www.arpnjournals.com

RESEARCH TOPICS RECOMMENDATION SYSTEM FOR NOVICE RESEARCHERS

Mohammad Mahbub Alam and Maizatul Akmar Ismail
Department of Information Systems University of Malaya, Kuala Lumpur, Malaysia
E-Mail: mahbubalam4545@gmail.com

ABSTRACT

Recently, recommender systems (RSs) are making great stride in academic arena for researchers. There exist some true potential for recommender systems to assist novice researchers in instigating their preliminary activities. Providing personalized recommendation on research topics is of great help for beginners to enter the area of their research interest. To select a suitable topic is one of the most common difficulties faced by novice researchers. As a result, in most of the cases, they ended up with futile endeavor after spending substantial amount of time and effort. In this paper, we present a model for RS that will recommend novice researchers a list of active topics in their chosen field of research. Two bibliometric measures – citation count and yearly publication rate are applied in the recommendation process. In the core of this proposed model is the usage of these two measures to identify active and inactive research topics. The ultimate goal of this study is to assist novice researchers in taking early decision on research topic by recommending them active research topics.

Keywords: bibliometric, citation, novice researchers, research topic, trend analysis

INTRODUCTION

Choosing a topic is one of the most challenging tasks for new researchers in producing a research paper. They face difficulty in selecting a new topic on the basis of specific criteria [1]. Most of the time, it has been observed that they select some topics either too specific to get enough information about the topics, or too broad in scope and struggle hard to narrow down the scope. Hence, they are most likely to be confronted with the lack of information or information overload. Most of the time, they turn down dead-end roads, wasting a great deal of time along the way. It is a kind of cold start problem for new researchers.

Novice researchers usually come across one of two situations when choosing a research topic. The first situation takes place when they choose one from a list of topics provided by their supervisors. In that case, they can get relief from the stress of having to decide upon a topic on their own. However, it is not always true. Sometimes the topics provided are too broad and may not fit with what researchers have in mind.

The second situation occurs when the researchers need to choose a topic based on some hints or general ideas. It requires high level of background knowledge about the particular field where they can easily limit the scope. Usually, the beginners do not have strong grasp of knowledge and expertise [2] as required to deal with the research scope. In this regard, they need to know about the current trend of research and the active topics before entering to the target research area.

In this paper, we aim to answer the question: How to design an approach to identify the "active research topics" for a novice researcher before embarking into actual research?

To answer the question, we intend to develop a recommender system to help novice researchers by suggesting active topics in their area of research interest.

With this aim we are targeting only computer science discipline. Here we use active research topics to denote those topics in computer science on which a considerable number of research papers are being published. We assume that today's hot topic which is top in the list may take position at the bottom in a couple of years and will become an inactive topic. The rate at which the papers on that topic are being published will also decrease gradually. The opposite is likely to happen to an active topic – the number of papers will be increased and average citation per paper will also be increased. Computer science is fast expanding field of research due to new challenges and opportunities [3].

The proposed recommender system will categorize a research topic as active or inactive by performing citation and trend analysis at article level. To the best of our knowledge, there is not any study, using similar method in providing recommendation on active research topics in Computer Science. The main contributions of this study will be:

(a) To formulate a framework to categorize active and inactive topics in a research field by exploiting average citation rate and understanding current trend;

(b) To implement a recommender system to help novice researchers taking decision on a research topic;

(c) To provide background information on a research topic;

(d) To assist novice researchers by providing keywords to search relevant documents on a particular topic;

The main focus of this paper is related to the first contribution. The layout of the rest of this paper is organized as follows:

We present related work about the definition of novice researchers, different approaches of RSs, current practice for selecting research, bibliometric indicators and their significance. Following that we describe the

proposed recommendation approach. Next, we provide an example from experimental view point and expected results. We conclude with a discussion of applicability and future expansion of the model.

