2010

Combining Coauthorship Network and Content for Literature Recommendation

San-Yih Hwang
National Sun Yat-sen University, syhwang@mis.nsysu.edu.tw

Chih-Ping Wei
National Taiwan University, cpwei@mx.nthu.edu.tw

Yu-chin Huang
National Sun Yat-sen University

Yun Tang
National Sun Yat-sen University, tiffany523@gmail.com

COMBINING COAUTHORSHIP NETWORK AND CONTENT FOR LITERATURE RECOMMENDATION

San-Yih Hwang, Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan, R.O.C., syhwang@mis.nsysu.edu.tw

Chih-Ping Wei, Department of Information Management, National Taiwan University, Taipei, Taiwan, R.O.C., cpwei@mx.nthu.edu.tw

Yu-chin Huang, Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan, R.O.C.

Yun Tang, Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan, R.O.C., tiffany523@gmail.com

Abstract

This paper studies literature recommendation approaches using both content features and coauthorship relations of articles in literature databases. Most literature databases allow data access (via site subscription) without having to identify users, and thus task-focused recommendation is more appropriate in this context. Previous work mostly utilizes content and usage log for making task-focused recommendation. More recent works start to incorporate coauthorship network for recommendation and found it beneficial when the specified articles preferred by authors are similar in their content. However, it was also found that recommendation based on content features achieves better performance under other circumstances. Therefore, in this work we propose to incorporate both content and coauthorship network in making task-focused recommendation. Three hybrid methods, namely switching, proportional, and fusion are developed and compared. Our experimental results show that in general the proposed hybrid approach achieves better performance than approaches that utilize only one source of knowledge. In particular, the fusion method tends to have higher recommendation accuracy for articles of higher ranks. Besides, the content-based approach is more likely to recommend articles of low fidelity, whereas the coauthorship network-based approach has the least chance.

Keywords: task-focused recommendation, coauthorship network, content-based recommendation, digital library.

1 INTRODUCTION

In the past few years, many recommender systems that provide effective customization and personalization have been employed by many online stores or websites. Types of targeted products include books, CDs, and other products at Amazon.com (Linden et al. 2003) and Epinions.com (Massa et al. 2004), and movies by MovieLens (Miller et al. 2003) and FilmTrust (Golbeck et al. 2006). Traditional recommendation techniques require either the explicit specification of users' interests or the implicit derivation from users' rating scores on sample items. However, for literature recommendation, most literature databases do not need users to identify themselves, and thus it is difficult to track an individual's long-term browsing behavior to derive his/her interest profile. These recommendation techniques therefore become unsuitable for recommending articles in databases involving academic literature.

A more appropriate recommendation approach for literature databases is the task-focused approach (Herlocker et al. 2001), by which a small number of articles recently viewed by a user form his/her task profile and other articles similar to articles in the task profile are recommended. Most existing methods use either content or usage log in defining article similarities (Mobasher et al. 2000, Srivastava et al. 2000, Hwang and Chuang 2004). More recently, Hwang et al. (2010) proposed to utilize coauthorship relations derived from articles in a literature database for task-focused recommendation. The proposed technique was shown to be more effective than the content-based technique when articles in a task profile are similar in their contents but is less effective otherwise. A preliminary hybrid method that switches between the coauthorship network-based and content-based techniques on the basis of the content coherence of a task profile was shown to achieve comparable or better recommendation effectiveness, when compared with the pure coauthorship network-based and content-based techniques.

In this work, we investigate the approach that combines content and coauthorship network for task-focused literature recommendation. We develop three hybrid methods, namely *switch*, *proportional*, and *fusion* methods, and compare them with pure content-based and coauthorship network-based approaches. Results of experiments based on a data set involving articles from prestigious data mining conferences and journals show that the three hybrid methods generally achieve better performance than their counterparts that utilize only one source of knowledge. Of the three proposed hybrid methods, fusion method is found to achieve higher recommendation accuracy for short recommendation list. In addition, the incorporation of coauthorship network in the design of recommendation methods contributes to the better fidelity of recommended articles.

The remainder of this paper is organized as follows: In Section 2, we review prior works relevant to this study. We describe the three hybrid literature recommendation methods in Section 3. We then detail our evaluation design and discuss important evaluation results in Section 4. We conclude in Section 5 with a summary of this study and some future research directions.

