IEEE Access
The journal for rapid open access publishing

Received August 25, 2015, accepted September 18, 2015, date of publication September 28, 2015, date of current version October 6, 2015.

Digital Object Identifier 10.1109/ACCESS.2015.2481320

Context-Based Collaborative Filtering for Citation Recommendation

HAIFENG LIU, XIANGJIE KONG, (Member, IEEE), XIAOMEI BAI, WEI WANG, TESHOME MEGERSA BEKELE, AND FENG XIA, (Senior Member, IEEE)
School of Software, Dalian University of Technology, Dalian 116620, China

Corresponding author: X. Kong (xjkong@ieee.org)

This work was supported in part by the Fund of Key Laboratory of Control Engineering of Henan Province, Henan Polytechnic University, Kaifeng, China, under Grant KG2014-11, in part by the Fundamental Research Funds for the Central Universities under Grant DUT15QY06 and Grant DUT15YQ112, and in part by the National Natural Science Foundation of China under Grant 51209036, Grant 61174174, and Grant 61203165.

ABSTRACT Citation recommendation is an interesting and significant research area as it solves the information overload in academia by automatically suggesting relevant references for a research paper. Recently, with the rapid proliferation of information technology, research papers are rapidly published in various conferences and journals. This makes citation recommendation a highly important and challenging discipline. In this paper, we propose a novel citation recommendation method that uses only easily obtained citation relations as source data. The rationale underlying this method is that, if two citing papers are significantly co-occurring with the same citing paper(s), they should be similar to some extent. Based on the above rationale, an association mining technique is employed to obtain the paper representation of each citing paper from the citation context. Then, these paper representations are pairwise compared to compute similarities between the citing papers for collaborative filtering. We evaluate our proposed method through two relevant real-world data sets. Our experimental results demonstrate that the proposed method significantly outperforms the baseline method in terms of precision, recall, and F1, as well as mean average precision and mean reciprocal rank, which are metrics related to the rank information in the recommendation list.

INDEX TERMS Citation recommendation, collaborative filtering, citation context, citation relation matrix, association mining.

I. INTRODUCTION

Recommender Systems (RSs), which aim to suggest items of potential interest for solving information overload, have attracted growing amounts of attention [1], [2]. They have been successfully applied in many fields such as e-commerce [3], movies [4], music [5], e-learning [6], mobile service [7], and so on [8]. In recent years, with the rapid proliferation of information technology, more and more research papers are published and shared in many free digital databases or personal websites. To a high extent, this phenomenon contributes to information overload in research and academia. Consequently, some researchers accordingly apply recommendation techniques to solve the problem of academic information overload [9].

Generally, in order to help readers understand a research paper well, the author needs to cite relevant and important previous work as references for the paper. As described by Sun et al. [10], relevant papers were found by researchers in the print age through library catalogs. Subsequently, literature search engines were developed specially (such as Google Scholar) or integrated into various websites of online digital libraries (such as ACM Portal, IEEE Xplore, Elsevier, etc.) to retrieve relevant papers in diverse research areas. Keyword-based query is one commonly used technique for information retrieval. But whether search results meet researchers' needs depends on keywords used by them to a great extent. Moreover, the search results are the same for the same search keywords whereas researchers' needs are different. To this end, citation recommendation aims to suggest relevant papers as references for satisfying researchers' personalized citation requirements.

Collaborative Filtering (CF) is a classical recommendation method and has been exploited by researchers in the area of academic recommendation to recommend citations [11].

In the method, citing papers and cited papers are regarded as users and items in e-commerce. Similar citing papers (users) are likely to cite the same cited papers (items). Citing papers with common cited papers are considered to be similar. The similarities between citing papers are computed based on their common cited papers. As shown in Fig. 1(a), citing papers $i1$ and $i2$ cited the same paper $j2$ simultaneously, so they are similar to some extent. In this paper, we consider that two citing papers are similar if they are co-occurring with the same citing papers. As shown in Fig. 1(b), there is no common paper cited by citing papers $i1$ and $i4$. However, $i1$, $i2$, and $i3$ cited the same paper $j1$ simultaneously, $i4$, $i2$, and $i3$ cited the same paper $j2$ simultaneously. That means $i1$ and $i4$ are co-occurring with the same papers $i2$ and $i3$ so they are similar to some extent. Furthermore, we propose a citation Context-based Collaborative Filtering method called CCF for citation recommendation. Using citation context, an association mining technique is employed to determine co-occurred citing papers. Each citing paper is represented by other citing papers. Then, pairwise paper representations are compared to compute similarities between citing papers. Finally, the citations of citing papers similar to a target paper are used to predict the target's citations. Our experiments on two real-world datasets demonstrate the effectiveness of the proposed method.

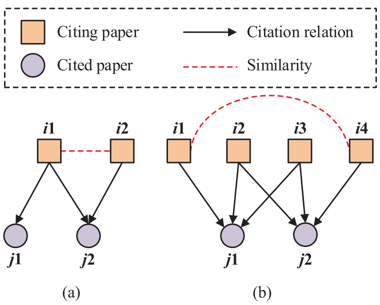


FIGURE 1. **Citing papers are similar based on (a) common cited papers or (b) co-occurred citing papers.**

The rest of the paper is organized as follows. Section II reviews related work on citation recommendation. Section III introduces the design detail of our proposed method. Section IV describes the experimental setup and elaborates our results in detail. Section V concludes the paper.

