2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

A Paper Recommendation System Based on User's Research Interests

Betül Bulut*, Buket Kaya+, Reda Alhajj and Mehmet Kaya*
*Department of Computer Engineering, Fırat University, Elazığ, Turkey
+Department of Electronics and Automation, Fırat University, Elazığ, Turkey
Department of Computer Science, University of Calgary, Calgary, AB, Canada
betulblt93@gmail.com, bkaya@firat.edu.tr, alhajj@ucalgary.ca, kaya@firat.edu.tr

***Abstract*— Researchers and scientists read articles to improve their studies. Researchers spend too much time and struggle to find the suitable article they are looking for. The purpose of article recommendation system is to reduce the time they spend and present to them the related articles they are not aware of. Classic article recommendation systems do not consider the user's information, they show the same results in the same sort for each researcher. In this study, an article recommendation system that takes into consideration the researcher's work field and the publisher's previous articles is presented. One of the most important innovations of this work is the use of TF-IDF and Cosinus similarity to make article recommendations taking user's past articles into consideration. As a result of the work, the users have been recommended articles, and the method we present has proved more successful results compared to equivalent methods according to f-Measure criterion.**

Keywords— Paper recommendation system, User's research interest, TF-IDF, Cosinus similarity

I. Introduction

Academicians, researchers use existing academic knowledge by examining the previous studies in their fields while conducting scientific studies. Researchers show great interest in digital resources and libraries, as their access to printed articles such as journals, articles, papers are faster and easier in the digital environment. The number of academic studies is increasing every year because of the improvements in technology, the increase of the number of universities every year and the rise of interest among people in knowledge and science. With the development of technology, the interest in digital resources for access to information has increased over time. Digital libraries such as Google Scholar, IEEE Digital Library, and ScienceDirect are the most commonly used digital libraries by researchers. With the increasing number of digital libraries and support of universities to utilize digital resources, access to knowledge from digital sources is increasing every day. A problem brought along with these developments that are positive for science is that researchers are having struggle to find the articles they seek from a large number of articles. Researchers make a lot of effort and lose a lot of time while finding the most related article to their field of study. Article recommendation systems have been developed to reduce the time lost by the researcher.

The aim of article recommendation systems is to try to find the most related article to the field from many articles. One of the most important user-oriented purposes is to reduce the time that researchers lose while searching for a article. Past studies on article recommendation generally compare articles according to their content. And, recommend similar articles for each other. The disadvantage of these previous studies is that they recommend the same articles to all researchers. This issue can be explained in more detail with an example. While listing the related studies that provide a customized recommendation on finding a referee and use link-based and random forest tree-based articles, RFT articles, link-based articles, and articles on finding a referee are listed. On the other hand, what the person that reads the study looking for is finding a referee and it would be a more successful recommendation method to show the link-based study that aims to appoint a referee (if there are any previous articles with a different method) instead of showing excessive amount of RFT articles. Articles of the researcher and field of research in previous studies are insignificant and all researchers are recommended the same things regardless of who they are. In this study, however, the proposed method is based on the characteristics of the researcher. The difference of our study from other studies is that each user is recommended the most related article to the field of work considering their characteristics.

The proposed method in this study has been tested in the IEEE Digital Library data set with the participation of volunteer researchers and successful results were obtained compared to previous studies.

The rest of the paper is organized as follows: Section II describes the main approaches to paper recommendation. In Section III, the system architecture is illustrated. In Section IV, data gathered through the study is explained, experimental results of the method are reported on real datasets and, the test results of study are evaluated in general terms. Section V concludes this paper.

II. Related Works

Expert systems are the generic names given to systems that make computer analysis of which resource makes a task better by using an information repository. Our study can be

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain
978-1-5386-6051-5/18/$31.00 © 2018 IEEE

considered as an expert system study as the problem that it was based on can be generalized as finding a specialist of a topic or finding the person most suitable for that topic.

The recommendation systems of academic articles have become more important with the internet becoming widespread, the increase in the number of academic articles and the popularization of online academic search and reading.

Although there have been many studies on labeling, the number of studies dealing with academic articles is relatively low. One of the outstanding studies on labeling and academic article recommendation method is the one proposed by Choochaiwattana [1] in 2010, and the other is Bahulikar's [2] work in 2017.