RELATED WORK

Given the actual need of our research, we have reviewed four types of related articles. We started with the articles focusing on novice researchers, then some classical papers on recommender systems. After that we examined recent literatures on selecting research topics and bibliometric measures.

New researchers gain pragmatic research skills after having years of experience in a particular field of their interest. A novice usually holds a position at the preliminary level of skill acquisition. According to the Dreyfus model of skill acquisition, one needs to pass through five levels of proficiency. These levels are: (i) novice, (ii) competent, (iii) proficient, (iv) expert and (v) mastery [4]. Eraut [5] has distinguished clearly between novice and expert based on Dreyfus model. He defined that novice is "rigid adherence to taught rules or plans". They do not exercise "discretionary judgment". Nevertheless, experts are said to have transcends reliance on rules, guidelines, and maxims "intuitive grasp of situations based on deep, tacit understanding" and have "vision of what is possible". They use "analytical approaches" in new situations or in case of problems. Some studies [6, 7] define graduate students specifically first-year Ph.D. students as the novice researchers. Therefore, we can define that novice researchers are those who have theoretical knowledge and do not have practical expertise in doing research. They lack skill required for efficient use of research tools and techniques. They also do not have sufficient knowledge about bibliometric indicators of scholarly articles like – citation, impact factor, Eigenfactor, h-index, etc.

Recommender Systems (RSs) have been piquing researchers' interest since the mid-1990s [8]. It has become an extensive area of research, and is being applied in a variety of application areas including e-commerce, e-health and online learning. Traditionally, the common feature of a recommender system is information filtering [9, 10] through which it filters out items that are likely to be interesting to the target users.

There are three classical approaches followed in developing recommender systems, such as Content-Based (CB), Collaborative Filtering (CF), and Knowledge-Based (KB) [11]. CB utilizes item features, whereas CF concerns about user similarity, and KB uses background knowledge about similar type of items. There is another type of recommendation approach emerged by combining two or all of these three approaches, called Hybrid [11].

CF is one of the most widely used methods in recommender systems. It recommends items to people based on what other similar people have previously liked [12]. However, CB recommends when it finds appropriate matching between item features and user profile as opposed to similarity between users' preference in collaborative filtering approaches [12]. On the other hand, the objective of CB approaches is to focus on finding correlations between content of items. In general, knowledge-based systems rely on detailed knowledge about item characteristics. There are two basic types of knowledge-based recommender systems, such as constraint-based [8] and case-based systems [14].

Numerous studies have been conducted on graduate research management and provided some guidelines for selecting research topics. Here we are discussing a few of them. One study suggested that one should be well-informed of the literature on a particular field in which he or she intends to do research, before selecting a topic. Substantial knowledge base together with personal interest of the researcher plays a significant role in choosing an appropriate topic. However, the most crucial concern when initiating any research should be that the chosen topic must bridge the research gap and leads to a new contribution to the field [15]. Furthermore, a suitable topic is one of the prerequisites of a successful research. It should be selected based on some practical considerations which are put together as FINER – Feasibility, Interest, Novelty, Ethics, and Relevance [14, 15]. For instance, some interesting topics may not be feasible in the context of time, resources and setting requirements. Therefore, it is wise to know the practical limits and difficulties early on, before wasting much time and effort along impracticable paths. While identifying the topic, it should be ensured that the topic will yield novel information and qualify institutional ethical standards. Moreover, it must be relevant to the research needs [16].

Some of the studies highlighted the common practices among researchers. It has been observed that researchers tempt to select a topic without doing any groundwork. Sometimes, they only consider the immediate practical outcome of the research [17]. It is quite natural that the outcome of a high-impact study may not have immediate practical application and may definitely have long-term implication. One should pick a topic considering the potential of completing, worthiness of effort and applicability [17].

Another study [18] discussed about the characteristics and motivation of research topics. A practicable and compatible research topic must be selected for the research to be successful. The topic should fit researcher's background and academic preparation. The study shows that real-world work environment is a motivating factor for research topic. One can have a good idea from real-world context targeting some achievable objectives. Most graduate research involves seeking out ideas and materials already developed or documented by others.

