Article Recommendation based on Common Author Relations and Author based Search Patterns

Mrs. Prajakta Vilas Sahane[1] Mrs. Anuradha Nawathe[2]
[1,2]Department of Computer Engineering
[1,2]Amrutvahini College of Engineering, Sangamner Savitribai Phule (Pune) University, Maharashtra India

Abstract— Research paper recommender frameworks are programming applications or frameworks that assistance singular clients to locate the most significant research papers to their requirements or tastes. Diverse scientists may have their own components and there may be relating strategies for them bringing about better proposals. In this paper, we propose a novel suggestion strategy which joins data on regular writer relations between articles (i.e., two articles with the same author(s)). The justification hidden our technique is that analysts frequently seek articles distributed by the same author(s). Since not all scientists have such writer based hunt designs, we introduce two elements, which are characterized in view of data about pairwise articles with regular writer relations and much of the time showed up writers, to decide target analysts for proposal. In commitment we had likewise grown new elements to investigate which analysts have creator based pursuit designs so that all scientists can acquire palatable proposals.

Key words: Collaborative Filtering, Common Author, Random Walk, Article Recommendation, Citation Recommendation

I. INTRODUCTION

The capacity of computers to provide recommendations was recognized fairly early in the history of computing. Grundy [1], a computer-based librarian, was an early step towards automatic recommender systems. Recommender systems have become extremely common in recent years. They are applied in a variety of applications. The most popular ones are probably movies, books, news, music, research papers, social tags, and products in general [2]. Most recommender systems typically produce a list of recommendations in one of the three ways- through collaborative-filtering technique, content-based technique, and hybrid algorithm [3]. Academic recommender systems aim to solve the information overload problem in big scholarly data such as finding relevant research paper, relevant publication venue, etc. Fig. 1 shows the corresponding recommendation tasks in above-mentioned scenarios, including (i) article recommendation [4], [5], [6], [7] for suggesting relevant articles to a researcher or an article for the purposes of reading or citation, (ii) reviewer recommendation [8], [9] for assigning a manuscript to the most appropriate reviewers (e.g., an expert in the same domain), (iii) venue recommendation [10], for suggesting a topic-relevant conference or journal to publish a new article, and (iv) collaboration recommendation [9], [10] for suggesting new partners to execute joint research (e.g., exploring cross-domain solution). There exist some interesting studies on these recommendation tasks. Gori and Pucci [1] built a citation relation graph and employed a random walk algorithm to compute ranking scores of each possible citation. Tayal et al. [6] assigned relevant weights to various factors which affect the expertise of the reviewer to create a fuzzy set and then compute the expertise. Yang and Davison [10] extracted features related to writing-style information for computing similarity between articles and then applied traditional collaborative filtering to recommend a venue for submission.

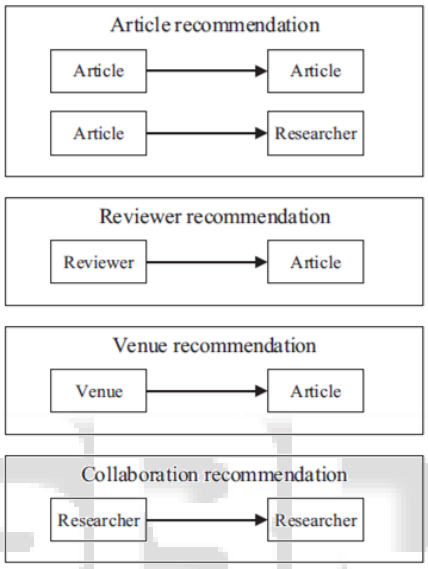

Fig. 1: Four academic recommendation tasks regarding to the entities in academia: researcher, article, venue, and reviewer.

