approach

Summing keyword word vectors as a representation vector of textual information is a widely adopted class of methods. Suppose the set of keywords in document p is $\{k_1, k_2, k_3, ..., k_n\}$, the set of keyword corresponding word vectors is $\{\vec{v_1}, \vec{v_2}, \vec{v_3}, ..., \vec{v_n}\}$. Three types of word vector combinations are proposed.

- The keyword vector is directly summed as the document's representation vector $\vec{V_p} = \Sigma\vec{v_i}$.
- The keyword vector is weighted according to TF-IDF value and then summed to form the document's representation vector $\vec{V_p} = \Sigma\vec{v_i} * t_i$, where t_i is the TF-IDF value corresponding to the keyword k_i.
- Sort the keywords according to TF-IDF values from highest to lowest $[(t_{i_1}, \vec{v_{i_1}}]), (t_{i_2}, \vec{v_{i_2}}), ...(t_{i_r}, \vec{v_{i_r}})]$. Take the first r/2 word vectors and represent them as vectors of the paper by weighting and summing them according to TF-IDF values.

Figure 4 is a visualization of the paper vectors generated in three different ways above projected into two dimensions. Red dots and blue dots are two different types of papers. It can be seen that the first way is to sum the vectors directly, and the resulting paper vectors do not better distinguish between two types of scholarly papers in different fields. The second way is that the two types of nodes become linearly distinguishable. However, some nodes appear clustered. The third way distinguishes the two types of nodes better. The distance between nodes of the same class is closer while the distance between nodes of different classes is farther. Consider that the keywords of the scholarly paper differ in their importance to the paper itself, so the third way is used to represent the paper vector.

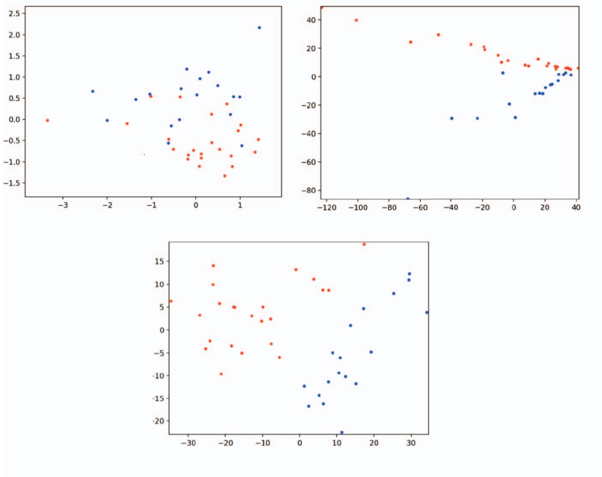

Fig. 4. Three ways to generate a paper vector

Algorithm 2 gives a clustering method based on the TF-IDF selection of keywords to represent paper. Experiments are conducted using papers from two fields. One is 115 papers in the field of network security, and the other is 92 papers in the field of speech recognition. The paper nodes are represented by the above method, and then clustering analysis is performed using K-means.

TABLE III
CLUSTER ANALYSIS BASED ON TF-IDF WEIGHTED REPRESENTATION

data sizes	**20%**	**40%**	**60%**	**80%**	**100%**
precision	0.82	0.84	0.83	0.86	0.84
recall	0.81	0.88	0.78	0.82	0.82
F1-score	0.81	0.86	0.78	0.82	0.82

Table 3 shows the results for clustering accuracy, recall, and F1-score for data sizes from 20% to 100%. It can be seen that the representation in Algorithm 2 can better represent the theme of the paper. Since the number of categories is unknown

Algorithm 2 Clustering method based on TF-IDF selection of keywords to represent paper

```
Input: The paper nodes to be clustered
Output: The paper clustering results
wordIDF[word]=TF-IDF value of the word associated with the papers
word2vec[word]=word vector of words
for line in papers do
    for word in line do
        wordSelect.append((wordIDF[word],word2vec[word]))
    end for
    wordSelect.sort()
    paperVec.append(x) //x is the sum of the first half of the wordSelect vector.
end for
for cluster in range(1,K) do
    Clustering papers into cluster classes using KMeans.
    Calculation of the Calinski-Harabasz Index for classification results.
end for
return The highest clustered results for Calinski-Harabasz Index
```

in the actual recommendation process, the choice of k-value is more important for the K-means algorithm. The Calinski-Harabasz Index can be used to evaluate the clustering effect to determine optimal k-value.