Most of the studies on research topics emphasized the importance of background study and some precautious measures to guarantee the success of research. Instead of carrying out background study on multiple topics, it is imperative to carry out study on some selective topics in a hierarchical order. Thus, researchers can reduce

the time to be spent on groundwork, prior to deciding on a particular topic.

As mentioned earlier, our proposed algorithm has been developed based on bibliometric measures. Bibliometrics indicators, such as the number of publications and citations are commonly used to gauge research performance [19]. Authors usually do bibliographic citations to acknowledge prior relevant work. Nonetheless, an article can be cited for many reasons. One may cite simply because he/she wants to make the reference list longer. Or one may cite a paper to dispute results or theories that can be considered as negative citation [20]. Despite all the criticism of citation, it is still being considered as the atomic component for evaluating research performance in reproductive studies. However, citation data can be used in many ways, for a variety of purposes. The outcomes that citation analysis generates are statistically significant because of massive and systematic data collection [21]. Common usage of citation analysis is to calculate h-index, i10-index, Impact Factor, Eigen factor, university ranking and the like. However, Citation pattern varies widely depending on two parameters – (i) research field and (ii) time span. Different disciplines show different pattern of citation for the same time frame. For example, the articles in a leading biology journal might receive 20 citations on average within two years, while the articles in leading mathematics journal would receive 2 citations over the period of time [21]. Given the scope of the study, we consider only the Computer Science discipline.

In the proposed model, we introduced two-pronged approaches – weighted quantitative analysis of citation counts and trend analysis by tracking publishing output. The main task of quantitative analysis is to benchmark the average citation per paper on a topic against the field average. The citation per paper score is a good statistic since a big number of publications tends to receive a higher number of citations [21].

PROPOSED RECOMMENDATION APPROACH

In this section, we introduce the method to be used to model the proposed recommender engine reflecting the objectives envisaged in the study. The model adopted an approach based on bibliometric indicators and some statistical facts and figures. Our recommendation procedure has been divided into three phases.

Phase I is the preprocessing of data. Here, the background data can be collected from scholarly digital library such as web of science, IEEE, ACM, etc., Two types database need to be formed based on these data - one for implementing a classification system of Computer Science discipline and another one for storing bibliometric data.

Phase II is identifying active and inactive research topics based on average citation per paper. Then a candidate set of active topics will be formed with proper ranking using the proposed algorithm.

Phase III is optimization of the candidate list of topics. In this phase, the topics that have declining trend in

the number of yearly publication are deleted. Here, trend analysis is performed for each of the topics in the candidate list. The optimal set of topics will be recommended for the target users. All the three phases are explained elaborately in the following sections:

PHASE I: Preprocessing

The classification system is the basic component in our recommendation engine. It receives user preference as input and generates all the linked topics which will later be used as input for the next phase. For example, a user inputs the term "Recommender System", then the system will output all the terms connected under the term "Recommender System". The succeeding phases will process data based on these terms. We will hereafter refer these terms as topics. To build the system we adopt a poly-hierarchical ontology as implemented in ACM Computing Classification System for integrating into the search capability of the Digital Library. Different levels in the ontology hierarchy represent the relationships amongst branches, fields, sub-fields and topics of Computer Science (CS) discipline as mentioned in Figure-1.
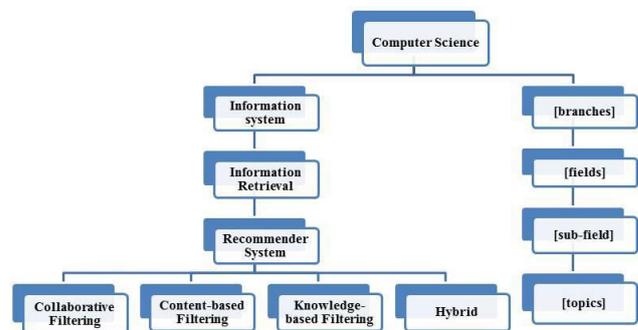

Figure-1. Hierarchical structure of computer science discipline.