Contrary to the methods known by Ravi et al. [3], a model has been proposed that analyzes text using repetitive neural networks and recommends academic articles accordingly. In addition, ontology-based academic article recommendation studies have also been conducted [4][5].

There are also hybrid studies within academic article recommendation studies. While Lee et al. [6] proposed a new model combining 2 different approaches, content-based and graph-based; Bancu et al. [7] proposed one that combines content-based and collaborative filtering methods. Apart from these, there are also approaches that combine two sets of data and use it as an academic article recommendation method, as was done by Zhao et al. [8].

West et al. [9] indicated the references with nodes. They sorted the references with the Eigenfactor algorithm and clustered the knots using MapEquation. They found the most important node in the different steps of the hierarchy of these clusters. They used an algorithm similar to PageRank to find this node. They recommended the article of the most important node they found.

Xia et al. [10] conducted a study that recommended articles with same authors for articles containing more than one author.

Zhou et al. [11] tried to list the confidence index of each academic article by suggesting a paper rank algorithm similar to the page rank algorithm, different from all studies and constructed the recommendation system on the data obtained from this.

Chen et al. [12] have developed a recommendation system based on references from academic articles. The main reason why the study of Chen et al. became prominent is that it scores the references by looking at information such as from how many different sources and how many different scholars it comes from instead of the number of references.

Finding similarity of different articles with the keyword can be called a label-based similarity computation algorithm. Although researchers have suggested many methods in past studies, the most simple and easily integrated TF-IDF method was used for this method. There are many other studies [13][14][15][16][17] that use this method for the same purpose with us.

Many social network analysis methods (such as degree distributions, diameter, and clustering coefficient[18]–[21]) are intended to find user-based impact values at the macro-level. The majority of studies on social impact are also seeking a qualitative value[22], [23]. One of them is the Topical Affinity Propagation (TAP) method proposed by Tang et al. in their articles [24] named "Social Influence Analysis in Large-scale Networks". Tang et al. looks for the most influential individual in a particular subject or group and provide a model that allows a quick search in dynamic social networks for the most knowledgeable, in other words, the most influential expert on community.

There are also similar recommendation systems [25][26][27] using bipartite graphs. Ohta et al. proposed an academic article suggestion model using bipartite graphs. [28]

Watanabe et al. also collected the information like the number of times an academic article was read except for it's metadata and added them to the system of recommendation. [29] Similar to our study, the academic article recommendation model of Cui et al. [30] that also deals with the relationship among the users is one of few studies on this topic.

One of the most similar studies to our study belongs to Hocheol Jeon and Changho Jeon [31] . The authors have proposed a user profile-based model named PRPRS (Personalized Research Paper Recommendation System) in these studies. In this model, an algorithm was designed to determine the key word and to sort results by the subject and the frequency of the keyword for each user profile. They used cosine equivalence to calculate the relevancy of the articles to the subject.

Xue et al. and Chen et al. suggested one of the most related studies to our study because they use both online databases and analyze the user's published studies to determine the user's interest. [32] [33] The most original feature of the study of Xue et al. is that it implements the algorithm in a real online academic database. Chen and his colleagues have suggested that researchers should work with the vision that they can specialize in more than one field because of the fact that there is not such a limitation for them to specialize in just one field.

III. SYSTEM ARTHIITECTURE

In this study, the most frequently used digital databases were utilized by the researchers in Computer Sciences, Electrical and Electronics Engineering and Mechanical Engineering to create the data set. Academic databases used as data set are:

1. ACM Digital Library (Association for Computing Machinery)
2. IEEE Xplore Digital Library
3. The DBLP Computer Science Bibliography

ACM Digital Library was used for the metadata such as title, author name, keyword, article date and abstract. IEEE Xplore Digital Library was used for the metadata such as title, author name, keyword, article year and abstract. The DBLP Computer Science Bibliography digital library was used for the metadata such as article year, author name, article title. IEEE Xplore Digital Library and ACM Digital Library data has never been stored in our system. The public information of the selected article, the articles of the user and articles containing the same key words as the selected article were received like a user via http and not kept after the final recommendation when we make user-based advice for an article.