II. RELATED WORK

A. Article-Article Recommendation

Article-article based recommendation, i.e., citation recommendation, includes global citation recommendation [4], and local citation recommendation [5]. Global citation recommendation aims to recommend a list of citations for a given query article. Strohman et al. linearly combined text features and citation graph features to measure the relevance between articles. They conducted relevant experiments with their proposed citation recommender system and concluded that similarity between bibliographies of articles and Katz distance are the most important features. Gori and Pucci [11] used citation relations between articles to built a citation graph and applied a random walk algorithm in the graph to compute ranking scores of each article as a reference of a target article. Bethard and Jurafsky incorporated a wide variety of features (including author impact, author citation habits, citation count, and publication ages) to build a retrieval model for literature search. Meng et al. incorporated various types of information including content, authorship, and collaboration network to build a unified graph-based model for personal global citation recommendation. Ren et al. proposed to cluster article citations into interest groups to determine the significance of different structural relevance

features for each group while deriving an articles relative authority within each group. Liu et al. employed the pseudo relevance feedback (PRF) algorithm to determine important nodes like authors and venues on a heterogeneous bibliographic graph. Then, a random walk algorithm was run to compute the ranking scores of an article.

B. Article-Researcher Recommendation

Article-researcher recommendation is our focus in this paper. Most existing studies compute similarities among researchers and articles based on articles contents or tags in social tagging systems and then apply traditional collaborative filtering to generate recommendations. Sugiyama and Kan examined the effect of modeling a researchers past works in scientific article recommendation to the researcher. A researchers profile was derived from his past works and other works which are the references or citations of those works. Apart from previous explicit citations, additionally took into account implicit citations. Potential citation articles were discovered using collaborative filtering and then combined with previous.

III. LITERATURE SURVEY

Sr. No	Author	Demerits
1.	Gori and Pucci	Built a citation relation graph and employed a random walk algorithm to compute ranking scores of each possible citation.
2.	Tayal et al.	Assigned relevant weights to various factors which affect the expertise of the reviewer to create a fuzzy set and then compute the expertise.
3.	Yang and Davison	Extracted features related to writing-style information for computing similarity between articles and then applied traditional collaborative filtering to recommend a venue for submission.
4.	Xia et al.	Considered three academic factors (i.e., co-author order, collaboration time, and number of collaboration) to define link importance, and then employed a random walk algorithm to compute rankings of potential collaborators.

Table 1: Literature Survey Table

Existing studies [6], [11] generally compute the content similarity between articles to find articles which are similar to the targets articles of interest, or compute the similarity between the targets profile and an new articles content to find matches. However, content extraction is not such simple because an article includes too many words. In this paper, we extract only author information to build relations between articles, i.e., common author relations. Then, these relations and researchers' historical preferences are used together to build a heterogeneous graph for article ranking.

The rationale of incorporating common author relations is that, the continuous development of internet technology enables researchers to easily build personal websites and share publications with others, which makes it more convenient to search articles published by the same author(s) for researchers who have such a search preference based on authors (we call it author-based search pattern). In addition, most studies ignore the fact that there exist different recommendation methods suitable for different targets. Therefore, we define features to find relevant target researchers who have author-based search patterns by analyzing information on common author relations existing in a researchers historical preferences. In summary, we propose a novel Common Author relation-based Recommendation method (CARE) for specific target researchers with author-based search patterns

IV. PROPOSED SYSTEM

Our Proposed system mainly includes two components: (i) researcher selection module and (ii) graph based article ranking module. The first component is responsible for extracting relevant features from researchers historical preferences and then selecting researchers with author-based search patterns as recommendation targets. The second component is responsible for incorporating common author relations to build a graph and generating article ranking list through a graph based random walk algorithm. In the domain of recommender systems, random walk-based ranking is a classical technique for recommendation.

Algorithm 1 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** $v \in V_R \bigcup V_A$ **do**
3: $ScoreAll(v) = 0$; //initial ranking scores are 0
4: **end for**
5: $ScoreAll(v_0) = 1$;
6: **for** $step = 0; step < maxStep; step++$ **do**
7: **for each** $v \in V_R \bigcup V_A$ **do**
8: $tmpScore(v) = 0$; //initial values are 0
9: **end for**
10: **for each** $v_x \in V_R \bigcup V_A$ **do**
11: **for each** $v_y \in V_R \bigcup V_A$ **do**
12: $tmpScore(v_y) = \alpha \times ScoreAll(v_x) \times T(v_x, v_y) + tmpScore(v_y)$;
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** $ScoreArticle(1:m)$;

Fig. 2: Proposed system algorithm

A. Target Researcher Selection

For researchers who find articles of interest by searching article written by the same authors, in their online article libraries, possibly there are lots of articles which are mainly written by one or several authors. Therefore, we define two

features which are relevant to common authors between any two articles to help determine target researchers.