II. RELATED WORK

In academia, there exist various kinds of entities, such as researchers, authors, papers, readers, reviewers, venues, etc. Different academic recommendation problems always involve in one or several entities, such as venue recommendation (venue and researchers) [12], collaboration recommendation (researchers) [13], [14], scientific article recommendation (papers and researchers) [15], reviewer recommendation (reviewers and papers) [16]. In this section, we review some related works on citation recommendation. There are two classes of recommendation tasks in this area: global citation recommendation for a manuscript and local citation recommendation for specific content of a manuscript (e.g., a word, a phrase, a sentence, etc).

A. LOCAL CITATION RECOMMENDATION

Local citation recommendation aims to recommend citations for specific context such as a word of each place where a citation should be made in a paper, which is also called context-aware citation recommendation. Tang and Zhang [17] proposed a two-layer restricted Boltzmann machine model for modeling article contents and citation relationships. After obtaining topic representation of a citation context and learned model parameters, they calculated the probability of each paper being the reference paper for the citation context and then recommend top N ranked papers. He et al. [18] proposed a context-aware citation recommendation method. Given a query manuscript and citation context, they presented a probabilistic model to measure the relevance between articles as well as the relevance between citation context and articles. He et al. [19] extended their work in the context of lacking a bibliography of the target manuscript by utilizing 4 different models such as language model to find citation contexts. Lu et al. [20] proposed to recommend citations using a translation model that defines the probability of translating one word in one language into a word in another language. They assumed that the languages used in citation contexts and in article's content are different. They therefore translated one word in context to one word in citation. Huang et al. [21] also proposed to exploit a translation model for citation recommendation. They regarded an article as new 'words' in another language and directly estimated the probability of citing an article given a citation context. Tang et al. [22] proposed a cross-language context-aware citation recommendation method. This method introduced a joint embedding model to recommend English citations for a given context of the place where a citation was made in a Chinese article.

B. GLOBAL CITATION RECOMMENDATION

Global citation recommendation aims to recommend a list of citations for a given query article. McNee et al. [11] proposed to use the citation web between papers to create the raring matrix and then to apply the traditional collaborative filtering method to recommend citations. Gori and Pucci [23] proposed a random walk-based citation recommendation method. They built a citation graph based on citation relations between papers and then employed a random walk algorithm (PageRank) in a citation graph to compute a preference score of target on each reference for citation ranking. Strohman et al. [24] proposed to take paper content and author information into their evaluation model and to use bibliography similarity and Katz measurement to rank candidate citations. Nallapati et al. [25] proposed to jointly model the text and the citation relationship under a framework of topic model using two models. The first model generates the presence or absence of a citation represented by a Bernoulli random variable. The second model models a citation as a multinomial sampling process of the target document. Bethard and Jurafsky [26] proposed a retrieval model for literature search. They incorporated a wide variety of features (e.g., author impact, author citation habits, citation count, and publication age) and learn weights of these features to predict unknown citation by training the model. Meng et al. [27] proposed a unified graph-based model with random walk. They incorporated various types of information (e.g., content, authorship, citation and collaboration network) into the model to provide personal global citation recommendation. Ren et al. [28] proposed a cluster-based citation recommendation framework in the context of heterogenous bibliographic networks. They assumed that citations tend to be softly clustered into interest groups based on multiple types of relationship in the network, and then predict each query's citations based on related groups. Liu et al. [29] proposed to employ the pseudo relevance feedback (PRF) algorithm to find seed papers given a target query. They located important seed nodes such authors, citations, topics, and venues related to seed papers on a heterogeneous bibliographic graph, and then employed random walk to compute papers' rankings after extracting various meta paths.

Most of these works [24]–[29] extract some information (e.g., keywords, authors, venues, topics, etc.) from papers' content to build associations among various objects (e.g., papers, authors, venues) for citation recommendation. Different from them, in this work, we propose a global recommendation method which does not use papers' content information. In addition, unlike some works [23], [27], [29] which employ graph model-based methods, we utilize neighbor-based collaborative filtering technique to predict citation preference. McNee et al. [11] also applied collaborative filtering in citation recommendation, but we exploit citation context rather than only citation relation matrix to find similar neighbors. Experimental results have verified the effectiveness of proposed method.

III. DESIGN OF CCF

In citation relation matrix C, if a paper i cited a paper j, $C_{i,j} = 1$; on the contrary, $C_{i,j} = 0$. This matrix can be viewed as a binary rating matrix. Here, we call the value of $C_{i,j}$ *citation score*. Whether a paper j should be recommended to a paper i as reference, depends on the predicted citation score of i on j. In this section, we present a citation context-based citation recommendation method for suggesting relevant papers as references of another paper. Our proposed method mainly includes the following several steps, as shown in Fig. 2. First, based on citation context, the associations between citing papers are mined to transform citation relation matrix into association matrix between citing papers, i.e., building each citing paper's representation with other citing papers. Then, these paper representations are compared to calculate pairwise papers' similarity for determining neighbor papers. Next, the citation scores of neighbor papers are used to predict

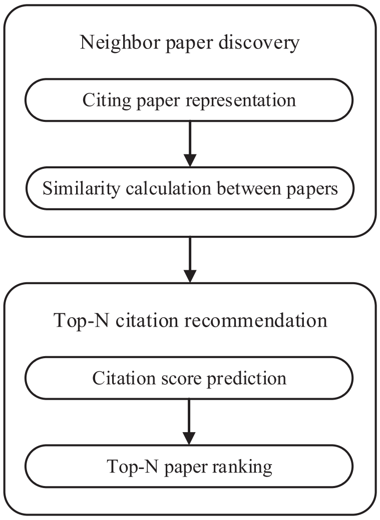


FIGURE 2. The process of proposed recommendation method.

