Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design

Time-aware Neural Collaborative Filtering with Multi-dimensional Features on Academic Paper Recommendation

Yibo Lu
School of Computer Science
South China Normal University
Guangzhou, China
luyibo@m.scnu.edu.cn

Yi He
School of Computer Science
South China Normal University
Guangzhou, China
yihe@m.scnu.edu.cn

Yixiang Cai
School of Computer Science
South China Normal University
Guangzhou, China
choisound@qq.com

Zelin Peng
School of Computer Science
South China Normal University
Guangzhou, China
zelin@m.scnu.edu.cn

Yong Tang *
School of Computer Science
South China Normal University
Guangzhou, China
ytang@m.scnu.edu.cn

***Abstract*—In modern academic social network, it is very difficult for scholars to find academic papers consistent with their research direction. Time is a critical factor in paper recommendation. As time goes on, the impact of an academic paper would gradually fade. Likewise, the research interests of users may also change. Therefore, we propose a temporal perceptual neural collaborative filtering model that integrates the multi-dimensional features of papers. We conducted our experiments on the dataset from CiteULike, comparing the recommended results by using four time-decay functions and evaluating our model with multiple evaluation indicators. The satisfactory results show that our model is effective in filtering out the expired papers by considering the characteristics of papers and the changes of scholars' interests.**

Keywords—Academic Social Network, Paper Recommendation, Temporal Perceptual Neural Collaborative Filtering, Multi-dimensional Features, Time-decay Functions

I. INTRODUCTION

With the rapid development of social network, the communication between scholars and scholars are increasing constantly. However, in the era of information explosion, how to dig out significant information from those tremendous data is a major challenge. Recommender system play a pivotal role in alleviating information overloaded, having been widely applied for various online services, including the social network. Thus, Scholars communicate with each other more frequently, which gives rise to a large number of academic papers. It is the key of modern recommendation system to find out what people are interested or is in line with theirs research direction from those various papers.

At present, academic paper recommendation algorithms mainly include recommendation based on data mining and information fusion, content-based filtering (CBF), collaborative filtering algorithm (CF), community discovery algorithm based on graph, knowledge-based recommendation, group recommendation, and hybrid algorithm combining two or more of them to overcome the defects of a single algorithm. Among them, CF is one of the most widely used algorithms with the best recommended effect. It has gained a great deal of attention in the last decade. Traditional collaborative filtering algorithm based on neighborhood. According to user's behavior data, the history of computing the similarity between the user or item. Then divides the users or items with great similarity into the same set of neighbors. In the end, the neighbors recommend each other. This algorithm can use the feedback information of similar users to find the potential interests of users, but it also has problems such as cold-start of data, sparseness, and expired information.

Among the whole collaborative filtering methods, matrix factorization (MF) is an excellent one, which projects users and items into a shared latent space and applies a vector of latent features to represent a user or an item. Thereafter a user's interaction on an item is modeled as the inner product of their latent vectors. But this can be hindered by the simple choice of the interaction function—inner product. In recent years, deep neural networks also have yielded some success on recommender system but not enough. Therefore, neural collaborative filtering (NCF) has been put forward by He et al. [1] through combining both of them. In this Framework, it generalized matrix factorization (GMF) and multi-layer perceptron (MLP) to enhance recommendation effect. As the name implies, GMF adopts the neural network structure to simulate the potential characteristics of users and items, and MLP is used to learn the interaction of them. On this basis, Bai et al. [2] introduce the neighborhood information in recommendation. In order to characterize neighborhood information, they proposed to use a community-based algorithm based on the interaction network. Li et al. [3] adopt the usage of the NeuMF（another abbreviation for NCF）, a neural collaborative filtering method in commercial site recommendation. Then they propose neural collaborative filtering recommendation system（NeuMF-RS）based on NeuMF method. Harada et al. [4] introduce context-aware recommendation for game apps that combines neural collaborative filtering and item embedding. It has the innovation of embedding items in the context of the game. However, all of these works ignore a significant aspect, the temporal attribute.