2 LITERATURE REVIEW

Recommender systems have become an important research area since the appearance of the first paper on collaborative filtering in the mid-1990s, and they typically suggest items, e.g., information, products or services, that are of interest to users based on customer demographics, features of items, and/or user preferences, e.g., ratings or purchasing history (Adomavicius & Tuzhilin 2005). There are two ways to generate interest profiles from users: explicit and implicit relevance feedback. Explicit relevance feedback includes ratings explicitly provided by users to indicate their preferences on some items. On the other hand, implicit relevance feedback derives users' ratings on items by observing their behaviours. Recommender systems can utilize these interest profiles to estimate the ratings of unrated items for users. Two most popular recommendation approaches are content-based recommendation and collaborative filtering. An excellent survey of the various recommendation techniques can be found in (Adomavicius & Tuzhilin 2005).

Most existing recommendation techniques concentrate on meeting users' *long-term* information need; however, there are cases when users need information that is specific to a task at hand, which may or may not be relevant to his/her long term interest. Recommendation under such a scenario is called task-focused recommendation (Herlocker et al. 2001). Task-focused recommendation requires a user to specify a task profile, which includes a small set of documents S that the user recently accessed, and the goal is to recommend documents whose contents are similar to and/or that are often accessed together with the documents in S (Srivastava et al. 2000, Mobasher et al. 2000, Yang et al. 2001, Hwang et al. 2004).

Social networks embody human interactions with numerical formulae. The basic elements of social networks include node, relation, content, direction and strength (Hanneman & Riddle 2005). Node is the basic element in a social network and is also called an actor, a representative of a person in the network. Relation is represented by edges connecting nodes and each edge can be characterized by content, direction and strength. Content is the cause of relationship between two actors. Wasserman & Faust (1994) classified the content into eight sorts: kinship (e.g., *brother of* and *father of*), social roles (e.g., *boss of*, *teacher of*, and *friend of*), affective (e.g., *likes*, *respects*, and *hates*), cognitive (e.g., *knows*, and *views as similar*), actions (e.g., *talks to*, *has lunch with*, and *attacks*), flows (e.g., *number of cars moving between*), distance (e.g., *number of miles between*) and co-occurrence (e.g., *is in the same club as* and *has the same color hair as*).

Approaches that take social relationships into account when building recommender systems have been investigated recently. The study by Lam (Lam 2004) incorporates social network into collaborative filtering systems. His approach defines the similarity between two users by looking at the strength of their social closeness as well as the similarity of ratings given to co-rated items. The preference of an un-rated item is estimated for an user based on the known ratings of his or her nearest neighbours for the target item and their similarities to the user. The experimental results show that the collaborative filtering systems with social network elements outperform the traditional ones. The recent emergence of Web 2.0 technologies has resulted in many social networking sites that allow users to express their web of trust. The usefulness of trust has been demonstrated in many computation related fields, for example, security/authentication, P2P system, multi-agent systems, and more recently recommender systems (Golbeck and Hendler 2006). The trustees of a user may serve as his/her recommendation partners. To remedy the problem that each user may specify only a limited number of trustees, many studies focus on how to infer the strengths of unspecified trust relations (Golbeck and Hendler 2006, Richardson et al. 2003, Ziegler and Lausen 2004). However, these studies all are intended to match users' long-term interests, rather than the task-focused recommendation as noted in this study.

In academia, co-authoring relationships are perhaps one of the most important types of connections between scholars. Researchers have shown great interests in analyzing coauthorship networks specific to their research communities to shed light on the collaboration characteristics of their communities (Newman 2001, Barabasi 2002, Acedo et al. 2006). Several generic properties concerning coauthorship networks in various fields have been identified. Hwang et al. (2010) observed that an article may be of interest to a user if the authors of the article are professionally close to articles in his/her task profile. Following this observation, they proposed a novel task-focused literature recommendation technique that utilizes a coauthorship network to recommend articles, with respect to an active user's task profile. Specifically, three methods to define the closeness between two scholars based on coauthorship network were proposed. Each method has a unique way in defining closeness matrix (C^*), which represents the strengths of the relationships between authors. Subsequently, authors of article a_i are extended by combining $\vec{a_i}$ and C^*, resulting in an extended author vector $\vec{a_i^*} = \langle r_{i1}^*, r_{i2}^*, \ldots, r_{in}^* \rangle$, where r_{ij}^* (i.e., the extended degree of authorship of scholar s_j for article a_i) is defined as $r_{ij}^* = \underset{1 \le k \le n}{Max}(r_{ik} \times cs_{kj}^*)$. After the extended author vectors are derived for all articles in the literature database, the coauthorship similarity of any two articles, a_i and a_j, is defined by the cosine similarity measure:

$$sim_{coauthorship}(a_i, a_j) = \frac{\vec{a_i^*} \bullet \vec{a_j^*}}{\left|\vec{a_i^*}\right| \times \left|\vec{a_j^*}\right|},$$

where $\left|\vec{a_i^*}\right|$ is the length of $\vec{a_i^*}$.

