



Analysis of Scientific Paper Recommendation System Using Article Side Information

Ms. Rachana R. Jadhav^{#1}, Prof. N. R. Wankhade^{*2}

[#] Masters of Computer Engineering,
Late G. N. Sapkal COE, Nashik, Maharashtra, India

¹ rachana1.jadhav@gmail.com

² nileshrw_2000@yahoo.com

^{*} Deptt. Of Computer Engineering
Late G. N. Sapkal COE, Nashik, Maharashtra, India

Abstract— Retrieving scientific article recommendation system plays important role for research in the field of academia. Most of the existing systems have designed integrated techniques for all target researchers therefore some of the algorithm generates recommendations for the researchers. In this paper we proposed different method to generate recommendation which is based on side information available in the papers such as citations and contents. The contents will be the abstracts of that paper. We use various similarity measurements like Euclidean Distance, cosine distance, Jaccard Coefficient, dice coefficient and SMTP for text processing distance measures calculating similarity between contents in the document. As a result we found that SMTP works better than other four similarity measures. We proposed a content based recommendation along with collaborative filtering which considers contents or unique words from the paper to recommendation purpose where top k recommendations were more closely similar to the target researcher's document. We perform an experiment on real world dataset and compared results demonstrate that proposed system gives more accurate results than baseline system.

Keywords— Content based recommendation, SMTP, Unique words, text processing, Collaborative Filtering, Article Recommendation.

I. INTRODUCTION

With the fast evaluation of big scholarly data, digital form plays important role and act as a major tool to store the knowledge. Further, it can be obtained any time anywhere all over the world. This digitized archival material has been made available online to the people free of cost or by paying a fee. This creates the information overload problem to academia while accessing article for more knowledge. Researchers in academia needs to find idea and article of interest for their research, information of article citation, topic of manuscript relevant to journal, needs to find expert, and reviewer in the domain which the manuscript belongs to. These paper related activities laborious task for researchers to locate appropriate journals, article and reviewers for afore mentioned work. Feng Xia *et.al* present two features, which are defined based on information about pairwise articles with common author relations and frequently appeared authors, to determine target researchers for recommendation [1]. Many existing system fails to deep analyse the information therefore J. Sun, [2]

analyse 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 which can suggest relevant articles.

The aim of academic recommender system is to tackle and solved the problem of information overload in big scholarly data as getting right research paper publication venue and concern recommendation tasks in above stated scenarios including (i) Suggestion about relevant articles to researchers for the various purpose like reading or citation [3] or article recommendation for the selection of relevant articles. (ii) The section of most appropriate reviewer for assigning a manuscript through reviewer recommendation [4]. (iii) Venue recommendation [5] to find a topic related conference or journal to publish a new article, paper or manuscript. (iv) Collaboration recommendation [6]. To get new partners to implement joint research (e.g. cross or multi-domain solution) and to have significant advantages that allows researchers to access more knowledge easily. An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

Existing system generally evaluate only content similarity between document to find the related article. In this paper we focus on common author searching pattern of target researcher along with content similarity.

II. EXISTING WORK

In existing system for target researcher's selection collaborative filtering method is used [1]. Two kind of links are exists in this researcher article and author. First link represents researcher-article relationship, where second link represents article-author relationship. In the recommendation method second link is ignored in collaborative filtering still these information may influence recommendation quality, as a result it is necessary to incorporate the second link to propose recommendation. Existing article recommendation techniques does not success to perform a detail analysis of the information. For solve this problem author proposes a novel approach to recommend scientific articles to researchers by leveraging content and connections [2]. This problem also tackles by Citation-based scientific Article Recommendation

(CAR) [3]. It combines information of researcher's historical preferences and citation relations between articles. In the method is proposed, weak citation relations are first filtered out through an association mining technique using researcher's historical preferences. After that these filtered citation relations are includes into a graph-based article ranking method for improve recommendation quality.

III. PROPOSED SYSTEM

To propose a novel recommendation method, this incorporates common author relations between articles to generate better recommendations for relevant target researchers using side-information. Such side-information may be of different kinds, such as document provenance information, the links in the document, user-access behaviour from web logs, or other non-textual attributes which are embedded into the text document.

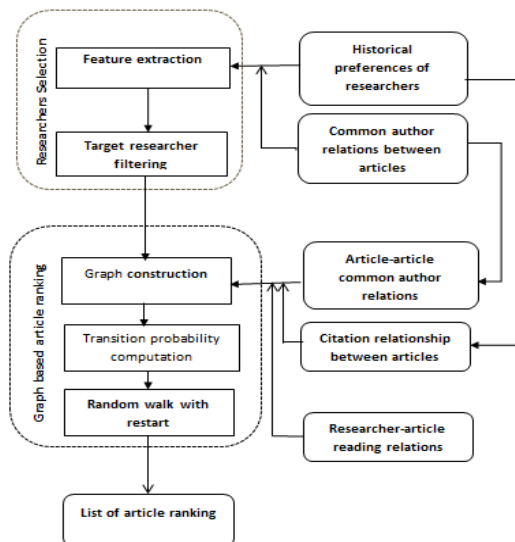


Fig.1 Proposed System

Graph based article ranking algorithm is used to rank the documents from the given graph,

Algorithm : Graph-based article ranking.