A poly-hierarchical ontology has been used because it allows a topic to have relationship with multiple parent topics. For example, the Hybrid recommender system has three parent terms – Collaborative Filtering, Content-based Filtering and Knowledge-based Filtering. In CS domain, one branch may be linked with other branches. Research topics and sub-topics stay at the bottom two levels. These two levels are dynamic as new topics are emerging with the advancement of fields and sub-fields in CS. That is why it is necessary to update these two levels periodically by feeding new topics and sub-topics. The updating task needs to be performed manually by keeping track of new keywords. Most of the scholarly digital libraries have rich repository of keywords such as Springer Real-time keywords, ACM Keyword List in Computer Science, Google Keyword Tools/ Trend and the like. Perhaps, it can be done by an agent-based system by applying graph-based database model [22]. Another way is to extract key-words from article title and abstract [23] and create local dictionary of keywords.

Another important component of the recommendation engine is the bibliometric database. It

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

contains historical data for citation and publication of corresponding research topics. The database also needs to be updated regularly as citation counts are always changing. By using these data, we derive the average citation per paper for a particular topic as well as for the entire Computer Science discipline. The derived information is the key measure for both citation rate analysis and trend analysis in the recommendation process. Therefore, preprocessing is required to update the Classification System and the bibliometric database as well.

PHASE II: Citation rate analysis

Citation analysis is the major task to be performed in this phase. Applying the result from the analysis, a candidate list of research topics will be formed. The system will calculate average citation rate for each of the topics found in Phase I. It will also compute average citation rate for the entire discipline. Then it will compare each of the topics against the discipline average by performing ratio analysis. Finally, all the topics will form a rank set of candidate list according to the results of ration analysis. The underlying logic behind these calculations is governed by the following assumptions:

i. If the average citation per paper of a topic in a sub-field is greater than the average citation count of the entire CS discipline, then that topic is to be considered as an active topic.

ii. If the average citation of a topic is equal or less than the average citations of the entire CS discipline, then that topic is to be considered to be as less active or inactive topic.

The method for recommending the research topic has been explained with mathematical notations below:

Let's say, there are N number of papers $x_1, x_2,..., x_N$ published in computer science denoted by X and their corresponding citation $c(x_1), c(x_2), ... ,c(x_N)$. Take TC be the total citation count and AC be the average citation received by each paper.

Step 1: Calculate total citations for the entire CS discipline as:
$$TC = c(x_1) + c(x_2) + \cdots + c(x_N)$$

Step 2: Estimate the average citation per paper with respect to the entire CS discipline X as:
$$AC = \frac{TC}{N} = \frac{\sum_{i=1}^{N} c(x_i)}{N}$$

Step 3: Similarly, we calculate average citation for a topic. If there is M number of papers found in the database for a query on a topic, then we can define the average citation per paper with respect to the topic as:
$$AC_{topic} = \frac{TC_{topic}}{M} = \frac{\sum_{i=1}^{M} c(x_i)}{M};$$
$$Where, TC_{topic} = c(x_1) + c(x_2) + \cdots + c(x_M)$$

Step 4: Based on the user's query, we extract all the relevant topics from the classification system.

