A Hybrid Framework for Article Recommendation using Citation Network and Author based Search Pattern

Vruttika V. Mogal

Department of Computer Engineering
Matoshri College of Engineering And Research Center, Nasik.

Abstract: With the rapid growth in the information technology and the growing amount of the scholarly data, the recommendation techniques are becoming more popular. The main aim of the recommender system is to automatically suggest the users with the items of the potential interest. In the recent years they are being employed in the digital libraries for the researcher to get the required article. But this search tool gives the list of the relevant articles based in the keyword based queries. If any researchers have the same set of keywords then the obtained result would also be same. It is not feasible to describe the searchers need depending on the several limited keywords. In the proposed system, a novel recommendation method is being introduced in which it incorporates common author relations between the articles to help generate better recommendations for relevant target researcher. In this system the citation network is being constructed in order to give the relevant output to the researcher. Also the hybrid framework is being considered for the implementation. In order to determine the relevant researchers with author-based search patterns, the ratio of pairwise articles with common author relations and the ratio of the most frequently appeared author.

Keywords: Citation Network, Random walk with Restart, Author based Search Pattern, Hybrid Framework.

1. Introduction

With the explosive growth in the amount of data, it could be observed that tremendous amount of information could be stored and accessible to anyone around the world. Last decade have witnessed the tremendous increase in the digital information with the evolution of the Internet. On the Internet, where the number of choices for the user is very Large, there is need to prioritize, filter and efficiently deliver the needed and the relevant information to reduce the problem of the information overload, which has created a problem to many users. This problem is solved by the system know as the recommendation system. Recommendation system solves this problem by searching through large volume of dynamically generated information to provide users with relevant and proper content. These are the systems that deal with the problem of information overload by filtering vital information fragment out of large amount of dynamically generated information according to users preferences, interest or observed behaviour about content. They the ability to predict that whether the user would prefer a content or not based on the users profile .In scientific libraries, recommender system support users by allowing them to move beyond the catlog searches. For example, a researcher in academia needs to find articles of interest to read for generating a research idea or citing an article related to the article he is writing, an author needs to submit his manuscript to a certain journal of which the topic is relevant to the manuscript, an editor needs to assign a manuscript to a reviewer who is an expert in the domain which the manuscript belongs to, or a researcher in a domain needs to collaborative with another researcher in another domain. The result of these activities would involve a large number of articles, reviewers and researchers. As a result, it would become quite difficult for the researchers to track the relevant articles, journals, researchers and reviewers for the required purposes. Referential search is the most important task, which is being daily performed by the thousands of researchers around the world. This task would consists of finding and analyzing many scientific papers in order to find the work that could be helpful to develop a specific research.

A traditional approach which is usually used by the researchers is to follow the list of references in papers which they already posses. But this is a quite limited approach since there is no guarantee of full coverage and important references may not be cited, besides the fact that only papers published before the current paper can be traced. In order to find the articles of interest for target researchers in the context of the big data the proposed system focuses

on a novel recommendation system, which incorporates common author relations between the articles to help generate better recommendations for relevant target researchers. The proposed system present two features including the ratio of pairwise articles with common author relations and the ratio of the most frequently appeared author, to help determine relevant researchers with author-based search patterns. It also proposes a new guideline for patent mining using author and articles side attribute.

2. Literature Survey

M. Gori and A. Pucci [1] proposed a system a random-walk based scoring algorithm, which can be used to recommend papers according to a small set of user selected relevant article. This particular recommendation system does the work of the article recommendation. It does the work of suggesting the relevant articles to a researcher or an article for the purpose of reading or citation. They built a citation relation graph and employed a random walk algorithm to compute ranking scores for each possible citation. J. Sun, J. Ma, Z. Liu, and Y. Miao[2] proposed a novel approach to recommend scientific articles to researchers by leveraging content and connections. In this approach, they first analyzed the semantic content of the article by keyword similarity calculation and then extract online users' connections to support article voting and finally employ a two-stage recommendation process to suggest relevant articles. There system had three main contributions. First, the relevance of articles is analyzed from a novel and comprehensive perspective. Semantic content and online connections are deeply mined and aggregated to improve recommendation quality. Secondly, proposed semantic social aggregation approach works in terms of improving the recommendation performance. Thirdly, a recommender system is developed to assist article recommendation in a social network environment. This system also had some limitations in it. They compute the content similarity between articles to find articles which are similar to the target's articles of interest, or compute the similarity between the target's profile and an new article's content to find matches. Also the factors of time and quality are not being considered. The system does not include of complex data fusion techniques.