TABLE IV
SELECTION OF K-VALUES VIA CALINSKI-HARABASZ

K-value	2	3	4	5
Calinski-Harabasz Index	21.49	17.05	16.09	15.13

As can be seen in Table 4, clustering works best when the k-value is 2. This is consistent with the use of the paper data for two different types of domains.

B. Generate a vector of research preferences by combining historical operational information

Researchers' historical manipulation of the paper is divided into three categories: downloading, collecting, and browsing. Both downloading and collecting imply that the researcher has a greater interest in the paper. In the generation of research staff research preferences, different weights should be given to the paper for the different operations. A better vector representation of the paper is given in the previous section by validating the three vector combination approaches through paper classification experiments. This section designs the generation of the user's research preference vector based on this combination approach, as shown in Figure 5.

The list of keywords corresponding to $p_i|i \in [1,n]$ in the paper $\{p_1,p_2,...,p_n\}$ operated by the researcher u is sorted as $\{k_{i_1},k_{i_2},...,k_{i_{r/2}}\}$ according to TF-IDF. For $p_i|i \in [1,n]$ choose $\{k_{i_1},k_{i_2},...,k_{i_{r/2}}\}$ to get the important keywords $\{k_1,k_2,..,.k_m\}$ corresponding to the researcher u. Weight voting for the keyword $k_i|i \in [1,m]$is performed according to the historical operation of researcher u for the paper p_i. As the k_4 keyword node in Figure 4, the researcher associates k_4 with a paper collection and a download of paper. Then its weights are $W = w_3+w_1$, where w_3 is the collection operation weight, and w_1 is the download operation weight. The first $[m/2]$ keywords are selected according to their weight values, assuming $\{k_1,k_2,...,k_{[m/2]}\}$. Its corresponding word vector is $\{\vec{v_1},\vec{v_2},...,v_{\vec{[m/2]}}\}$, and the corresponding TF-IDF value is $\{t_1,t_2,...,t_{[m/2]}\}$. So the researcher's research preference vector is $\vec{V_u}=\Sigma_{i=1}^{[m/2]}t_i\vec{v_i}$.

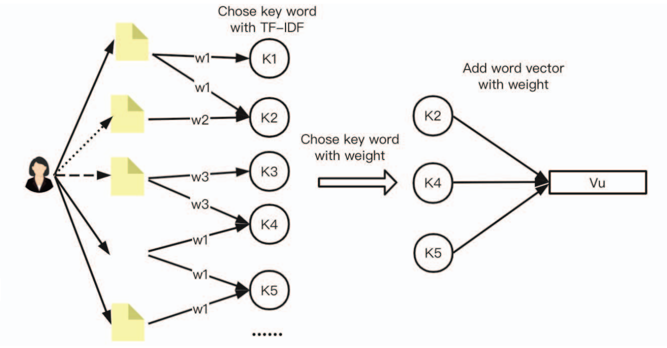

Fig. 5. The process of generating a vector of research preferences of scientific researchers

V. RECOMMEND SCHOLARLY PAPERS BASED ON PREFERENCES AND RELATED PATH INFORMATION

Assume that the list of the paper downloaded, collected, and viewed by the researcher u history is $\{p_1,p_2,...,p_n\}$. Select documents $p_i|i \in [1,n],p_j|j \in [1,n]$ with similar operation times, where the operation time of the paper p_i is earlier than that of the paper p_j. It can be assumed that the researcher u may have queried the paper p_j of his interest from the paper p_i based on the critical path. The path information p_i to p_j in the knowledge graph is used to learn the habits of the researcher u in querying the paper and is used to recommend new papers for u. The KPRN [7] recommendation model is a path-based recommendation method. This model uses LSTM to extract information from each path. Then, it uses a pooling layer to synthesize the information from multiple paths to get the user's recommendation score for the candidate. Therefore, in this paper, we improve the KPRN model based on the research preferences of researchers.

Figure 6 shows an example of the path from the paper p by the researcher u to the target paper in the scientific and technical service knowledge graph operated. The delayed extended bi-directional BFS path generation method given above can quickly find out the path association information between two documents in the scientific and technical service knowledge graph.