TABLE 1. Citation relation data.

	i_1	i_2	i_3	i_4	i_5
j_1	1	1	1	-	1
j_2	-	1	1	1	-

TABLE 2. Paper vectors.

	i_1	i_2	i_3	i_4	i_5
i_1	0	1	1	0	1
i_2	1	0	1	1	1
i_3	1	1	0	1	1
i_4	0	1	1	0	0
i_5	1	1	1	0	0

citation scores of a target on relevant papers. Finally, these scores are ranked and top-N papers are recommended to the target. We illustrate the detail of each step in the following subsections.

A. CITING PAPER REPRESENTATION

In this section, we introduce how to obtain citing paper representation based on the citation context. In citation relation, there are two roles of papers: citing papers which cited other papers and cited papers which are cited by other papers. In order to obtain citing paper representation, the original double-role relation matrix is transformed into a single-role (only citing papers) association matrix by computing the most significant co-occurrences. We consider two citing papers to be co-occurred if they cited the same cited paper/papers. Let's take the following example for the purpose of illustration. Table 1 shows the citation relations between citing papers i_1, i_2, i_3, i_4, i_5 and cited papers j_1, j_2. Here, for simplicity, we consider two citing papers to be significantly co-occurred if at least one cited paper was cited by them simultaneously. In addition, a binary value of 1 or 0 is used for stating whether two citing papers are co-occurred or not. Table 2 shows the paper vectors describing the five citing papers calculated in this manner. Citing papers i_2 and i_3 are described by the same paper vectors, therefore they are highly related. Actually, significant co-occurrences of two citing papers depend on the their citation context. Therefore, we calculate a significance score based on contingency tables commonly-used in statistical theory.

TABLE 3. Contingency table.

	$paper_{i2}$	$\neg paper_{i2}$
$paper_{i1}$	N_{11}	N_{12}
$\neg paper_{i1}$	N_{21}	N_{22}

Table 3 shows the contingency table for two citing papers $paper_{i1}$ and $paper_{i2}$, which cited N_{11} papers simultaneously. Additionally, $paper_{i1}$ cited N_{12} papers which were not cited by $paper_{i2}$, $paper_{i2}$ cited N_{21} papers which were not cited by $paper_{i1}$, and N_{22} papers were not cited by them at all. The χ^2 test was then used to measure the association between the two citing papers, as shown in Equation 1. Based on the contingency table, the χ^2 test sums the squared z-scores for each cell in the contingency table and is correlated to the expected frequencies. Based on the cumulative distribution function of χ^2 as shown in Equation 2, a larger value of χ^2 generates a larger value of association possibility *prob*. The larger the value of *prob* is, the more significant the co-occurrence between two citing papers is. We use a threshold *ts* to determine whether the co-occurrence is significant or not. For two citing papers, if *prob* is larger than *ts*, they are regarded to be significantly co-occurred, and vice versa. In the obtained association matrix, there are only binary values (1 for significant co-occurrence and 0 for others). Each row of association matrix represents a citing paper and they are used for similarity calculation.

$$\chi^2 = \frac{(|N_{11} \cdot N_{22} - N_{12} \cdot N_{21}| - \frac{N}{2})^2}{R_1 \cdot R_2 \cdot C_1 \cdot C_2} \tag{1}$$

where $R_1 = N_{11} + N_{12}$, $R_2 = N_{21} + N_{22}$, $C_1 = N_{11} + N_{21}$, $C_2 = N_{12} + N_{22}$, and $N = C_1 + C_2 = R_1 + R_2$. As stated above, N_{11}, N_{12}, N_{21}, and N_{22} are the number of corresponding papers.

$$prob = F(x|v) = \int_0^x \frac{t^{v/2-1} \cdot e^{-t/2}}{2^{v/2} \cdot \Gamma(v/2)} dt \tag{2}$$

where v is the freedom degree and equals 1 for the contingency table here, $\Gamma(\cdot)$ is the Gamma function, x is the value of χ^2 computed using Equation 1.

B. SIMILARITY CALCULATION

The traditional collaborative filtering method calculates similarities between citing papers based on the original citation relation matrix, where each citing paper is represented by other cited papers. In this work, based on the obtained association matrix, we compute the cosine similarity of paper vectors as the similarity for pairwise citing papers using Equation 3.

$$sim_{i1,i2} = \frac{V_{i1} \cdot V_{i2}}{|V_{i1}| \cdot |V_{i2}|} \tag{3}$$

where V_{i1} and V_{i2} are the paper vectors of citing papers $i1$ and $i2$.