E. Medvet, A. Bartoli, and G. Piccinin,[4] propose a venue recommendation system which requires only title and abstract, differently from previous works which require full-text and reference list: hence, the system can be used even in the early stages of the authoring process and greatly simplifies the building and maintenance of the knowledge base necessary for generating meaningful recommendations. They propose a topic matching procedure that can form the basis of a recommendation system for scientific paper submission. The best performing existing proposals require the full-text of the paper to be examined, including the list of references and of authors, while their approach requires only title and abstract. Their analysis suggests that recommendations built upon long established n-gram based text classification methods may be highly effective, while recommendations based on generative and probabilistic topic models may lead to unsatisfactory results. The proposed system is feasible also from a performance point of view, as the learning phase requires a few minutes while a recommendation may be generated in a couple of seconds. This particular system was based on to suggest a topic-relevant conference or journal to publish a new article.

H. Liu, Z. Yang, I. Lee, Z. Xu, S. Yu, and F. Xia[3] proposed a novel article recommendation method called Citation-based scientific Article Recommendation (CAR). CAR combines the information of researchers' historical preferences and citation relations between articles. They have considered the fact that, not all pairwise articles with citation relations are highly relevant although researchers generally find articles of interest by searching citations. Therefore, in the proposed method, weak citation relations are first filtered out through an association mining technique using data on researchers' historical preferences. Then, these filtered citation relations are incorporated into a graph-based article ranking method for enhancing recommendation quality. J. Tang , G.-J. Qi, L. Zhang, and C. Xu[5] proposed a novel cross-space affinity learning algorithm over different spaces with heterogeneous structures. Unlike most of affinity learning algorithms on the homogeneous space, they constructed a cross-space tensor model to learn the affinity measures on heterogeneous spaces subject to a set of order constraints from the training pool. The basic idea is to first construct cross-space tensors (CSTs) from these heterogeneous spaces to represent the correlation between them, and the cross-space affinity can then be learned by exploiting a set of order constraints on the affinity from a training pool. C. Nascimento, A. H. F. Laender, da A. S. Silva, and M. G. Goncalves[6] propose a novel source independent framework for research paper recommendation. The framework requires as input only a single research paper and generates several potential queries by using terms in that paper, which are then submitted to existing Web information sources that hold research papers. Once a set of candidate papers for recommendation is generated, the framework applies content based recommending algorithms to rank the candidates in order to recommend the ones most related to the input paper.

This is done by using only publicly available metadata that is the title and abstract.

3. System Model

With the rapid growth in the information technology and the growing amount of the scholarly data, the recommendation techniques are becoming more popular. The main aim of the. recommender system is to automatically suggest the users with the items of the potential interest. It is well known as the tool for effectively solving information overload problem. As a result it has been implemented in multiple domains such as movies ,music, e-commerce ,e-learning, traffic, news etc. In recent years, web search tools employed by scientific digital libraries like IEEE Xplore, and literature search engines like Google Scholar, can retrieve a list of relevant articles in diverse technological fields using keyword-based queries. However there were certain limitations that were being present in these systems. These systems made the search for the researchers based on the keywords being entered by the researchers. Also the obtained results for the researchers used to be same if the keywords given by them would be same .along with it was not feasible for the researchers to search the articles if the searcher had no ideas of what they are looking for. Existing systems did the computation based on the content similarity between the articles. That is to find the articles that are similar to the target articles of interest or compute the similarity between the targets profile and the new article content to find the matches. But this work of the extraction of the content is not an easy task as the numbers of words in article are many.

The proposed system does the recommendation work based on the citation relation and the author based search pattern. The system consists of six main modules in it. The first module is the Article Preprocessing module. This module consists of set of all the research articles. In this module all the stop words present in each document are being removed. Here document is nothing but the research article. To remove the stop word the porter stemming algorithm is being used. The second module is the Article Feature Extraction module. Each algorithm work on numerical data, so in this module a feature vector of each research article is being created using the tf/idf calculations. All feature vector are store in document.
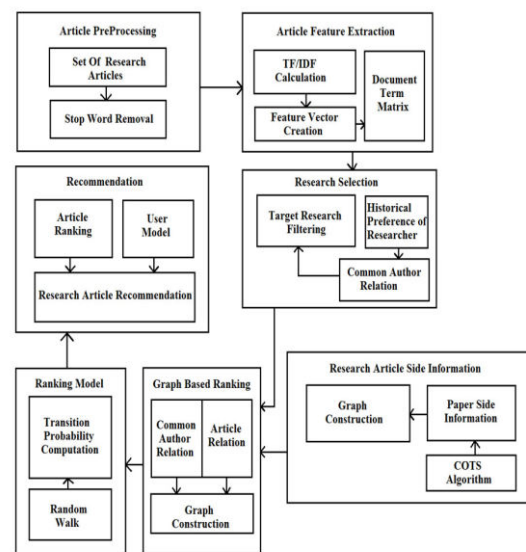
Fig1. Block Diagram of Proposed System