Time is really momentous factor. As we all know, if an academic paper is published decades ago, its impact would be feeble, which means less reference value (of course, there are several classic technologies and papers, but we don't consider them here). Meanwhile, the users' interests also changed over time. Some previous researches [5] showed that the evaluations which ignore time's impact promised more unreliable than the time-aware evaluation. In the paper of Alzogbi [7], it defined the concept-drift as user interest. By computing this score, the user's research interest has been confirmed to improve the recommended results. In this paper, we put forward a modified NCF framework named TGMF-FMLP. The time is regard as a dynamic weight multiplies the matrix in our model. In addition, in order to gain the best experiment results, we apply four time-decay functions to compare the effect. At the same time, the features of papers are trained to improve the recommended results. The remainders of this paper are as follows; in section 2, we will introduce some previous studies of academic paper recommendation. Then we will introduce our TGMF-FMLP method, the time-decay function and the technique of topic extraction in Section 3. In Section4, the experiments and results are shown to evaluate our model. Finally, we summarize our work and do a outlook for future research in Section 5.

II. Related work

Today, social network has expanded from serving the public life to sharing professional knowledge. SCHOLAT as a professional social network platform in the field of academic research has been widely concerned by scholars. Scholars can share research achievements, learn about others, and track the latest research progress in the field, which has changed the traditional research exchange and cooperation mode. However, due to the explosion of academic achievements, it is difficult for scholars to find academic papers which they interested. Therefore, recommending academic papers to scholars what they are interested has become a key function of the current academic social network. Before this, many scholars have conducted relevant research on paper recommendation. Based a heterogeneous graph in which both citation and content knowledge are included, Pan et al. [17] proposed a novel recommendation. They formally presented three definitions of the multi-layer graph: Papers Citation Graph, Key-Terms Graph and Connectivity between Paper Citation Graph and KeyTerms Graph. The experimental results demonstrate that their approach outperforms traditional methods. Rahdari et al. [18] put forward paper tuner, a user-controlled interface for recommending papers and presentations at a research conference. The availability of multiple sources of information about user interests makes hybrid recommendation approach attractive in a conference context. Sugiyama and Kan [19] identified "potential citation papers" through the use of collaborative filtering. Amami et al. [20] adopted LDA (Latent Dirichlet Allocation) to extract the theme of the paper.

But none of the above studies considered the important influence of time on the reference value of the paper. In our algorithm, we take the time into account and our TGMF-FMLP model is proposed by integrating the features of the paper and the tags marked by user.

III. Theory preparation and experimental basis

In this section, we describe our model in detail and the technology what we use.

A. Time-decay Function

Users' interests may change with the migration of time, and the influence of academic papers can gradually decrease with the change of time. That is to say, the initial research interest of scholars not only may become stable through the deepening of research content, but it may also turn to other research fields. Therefore, some of the user's recent behavior is most likely representative of their recent research interests. At present, there are several kinds of time-decay functions proposed.

1) *Linear time-decay* [8]*:* The decay of time is linear.

$$W(t_p, t) = \frac{1}{1+\tau|t-t_p|} \tag{1}$$

Where t is the current time, t_p represents the interaction time between the user and the item.

2) *Logistic time-decay* [9]*:* Originally, it was used as a curve for the growth rate of human population. The initial state of the function growth was the state of exponential growth, and then the growth would reach a relatively stable state, close to the end.

$$W(t_p, t) = \frac{1}{1+e^{\mu*(t-t_p)}} \tag{2}$$

μ is also a weight coefficient.

3) *Exponential time-decay* [10]*:* That is, time decays exponentially.

$$W(t_p, t) = e^{-\gamma*(t-t_p)} \tag{3}$$

γ is attenuation weight. Distinctly, the γ larger, the faster t decays.

4) *Ebbinghaus time-decay* [11]*:* A study by h.ebbinghaus, a German psychologist, has described how the human brain forgets new things.

$$W(t_p, t) = 0.2 * e^{\overline{\frac{0.42}{(|t-t_p|+0.00255)^{0.225}}}} \tag{4}$$

The calculation in the curve is based on days.

In the Section 4, the recommendation results of these four functions will be displayed.