Finally, articles that have the highest average coauthorship similarity to the articles in the target task profile are recommended.

3 COMBINING CONTENT AND COAUTHORSHIP NETWORK FOR RECOMMENDATION

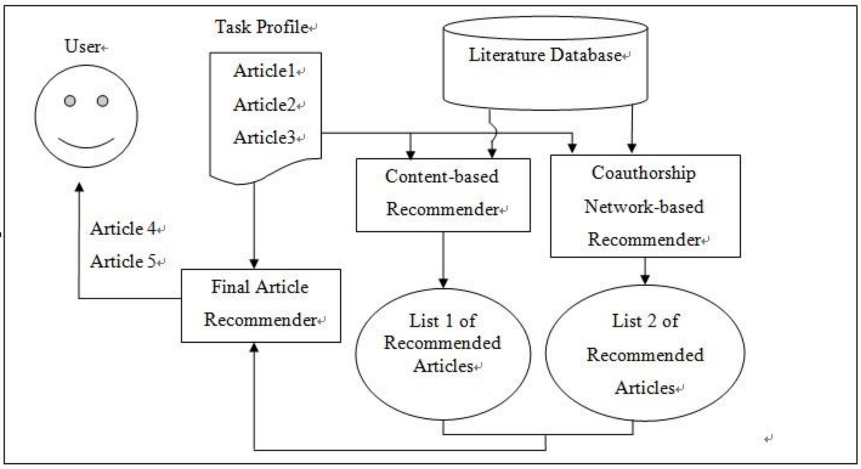


Figure 1. Architecture of the proposed recommendation framework

The system architecture of the proposed recommendation framework is shown in Figure 1.When users log onto the literature database, they may specify a number of articles that are of interest to them, referred to as a task profile. A screenshot of selecting articles of interest in SDOS literature database published by Elsevier Inc. is shown in Figure 2. The Content-based Recommender system extracts the content features of each article using TF-IDF measures (Salton et al. 1986) and recommends a list of articles that are similar in content to the articles in the task profile. The Coauthorship Network-based Recommender system constructs a coauthorship network for scholars (co)authored at least one article in the target literature database and subsequently suggests a list of articles that are close to the articles in the task profile in terms of coauthorship closeness. Finally, the Final Article Recommender system will select a number of unseen articles by properly combining the two lists of articles recommended by Content-based Recommender and Coauthorship Network-based Recommender systems.

Figure 2: A Screenshot of Article Selection in SDOS

We first describe how the content-based recommender and the co-authorship network-based recommender systems work and then present the three proposed methods for combining the two lists of recommended articles.

The Content-based Recommender System

We adopt the vector model and represent each article as a vector using the popular TF×IDF measure. Specifically, 4,000 terms derived from the content of the articles in the target literature database and having the highest average TF×IDF values are selected. Each article in the literature database is then represented as a 4,000-dimensional vector using the TF×IDF document representation scheme. Each article included in a given task profile is also represented, using the TF×IDF document representation scheme, in the same term space. The content similarity of two articles is determined using the cosine similarity measure. Let $\vec{d}_i$ and $\vec{d}_j$ be the term vectors of articles a_i and a_j, respectively. The content similarity between a_i and a_j is then defined as:

$$sim_{content}(a_i, a_j) = \frac{\vec{d}_i \bullet \vec{d}_j}{\left|\vec{d}_i\right| \times \left|\vec{d}_j\right|},$$

where $\left|\vec{d}_i\right|$ is the length of $\vec{d}_i$.

Following the previous work (Hwang and Chuang 2004), we estimate the content similarity between a user's task profile P and an article a_j in the literature database using the average method:

$$sim_{content}(P, a_j) = \frac{\sum_{a_p \in P} sim_{content}(a_p, a_j)}{|P|},$$

where $|P|$ is the number of articles in the task profile P. Given a task profile P, the content-based recommender system recommends a list of articles that have the highest content similarity to P.

The Coauthorship Network-based Recommender System

The coauthorship network involves all scholars who (co)author at least one article in the target literature database, and the coauthorship strength cs_{ij} from scholar s_i to scholar s_j is defined below:

$$cs_{ij} = \frac{|A_i \cap A_j|}{|A_i|},$$

where A_i and A_j denote the sets of articles (co)authored by s_i and s_j, respectively, and $A_i \cap A_j$ is the set of articles of which both s_i and s_j are coauthors. The relation $0 \le cs_{ij} \le 1$ reflects the direct coauthorship relationship from s_i to s_j. When cs_{ij} is large, we can infer that scholar s_j covers much of scholar s_i's professional expertise. Thus, if a user is interested in scholar s_i's work, it is likely that the user will be interested in scholar s_j's work.