Step 5: we calculate the average citations per paper for all the relevant topics. Let's say, there are p

number of relevant topics found. Then we repeat Step 3 p times.
$$AC_{topic\,1} = \frac{TC_{topic\,1}}{M_1} = \frac{\sum_{i=1}^{M_1} c(x_i)}{M_1};$$
$$\vdots$$
$$AC_{topic\,p} = \frac{TC_{topic\,p}}{M_p} = \frac{\sum_{i=1}^{M_p} c(x_i)}{M_p};$$

Step 6: Find the ratio for each of the topics' average citation between a topic and the entire CS discipline. Repeat the process for all the topics.

$$D_1 = \frac{AC_{topic1}}{AC};$$
$$\vdots$$
$$D_n = \frac{AC_{topicp}}{AC};$$

Step 7: Arrange the topics in ascending order and form a rank set of candidate list
$$R= \{topic_1, topic_2, ..., topic_p\}$$

Thus the set R represents a candidate list of topics. This is not the final list to be considered as an effective recommendation to novice researchers. Out of these topics, some have high citation rate. However, they may have declining trend in yearly publication rate. Nonetheless, some emerging topics may have just below the entire discipline average. At this stage, we feel the need to perform trend analysis to ascertain about the real status of each topic. So that we can accept the following assumptions classify active and inactive topics:

iii. If a topic has both high citation rate and increasing trend in yearly publication, then that topic will be considered to be active topic.

iv. If a topic's citation rate is close to the discipline average and if it shows increasing trend in yearly publication, then that topic will be considered to be active topic.

After the trend analysis, the system will identify inactive topics. If any inactive topics are found, then it will prune the rank set R to bring accuracy in generating recommendation. Finally, the set R will be re-ranked to form a new set R`.

PHASE III: Research trend analysis

Trend analysis will be performed in this phase on each of the research topics in the candidate list created in the previous phase. The present trend of research on each of the topics will help to identify whether a topic is active or not. The main purpose of the analysis is to prune the set R by removing inactive topics, if found according to the assumption III and assumption IV. As we assume that some of the topics in set R are not in active research in that they are supposed to be saturated and the researchers find them hard to explore further. Therefore, it is likely to impact a declining trend on the number of papers being published per year. We consider these inactive topics are unsuitable for novice researchers and hence, they need to be removed from the recommendation list.

www.arpnjournals.com

To carry out the trend analysis a few possible measures could be used, such as slope or gradient (rise over run ratio), Spearman's rank correlation coefficient and Pearson Correlation Coefficient (PCC). The slope is calculated using two points passing through a trend line which means that only two years of data could be taken into account. The difficulty with this method is to select appropriate data. Besides, this formula cannot be applied for a vertical line, parallel to the y axis where the slope can be taken as infinite. Spearman's rank correlation coefficient and Pearson correlation coefficient have similar properties. They also have shortcomings similar to the slope formula. They fail as a horizontal line parallel to x axis. Since, our concern is only negative correlation, we can ignore this limitation. Although a plenty of statistical tools and techniques are available to perform trend analysis, here, we use sample Pearson correlation coefficient (PCC) for simplicity and widespread usage. It shows relationship between the number of articles and years of publication.

For analyzing the trend, we form a dataset from the last five years of data. We assume that the lifespan of any topic in CS is greater than five years which means a topic would be active for research at least for five years.

We create one dataset $\{x_1, ..., x_n\}$ having n values for publication years and another set of data $\{y_1, ..., y_n\}$ holding n values for corresponding number of articles published, then we can obtain the following formula in line with the PCC :

$$r = r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2}\sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

The value of the coefficient ranges from +1 to -1 inclusive, where 1 is total positive correlation, 0 is no correlation, and -1 is total negative correlation. If the dataset of a topic shows negative correlation, then that topic has declining trend as shown in Figure-2. If the coefficient value for a given topic shows positive (r > 0) correlation, then the publication trend of that topic is upward, meaning that yearly publication on that topic is increasing. That is the positive notion for the topic to be considered for recommendation. After calculating the PCC, r the system will perform pruning if required. Then it will generate rank score by multiply the ratio value, D_i with the PCC value, r_i for each of the topics. The step by step procedures are mentioned below:
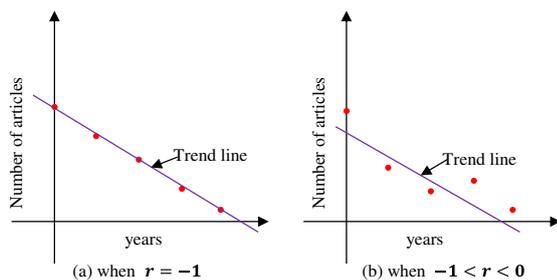

