

Heuristic Approach for Recommending Research Articles Using Collaborative Tagging

Atul Kumar

*Department Of Computer Science
Chandigarh University, Mohali, Punjab, India*

Sumit Kaur

*Department Of Computer Science
Chandigarh University, Mohali, Punjab, India*

Amritpal Singh

*Department Of Computer Science
Thapar University, Patiala, Punjab, India*

Abstract—This paper introduces Research Article Recommender System (RAR_System), a robust recommender system for research articles, which can be the alternative to current systems which mainly rely on searching the specific keywords in the database. But due to synonyms of the keywords and unclear nomenclatures, current systems provide unsatisfied results. RAR_System is based on content filtering and collaborative tagging. RAR_System provides the bottom up approach to classify research articles i.e. this system allows users to tag their papers with their own keywords so that classification of research articles is not only done by system itself. RAR_System improves the collaborative tagging approach by combining it with author analysis, publishing date analysis and name analysis. This system also optimized the retrieved results by using proposed heuristic approach. The proposed system has also taken care of the cold start problem for new articles.

Keyword -Recommender system, collaborative tagging, content filtering, heuristic approach, cold start problem.

I. INTRODUCTION

Recommendations basically are the suggestions we came across from others in our day to day life. Actually we all depend on suggestions from others to do our daily work for example researcher during their academic research work seek for the suggestions from their guide to select the topic for their research .And the best recommender systems are those which provide the related recommendations to the users which are beneficial for them for example if we are watching cartoon video then ads or recommendations should be related to the kids. Recommender Systems (RSs) are software tools and techniques providing suggestions for items to be of use to a user and the suggestions relate to various decision-making processes, such as what items to buy, what music to listen to, or what online news to read [1].These days recommender systems play an crucial role in highly rated e-commerce and Internet sites like Amazon.com, Flipkart.com, Myntra.com, Jabong.com, YouTube, Yahoo, Netflix. As more and more research papers are publishing exponentially yearly, hence it is very difficult for the researcher to find the related research paper. This difficulty and time taken is partly caused by the increase in publications and also due to the inefficiency of the recommender systems which failed to provide related recommendations to the researchers.

This paper presents a RAR_System, a research articles recommender system which is based on content filtering and collaborative tagging. In content based recommender system recommendations are provided on basis of description of items and the user profile, keywords are used to describe the items in content based recommender systems. Set of terms describes the content of each item, usually the words that occur in the document. Recommendations are made when user profile matches with the same terms. There are several issues associated with implementation of a content-based filtering system. First, terms can either be assigned automatically or manually. When terms are assigned automatically a method has to be chosen that can extract these terms from items. Second, the terms have to be represented such that both the user profile and the items can be compared in a meaningful way. Third, a learning algorithm has to be chosen that is able to learn the user profile based on seen items and can make recommendations based on this user profile [2]. Content filtering approach is mostly used with the text documents.

The proposed system is using the collaborative tagging technique that follow the bottom up approach for the recommender system which simply means that tagging a research articles is not done by the system administrator only, user or author of the research articles have authority to tag their own research articles. Collaborative tagging is a powerful approach that allows user to index or tag their items.

All users can participate and contribute their own personal vocabularies (keywords) to generate a collaboratively built 'bottom-up' system which more accurately reflects users' thinking of the world around them [3]. Also implementing the collaborative tagging approach is very economical. The proposed system search or provide the recommendations on the basis of tag assigned by the user itself.

Heuristic approach is used to optimize the results. Heuristic approach is used provide the accurate, precise and quick results for the particular problem i.e. recommending research articles.

The proposed system also provides solution to the cold start problem for the new items i.e. research articles. Cold start problem is very common in recommender systems. Cold start problem occur when new item or new user is added i.e. we do not know how to recommend that new item or what to recommend to that new user added to the system. The proposed system also has taken care of this problem.

The paper itself is structured as follows. Section II provides the related work of content based recommender systems and work done in field of collaborative tagging. Section III provides methodology used by us to recommend the research articles along with the result obtained. Section IV provides discussion and future work. Finally, Section V summarizes the results and points out future work.