P1=[Alice–Download→SRCNN–is_written→Dong Chao–write→FSRCNN]

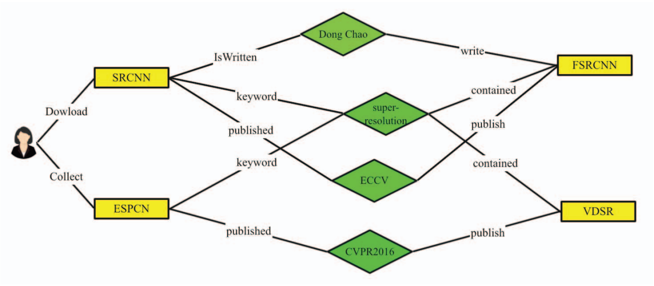

Fig. 6. Example of user-to-item related path

P2=[Alice–Download→SRCNN–keyword→sr–contained→FSRCNN]

P3=[Alice–Download→SRCNN–published→ECCV–publish→FSRCNN]

P4=[Alice–Collect→ESPCN–keyword→sr–contained→FSRCNN]

The four paths above are the corresponding paths of the researcher u through his historical operation of the paper to the paper node FSRCNN in the knowledge graph of scholarly papers. These paths consist of nodes and the relationships between nodes.

We use the user preference generation method described previously to obtain the research preference vector $\vec{V_k}$ for researcher u. Of the four paths from the researcher node u to the paper node FSRCNN, path P_2 passes through the paper node SRCNN. Then we use the paper representation to obtain the representation vector $\vec{V_i}$, and take $cos(\vec{V_k}, \vec{V_i})$ as the degree of preference of u for this paper. In this paper, we propose a scholarly paper recommendation model based on preferences and related information, as shown in Figure 7.

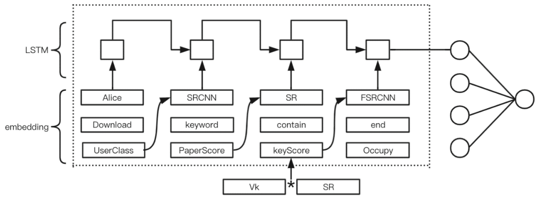

Fig. 7. Scholarly paper recommendation model based on preferences and related information

For the four paths from node u to the target node FSRCNN, the user preference vector yields the preference of u for the nodes in the paths. Summarize path information and preference information using LSTM, respectively, and synthesize the information extracted from the four paths to generate u for the target node FSRCNN scoring. The network is divided into an embedding layer and an LSTM layer. In the embedding layer, the nodes, relationships, and node attributes are encoded. Take path P2 as an example, where UserClass represents the attribute characteristics of u, and PaperScore represents the PaperNode SRCNN scoring. For the keyword node as well as the journal node, the preference vector $\vec{V_u}$ for user u is derived by the researcher's research preference extraction method. Take the point multiplication result of $\vec{V_u}$ with the word vector of keyword SR as the keyScore of SR. Connect the results of embedding $\vec{V_1}, \vec{V_2}, \vec{V_3}$ to $(\vec{V_1}, \vec{V_2}, \vec{V_3})$, and the paths $(r_1, r_2, r_3, ..., r_n)$ are processed separately by LSTM to obtain $(P_1, P_2, P_3, ..., P_n)$. The rating of the target paper is $S = sigmoid(\Sigma_{i=1}^{n} P_i W_i + b)$.

VI. Experiments

A. Experimental data

In this paper, the DBLP dataset [8] and the AMiner Academic Citation Dataset[1] are used to construct a knowledge graph of the scholarly paper. Since the DBLP dataset itself does not provide citation information, this portion of the information was obtained from the AMiner Academic Citation Dataset on kaggle and fused into the constructed knowledge graph. The resulting knowledge graph consists of 223,431 author nodes, 337,561 articles, 5578 journal and conference nodes, 1179 keyword nodes, and 16,328,642 citation relationships. To train the model and recommend papers for researchers through this model, it is necessary to obtain data on researchers' preference scores for candidate paper. In this paper, we collect 5,000 papers published by researchers as well as the paper cited by this paper, which, to some extent, can reflect the researchers' preference model. Figure 8 shows how experimental data are generated.

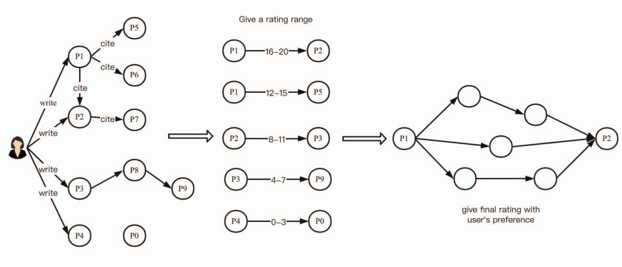

Fig. 8. Experimental data generation method

As shown in Figure 8, the author publishes documents P_1, P_2, P_3, P_4, where P_1 cited document P_2. Then the author is likely to have queried the paper P_2 via P_1, and the related paths between P_1 and P_2 reflect the author's querying habits. Along this line, the experimental data are classified into five categories $\{(P_1, P_2), (P_1, P_5), (P_2, P_3), (P_3, P_9), (P_4, P_0)\}$, where P_1 cites P_2 and both are published by the author, P_1 cites P_5 and P_5 is not published by the author, P_2, P_3 are published by the same author, P_3 and P_9 are related within a certain number of steps, P_0 is a randomly selected paper. It can be considered that these five types of data, the correlation between the paper gradually decreases. Different scoring ranges are assigned to the above five categories of data, and the final score is determined based on the author's preference for the nodes in the paper search.