Input:

Graph, G;

Random walk probability, α ;

Target researcher vertex, v_0 ;

Maximum step length of iteration, maxStep;

Transition probability matrix, T;

Output:

Ranking scores of all article vertices, ScoreArticle(1:m);

// m article vertices

1: Define ranking scores of all vertices, ScoreAll(1 : n + m); // n + m vertices

2: for each $v \in V_R \cup V_A$ do

```

3:   ScoreAll(v) = 0; //initial ranking scores are 0
4: end for
5: ScoreAll(v0) = 1;
6: for step = 0; step < maxStep; step ++ do
7:   for each  $v \in V_R \cup V_A$  do
8:     tmpScore(v) = 0; //initial values are 0
9:   end for
10:  for each  $v_x \in V_R \cup V_A$  do
11:    for each  $v_y \in V_R \cup V_A$  do
12:      tmpScore(v_y) =  $\alpha \times \text{ScoreAll}(v_x) \times T(v_x, v_y) +$ 
      tmpScore(v_y);
13:    end for
14:    if  $v_x = v_0$  then
15:      tmpScore(v_x) = tmpScore(v_x) + 1 -  $\alpha$ ;
16:    end if
17:  end for
18:  ScoreAll = tmpScore;
19: end for
20: ScoreArticle(1 : m) = ScoreAll(n + 1 : n + m); //
select ranking scores of article vertices
21: return ScoreArticle(1 : m);

```

In our proposed system we used SMTP similarity measures along with graph base article ranking on the content base filtering method. In this we find the unique words in the document, calculate the term matrix, and then similarity measure is calculated for document of target researcher with all documents present in the database next step is to apply random walk algorithm to display recommendation list.

IV. EXPERIMENTAL RESULTS

We evaluate the recommendation quality of our proposed method with existing collaborative filtering method. Before that we experiment on similarity measures different method in which we found that SMTP gives more accurate results than other similarity measures where X axis contains Top K Recommendation and Y axis contains accuracy.

The proposed technique satisfies the symmetry property as $S_{SMTP}(d_i, d_j) = S_{SMTP}(d_j, d_i)$.

Given two documents $d_i = (d_{i1}, d_{i2}, \dots, d_{im})$ and $d_j = (d_{j1}, d_{j2}, \dots, d_{jmi})$ the S_{SMTP} similarity measure is given by the Eq. (1),

$$S_{SMTP}(d_i, d_j) = \frac{F(d_i, d_j) + \lambda}{1 + \lambda} \quad (1)$$

Where the function F is given by Eq. (2),

$$F(d_i, d_j) = \frac{\sum_{k=1}^m N_{*}(d_i, d_j)}{\sum_{k=1}^m N_{\cup}(d_i, d_j)} \quad (2)$$

Where

$$N^*(d_{ik}, d_{jk}) = \begin{cases} 0.5 \times \left(1 + \exp \left\{ \left(\frac{d_{ik} - d_{jk}}{\sigma_k} \right)^2 \right\} \right) & \text{if } d_{ik} d_{jk} > 0 \\ 0, & \text{if } d_{ik} = 0 \text{ and } d_{jk} = 0 \\ -\lambda, & \text{otherwise} \end{cases} \quad (3)$$

Where s_k is the standard deviation of all the non-zero values for the feature w_k .

$$N_U(d_{ik}, d_{jk}) = \begin{cases} 0, & \text{if } d_{ik} = 0 \text{ and } d_{jk} = 0 \\ 1, & \text{otherwise.} \end{cases} \quad (4)$$

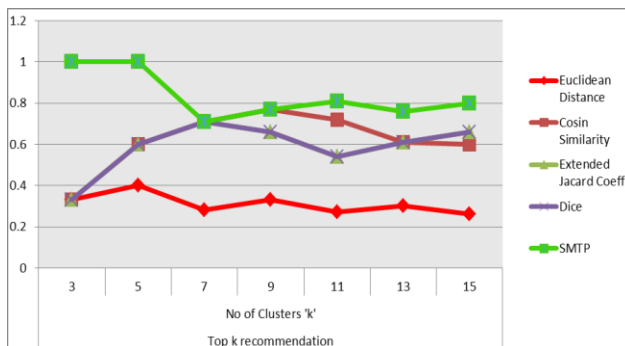


Fig 2. Content based recommendation proposed approach vs previous approach

We employed three different metrics, namely, Precision, Recall, and F-measure. Where Precision represents probability of selected article appeared in the recommendation list.