II. RELATED WORK

This section of this paper has highlighted some previous work done in the field of recommender systems and collaborative tagging. However in this section complete list of related work is not presented. There is not much work is done in the field of recommender systems for research articles. Different authors suggest the use of content and collaborative filtering for recommendations. The strength of currently used academic search engines lies in finding documents containing specific keywords. Due to synonyms and unclear nomenclatures, this approach delivers in practice, often unsatisfying results [4]. When we do text based analysis there are many problems pop like unclear nomenclatures, synonyms or context depending on the meanings of words. If different words or terms are used then it is difficult to relate similar research articles.

When using collaborative filtering for the recommending research articles then there can be cold start problem for new research articles because it is highly rely on the rating from different users, but if there will be no user or user do not rate any item then there will cold start problem for sure. Collaborative filtering used in the field of recommender system is criticized for various reasons. Some authors claim that collaborative filtering would be ineffective in domains where more items than users exist. Others believe that users would not be spending time for explicitly rating research papers [4].

Several authors via research papers documented the benefits of use of collaborative tagging. Vander Wal [8] and Mathes [9] have discussed the potential benefits of tagging for personal information management. Vander Wal [8] has observed that in tagging systems there exists a powerful tool, allowing users to index their information resources with their

own keywords [3]. Collaborative tagging technique follow the bottom up approach for the recommender system which simply means that tagging a research articles is not done by the system administrator only, user or author of the research articles have authority to tag their own research articles.

Nan Zheng, Qiudan Li [5] proposed the recommender system based on tags and time information. They show empirically using data from a real-world dataset that tag and time information can well express user's taste and we also show that better performances can be achieved if such information is integrated into CF [5].

Another Collaborative filtering recommender system based on tag information is proposed by author [6]. This paper explores the utilization of tagging information to provide the related recommendations. This is based on the distinctive three dimensional relationships among users, tags and items, a new user profiling and similarity measure method is proposed. Author experiments suggest that the approach proposed is better than the traditional collaborative filtering recommender systems using only rating data [6].

Several author also proposed the incorporation of different filtering techniques with the meta-heuristic techniques to provide accurate and precise results. Punam Bedi, Ravish Sharma [7] has proposed the trust based recommender system using ant colony optimization. They incorporated the collaborative filtering technique with the ant colony optimization.

From the literature survey we concluded that there are few recommender systems exist for recommending research articles. Those who exist currently using content and collaborative filtering techniques and provide insufficient results. Both the techniques have their own limitations. Content filtering heavily depends upon the keyword and the description of the items. Content based recommender system becomes more complex due to semantic analysis and also in this user can get the recommendations similar to the items already rated by him in the past. Collaborative filtering based recommender systems suffer from sparsity, scalability and cold start problem. Also from the literature survey we found that the accuracy and precision of the results can be improved by the use of collaborative tagging and heuristic techniques.

III. PROPOSED WORK

This paper proposes the RAR_System, a recommender system for research articles based on content filtering and collaborative tagging. In addition, RAR_System provides a simple GUI to handle all the user queries. The methodology to recommend research articles follows steps mentioned and explained as follows.

- (A) RAR_System facilitates user to upload their research articles i.e. user can upload new research articles to the system.
- (B) RAR_System parses the uploaded research article before storing it to the database and basic text mining techniques are applied i.e. tokenization, stop word removal.
 - a) Tokenization: This is done to explore the words in the sentence by removing punctuation marks and other characters like hyphen and brackets.

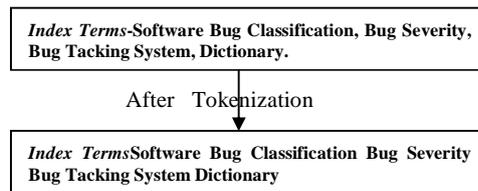


Fig. 3.1 Tokenization

- b) Stop word removal-This process is applied so that textual data is reduced by removing unnecessary words

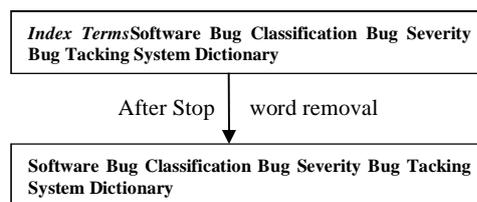


Fig. 3.2 Stop word removal

- (C) RAR_System extracts the useful information from the uploaded research article like name of the document, author name, keywords. Also RAR_System facilitates user to tag their research and add more information regarding that research article.
- (D) Now user can search the particular research articles by the name of the article, author name or by the date of upload.