[1] https://www.kaggle.com/kmader/aminer-academic-citation-dataset

B. Experimental evaluation criteria

In this paper, HR (Hit Ratio) and NDCG (Normalized Discounted Cummulative Gain) are used to evaluate the goodness of the recommendation list given by the recommendation algorithm. In top-k recommendations, HR is a common measure of recall. Also, the recommendation results should be measured by considering two factors: the greater the relevance of the recommendation results, the higher the score should be; a good relevance ranking at the top of the recommendation list indicates that the recommendation results are better, and the score should be higher. NDCG can describe these two points very well.

The training model is used to predict the top-k list of users' preferences for paper. The K-value is taken from 1 to 15, and the HR and NDCG metrics are used to compare the effectiveness of the method proposed in this paper with the KPRN recommendation model on the paper recommendation.

C. Experimental results and analysis

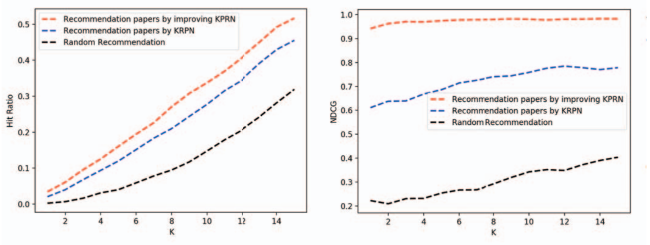

Fig. 9. Example of user-to-item related path

Figure 9 shows a comparison between the HR and NDCG evaluation criteria using the original KPRN model and the scholarly paper recommendation method proposed in this paper which combines user preferences and related path information. The horizontal coordinate is K = [1,2,3.... .15], representing the Top-K task recommended to the user. The vertical coordinates are the HR and NDCG values. Under the HR evaluation criterion, it can be seen that the improved KPRN model has some improvement relative to the original model when the K value is relatively small. As the K value gradually increases, the two effects gradually converge. However, the method proposed in this paper always has some improvement over the KPRN model. It shows that the model proposed in this paper is more accurate in predicting the researchers' rating of scholarly papers.

Under the NDCG evaluation criteria, it is clearer to see that the recommended method proposed in this chapter is better than the original KPRN model. And the proposed method in this paper also achieves a higher accuracy at K=1. This is because this paper uses certain rules to generate the data by which the rules formulate the researchers' research preferences and score the paper. Such accuracy will not be achieved when the actual recommendations are made. However, it can be shown that the proposed method can effectively exploit user preferences to outperform the original KPRN model when the actual data includes the preferences of different researchers for different papers.

In summary, it can be seen that the extracted user preferences are effective in improving the accuracy of the recommendation results. The model proposed in this paper outperforms the KPRN model under both HR and NDCG evaluation criteria.

VII. Related works

The recommendation system is usually divided into two steps: recall and ranking. The traditional method of recall of the recommendation system uses multiple recalls. [9] models the combinations between features by decomposing the parameters. [10] uses two-tower neural networks for modeling, one tower encodes item content features, and the other tower encodes user-side features and recalls items according to their similarity. With the development of the knowledge graph, there have been an increasing number of studies on integrating the knowledge graph into recommendation systems. [11] proposed a knowledge graph attention network that propagates the embedding vectors of neighbor nodes using recursion and other methods, and gives different weights to the neighbors using an attention mechanism. [12] proposed the Ripple Network, which takes the user's operations on nodes in the knowledge graph as a starting point and spreads the interest values to the surrounding neighbor nodes to form user preferences. There has also been a great deal of research on the ranking aspect of recommendation systems. [13] proposed a model that can dynamically learn feature importance and fine granularity. [14] presents the challenges of scalability and cold start for Facebook Marketplace recommendations and solves them by building a deep learning retrieval system based on collaborative multi-modality. [15] constructs the long-term interests of news users by extracting keywords through attentional networks and short-term user interests from users' recent browsing action records through GRU networks.