$$\text{Precision} = \frac{1}{n} \sum_{i=1}^n \text{Precision}_i \quad (5)$$

Recall represents the probability that the recommended articles appeared in researcher's collected list.

$$\text{Recall} = \frac{1}{n} \sum_{i=1}^n \text{Recall}_i \quad (6)$$

and F Measure assigns equal weight for precision and recall, is defined as

$$\text{F-Measure} = \frac{2(\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}} \quad (7)$$

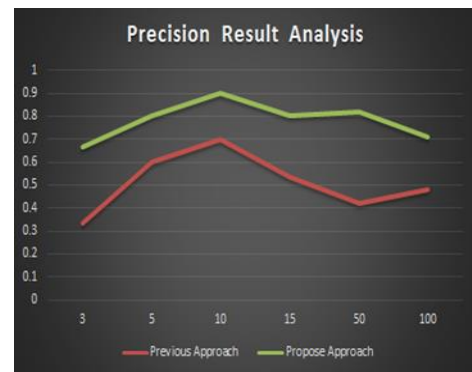


Fig. (a) Precision

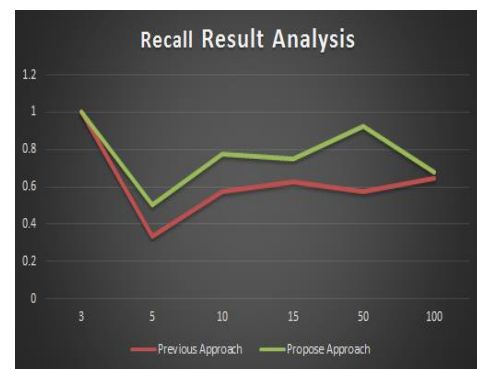


Fig. (b) Recall

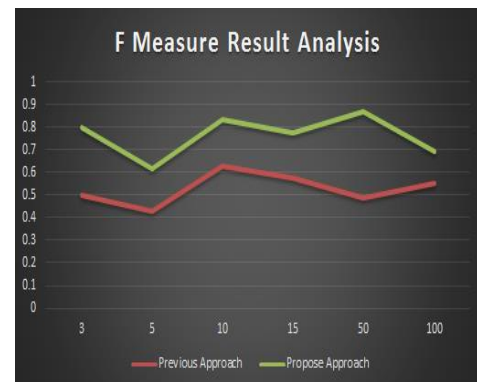


Fig. (c) F-Measure

Fig 3: Comparison of precision, recall, and F1 of proposed system with previous system for different probability.

We can see that, precision is 0.33 in baseline where proposed gives 0.66 for top 3 recommendations, while number of recommendation is increased from 3 to 5 recall is 0.26 in baseline and 0.46 in proposed system, F measure is 0.5 for 3 recommendation where 0.8 for proposed method. It means that, proposed method performs better than Baseline method for relevant researchers filtered. This indicates that, incorporating common author relations is able to help generate accurate recommendations for relevant researchers rather than all researchers.

V. CONCLUSIONS

In this paper we proposed a content based recommendation approach with side information in the article. Method that exploits information related to common author relations and historical preferences has been proposed to recommend articles of interest for specific researchers with content based search pattern. The information on common author's relations and contents from documents was incorporated to build a graph based article ranking algorithm for generating a recommendation list. The experimental results demonstrated that, for relevant targets determined by features, our proposed method performs better than the Baseline method.

REFERENCES

- [1] Feng Xia, Haifeng Liu, Ivan Lee, Longbing Cao, "Scientific Article Recommendation: Exploiting Common Author Relations and Historical Preferences", IEEE Transactions on Big Data Vol. 2, no. 2, pp. 101–112, 2016.
- [2] J. Sun, J. Ma, Z. Liu, and Y. Miao, Leveraging content and connections for scientific article recommendation in social computing contexts, *The Computer Journal*, vol. 57, no. 9, pp. 1331–1342, 2014. S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.
- [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] D. K. Tayal, P. Saxena, A. Shama, G. Khanna, and S. Gupta, "New method for solving reviewer assignment problem using type-2 fuzzy sets and fuzzy functions," *Applied intelligence*, vol. 40, no. 1, pp. 54–73, 2014. (2002) The IEEE website. [Online]. Available: <http://www.ieee.org/>
- [5] Z. Yang and B. D. Davison, "Venue recommendation: Submitting your paper with style," in *2012 11th International Conference on Machine Learning and Applications*, vol. 1, 2012, pp. 681–686.
- [6] 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. 1004–1010.
- [7] Naresh Kumar Nagwani, "A Comment on "A Similarity Measure for Text Classification and Clustering" in IEEE transaction on knowledge and data engineering, vol. 27, no. 9, 2015, 2589-2590.