In (Hwang et al. 2010), three schemes were proposed to derive the closeness from one scholar to another based on the coauthorship network. However, it was found that the *nontransitive closeness* scheme, which considers only direct coauthorship relations, achieves the best recommendation accuracy under most scenarios. Therefore, in this paper, we use the nontransitive closeness scheme for the coauthorship network-based recommender system. Each article is then represented as an extended author vector $\vec{a}_i^{\,*} = \langle r_{i1}^*, r_{i2}^*, \ldots, r_{in}^* \rangle$, where r_{ij}^* (i.e., the extended degree of authorship of scholar s_j for article a_i) is defined as $r_{ij}^* = \underset{1 \le k \le n}{Max}(r_{ik} \times cs_{kj}^*)$. The coauthorship similarity of any two articles, a_i and a_j, is defined using cosine similarity measure:

$$sim_{coauthorship}(a_i, a_j) = \frac{\vec{a}_i^{\,*} \bullet \vec{a}_j^{\,*}}{|\vec{a}_i^{\,*}| \times |\vec{a}_j^{\,*}|}.$$

Given a task profile P, the coauthorship network-based recommender system recommends a list of articles that have the highest coauthorship similarity to P.

Combining Two Recommendation Lists—The Switching Method

This method, originally described in (Hwang et al. 2010), was motivated by the observation that the coauthorship network-based approach is more effective than the content-based technique when articles in a task profile are similar in their contents but is less effective otherwise. Therefore, the switching method adopts coauthorship network-based technique when the articles in a target task profile are similar in their content and switches to the content-based technique otherwise. Specifically, we define the content coherence of a task profile as the pair-wise content similarity of the articles in the task profile and determine whether to switch to the content-based technique on the basis of a similarity threshold, α. α was set to 0.1 for which the best recommendation accuracy was achieved (Hwang et al. 2010).

Combining Two Recommendation Lists—The Proportional Method

Unlike the switching method which only choose one recommendation list for a given task profile, the proportional method selects a ratio of articles from each recommendation list in proportion to the difference between the content coherence of the task profile and the threshold α. When α is equal to the content coherence of the task profile, each recommendation list has equal ratio (i.e., 1/2) and thus contribute the same number of the recommended articles. If the content coherence of the task profile is high, the coauthorship network-based approach receives a higher ratio, and vice versa. Specifically, let c be the content coherence of a task profile and r_a be the ratio of articles extracted from the recommendation list of the coauthorship network-based approach. r_a can be computed as follows:

$$r_a = \begin{cases} \dfrac{c}{2\alpha} & if \quad c \le \alpha \\ \dfrac{1}{2} + \dfrac{c-\alpha}{2(1-\alpha)} & otherwise \end{cases}.$$

Let r_c be ratio of articles extracted from the recommendation list of the content-based approach. Obviously, $r_c = 1 - r_a$.

When c is close to 0, i.e., articles in the task profile are very diversified in their content, r_a=0 (or r_c=1). Therefore, the proportional method simply adopts the recommendation list given by the content-based recommender system. On the other hand, when c is close to 1, i.e., articles in the task profile are very similar in their content, r_a=1 (or r_c=0). In this case, the proportional method is reduced to the coauthorship-based recommender system. Finally, when $c = \alpha$, Both the content-based and coauthorship network-based recommender systems are equally well, and each recommends the same number of articles.

Combining Two Recommendation Lists—The Fusion Method

The last method for combining the two recommendation lists is called fusion, originally proposed by Torres et al (2004) for combining items recommended from different sources. The generation of the final recommendation list is as follows: every article receives a score that is the summation of the ranks in their original recommendation lists. The final recommendation list is sorted based on these scores in ascending order. However, such an approach requires each of the content-based and coauthorship network-based approaches sort all the articles based on their respective (content- or social-similarity) scores, which could be time-consuming. To remedy this problem, we modify the fusion method by requiring each of the content-based and social network-based approaches report a top-K' recommendation list, where K' will be empirically determined. For the fusion method, we give priorities to the articles that appear in the both recommendation lists, listed in ascending order of their scores. When there are rooms for more articles, we append to the recommendation list the articles with low scores that appear in only one list.