C. SCORE PREDICTION AND RECOMMENDATION

We predict the citation score $\hat{S}core_{i_0,j}$ of target citing paper i_0 on a paper j by averaging the scores of neighbor papers of i_0 while each score is weighted by the corresponding similarity $sim_{i_0,i}$, see Equation 4.

$$\hat{S}core_{i_0,j} = \frac{\sum (sim_{i_0,i} \times Score_{i,j})}{\sum sim_{i_0,i}} \tag{4}$$

where i is one of neighbor citing papers of i_0.

IV. EXPERIMENTS

A. DATASET

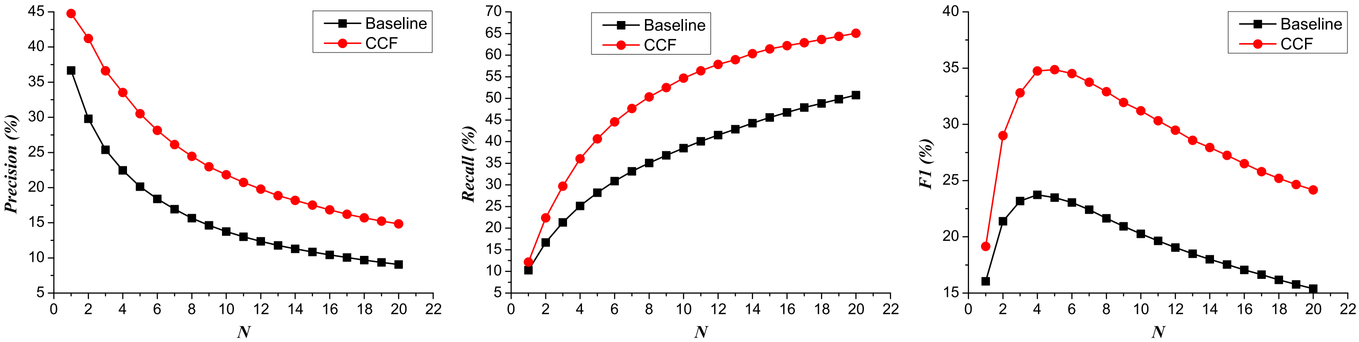
We utilized two datasets provided by the 2003 KDD Cup [30] in our experiments: the HEP-PH (high energy physics phenomenology) and the HEP-TH (high energy physics theory). These datasets were extracted from the e-print arXiv.org website and include 421578 citation relations between 34546 papers as well as 352807 citation relations between 27770 papers, respectively. We removed papers which cited less than 20 papers or which were cited by less than 5 papers. The distributions of the preprocessed datasets are shown in Table 4. Like most datasets in the area of recommender systems, citation relations in the two datasets are very sparse (0.9971 for HEP-PH and 0.9960 for HEP-TH), i.e., data sparsity. The sparsity indicates the ratio of the difference between numbers of all possible relations and existing citation relations to the number of all possible relations.

TABLE 4. Data statistics.

Datasets	HEP-PH	HEP-TH
Number of citing papers	6580	6263
Number of cited papers	10760	7928
Number of citation relations	205406	199298
Sparsity of citation relations	0.9971	0.9960

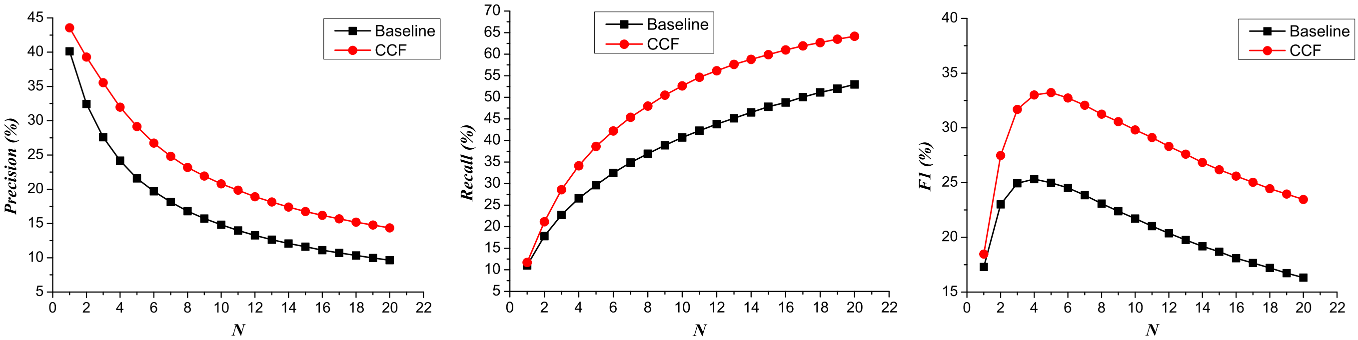
B. EXPERIMENTAL SETUP

To evaluate the quality of recommendations, we divided each dataset randomly into a training set and a test set using the following procedure. For each citing paper, we randomly select cited papers into the test set at the ratio of 20%. We employed three commonly used evaluation metrics in our experiments: a) Precision, is the ration of the number of cited papers in the top-N recommendation list to the length of the same list; b) Recall, is the ratio of the number of cited papers in the top-N recommendation list to the total number of all cited articles; c) F1, is a harmonic mean of precision