B. Topic extraction

Academic abstract clustering is an effective text categorization method, which uses the concept of similarity to divide the text into several meaningful clusters according to the topic. And all documents in the same category can share the same topic. However, there is a problem of sparse data in the text of academic abstract, and the connotation of words in different contexts is quite different. Therefore, traditional clustering methods can't achieve satisfactory results for multi-text abstract clustering. Nowadays, there are mainly the following three methods: text representation based on spatial vector model, text representation based on thematic model and neural network model. The spatial vector model represents each text with one vector in the spatial vector, and each one-dimensional vector corresponds to each different word in the corpus. The length of the vector is determined by the position of the word in the document, and the vector dimension is determined by the number of words in the whole document set after word segmentation. But it ignores the context meaning relationship of the words in the document. The typical representatives of the topic model are Term frequency-inverse Document Frequency (TF-IDF) and LDA. TF-IDF can get the frequency of words in the document, and then get the subject words. LDA, which is used to extract implicit topics from a large number of documents, is a typical word bag model. It believes that words in documents exist independently from each other. Meanwhile, LDA is also a double sparse model which divides documents into as few different topics as possible. Each topic is represented by as few words as possible. However, getting sparse document vector is his shortcoming. Therefore, Doc2Vec came into being.

Doc2Vec was put forward by Le and Mikolov [12], an unsupervised algorithm that learns fixed-length feature representations from variable-length pieces of texts, such as sentences, paragraphs, and documents. Compared with Word2Vec, it overcame the weaknesses of bag-of-words models: they lose the ordering of the words and they also ignore semantics of the words, thus the accuracy was raised. To sum up, we choose Doc2Vec to convert documents of the dataset into vectors. The main steps are as follows:

- Find out the documents which were marked.
- Convert documents into vectors. (The paragraph vectors are inferred by fixing the word vectors and training the new paragraph vector until convergence).
- Concatenate the paragraph vector with several word vectors from a paragraph and predict the following word in the given context.

C. TGMF

After the above technical explanation, we will introduce our model in detail. MF is one of the most effective recommendation models for model-based collaborative filtering. It can effectively alleviate the problem of data sparsity and scalability by using dimensionality reduction technology. NCF generalizes it, and a user and an item are as feature vectors input the first layer of the neural network (in one-hot coding to solve the cold-start [13] problem), namely the input layer. Then the sparse representation of the input layer is mapped to a dense vector through a full connection layer, i.e. connect the two vectors. But many other directions are worthy of exploration and research, because it is a pure collaborative filtering model. Hence, for considering the time attribute, we put forward the model – TGMF. The specific detail of it is shown in Fig. 1.

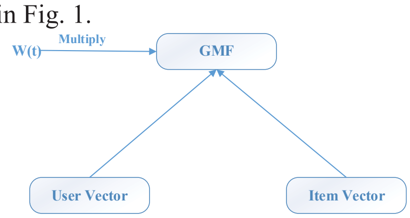

Fig. 1. Time-aware generalized matrix factorization module

The ID of the user and the project are regarded as potential vectors and then assigned different values based on the time differences, which are mapped to the output layer of TGMF. The formulation of this process is shown in (5).

$$\hat{y}_{ui} = a_s T * (c_u \odot c_i) \tag{5}$$

Where $\hat{y}_{TGMF}$ represent the output of TGMF, T is the value of time-decay function, c_u, c_i are the potential vector of the user and the item respectively. In mathematics, $\odot$ is the entry wise product, which means a binary operation. It takes two matrices of the same dimensions and produces another matrix of the same dimension as the operands where each element i, j is the product of elements i, j of the original two matrices. The Sigmoid is still the activation function of the output layer, i.e. a_s.

D. FMLP

How to measure the value of an academic paper? There are many influence factors which represent better for academic paper recommendation. In our model, some attributes of them are embedded in input layers as characteristics. Table. 1 shows the features we used.

TABLE I. THE INFORMATION OF FEATURES

Features of the category		
Interaction feature	***Attribute feature***	***Text feature***
timestamp	journal	Paper's title
tag	type	Paper's abstract

1) Interaction feature: There is a part of dataset from CiteULike which records for each user (user_hash): the papers (paper_id) he or she added to his or her library along with the timestamp and the tag. The tag is a single-word keyword and the user decides to associate to a paper in his or her library.

2) Attribute feature: There are many attributes of the academic paper, such as type, journal, publisher etc. In our module (Fig. 2), journal and type are embedded into the input layer.

3) Text feature: Title and abstract often contain the topic of a paper, especially the abstract, including the whole process from raising the problem to analyzing, experimenting and solving it in the end. Because of this, we obtain the paper's abstract and adopt the Doc2Vec model to extract the topic of each paper.