In order for metadata not to be confused, article year, article title and abstract metadata were placed in separate sections in our temporary data repository according to the articles they belonged. As author name is the most important point in our article recommendation system, it was shown both in the metadata table and a separate table.

According to user profile information, the article recommendation system consists of four basic structures. The data from the datasets goes to the dataset interpreter. The dataset interpreter separates the user information data from the ACM Digital Library, the IEEE Xplore Digital Library, and The DBLP Computer Science Bibliography datasets. It parses the metadata of each of the articles. It also parses and rearranges the metadata of the articles that is recommended as a result of the searched keyword. The parsed data are kept in Datasep Interpreter. Figure 3.1 shows the structure of the system architecture.

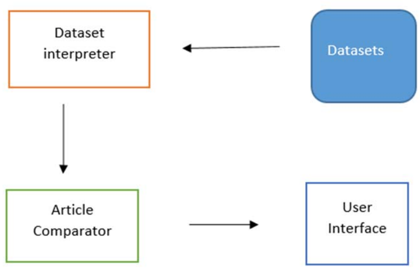

Figure 3.1: System Architecture

In the system that recommends article according to the user profile information, we can basically divide the data from the datasets into two parts:

1. A set of metadata in the researcher's profile including year, title, abstract, keywords of each of the articles that they have searched so far.
2. A set of metadata that include year, title, abstract, keywords of the articles related to the researcher's field.

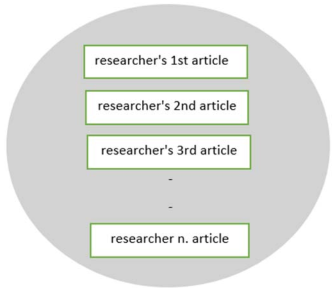

Figure 3.2: Creating a profile from author's articles

In Figure 3.2, the metadata of each of the articles previously written by the researcher are combined. In this way the user profile is created.

$$info_n = title_n + year_n + authors_n + abstract_n + keywords_n \quad \text{Formula (1)}$$

Formula (1) represents the sum of the metadata information of article number n. The formula that contains the metadata information of all the author's articles is as follows:

$$profile\_info = \sum_{k=0}^{n} info_n \quad \text{Formula (2)}$$

Formula (2) is a string that is the sum of metadata information for all articles of the author. In other words, it is the profile information of the user. The profile_info value is the profile of the author whose article is to be recommended. $info_n$ value is the metadata information calculated for each article of the author expressed in formula (1).

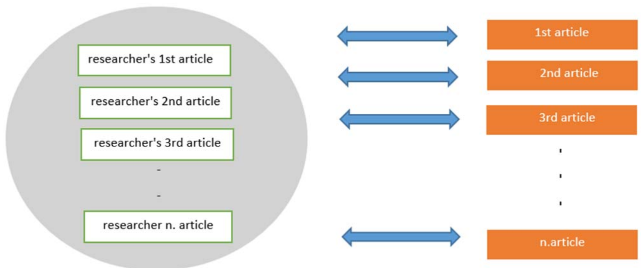

Figure 3.3: Author Profile and Article Comparison

The user's profile is created by combining the information of the n articles written by the researcher. The created profile and each of the articles are compared one by one. Figure 3.3 represents the comparison processes.

The processed data of the author's information and the results from the keywords are compared to the article. It compares the articles using comparative cosine similarity and TF-IDF methods. The article comparator, compares the author information ($profile\_info$) and each of the articles ($info_n$). The comparison result score of each article is determined. According to the comparison results from the comparator, the most related ten articles are presented to the user through User Interface.

IV. Experimental results

There are only a few studies that make article recommendations by paying attention to the information of the articles in the user's profile. In this study, what we tried is to determine the success criterion of the proposed method by using a few of the previously proposed methods.

In the first method, random email addresses and names of researchers were determined as experimental groups. The system sent the links and titles of ten articles that may interest the researchers based on their articles. The number of emails sent is ten, because it was anticipated that the researcher may not click on all links when a lot of articles are sent.

Table 1: Experimental Results

Number of emails sent to the researcher	Number of emails read by the researcher	Number of articles the researcher has clicked on	Number of articles by the researcher
10	7	6	52
10	3	1	18
10	8	5	26
10	5	4	15
10	4	2	20
10	3	2	15
10	6	3	31
10	1	1	19
10	7	4	37
10	3	1	17

Table 1 shows the number of mails taken by the researcher, the number of mails read by the researcher, and the number of the researchers' own articles. The researchers read 47% of the mails and read 29% of the articles.