(a) Precision (b) Recall (c) F1

FIGURE 3. **Comparison of precision, recall, and F1 for Baseline and CCF on HEP-PH.**



(a) Precision (b) Recall (c) F1

FIGURE 4. **Comparison of precision, recall, and F1 for Baseline and CCF on HEP-TH.**

and recall.

$$Precision = \frac{\sum_{p \in OUT(P)} |R(p) \bigcap T(p)|}{\sum_{p \in OUT(P)} |T(p)|} \tag{5}$$

$$Recall = \frac{\sum_{p \in OUT(P)} |R(p) \bigcap T(p)|}{\sum_{p \in OUT(P)} |R(p)|} \tag{6}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{7}$$

where P is the set of all papers and $OUT(P)$ is the set of citing papers. $R(p)$ is the list of papers recommended to citing paper p, and $T(p)$ is the set of papers selected into the test set for p.

In order to demonstrate the effectiveness of our proposed method, we compared the recommendation results of the following methods.

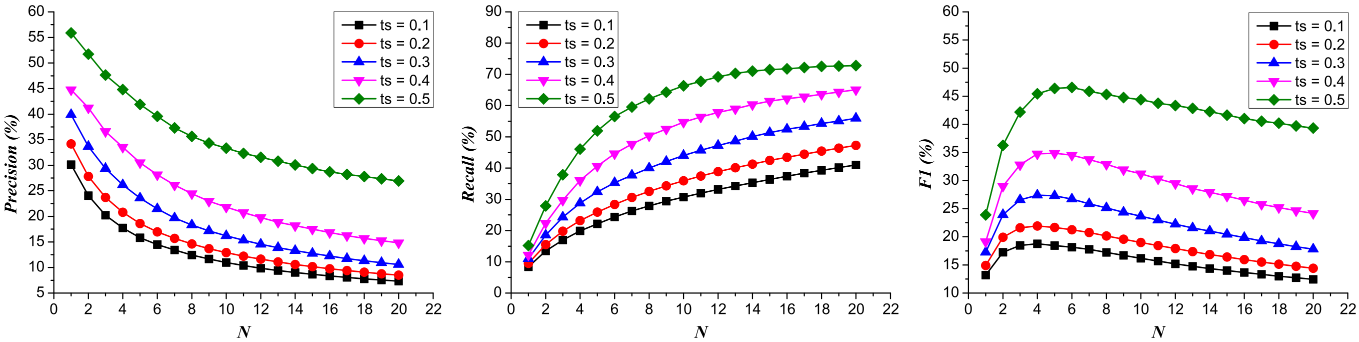
- **Baseline**: method, which regards citation relation matrix as rating matrix and generates top-N citation recommendations for a target paper. Similar neighbor papers are obtained by comparing cited papers of each neighbor with those of the target to compute their similarities.
- **CCF**: This is our proposed method.

C. PERFORMANCE COMPARISON

In this section, we compare the proposed method with Baseline method in terms of the citation recommendation performance. Fig. 3 shows the comparison results of Baseline and CCF (when *ts* equals 0.4) on HEP-PH dataset. As can be seen from this figure, CCF achieves much larger values of precision, recall, and F1 than Baseline for different top-N recommendations. In particular, CCF significantly obtains high improvement (about 10% in precision, 16% in recall, and 10% in F1), when N is larger than 4. In addition, we compare experimental results of Baseline and CCF on HEP-TH dataset, as shown in Fig. 4. From this figure, we can also see that, as the value of N is increasing, CCF always achieves large values of precision, recall, and F1 than Baseline. These experimental results on the two datasets demonstrate that, our proposed CCF significantly outperforms Baseline in terms of the three evaluation metrics. Furthermore, this indicates that similarity calculation based on citation context is able to help generate more accurate recommendations.

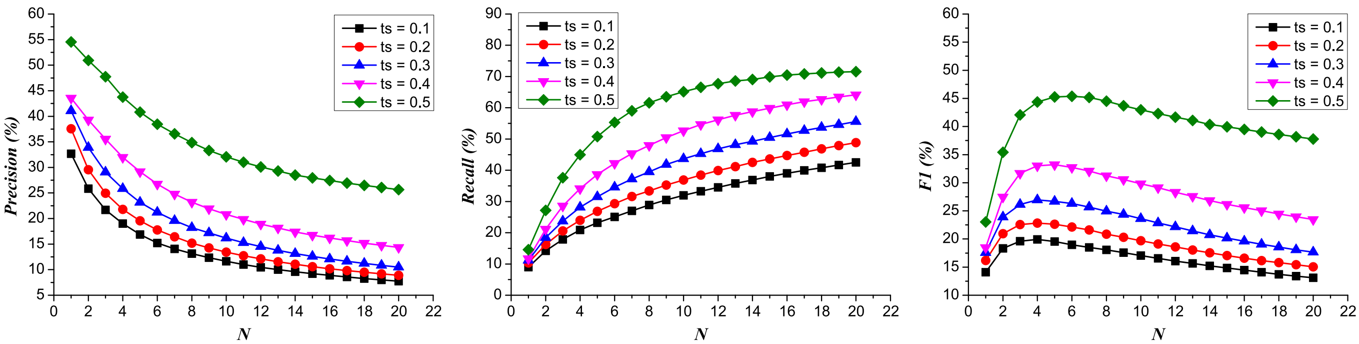
D. PARAMETER STUDY

As aforementioned, the function of the threshold *ts* is to determine whether two citing papers are co-occurred or not.