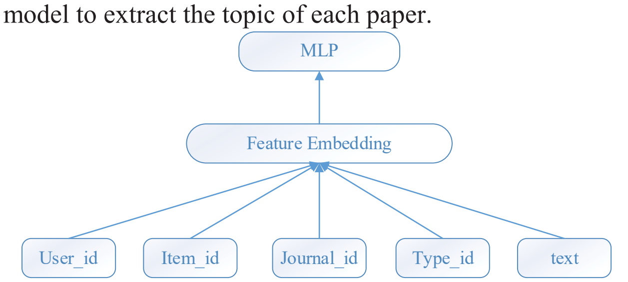

Fig. 2. Multidimensional feature module

We use the standard MLP to learn about the interaction between the user and the item's potential features, which gives the model a high level of flexibility and non-linear modeling capabilities. Rather than simply using the element-by-element inner product to describe the potential interaction features between the user and the project, as GMF does. The function of the feature embedding is mapping the high-dimensional sparse binary vector to the corresponding low-dimensional dense vector.

For a user, we get its binary vector by one-hot coding as $c_u = [0, 1, \ldots, \ldots, 0] \in R^M$, m is the number of users. Each user corresponds to a unique one-hot coding. There is an embedded matrix— $W^U = [W_1^U, W_2^U, \ldots W_M^U] \in R^{D\times M}$ for the whole users, where D is the dimension of the output embedded vector. The binary vector of u can be transformed into a low-dimensional dense embedding vector in this matrix.

The calculating process of this structure is shown in (6), (7) and (8).

$$Z_1 = [U_{id}; I_{id}; J_{id}; Tp_{id}; Text] \quad (6)$$

$$Z_{L+1} = a_L(W_L Z_L + b_L) \quad (7)$$

$$\hat{y}_{ui} = \sigma_r(Z_{L+1}) \quad (8)$$

L is the number of FMLP, W and b are respectively the weight matrix and the bias of current layer. Because of its advantage which is fit for sparse activations to prevent data's overfitting, rectifier linear units (ReLU) is used as the activation function—a_L.

E. TGMF-FMLP

So far, we have covered our two modules. We take the time factor into account on the original basis and multiply it to GMF as a dynamic weight to measure the impact of the paper and the interest of users in TGMF. In MLP, the journal and category attributes of the paper are extracted, Then, gain the topic of the paper by Doc2Vec. Finally, the three attributes above are embedded into MLP as vectors to train together with ID of the user and item. The whole model we put forward has shown in Fig. 3.

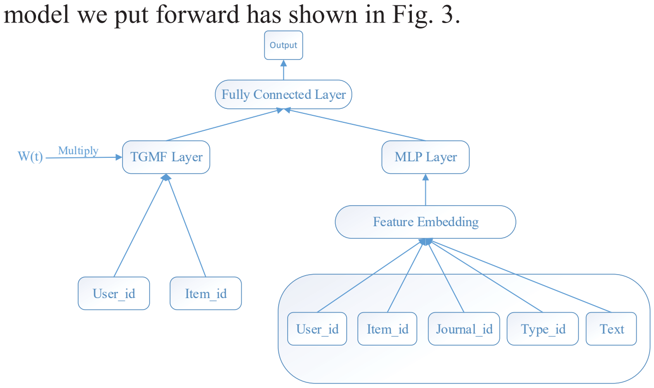

Fig. 3. TGMF-FMLP model

We still have TGMF and FMLP embedded separately, just like the original NCF. The low-dimensional dense embedded vectors are merged and sent to the fully connected layer of the neural network through forward. In the fully connected layer, we combine the output of both sides of model, and the joint prediction function is defined as shown in (9).

$$\hat{y}_{ui} = \sigma\left(W(t)\begin{bmatrix}\hat{y}_{ui}^{TGMF}\\ \hat{y}_{ui}^{FMLP}\end{bmatrix}\right) \quad (9)$$

The $\hat{y}_{ui}^{TGMF}$ is the output of TGMF module, it adopts a linear kernel to simulate potential feature interactions between users and papers. The $\hat{y}_{ui}^{MFP}$ is the output of FMLP module that learns the potential nonlinear relationship between users and items' features. $W(t)$ is the function of weight time.

IV. EXPERIMENTS AND RESULTS ANALYSIS

The TGMF-FMLP model is implemented by Tensorflow and we compare the influence of four time-decay functions on the results, and evaluate the results by HR (Hit Rate), MRR (Mean Reciprocal Rank) and NDCG (Normalized Discounted Cumulative Gain). In this section, we introduce the used dataset and the experiment. Then, we will analyze the results.