4 EVALUATIONS

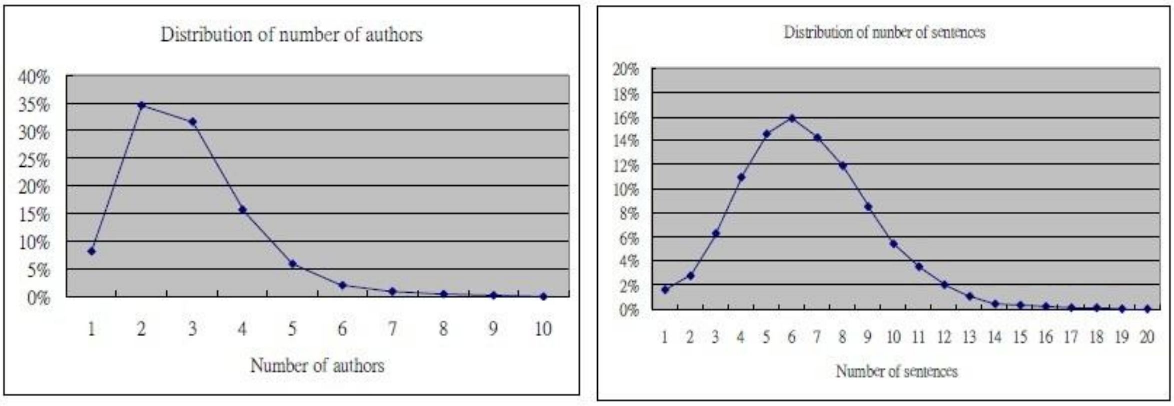
We based our experiments on a data mining data set that contains articles from seven major data mining conferences and three top data mining journals between 1996 to 2005. We also retrieved a total of 64 articles that were published in the aforementioned journals between Jan. 2005 and June 2006 and cite at least eight of our collected articles. The 64 articles will serve as a source for making up task profiles and are thus called test articles. We further expand the data mining data set by adding articles cited by the 64 test articles. Finally, we included some synthetic articles that are used to evaluate the fidelity of recommendation, which will be discussed later. Table1 lists the number of articles from each forum in our collection.

Data Mining Related Conferences (1996 - 2005)	**Number of Articles**
ACM CIKM	738
IEEE ICDM	565
ACM SIGKDD	624
SIAM International Conference on Data Mining	155
ACM SIGIR	716
ACM SIGMOD	494
VLDB Conf.	629
Subtotal	**3921**

Data Mining Related Journals (1996 - 2005)	Number of Articles
Data Mining and Knowledge Discovery	158
IEEE Transactions on Knowledge and Data Engineering	948
Machine Learning	434
Subtotal	**1540**
Expanded Cited Articles	**1061**
Expanded Synthetic Articles	**1000**
Total without Synthetic Articles	**6522**
Total with Synthetic Articles	**7522**

Table 1. The summary of collected articles

One of the major weaknesses with content-based approach is the potentially poor quality of the recommended articles because articles similar to a task profile may not be of high quality. On the other hand, coauthorship network-based approaches can alleviate the fidelity problem because it makes recommendation based on article authors. We added 1000 synthetic articles into the article set to serve as the low-quality articles. The authors and abstract of each synthetic article are determined in the following manner. To determine the set of authors for each synthetic article, we follow the same trend exhibited in our collected articles. Specifically, we first randomly determine the number of authors for a synthetic article by following the distribution on the number of authors of the articles in our data set which is shown in Figure 3(a). Then we randomly choose an author in inverse proportion to the number of articles published by each author in our collected articles. The rationale is that people who published more articles in our data set, which appear in prestigious conferences or journals, tend to have lower chance of writing low-quality papers. To determine the abstract content of the articles, we first randomly determine the number of sentences *N* by following the distribution on the number of sentences in the article abstracts, as shown in Figure 3(b).



(a) *(b)*

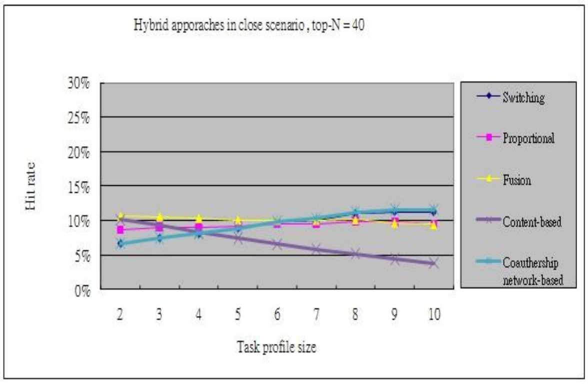
Figure 3. (a) Distribution on number of authors in our article set (b) Distribution of number of sentences