The paper recommendation is an important application of recommendation systems, and applying recommendation systems to the paper can save researchers, scientific and technical practitioners a great deal of time and cost. [1] uses extended activation techniques to fuse text and citation information to provide readers with customized recommendations. [2] proposed a subspace clustering method to handle the situation where the user space is too large and the paper space is relatively too small. Making recommendations for users through some frequent pattern mining algorithms is also a common means of recommendation based on collaborative filtering methods. [3] uses graph structure fusion between content and collaborative filtering based recommendation methods. [4] proposed a multivariate heterogeneous network BG (Bi-Relational Graph), which introduces the similarity relationship between documents and authors. [5] proposed a citation network, the basic idea of which is that if two documents have the same reference or are cited by the same paper, they are considered to be similar.

VIII. Conclusion

To address the scholarly paper recommendation problem, this paper proposes a scholarly paper recommendation method based on KPRN that combines user preferences and knowledge graph path information. The method generates user preferences from historical document operation data, extracts the effective paths in the knowledge graph in a short period by using the delayed expansion bi-directional BFS path generation method, combines the two information with the path information and the researcher's preference information by using LSTM cyclic neural network, and then fuses the information of multiple paths to give a ranking of user preferences for the target scholarly paper. The experimental results show the validity and good interpretability of this method.

IX. Acknowledgment

Research in this paper is partially supported by the National Key Research and Development Program of China (2018YFB1402901) and the National Science Foundation of China (61772155, 61832004, 61832014).

References

[1] A. Woodruff, R. Gossweiler, J. Pitkow, E. H. Chi, and S. K. Card, "Enhancing a digital book with a reading recommender," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2000, pp. 153–160.

[2] N. Agarwal, E. Haque, H. Liu, and L. Parsons, "Research paper recommender systems: A subspace clustering approach," in *International Conference on Web-Age Information Management*. Springer, 2005, pp. 475–491.

[3] Z. Huang, W. Chung, T.-H. Ong, and H. Chen, "A graph-based recommender system for digital library," in *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, 2002, pp. 65–73.

[4] G. Tian and L. Jing, "Recommending scientific articles using bi-relational graph-based iterative rwr," in *Proceedings of the 7th ACM conference on Recommender systems*, 2013, pp. 399–402.

[5] H. Liu, X. Kong, X. Bai, W. Wang, T. M. Bekele, and F. Xia, "Context-based collaborative filtering for citation recommendation," *IEEE Access*, vol. 3, pp. 1695–1703, 2015.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[7] X. Wang, D. Wang, C. Xu, X. He, Y. Cao, and T.-S. Chua, "Explainable reasoning over knowledge graphs for recommendation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 5329–5336.

[8] M. Mezzanzanica, F. Mercorio, M. Cesarini, V. Moscato, and A. Picariello, "Graphdblp: a system for analysing networks of computer scientists through graph databases," *Multimedia Tools and Applications*, 2018. [Online]. Available: https://doi.org/10.1007/s11042-017-5503-2

[9] S. Rendle, "Factorization machines," in *2010 IEEE International Conference on Data Mining*. IEEE, 2010, pp. 995–1000.

[10] X. Yi, J. Yang, L. Hong, D. Z. Cheng, L. Heldt, A. Kumthekar, Z. Zhao, L. Wei, and E. Chi, "Sampling-bias-corrected neural modeling for large corpus item recommendations," in *Proceedings of the 13th ACM Conference on Recommender Systems*, 2019, pp. 269–277.

[11] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, "Kgat: Knowledge graph attention network for recommendation," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 950–958.

[12] H. Wang, F. Zhang, J. Wang, M. Zhao, W. Li, X. Xie, and M. Guo, "Ripple network: propagating user preferences on the knowledge graph for recommender systems. corr abs/1803.03467 (2018)," 1803.

[13] T. Huang, Z. Zhang, and J. Zhang, "Fibinet: combining feature importance and bilinear feature interaction for click-through rate prediction," in *Proceedings of the 13th ACM Conference on Recommender Systems*, 2019, pp. 169–177.

[14] L. Zheng, Z. Tan, K. Han, and R. Mao, "Collaborative multi-modal deep learning for the personalized product retrieval in facebook marketplace," *arXiv preprint arXiv:1805.12312*, 2018.

[15] M. An, F. Wu, C. Wu, K. Zhang, Z. Liu, and X. Xie, "Neural news recommendation with long-and short-term user representations," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 336–345.