B. Graph-based Article Ranking

Graph Construction As aforementioned, in field of academic recommendation, there are many entities such as researchers, articles, conferences, journals, and so on. In this paper, we consider some of them and then design a method for recommending scientific articles. Scientific article recommender systems include a set of n researchers R = {R1; ; Rn} and a set of m articles A = {A1; ;Am}.

C. Transition Probability Computation

We subsequently utilize the matrices to build a transition matrix, of which each element represents the transition probability between two corresponding vertices (article to article, article to researcher, and researcher to researcher). After obtaining the transition probability matrix, a random walk with restart method is employed to compute articles' rankings. Generally, the algorithm finds articles of interest based on the meta path: researcher-article-researcher-article. This means, a researcher is likely to be interested in an article which another researcher who has similar historical preferences expressed interest to. We incorporate common author relations between articles and then add another Meta path: researcher-article-article. This means, a researcher is likely to be interested in an article which is similar to another article which another researcher has expressed interest to. Our algorithm considers the two meta paths. Starting from a source vertex v0 (target researcher), we perform random walk with restart in the graph built in previous section. After walking to any vertex vx, we continue the next random walk with probability and walk to another vertex vy which links to vx with transition probability T(vx; vy). With probability 1-ǎ we return to source vertex v0.

V. EXPERIMENTAL RESULT

We defined increase rate to represent proposed improvement ratio to Baseline for different thresholds of FE1 and FE2. Note that increase rate is the same for precision. Fig. above shows the increase rate when the thresholds of FE2 is 0.15, respectively. It can be observed that, the increase rate is positive for these four situations. This demonstrates that, for the same researchers who are filtered using the threshold of FE2, Proposed system performs better than the current method.

Sr. No	Size of Data updates	Time to detect
0	32	34
2	24	26
4	16	20
6	14	18
8	13	16
10	12	15
12	11	13
14	11	12
16	9	11
18	8	9
20	9	8
22	6	8

Table 2: Graph Table

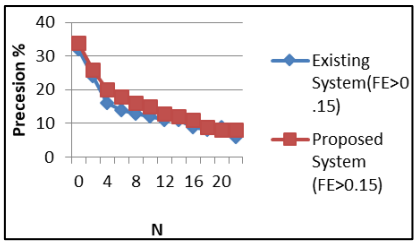

Fig. 3: Comparison of precision for all researchers and relevant researchers.

VI. CONCLUSIONS

A novel method that exploits information pertaining to common author relations and historical preferences has been proposed to recommend articles of interest for specific researchers with author-based search patterns. In order to determine specific targets, we defined two features (i.e. FE1 and FE2) which are relevant to common author relations between articles. We had also developed new features to explore which researchers have author-based search patterns. Then, the information on common author's relations was incorporated to build a graph based article ranking algorithm for generating a recommendation list.

REFERENCES

[1] F Xia, H.Liu,I.Lee "Scientific Article Recommendation: Exploiting Common Author Relation and Historical Preferences" IEEE Transaction on Big Data, 2016
[2] P. Gupta, A. Goel, J. Lin, A. Sharma, D.Wang, and R. B. Zadeh WTF; the who-to-follow system at Twitter, Proceedings of the 22nd international conference on World Wide Web
[3] H. Jafarkarimi, A.T.H. Sim and R. Saadatdoost A Naïve Recommendation Model for Large Databases, International Journal of Information and Educa tion Technology, June 2012.
[4] M. Gori and A. Pucci, " Research paper recommender systems: A random-walk based approach," in 2006 IEEE/WIC/ACM International Conference on Web Intelligence, 2006, pp. 778– 781.
[5] J. Tang and J. Zhang, " A discriminative approach to topic-based citation recommendation," in PAKDD' 09 Proceedings of the 13th Pacific-Asia Conferences on Advances in Knowledge Discovery and Data Mining, 2009, pp. 572– 579.
[6] 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.
[7] 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.
[8] T. Kolasa and D. Krol, A survey of algorithms for paper-reviewer assignment problem," IETE Technical Review, vol. 28, no. 2, pp.123– 134, 2011.
[9] D. K. Tayal, P. Saxena, A. Sharma, 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.

[10] 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. 681686.
[11] E. Rich,: User Modeling via stereotypes, Cognitive science, vol. 3, No.4, pp.329-354, October, 2010