A. Dataset

CiteULike (In February 2019, CiteULike announced that it would be ceasing operations as of March 30, 2019.) was a web service which allowed users to save and share citations to academic papers. CiteULike is generally used to help you store, manage, and share academic papers you are reading. You can also classify papers and label individual papers or groups of papers. Even more exciting, CiteULike offers social features. This allows you to share your library of papers to see who is as interested in writing a paper as you are, and along the way, you may find more useful literature in your field, who is doing research similar to yours, what their concerns are, etc.

In our dataset, it contains 210137 papers, 3039 users and 284960 ratings which represent the interaction between users and papers from 11, 2004 to 12, 2007. The description of it is shown in Table II.

TABLE II. DATASET DESCRIPTION

Description	**Files of Dataset**			
	Users.dat	***Field.csv***	***Papers.dat***	***Ratings.csv***
content	users_id, u_map_id	f_term	p_id, p_type	r_text, r_tag[a], r_timestamp

[a] the tag for paper edited by user

B. Experiment Design

We divided the dataset into training set and test set, with 100 samples in each batch. There are 4 negative samples of every 100 samples in the training set. The test set contains 99 negative samples, i.e. only one is correct. The user and the paper were embedded as 32-dimensional vectors, and the loss rate was initialized to 0.2. The Sigmoid was used as the activation function in TGMF module, and the FMLP module used ReLU as the activation function, and the number of initialization iterations was set as 50. In terms of the optimizer, we chose Adam, which has the following advantages combined with the current scenario of our experiment.

- Simple implementation, efficient calculation, less memory requirements.
- The amount of data is relatively large, and Adam is more suitable for the current scene.
- The time-decay function is uncertain. Four kinds of time-decay functions are adopted, and Adam is used to make the update of parameters not affected by the gradient expansion transformation.

C. Results analysis

We performed the same experiment with four different time-decay functions and evaluated them separately using HR, RR and NDCG. The best ten results are shown in Table III.

TABLE III. RESULTS OF DIFFERENT TIME-DECAY FUNCTIONS

Top-10	Time-decay Functions			
	Exponential	***Linear***	***Logistic***	***Ebbinghaus***
HR	0.62605	0.49352	**0.75843**	0.54189
MRR	0.47426	0.39740	**0.62650**	0.41826
NDCG	0.46438	0.34275	**0.66218**	0.46996

As we can see from Table III, different time-decay functions correspond to different experimental results. Among them, Logistic time-decay function has the best effect, with a hit rate as high as 75.84%. MRR and NDCG also reached 62.65% and 66.21% respectively. Therefore, we chose Logistic as the time-decay function and further evaluated the results of top-5, top-15 and top-20 in the test set. All of these results are shown in Fig. 4.

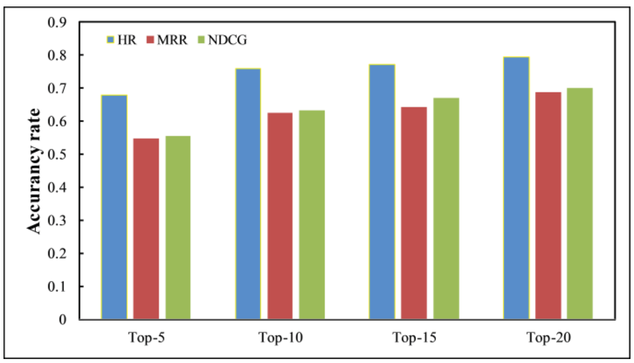

Fig. 4. Evaluation of the result

In Fig. 4, we find that with the increase of recommendation results, the accuracy of HR, MRR and NDCG evaluation methods is also gradually increasing. The above shows that after considering the time factor and some features of the paper, the recommended results are satisfactory.