Figure-2. Negative correlation for declining trend

Step 1: Calculate the coefficient r for each the topics applying the PCC equation as mentioned before.

$$r_1 \rightarrow for\ topic1$$
$$\vdots$$
$$r_p \rightarrow for\ topicp$$

Step 2: Identify the topics in the range of $-1 \leq r < 0$ delete them from the set R.

Step 3: Multiply PCC value with the ratio found in phase II to calculate the rank score for the remaining topics in set R.

$$D_1 \times r_1 = N_1$$
$$\vdots$$
$$D_p \times r_p = N_p$$

Step 4: Finally, we can obtain the new rank set $R`$ of active topics and to be presented to the users as their personalized recommendation list. The topics will be represented in descending order. The high scored topic will become top in the list.
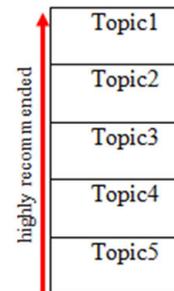

Figure-3. List of recommended active topics

After pruning, if the rank set becomes a null set i.e., R = ∅, then the system will not recommend any topic and the recommendation process will stop that point. That means no suitable topic is found to be recommended.

The researchers can use these topics as keywords to search relevant articles in the scholarly digital libraries like - IEEE Explore, ACM, Web of Science, and Springer Link. Additionally, the recommender will display time span of each of the topics in the recommended list, so that the novice researchers may know how long the topic exists as research interest.

SIMULATION OF EXAMPLE

We are currently in the process of implementing the recommender system within the proposed model. The model has been illustrated here with simple example for better understanding. The illustration is based on some dummy datasets. Basically, we discuss all the three Phases of the model.

Let's say, a novice researcher from Computer Science background majoring in Information Systems intends to do his/her thesis on recommender systems. He enters the keyword "Recommender System" into the RS. Then the system will perform all the three Phases in our proposed model sequentially.

In Phase I, the system will retrieve all the related topics under the recommender systems as well as

www.arpnjournals.com

associated datasets from the database. For example, there are three relevant topics found, such as Collaborative Filtering (CF), Content-Based (CB) and Knowledge-Based (KB) and each of the topics has five different articles. The Figure-4 shows citation network map for these topics. The circles (nodes) represent articles and the edges (arrows) indicate citations from one article to another. The articles on CF, CB and KB are denoted by letters A to E; V to Z and P to T respectively. Figure (a) shows highly connected citation network as all the nodes are connected in a mesh pattern. However, Figure (b) is moderately connected and (c) is weakly connected for less citation counts.
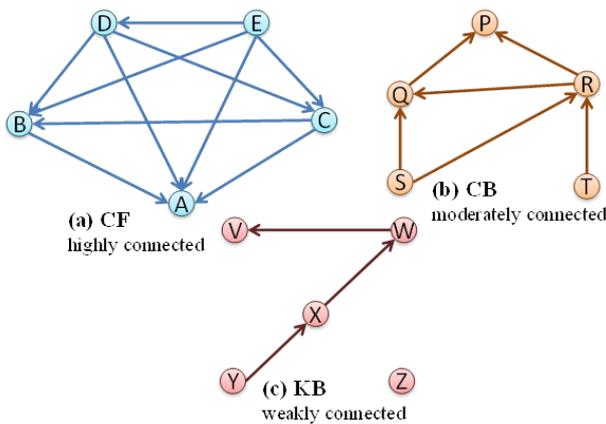

Figure-4. Citation network map.

The following table shows distribution of citations received by the articles on different topics.