(a) Precision (b) Recall (c) F1

FIGURE 5. Precision, recall, and F1 of CCF for different threshold *ts* on HEP-PH.



(a) Precision (b) Recall (c) F1

FIGURE 6. Precision, recall, and F1 of CCF for different threshold *ts* on HEP-TH.

If their association degree is beyond ts, they are considered to be co-occurred, and vice versa. Obviously, similarity between papers can be influenced by ts. In this section, in order to understand the impact of ts on recommendation quality, we compare the experimental results of CCF for different values of ts on the two datasets.

Fig. 5 shows the comparison results of CCF on HEP-PH when ts is equal to 0.1, 0.2, 0.3, 0.4, and 0.5, respectively. From this figure, we can see that, as the value of ts is increasing, CCF achieves larger values of precision, recall, and F1. When ts is equal to 0.1 (i.e., the smallest value), the result is the worst. When ts is equal to 0.5 (i.e., the largest value), the result is the best. This is due to the fact that, the larger value of ts generates more accurate estimation of co-occurrence between two papers. Subsequently, these accurate co-occurrences are able to help accurately calculate similarities between papers. The comparisons of CCF for different values of ts on HEP-TH have also been made as shown in Fig. 6, and the experimental results demonstrate the same variation trend of precision, recall, and F1 of CCF as the value of ts increases. Note that only a enough large value of ts can generate better recommendation quality. Actually, when ts is equal to 0.1 or 0.2, CCF does not outperform Baseline, and when ts is larger than 0.2, CCF begins to perform better than Baseline. Therefore, a suitable value of ts is extremely important. This is why we discuss the impact of ts on recommendation quality of CCF here.

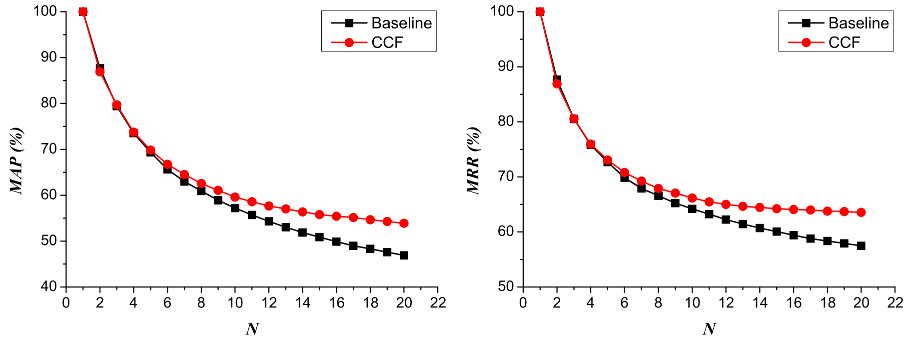
E. COMPARISON RESULTS ON OTHER METRICS

In addition to the three commonly used metrics stated in the above section, we also evaluate two other metrics: Mean Average Precision (MAP) and Mean Reciprocal Rank (MRR). Unlike previous three metrics which evaluate the ability to return overall cited papers, these two metrics take the rank information of relevant cited papers in the recommendation list into consideration. They are defined as follows:

$$MAP = \frac{1}{I}\sum_{i\in I}\frac{1}{n_i}\sum_{k=1}^{N}P(R_{ik}) \tag{8}$$

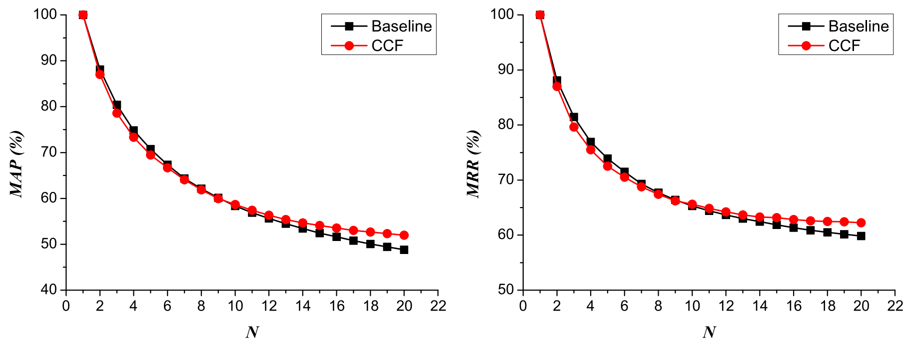
$$MRR = \frac{1}{I}\sum_{i\in I}\frac{1}{rank(i)} \tag{9}$$

where I denotes the set of citing papers, n_i is the number of relevant cited papers in the recommendation list of citing paper i, N is the length of recommendation list, $P(R_{ik})$ represents the precision of retrieved results from the top result until paper k is reached, $rank(i)$ represents the rank of the first relevant cited paper in the recommendation list of citing paper i.



(a) MAP (b) MRR

FIGURE 7. Comparison of MAP and MRR for Baseline and CCF on HEP-PH.