term matrix where row represent unique word and column is document-id. The third module is the Research selection .In this module historical preference of researcher and common author relation that is author which share same interest in research topic or publish same paper are present as the input. In research history there is each researchers domain of interest. The fourth module is the Graph based ranking module. In this module the common author relation is the number of authors presented same article. Researcher, who is a registered user in CiteULike website,has read one or more articles he is interested in. Traditional collaborative filtering methods utilize the first kind of link to generate recommendation. The rational underlying these methods is that, two researchers who are interested in the same articles are similar and then the taste of similar researchers are used to predict those of target researchers. Here the graph is being constructed to generate the article ranking list. The graph construction is based on the random walk algorithm In the domain of recommender systems, random walk-based ranking is a classical technique for recommendation. Based on the technique, many researchers have successfully applied it to various recommendation scenarios. The fifth module is the ranking module. In this module the ranking score of all the articles are being determined in order to help to calculate recommendation which is first problem and also to find target research i.e. researcher who is interested in finding similar citation or articles titles. The sixth module is the Recommendation module. In this module the target researchers is being give the relevant output that is the relevant article based on the article ranking which is being obtained from the graph. In this the relevant articles are being given the top ranks.

4. Conclusion

The proposed system provides a novel recommendation method which incorporates common author relations between articles to help generate better recommendations for relevant target researchers. In order to determine the relevant researchers with author based search pattern, two features are being introduced that is the ratio of pairwise articles with common author relations and the ratio of the most frequently appeared author. In addition to this an additional social relation is being incorporated such as the citation relationships to design a citation based recommendation method. The relevant targets are determined by analyzing the information on citation relations between articles .Due to the use of the hybrid framework, different recommendation methods which are suitable for different researchers are combined into one so that all researchers can obtain satisfactory recommendations.

5. Acknowledgements

References

[1] M. Gori and A. Pucci, Research paper recommender systems: A randomwalk based approach, in 2006 IEEE/WIC/ACM International Conference on Web Intelligence, 2006, pp. 778781.

[2] J. Sun, J. Ma, Z. Liu, and Y. Miao, Leveraging content and connectionsfor scientific article recommendation in social computing contexts, The Computer Journal, vol. 57, no. 9, pp. 13311342, 2014.

[3] H. Liu, Z. Yang, I. Lee, Z. Xu, S. Yu, and F. Xia, Car: Incorporating filtered citation relations for scientific article recommendation, in The 8th IEEE International Conference on Social Computing and Networking (SocialCom), Chengdu, China, Dec 2015.

[4] E. Medvet, A. Bartoli, and G. Piccinin, Publication venue recommendation based on paper abstract, in Proceedings of the 26th IEEE International Conference on Tools with Artificial Intelligence, 2014, pp.10041010.

[5] J. Tang, G.-J. Qi, L. Zhang, and C. Xu, Cross-space affinity learning with its application to movie recommendation, IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 7, pp. 15101519, 2013.

[6] C. Nascimento, A. H. F. Laender, da A. S. Silva, and M. G. Goncalves, A source independent framework for research paper recommendation, in Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries, 2011, pp. 297306.

[7] M. Qu, H. Zhu, J. Liu, G. Liu, and H. Xiong, A cost-effective recommender system for taxi drivers, in Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014, pp. 4554.

[8] H.-F. Wang and C.-T. Wu, A strategy-oriented operation module for recommender systems in e-commerce, Computers and Operations Research, vol. 39, no. 8, pp. 18371849, 2012.

[9] J. Wang and Y. Zhang, Opportunity model for e-commerce recommendation: right product; right time, in Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2013, pp. 303312.[20] J. Bobadilla, F. Serradilla, and A. Hernando99, pp. 91-121.

[10] T. Strohman,W. B. Croft, and D. Jensen, Recommending citations for academic papers, in Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2007, pp. 705706.

[11] K. Sugiyama and M.-Y. Kan, Scholarly paper recommendation via users recent research interests, in Proceedings of the 10th Annual Joint Conference on Digital Libraries, 2010, pp. 2938.

[12] M. Kaminskas, F. Ricci, and M. Schedl, Location-aware music recommendation using auto-tagging and hybrid matching, in Proceedings of the 7th ACM Conference on Recommender Systems,2013, pp. 1724.

[13] Y. Lu, J. He, D. Shan, and H. Yan, Recommending citations with translation model, in Proceedings of the 20th ACM International Conference on Information and Knowledge Management, 2011, pp.20172020.