4.1 Experiment Design

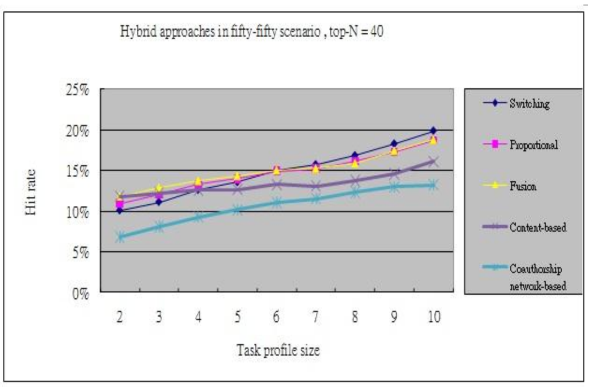
Each of the 64 test articles is treated as a subject, and the complete references of the test article are treated as the articles of interest to the subject. Let the referenced articles of a test article t_i be I_i. We split I_i into two sets: the task profile S_i and the prediction set P_i. In our experiments, we adopt each approach to recommend 40 articles for each S_i and evaluate how the set of recommended articles is close to P_i. Let the set of articles recommended by a method for task profile S_i be R_i. The hit rate of the method is defined as $\frac{|R_i \cap P_i|}{|P_i|}$. We adopt the average hit rate as the primary performance metric in our experiments. In addition we exercise three scenarios in our experiments, namely close scenario, diversified scenario and fifty-fifty scenario. In close scenario, each task profile comprises articles that have similar content to each other. On the other hand, in diversified scenario, articles in each task profile have low content similarity to each other. And the fifty-fifty scenario is the compromise of close and diversified scenario. It contains 50% task profiles for close scenario and another 50% task profiles for diversified scenario. We intend to examine how the different recommendation approaches perform under these three different scenarios.

4.2 Experimental Result

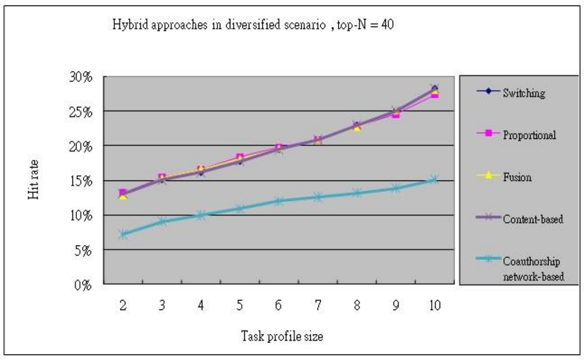
The average hit rates of the various methods, when applied to the data set without synthetic articles, are shown in Figure 4. In close scenario, the switching method is reduced to the coauthorship network-based method. The proportional and fusion approaches perform better than switching method when the task profile size is below 6, due to the benefit of combining recommendation lists of the two primitive methods. In fifty-fifty scenario, the three hybrid methods are equally well and outperform both primitive methods. As for the diversified scenario, the three hybrid methods and the coauthorship network-based approach are almost the same and outperform their content-based counterpart. To examine the performance differences of the three hybrid methods, we exercised different top-N values, and the results for top-N = 10 in diversified scenario is shown in Figure 5. It can see that the performance of fusion method is better than the other two hybrid methods. This is because fusion method is designed to give articles that are recommended by both content-based and cuauthorship network-based approaches higher ranks. For example, if an interesting article ranked 20th by content-based method in the recommendation list and ranked 25th by social network-based method, it may appear in the top-10 recommendation list of fusion because both primitive methods' lists have this article.



(a)



(b)



(c)

Figure 4. Hit rates of the three hybrid methods for recommending top-40 articles using content-based and coauthorship network-based approaches as benchmark under (a) close scenario (b) fifty-fifty scenario (c) diversified scenario