- (E) Recommendation to the user will be provide according to the tag of the research articles. Top five papers with similar tags will be recommended to the user.

Proposed Algorithm: Recommending Research Articles

$W_{Ri} = \{W_{R1}, W_{R2}, W_{R3}, \dots, W_{Rm}\} \dots$ Set of initial or update weights for each research articles

$U = \{U_1, U_2, U_3, \dots, U_n\} \dots$ Set of users

$R = \{R_1, R_2, R_3, \dots, R_m\} \dots$ Set of research articles

W_{min} = Minimum weight for the research article

W_{max} = Maximum weight for the research article

W_{Ri}^{New} = Updated weight of the research article

W_{Ri}^{Old} = Old weight of the research article

$f(U \succ R) = (u_i, r_j) \succ 1 < i < m, 2 < j < n \dots$ Mapping function of user and research articles

This function includes two types of users

1. Users who upload the research articles
2. Users who tagged the research articles

Step 1: Initialization

- (i) Initialize W_{min} and W_{max}
- (ii) $\sum_{i=1}^m W_{Ri} = W_{coldstartweight}$ where $W_{Ri} = 0$

Step 2: Updating

- (i) $W_{Ri}^{New} = W_{Ri}^{Old} + W^{Inc}$ where $W_{Ri}^{New} \leq W_{max}$
- (ii) $\sum_{j=1}^n W_{Rj}^{New} = W_{Rj}^{Old} + W^{Dec}$ where $W_{Ri}^{New} \geq W_{min}$

Step 3: Retrieval

Now all articles have weight

$$f(U, R) = (u_i \succ r_j) \\ = (u_i \succ r_j)_Y$$

This function recommends the top Y research articles sorted on the weight parameter for the user U_i

- (F) *Cold start problem*: Cold start problem is very common problem in recommender system where we do not know how to recommend that new item or what to recommend to that new user added to the system. This system provides equal chance to all new research articles to come to the recommendation list by updating the weight of the research articles in the database.

(G) *Performance measurement*- Performance of the system [3] [10] can be measure using two parameters.

- **RECALL** is the ratio of the number of relevant records retrieved to the total number of relevant records in the database. It is usually expressed as a percentage.

$$\text{Recall} = \frac{\text{No. of target research articles retrieved}}{\text{No. of target research articles}} \quad (3)$$

- **PRECISION** is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. It is usually expressed as a percentage.

$$\text{Precision} = \frac{\text{No. of target research articles retrieved}}{\text{No. of research articles retrieved}} \quad (4)$$

<p>X= No. of relevant records retrieved Y= No. of relevant records not retrieved Z= No. of irrelevant records retrieved</p> <p>RECALL= $\frac{X}{X+Y} * 100$</p> <p>PRECISION= $\frac{X}{X+Z} * 100$</p>
--

Figure 3.3: Precision and Recall

IV. EXPERIMENTAL RESULTS

Experiments are conducted to evaluate the proposed method in section III. Experiment is performed on dataset of user and research articles containing information related to research articles like tag information, author information, publish date information. To evaluate the accuracy of the recommendations recall and precision is used which is elaborated in section III.

This paper compares the proposed Tagged- based Recommender System (RSs) with the Traditional Recommender System (RSs).

Comparison of the two systems with recall and precision is illustrated in fig 4.1.

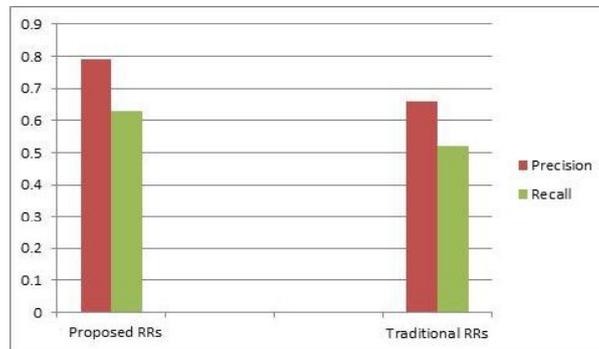


Figure 4.1: Comparison of Tag based and traditional recommender systems

The accuracy of the proposed system can be improved when more users tag the research articles.