Another method is to conduct a survey with a research group. Name and surname of the researchers who will participate in the survey were taken. Recommended articles created from our system were taken from the system. Our system has recommended 20 articles specific to each researcher. Each researcher who participated in the survey was presented the articles compiled for themselves in the form of a survey. In this two-choice survey, the researchers were asked to mark yes if the article interest them and mark no if not.

Table 2: Numbers of researchers and articles used in the experiment

Number of Researchers participating in the survey	Number of Articles Presented in the Survey	Number of Articles Marked As "Yes" by Researchers in the Survey	Number of Articles Marked As "No" by Researchers in the Survey
30	600	426	174

The researchers have marked the articles that interest and do not interest them in the survey. According to the results, 426 of the articles submitted to the researcher were marked as yes, 174 as no. 71% of the articles attracted the interest of the researchers, 29% did not.

V. Conclusions

The most popular databases in engineering field, ACM Digital Library, IEEE Xplore Digital Library and The DBLP Computer Science Bibliography, were used in the study. In contrast to classical article recommendation systems, the study recommends articles that researchers may be interested according to articles that the researcher has written. The results presented by a system that recommends articles by comparing user's articles and the articles that is shown as a result of search is promising.

VI. Acknowledgement

This study was partially supported Firat University, Scientific Research Projects Office under Grant No MF.18.31.

References

[1] W. Choochaiwattana, “Usage of tagging for research paper recommendation,” in *ICACTE 2010 - 2010 3rd International Conference on Advanced Computer Theory and Engineering, Proceedings*, 2010, vol. 2.

[2] S. Bahulikar, “Analyzing recommender systems and applying a location based approach using tagging,” in *2017 2nd International Conference for Convergence in Technology (I2CT)*, 2017.

[3] R. K. Mysore, M. Junichiro, and S. Ichiro, “Cross-Domain Academic Paper Recommendation by Semantic Linkage Approach Using Text Analysis and Recurrent Neural Networks,” in *Management of Engineering and Technology (PICMET), 2017 Portland International Conference on*, 2017.

[4] S. S. Weng and H. L. Chang, “Using ontology network analysis for research document recommendation,” *Expert Syst. Appl.*, vol. 34, no. 3, pp. 1857–1869, 2008.

[5] K. V. Neethukrishnan and K. P. Swaraj, “Ontology based research paper recommendation using personal ontology similarity method,” in *Electrical, Computer and Communication Technologies (ICECCT), 2017 Second International Conference on*, 2017.

[6] Y. C. Lee *et al.*, “Recommendation of research papers in DBpia:

A Hybrid approach exploiting content and collaborative data," in *2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings*, 2017, pp. 2966–2971.

[7] C. Bancu, M. Dagadita, M. Dascalu, C. Dobre, S. Trausan-Matu, and A. M. Florea, "ARSYS-article recommender system," in *Proceedings - 14th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2012*, 2012, pp. 349–355.

[8] W. Zhao, R. Wu, and H. Liu, "Paper recommendation based on the knowledge gap between a researcher's background knowledge and research target," *Inf. Process. Manag.*, vol. 52, no. 5, pp. 976–988, 2016.

[9] J. D. West, I. Wesley-Smith, and C. T. Bergstrom, "A Recommendation System Based on Hierarchical Clustering of an Article-Level Citation Network," *IEEE Trans. Big Data*, vol. 2, no. 2, pp. 113–123, 2016.

[10] F. Xia, H. Liu, I. Lee, and L. Cao, "Scientific Article Recommendation: Exploiting Common Author Relations and Historical Preferences," *IEEE Trans. Big Data*, vol. 2, no. 2, pp. 101–112, 2016.

[11] Q. Zhou, X. Chen, and C. Chen, "Authoritative scholarly paper recommendation based on paper communities," in *Proceedings - 17th IEEE International Conference on Computational Science and Engineering, CSE 2014, Jointly with 13th IEEE International Conference on Ubiquitous Computing and Communications, IUCC 2014, 13th International Symposium on Pervasive Systems,* , 2015, pp. 1536–1540.