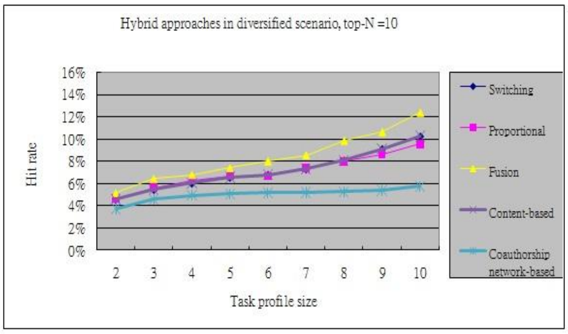


Figure 5. The result of various methods in diversified scenario when top-N = 10

We then examine the impact of including synthetic articles in the data set. Here we only show the results under the fifty-fifty scenario due to space limitation, though similar trend also exhibits in other scenarios. Figure 6 presents the drop of hit rate after incorporating synthetic articles. We notice that the coauthorship network-based approach incurs the lowest drop of hit rate, and the three hybrid methods, however, have the drop of hit rate close to that of the content-based method. We then examined the average rank of the first synthetic article recommended by the various approaches, using fusion method as the representative for the hybrid approach. Figure 7 shows that content-based approach tends to recommend synthetic articles within the top 10 of recommendation list. On the other hand, the first synthetic article recommended by coauthorship network-based approach has much lower rank (mostly after 40). The fusion method usually recommends the first synthetic articles at the ranks between 20 and 30. We further evaluate the fidelity rates of various approaches, which is defined as the ratio of recommended non-synthetic articles to the total recommended articles. For example, if we recommend 20 articles that contain two synthetic articles, the fidelity rate is 90%. The fidelity rates of the various methods are shown in Figure 8. Again, the coauthorship network-based approach shows the best recommendation fidelity, second by the fusion method.

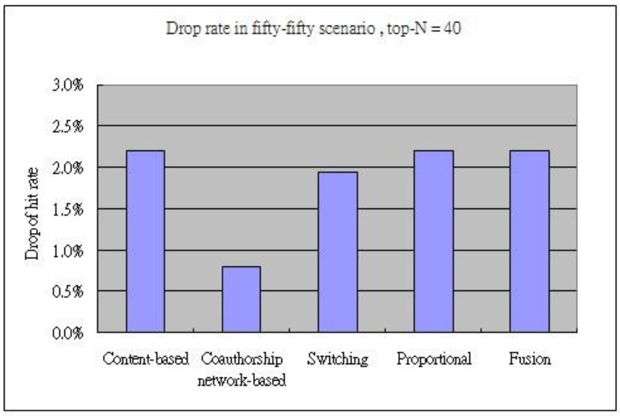


Figure 6. Drop of hit rate in fifty-fifty scenario

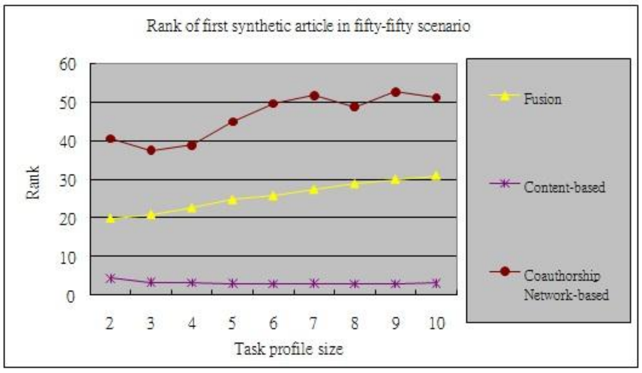


Figure 7. Rank of first synthetic article in fifty-fifty scenario

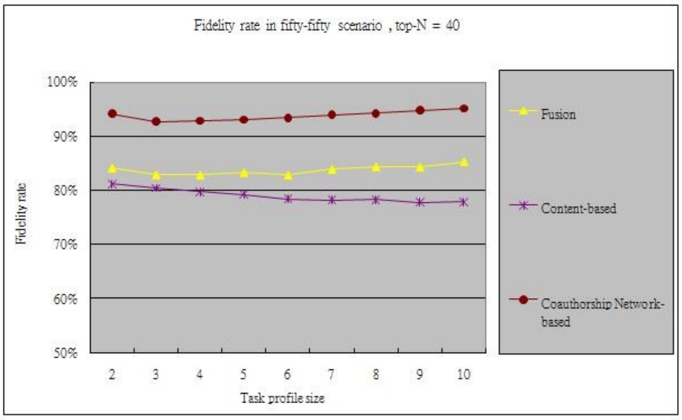


Figure 8. The fidelity rate of fusion in fifty-fifty scenario

5 CONCLUSIONS

In this study, we have investigated the hybrid approach that utilizes both content features and coauthorship network for making task-focused recommendation in literature databases. We then compare the various proposed methods using a data set that contains articles in prestigious data mining conferences and journals. The experimental results reach the following conclusions:

1. When a task profile is small and its articles exhibit high similarity in their contents, i.e., the close scenario, the proposed hybrid methods achieves the highest hit rate. When task profile size increases, coauthorship network-based approach becomes the best, though the hybrid methods exhibit only slightly inferior performance.
2. When articles of a task profile are dissimilar in their content, i.e., the diversified scenario, the hit rates of the three hybrid methods are almost the same and equally good as the content-based method. But fusion method performs the best when the recommendation list is short.
3. In the combination of close scenario and diversified scenario, i.e., the fifty-fifty scenario, the performances of the three hybrid methods are close and all better than content-based and coauthorship network-based approaches.
4. The content-based approach is more likely to recommend articles of low fidelity, whereas the coauthorship network-based approach has the least chance. The hybrid approach performs in between in terms of fidelity of recommended articles.