(a) MAP (b) MRR

FIGURE 8. Comparison of MAP and MRR for Baseline and CCF on HEP-TH.

In this section, we compare the proposed method with Baseline method in terms of these two metrics related to rank information. Fig. 7 demonstrates the comparison results of Baseline and CCF (when *ts* equals 0.4) on HEP-PH dataset. As can be seen from this figure, CCF achieves the almost same values of MAP and MRR to Baseline when N is less than 6, and CCF achieves the larger values of MAP and MRR than Baseline when N is larger than 6. With the value of N increases, their differences of both MAP and MRR are becoming larger and larger. In addition, we also compare the experimental results of Baseline and CCF on HEP-TH dataset, as shown in Fig. 8. We can see from this figure that, when N is less than 13, MAP and MRR of CCF are the almost same to those of Baseline, and when N is larger than 13, CCF achieves the larger values of MAP and MRR than Baseline. These results on the two datasets demonstrate our proposed CCF method performs better than Baseline method in terms of MAP and MRR. This also indicates that relevant citations appear earlier in recommendation list (i.e., the set of top-N recommended papers) generated by CCF when compared to Baseline.

V. CONCLUSION

In this paper, we have proposed a citation recommendation method (CCF) for recommending relevant papers as references of a target paper. Different from traditional collaborative filtering applied in citation recommendation, CCF computes similarities between citing papers by comparing pairwise paper representations which are obtained from citation context. The rationale underlying this similarity calculation is that, citing papers are considered to be similar if they are co-occurred with the same citing papers. Experimental results on two real-world datasets demonstrate that CCF outperforms Baseline method in terms of precision, recall, F1, MAP, and MRR. This indicates that the similarity calculation based on co-occurrence relation generates more accurate neighbor citing papers. In addition, our proposed method does not include content and other social relations such as co-authorship. We hope to include such information to improve citation recommendation in future work.

REFERENCES

[1] K. Verbert *et al.*, "Context-aware recommender systems for learning: A survey and future challenges," *IEEE Trans. Learn. Technol.*, vol. 5, no. 4, pp. 318–335, Oct./Dec. 2012.
[2] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowl. Syst.*, vol. 46, pp. 109–132, Jul. 2013.
[3] Z. Wu, J. Wu, J. Cao, and D. Tao, "HySAD: A semi-supervised hybrid shilling attack detector for trustworthy product recommendation," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 985–993.
[4] Y. Shi, M. Larson, and A. Hanjalic, "Mining mood-specific movie similarity with matrix factorization for context-aware recommendation," in *Proc. Workshop Context-Aware Movie Recommendation*, 2010, pp. 34–40.
[5] Y. C. Zhang, D. Ó. Séaghdha, D. Quercia, and T. Jambor, "Auralist: Introducing serendipity into music recommendation," in *Proc. 5th ACM Int. Conf. Web Search Data Mining*, 2012, pp. 13–22.

[6] W. Chen, Z. Niu, X. Zhao, and Y. Li, "A hybrid recommendation algorithm adapted in e-learning environments," *World Wide Web*, vol. 17, no. 2, pp. 271–284, 2014.

[7] F. Xia, N. Y. Asabere, A. M. Ahmed, J. Li, and X. Kong, "Mobile multimedia recommendation in smart communities: A survey," *IEEE Access*, vol. 1, pp. 606–624, Sep. 2013.

[8] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, "Recommender systems," *Phys. Rep.*, vol. 519, no. 1, pp. 1–49, 2012.

[9] F. Xia, N. Y. Asabere, H. Liu, N. Deonauth, and F. Li, "Folksonomy based socially-aware recommendation of scholarly papers for conference participants," in *Proc. Companion Pub. 23rd Int. Conf. World Wide Web Companion*, 2014, pp. 781–786.

[10] J. Sun, J. Ma, Z. Liu, and Y. Miao, "Leveraging content and connections for scientific article recommendation in social computing contexts," *Comput. J.*, vol. 57, no. 9, pp. 1331–1342, 2014.

[11] S. M. McNee *et al.*, "On the recommending of citations for research papers," in *Proc. ACM Conf. Comput. Supported Cooperat. Work*, 2002, pp. 116–125.

[12] E. Medvet, A. Bartoli, and G. Piccinin, "Publication venue recommendation based on paper abstract," in *Proc. IEEE 26th Int. Conf. Tools Artif. Intell.*, Nov. 2014, pp. 1004–1010.

[13] J. Tang, S. Wu, J. Sun, and H. Su, "Cross-domain collaboration recommendation," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 1285–1293.

[14] T. Huynh, K. Hoang, and D. Lam, "Trend based vertex similarity for academic collaboration recommendation," in *Computational Collective Intelligence. Technologies and Applications* (Lecture Notes in Computer Science), vol. 8083. Berlin, Germany: Springer-Verlag, 2013, pp. 11–20.

[15] M. S. Pera and Y.-K. Ng, "A personalized recommendation system on scholarly publications," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, 2011, pp. 2133–2136.

[16] T. Kolasa and D. Krol, "A survey of algorithms for paper-reviewer assignment problem," *IETE Tech. Rev.*, vol. 28, no. 2, pp. 123–134, 2011.