[12] C. H. Chen, S. D. Mayanglambam, F. Y. Hsu, C. Y. Lu, H. M. Lee, and J. M. Ho, "Novelty paper recommendation using citation authority diffusion," in *Proceedings - 2011 Conference on Technologies and Applications of Artificial Intelligence, TAAI 2011*, 2011, pp. 126–131.

[13] W. Zhang, T. Yoshida, and X. Tang, "A comparative study of TF*IDF, LSI and multi-words for text classification," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2758–2765, 2011.

[14] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," *Proc. first Instr. Conf. Mach. Learn.*, pp. 1–4, 2003.

[15] T. Kenter and M. de Rijke, "Short Text Similarity with Word Embeddings Categories and Subject Descriptors," *Proc. 24th ACM Int. Conf. Inf. Knowl. Manag. (CIKM 2015)*, pp. 1411–1420, 2015.

[16] S. Albitar, S. Fournier, and B. Espinasse, "An Effective TF/IDF-Based Text-to-Text Semantic Similarity Measure for Text Classification," *Web Inf. Syst. Eng. - Wise 2014, Pt I*, vol. 8786, pp. 105–114, 2014.

[17] C.-H. HUANG, J. YIN, and F. HOU, "A text similarity measurement combining word semantic information with TF-IDF method," in *Chinese Journal of Computers*, 2011.

[18] S. H. Strogatz, "Exploring complex networks.," *Nature*, vol. 410, no. 6825, pp. 268–276, 2001.

[19] M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Review*, vol. 45, no. 2. pp. 167–256, 2003.

[20] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On Power-Law Relationships of the Internet Topology," in *In SIGCOMM*, 1999, pp. 251–262.

[21] R. Albert and A. L. Barabasi, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 47–97, 2002.

[22] D. Krackhardt, "The strength of strong ties: The importance of Philos in organizations," *Networks Organ. Struct. Form, Action*, vol. 216, pp. 216–239, 1992.

[23] M. S. Granovetter, "The Strength of Weak Ties," *American Journal of Sociology*, vol. 78, no. 6. pp. 1360–1380, 1973.

[24] J. Tang, J. Sun, C. Wang, and Z. Yang, "Social influence analysis in large-scale networks," *Proc. SIGKDD'09*, pp. 807–816, 2009.

[25] A. A. Müngen and M. Kaya, "A Novel Method for Event Recommendation in Meetup," in *2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2017.

[26] E. Gündoğan, B. Kaya, and M. Kaya, "Prediction of Symptom-Disease Links in Online Helath Forums," in *IEEE/ACM International Conference on Advances in Social Networks and Mining (ASONAM, 2017)*, 2017.

[27] S. Aslan and M. Kaya, "Link prediction methods in bipartite networks," in *Computer Science and Engineering (UBMK), 2017 International Conference on*, 2017.

[28] M. Ohta, T. Hachiki, and A. Takasu, "Related paper recommendation to support online-browsing of research papers," in *4th International Conference on the Applications of Digital Information and Web Technologies, ICADIWT 2011*, 2011, pp. 130–136.

[29] S. Watanabe, T. Ito, T. Ozono, and T. Shintani, "A paper recommendation mechanism for the research support system Papits," in *Proceedings - International Workshop on Data Engineering Issues in E-Commerce, DEEC 2005*, 2005, vol. 2005, pp. 71–80.

[30] T. Cui, X. Tang, and Q. Zeng, "User network construction within online paper recommendation systems," in *Proceedings - 2010 IEEE 2nd Symposium on Web Society, SWS 2010*, 2010, pp. 361–366.

[31] H. Jeon and C. Jeon, "UserProfile-Based Personalized Research Paper Recommendation System," *Comput. Netw. Technol. (ICCNT), 2012 8th Int. Conf.*, pp. 134–138, 2012.

[32] H. Xue, J. Guo, Y. Lan, and L. Cao, "Personalized paper recommendation in online social scholar system," in *ASONAM 2014 - Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2014, pp. 612–619.

[33] J. Chen and Z. Ban, "Literature recommendation by researchers' publication analysis," in *Information and Automation (ICIA), 2016 IEEE International Conference on*, 2016.