Table-1. Citations received by articles

CF		CB		KB	
Article	Citations	Article	Citations	Article	Citations
A	4	P	2	V	1
B	3	Q	2	W	1
C	2	R	2	X	1
D	1	S	0	Y	0
E	0	T	0	Z	0

Then the system will proceed to Phase II and perform citation analysis based on the above datasets. Here, we show average citation rate calculation for each of the articles on the three topics with clusters of articles connected with each other in different patterns. For the citation analysis the system will follow the step-by-step procedures as mentioned in the previous section.

The following table shows calculation of average citations i.e. number of citations received per article on individual topic and the field (number of citations received by an article on average with respect to the entire Computer Science discipline).

Table-2. Average citations of different topics and entire CS discipline.

	Total No. of Articles	Total Citations	Average Citations	Ratio, D
CF	5	10	2	2:2=1.0
CB	5	6	1.2	1.2:2=0.6
KB	5	3	0.6	0.6:2=0.3
CS	200	400	2	

The calculation shows that the average citation of CF is higher than the entire discipline average and CB and KB have got fewer citations per article. As such, on the basis of the stated assumptions, we can say that CF is the most active among others. We form a candidate list of topics in the form of a rank set R.

Now, we need to consider the current trend of publications for each of the topics as described in Phase III. Table-3 shows yearly distribution of articles for the last five years. It is obvious that KB has got a declining trend, since it has negative correlation and the value of r is approaching -1. However, CF and CB show positive correlation, and thus has increasing trend. Therefore, KB can be removed from the candidate list because of lower citation rate and declining trend.

Table-3. Articles published in the last five years.

	Y1	Y2	Y3	Y4	Y5	r (PCC)	Score, N = r X D
CF	0	1	1	1	2	0.894	0.894
CB	1	0	1	2	1	0.447	0.2682
KB	2	1	1	1	0	-0.894	(pruned)

Consequently, we finalized the new list of topics for recommendation where CF is in the first rank and CB is in the second rank in the list. Now the novice researcher can use the name of these two topics as keywords for searching relevant articles.

DISCUSSION AND CONCLUSIONS

In this study, we explore a new approach of recommendation beyond the traditional information filtering methods for developing the proposed model. Nowadays, RSs are being developed hinged upon more complex algorithms. They are catering array of extensive and robust services to various types of users. RSs are becoming very popular among academic researchers. They find them conducive to expedite their research works. Academic Digital libraries are integrating RS in their search capabilities. For example, when anyone downloads a research paper from ACM digital library, he will be soon recommended relevant papers. Digital libraries have high potentials to provide more personalized services to different group of researchers. Both novice and experts are entitled to same service. Additionally, the optimal utilization of these libraries is heavily depending on users' experience in doing research. Our study reveals that it remains imperative to integrate recommender systems

with the existing digital libraries for providing personalized recommendation to individual users.

Recommendation of research topics will solve a real problem in practical setting for the novice researchers. Good recommendation result will assist them to take decision on a research topic. As long as, we implement the recommender system based on the proposed model, we will perform experiment with large scale real-world datasets to prove the capability and accuracy of the system. We expect that the desired result could contribute a lot for the academic research community. In the long run, the model will have a high-impact on the academic research community and the society as well. In the future, we plan to enhance the recommendation mechanism by adding more personalized features like context-awareness, generating recommendation on research materials.

ACKNOWLEDGEMENTS

REFERENCES

[1] S. Taskeen, A. Shehzadi, T. Khan and N. Saleem. 2014. Difficulties faced by novice researchers: A study of universities in Pakistan. International Journal of Art and Literature. Vol. 1, No. 1, pp. 1-4.

[2] M. A. Ismail and S. A. Kareem. 2011. Identifying how novice researchers search, locate, choose and use web resources at the early stage of research. Malaysian Journal of Library and Information Science. Vol. 16, No. 3, pp. 67-85.