[17] J. Tang and J. Zhang, "A discriminative approach to topic-based citation recommendation," in *Advances in Knowledge Discovery and Data Mining*. Berlin, Germany: Springer-Verlag, 2009, pp. 572–579.

[18] Q. He, J. Pei, D. Kifer, P. Mitra, and L. Giles, "Context-aware citation recommendation," in *Proc. 19th Int. Conf. World Wide Web*, 2010, pp. 421–430.

[19] Q. He, D. Kifer, J. Pei, P. Mitra, and C. L. Giles, "Citation recommendation without author supervision," in *Proc. 4th ACM Int. Conf. Web Search Data Mining*, 2011, pp. 755–764.

[20] Y. Lu, J. He, D. Shan, and H. Yan, "Recommending citations with translation model," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, 2011, pp. 2017–2020.

[21] W. Huang, S. Kataria, C. Caragea, P. Mitra, C. L. Giles, and L. Rokach, "Recommending citations: Translating papers into references," in *Proc. 21st ACM Int. Conf. Inf. Knowl. Manage.*, 2012, pp. 1910–1914.

[22] X. Tang, X. Wan, and X. Zhang, "Cross-language context-aware citation recommendation in scientific articles," in *Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2014, pp. 817–826.

[23] M. Gori and A. Pucci, "Research paper recommender systems: A random-walk based approach," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell.*, Dec. 2006, pp. 778–781.

[24] T. Strohman, W. B. Croft, and D. Jensen, "Recommending citations for academic papers," in *Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2007, pp. 705–706.

[25] R. M. Nallapati, A. Ahmed, E. P. Xing, and W. W. Cohen, "Joint latent topic models for text and citations," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2008, pp. 542–550.

[26] S. Bethard and D. Jurafsky, "Who should I cite: Learning literature search models from citation behavior," in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage.*, 2010, pp. 609–618.

[27] F. Meng, D. Gao, W. Li, X. Sun, and Y. Hou, "A unified graph model for personalized query-oriented reference paper recommendation," in *Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage.*, 2013, pp. 1509–1512.

[28] X. Ren *et al.*, "ClusCite: Effective citation recommendation by information network-based clustering," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 821–830.

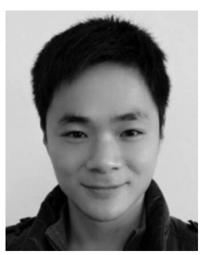
[29] X. Liu, Y. Yu, C. Guo, and Y. Sun, "Meta-path-based ranking with pseudo relevance feedback on heterogeneous graph for citation recommendation," in *Proc. 23rd ACM Int. Conf. Inf. Knowl. Manage.*, 2014, pp. 121–130.

[30] J. Gehrke, P. Ginsparg, and J. Kleinberg, "Overview of the 2003 KDD cup," *ACM SIGKDD Explorations Newslett.*, vol. 5, no. 2, pp. 149–151, 2003.

HAIFENG LIU received the B.Sc. and M.Sc. degrees in computer science from the Dalian University of Technology, Dalian, China, in 2009 and 2012, respectively, where he is currently pursuing the Ph.D. degree with the School of Software. His research interests include recommender systems, mobile computing, and big data.

XIANGJIE KONG received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2009. He is currently an Associate Professor with the School of Software, Dalian University of Technology, Dalian, China. He has authored over 30 scientific papers in international journals and conferences (with 10+ indexed by ISI SCIE). His research interests include big traffic data, social computing, and cyber-physical systems. He has served as the Workshop Chair or PC Member of a number of conferences.

XIAOMEI BAI received the B.S. degree in computer and application from the University of Science and Technology Liaoning, Anshan, China, in 2000, and the M.S. degree in software engineering from Jilin University, Changchun, China, in 2006. She is currently pursuing the Ph.D. degree with the Mobile and Social Computing Laboratory, School of Software, Dalian University of Technology, Dalian, China. Since 2000, she has been with Anshan Normal University, Anshan. Her research interests include computational social science, relation analysis, and influence analysis.



WEI WANG received the B.S. degree in electronic information science and technology from Shenyang University, Shenyang, China, in 2012. He is currently pursuing the Ph.D. degree in software engineering with the Dalian University of Technology, Dalian, China.

His research interests include big scholarly data, social network analysis, and computational social science.

TESHOME MEGERSA BEKELE received the B.Sc. degree in computer science from Hawassa University, Awassa, Ethiopia, in 2006, and the M.E. degree in software engineering from Chongqing University, Chongqing, China, in 2011. He is currently pursuing the Ph.D. degree with the Mobile and Social Computing Laboratory, Dalian University of Technology, Dalian, China. His research interests include big data analytics, cloud computing, data mining, and recommender systems.

FENG XIA received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China. He was a Research Fellow with the Queensland University of Technology, Brisbane, Australia. He is currently a Full Professor with the School of Software, Dalian University of Technology, Dalian, China. He has authored or co-authored two books and over 180 scientific papers in international journals and conferences. His research interests include social computing, mobile computing, and cyber-physical systems. He is a Senior Member of the Association for Computing Machinery. He is the (Guest) Editor of several international journals. He serves as General Chair, PC Chair, Workshop Chair, Publicity Chair, or PC Member of a number of conferences.