V. DISCUSSION AND FUTURE WORK

Finding related content is very difficult task in current scenario where there are huge amount of data is stored in the databases. Recommender systems are solution to this problem and attracting researcher's to explore this area in past few years. This paper also tries to solve this problem for recommending research articles. In this paper we propose the bottom-up approach i.e. collaborative tagging technique to tag research articles and provide recommendations on the basis of similar tags. This system provides auto-tagging as well as facilitates users to add new tags or make changes. But the wrong tag by users can affect the efficiency of the system. This can be easily integrated into the current systems. This paper also proposes the heuristic approach to provide accurate, precise, optimized results to the users and also taken care of the cold start problem for new items where we do not know how to recommend that new item or what to recommend to that new user added to the system. But it is also not eliminated completely; more techniques can be integrated to the system to solve this problem. Also this system can be further improved by combining collaborative and content filtering techniques with collaborative tagging.

VI. CONCLUSION

This paper proposed the system to recommend research articles to the user based on bottom-up approach i.e. collaborative tagging and content filtering. This paper also proposes the heuristic approach to provide accurate, precise and optimized results by eliminating the cold start problem for new items. This paper proposed the technique which can be easily integrated to the current systems. Also this system can be further improved by combining collaborative and content filtering techniques with collaborative tagging.

REFERENCES

- [1]F Ricci, L Rokach, B Shapira, "Introduction to recommender systems handbook," Springer US, 2011.
- [2]"Content based filtering" <http://recommender-systems.org/content-based-filtering/>
- [3]Bhatia, Kapil. "Collaborative Tagging for Software Reuse."Computer Science\ & Engineering Department, Thapar Institute of Engineering\ & Technology, Deemed University (2006).
- [4] Gipp, Bela, JöranBeel, and Christian Hentschel. "Scienstein: A research paper recommender system." In International Conference on Emerging Trends in Computing, pp. 309-315. 2009.
- [5]Zheng, Nan, and Qiudan Li. "A recommender system based on tag and time information for social tagging systems." Expert Systems with Applications 38, no. 4 (2011): 4575-4587.
- [6]Liang, Huizhi, YueXu, Yuefeng Li, and RichiNayak. "Collaborative filtering recommender systems using tag information." In Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on, vol. 3, pp. 59-62. IEEE, 2008.
- [7]Bedi, Punam, and Ravish Sharma. "Trust based recommender system using ant colony for trust computation." Expert Systems with Applications 39, no. 1 (2012): 1183-1190.

- [8] Vander Wal, T., "Explaining and Showing Broad and Narrow Folksonomies", Vanderwal.net, <http://www.vanderwal.net/random/entrysel.php?blog=1635>, 2005
- [9] Mathes, A., "Folksonomies – Cooperative Classification and Communication Through Shared Metadata", Adam Mathes.com, USA <http://adammathes.com/academic/computermediatedcommunication/folksonomies.pdf>, 2004
- [10] Basics of Search guide available at [http://www.frame.org.uk/Searching for Information/Basics.htm](http://www.frame.org.uk/Searching%20for%20Information/Basics.htm)
- [11] Bobadilla, Jesus, Fernando Ortega, Antonio Hernando, and Javier Alcalá. "Improving collaborative filtering recommender system results and performance using genetic algorithms." Knowledge-based systems 24, no. 8 (2011): 1310-1316
- [12] Beel, Joeran, Marcel Genzmehr, Stefan Langer, Andreas Nürnberger, and Bela Gipp. "A comparative analysis of offline and online evaluations and discussion of research paper recommender system evaluation." In Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation, pp. 7-14. ACM, 2013.
- [13] Mobasher, Bamshad, Robin Burke, Runa Bhattacharya, and Chad Williams. "Toward trustworthy recommender systems: An analysis of attack models and algorithm robustness." ACM Transactions on Internet Technology (TOIT) 7, no. 4 (2007): 23.
- [14] J. Delgado, "Content-based collaborative information filtering: actively learning to classify and recommend documents," IEEE 1999.
- [15] N.J. Belkin and W.B. Croft, "Filtering and information retrieval: two sides of the same coin," IEEE 1992.
- [16] Peter Aksenov, Astrid Kemperman and Theo Arentze., "Toward personalised and dynamic cultural routing: a three-level approach" 12th International Conference on Design and Decision Support Systems in Architecture and Urban Planning, DDS, ELSEVIER (257-269) 2014.