[3] A. Hoonlor, B. K. Szymanski and M. J Zaki. 2013. Trends in Computer Science Research. Communications of the ACM. Vol. 56, No. 10, pp. 74-83.

[4] S. E. Dreyfus and H. L. Dreyfus. 1980. A five-stage model of the mental activities involved in directed skill acquisition. California Univ Berkeley Operations Research Center.

[5] M. Eraut. 1994. Developing professional knowledge and competence. Psychology Press.

[6] T. J. Ellis and Y. Levy. 2010. A guide for novice researchers: Design and development research methods. In: Proceedings of Informing Science and IT Education Conference. InSITE.

[7] M. D. Ekstrand , P. Kannan, J. A. Stemper, J. T. Butler, J. A. Konstan and J. T. Riedl. 2010. Automatically building research reading lists. In: Proceedings of the fourth ACM conference on Recommender systems. pp. 159-166.

[8] A. Felfernig and R. Burke. 2008. Constraint-based recommender systems: technologies and research issues. In Proceedings of the 10th international conference on Electronic commerce. ACM. pp. 3.

[9] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon and J. Riedl. 1997. GroupLens: applying collaborative filtering to Usenet news. Communications of the ACM. Vol. 40, No. 3, pp. 77-87.

[10] M. J. Pazzani. 1999. A framework for collaborative, content-based and demographic filtering. Artificial Intelligence Review. Vol. 13, No. 5-6, pp. 393-408.

[11] G. Adomavicius and A. Tuzhilin. 2005. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. Knowledge and Data Engineering, IEEE Transactions. Vol. 17, No. 6, pp. 734-749.

[12] R. Torres, S. M. McNee, M. Abel, J. A. Konstan, and J. Riedl. 2004. Enhancing digital libraries with TechLens+. In: Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries. ACM. pp. 228-236.

[13] L. Liu. 2013. The implication of context and criteria information in recommender systems as applied to the service domain. Doctoral Dissertation. University of Manchester.

[14] R. Burke. 2000. Knowledge-based Recommender Systems. Encyclopedia of Library and Information Science: Marcel Dekker. Vol. 69.

[15] J. A. Bhatti, U. Akhtar, S. A. Raza and K. Ejaz. 2012. Selecting a research topic. J Pak Med Assoc. Vol. 62, pp. 184-6.

[16] S. B. Hulley, S. R. Cummings, W. S. Browner, D. G. Grady and T. B. Newman. 2001. Conceiving the research question. Designing Clinical Research. 2nd ed. Philadelphia, PA: Lippincott Williams and Wilkins, pp. 335.

[17] J. Bell. 2014. Doing Your Research Project: A guide for first-time researchers. McGraw-Hill Education. pp. 28-31.

[18] A. B. Badiru. 1996. Selecting a research topic. In: Project Management for Research. Springer Netherlands. pp. 37-53.

[19] B. Cronin and C. R. Sugimoto. 2014. Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact. MIT Press.

[20] M. H. Johnson, J. Cohen and G. Grudzinskas. 2012. The uses and abuses of bibliometrics. Reproductive biomedicine online. Vol. 24, No. 5, pp. 485-486.

[21] T. Reuters. 2008. Using bibliometrics: a guide to evaluating research performance with citation data. Philadelphia, Thomson Reuters.

[22] S. Jeon, Y. Khosiawan and B. Hong. 2013. Making a Graph Database from Unstructured Text. In: Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference. pp. 981-988.

[23] R. Bhowmik. 2008. Keyword extraction from abstracts and titles. In: Southeastcon, 2008. IEEE. pp. 610-617.

[24] J. L. Herlocker, J. A. Konstan and J. Riedl. 2000. Explaining collaborative filtering recommendations. In: Proceedings of the 2000 ACM conference on Computer supported cooperative work. pp. 241-250.

[25] N. De Bellis. 2009. Bibliometrics and citation analysis: from the science citation index to cybermetrics. Scarecrow Press.