V. CONCLUSION AND PROSPECTS

In this paper, we proposed TGMF-FMLP, a time-aware neural collaborative filtering with multi-dimensional features model. Firstly, we use four kinds of time-decay functions as weight coefficients to add them into the characteristic GMF model of analyzing the user and the potential linearity of the paper, so as to propose our TGMF module. From the CiteULike dataset at the same time, we obtain the interaction information between users and academic papers, namely the tags which were marked by users. Then, the interests of users are extracted. On the other hand, the type and abstract of papers are also extracted by us to analysis. The features which we have got are embedded in our model to be trained in the MLP, i.e. our FMLP module. Finally, the training results of the two modules are fully connected in the Fully Connected layer to obtain the final Top-10 recommended results. Therefore, we find that the Logistic time-decay function have the best effect. On this basis, we continue to carry out Top-5, Top-15 and Top-20 recommended experiments, and the results prove that the TGMF-FMLP model we proposed is very effective. But, our work is not enough. There are some other features of academic papers such as journal, book's title and publisher that we don't use yet. We are considering whether we can make a ranking for the journals of the papers, and then assign different importance weights to them in the future.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China under grant number U1811263, by National Natural Science Foundation of China under grant number 61772211.

REFERENCES

[1] He X, Liao L, Zhang H, et al. Neural Collaborative Filtering[J]. 2017.

[2] Ting Bai, Ji-Rong Wen, Jun Zhang, Wayne Xin Zhao: A Neural Collaborative Filtering Model with Interaction-based Neighborhood. CIKM 2017: 1979-1982.

[3] Nuo Li, Bin Guo, Yan Liu, Yao Jing, Yi Ouyang, Zhiwen Yu: Commercial Site Recommendation Based on Neural Collaborative Filtering. UbiComp/ISWC Adjunct 2018: 138-141.

[4] Shonosuke Harada, Kazuki Taniguchi, Makoto Yamada, Hisashi Kashima: Context-Regularized Neural Collaborative Filtering for Game App Recommendation. RecSys 2019: 16-20.

[5] Campos, P.G., D´ıez, F., Cantador, I.: Time-aware recommender systems: A comprehensive survey and analysis of existing evaluation protocols. User Modeling and User-Adapted Interaction 2014, 24(1-2): 67-119.

[6] Matuszyk, P., Vinagre, J., Spiliopoulou, M., Jorge, A.M., Gama, J.: Forgetting techniques for stream-based matrix factorization in recommender systems. Knowledge and Information Systems 2018, 55(2): 275-304.

[7] Anas Alzogbi: Time-aware Collaborative Topic Regression: Towards Higher Relevance in Textual Item Recommendation. BIRNDL@SIGIR 2018: 10-23.

[8] Wu D, Yuan Z, Yu K, et al. Temporal social tagging based collaborative filtering recommender for digital library[C]//International Conference on Asian Digital Libraries. Springer, Berlin, Heidelberg, 2012: 199-208.

[9] Lee T Q, Park Y, Park Y T. A time-based approach to effective recommender systems using implicit feedback[J]. Expert systems with applications, 2008, 34(4): 3055-3062.

[10] Wang X, Zhao Y L, Nie L, et al. Semantic-based location recommendation with multimodal venue semantics[J]. IEEE Transactions on Multimedia, 2015, 17(3): 409-419.

[11] Zhang X L , Tak Man Desmond Lee…. Securing Recommender Systems Against Shilling Attacks Using Social-Based Clustering[J]. Journal of Computer Science and Technology, 2013, 28(4):616-624.

[12] Le Q, Mikolov T. Distributed representations of sentences and documents[C]//International conference on machine learning. 2014: 1188-1196.

[13] Andrew I. Schein, Alexandrin Popescul, Lyle H. Ungar, David M. Pennock:

[14] Methods and metrics for cold-start recommendations. SIGIR 2002: 253-260.

[15] Sharon Givon, Victor Lavrenko:Predicting social-tags for cold start book recommendations. RecSys 2009: 333-336.

[16] Seung-Taek Park, Wei Chu:Pairwise preference regression for cold-start recommendation. RecSys 2009: 21-2.

[17] Linlin Pan, Xinyu Dai, Shujian Huang, Jiajun Chen:Academic Paper Recommendation Based on Heterogeneous Graph. CCL 2015: 381-392

[18] Behnam Rahdari, Peter Brusilovsky: User-controlled hybrid recommendation for academic papers. IUI Companion 2019: 99-100.

[19] Sugiyama K , Kan M Y . Exploiting potential citation papers in scholarly paper recommendation[J]. 2013.

[20] Amami M , Pasi G , Stella F , et al. An LDA-Based Approach to Scientific Paper Recommendation[C]// International Conference on Applications of Natural Language to Information Systems. Springer International Publishing, 2016.