Overall we conclude that the proposed hybrid methods in general yield equal or better recommendation accuracy than the best of the primitive approaches under all circumstances. Of the three hybrid methods, fusion method performs the best for short recommendation list. In terms of the fidelity of the recommended articles, the hybrid methods are second to the coauthorship network-based approach but better than the content-based approach.

The coauthorship network constructed in this work does not consider overlapping research areas between coauthors. Such information, when incorporated into the recommendation mechanism, may further improve recommendation accuracy. Besides, some of the previous works in task-focused recommendation make use of Web usage log. Future works may consider including usage log, in addition to content and coauthorship network, in the design of task-focused recommendation for literature databases.

References

Acedo, F.J., Barroso,C., Casanueva, C. and Galán, J..L. Co-authorship in management and organizational studies: An empirical and network analysis. *Journal of Management Studies*, 43, 5, 2006, 957-983.

Adomavicius, G., & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734-749.

Barabasi, A. L. *Linked: the new science of networks.* 1st edition, Basic Books, 2002.

Golbeck, J. and Hendler, J. (2006) Inferring binary trust relationships in web-based social networks. *ACM Transactions on Internet Technology*, 6(4), 497-529.

Hanneman, R. A., & Riddle, M. (2005). Introduction to Social Network Methods. (http://faculty.ucr.edu/~hanneman/nettext/)

Herlocker, J. & Konston, J. (2001). Content-Independent Task-Focused Recommendation, IEEE Internet Computing, 5(6), 40-47.

Hwang, S. Y. and Chuang, S. M. (2004). Combining article content and web usage for literature recommendation in digital libraries. Online Information Review, 28, 4, 260-272.

Hwang, S.-Y., Wei,C.-P. and Chen,L.-S. (2010). Coauthorship Networks and Academic Literature Recommendations. To appear in Electronic Commerce Research and Applications, *available at* http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6X4K-4Y3KSJX-1&_user=1196423&_coverDate=01/06/2010&_rdoc=1&_fmt=high&_orig=search&_sort=d&_docanchor=&view=c&_searchStrId=1221775117&_rerunOrigin=google&_acct=C000051947&_version=1&_urlVersion=0&_userid=1196423&md5=f6da197e2f0f45cd1d2070014551d22f

Lam, C. (2004). SNACK: Incorporating Social Network Information in Automated Collaborative Filtering. Proceedings of the 5th ACM Conference on Electronic Commerce, 254-255.

Linden, G., Smith, B., & York, J. (2003). Amazon.com Recommendations: Item-to-Item Collaborative Filtering. IEEE Internet Computing, 7(1), 76-80.

Miller, B. N., Albert, I., Lam, S. K., Konstan, J. A., & Riedl, J. (2003). MovieLens Unplugged: Experiences with an Occasionally Connected Recommender System. Proceedings of International Conference on Intelligent User Interfaces.

Mobasher, B., Dai, H., Luo, T., Sung, Y., & Zhu, J. (2000). Integrating Web Usage and Content Mining for More Effective Personalization. Proceedings of the International Conference on E-Commerce and Web Technologies, 165-176.

Newman, M. E. J. (2001). Scientific Collaboration Networks. I. Network Construction and Fundamental results. American Physical Society Journals, 64(1), 016131.

Richardson, M., Agrawal, R. and Domingos P.(2003) Trust management for the semantic web, *Proceedings of 2nd International Conference on Semantic Web*. Sanibel Island, FL., 351-368.

Salton, G. (1986), Another look at automatic text-retrieval systems. Communications of the ACM 29(7) ,648-56

Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. N. (2000). Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. *Proceedings of the ACM SIGKDD Explorations, 1*(2), 12-33.

Torres, R., McNee S. M., Abel, M., Konstan, J. A., & Riedl, J. (2004). Enhancing digital libraries with TechLens+. *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, 228-236.

Wasserman, S., & Faust, K. (1994). *Social network analysis: methods and applications*. New York: Cambridge University Press.

Yang, Q., Zhang, H. H., & Li, T. (2001). Mining Web Logs for Prediction Models in WWW Caching and Prefetching. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 473-478.

Ziegler, C. and Lausen., G. Spreading activation models for trust propagation. *Proceedings of IEEE International Conference on e-Technology, e-Commerce, and e-Service*, Taipei, Taiwan, 2004.