RECOMMENDING DOCUMENTS VIA KNOWLEDGE FLOW-BASED GROUP RECOMMENDATION

Chin-Hui Lai[1], Duen-Ren Liu[2] and Ya-Ting Chen[2]
[1]*Department of Information Management, Chung Yuan Christian University, Chung-Li, Tao Yuan County, Taiwan*
[2]*Institute of Information Management, National Chiao Tung University, Hsinchu, Taiwan*

Keywords: Collaborative filtering, Group recommendation, Document recommendation, Knowledge flow.

Abstract: Recommender systems can mitigate the information overload problem and help workers retrieve knowledge based on their preferences. In a knowledge-intensive environment, knowledge workers need to access task-related codified knowledge (documents) to perform tasks. A worker's document referencing behaviour can be modelled as a knowledge flow (KF) to represent the evolution of his/her information needs over time. Document recommendation methods can proactively support knowledge workers in the performance of tasks by recommending appropriate documents to meet their information needs. However, most traditional recommendation methods do not consider workers' knowledge flows and the information needs of the majority of a group of workers with similar knowledge flows. A group's needs may partially reflect the needs of an individual worker that cannot be inferred from his/her past referencing behaviour. Thus, we leverage the group perspective to complement the personal perspective by using a hybrid approach, which combines the KF-based group recommendation method (KFGR) with the user-based collaborative filtering method (UCF). The proposed hybrid method achieves a trade-off between the group-based and the personalized method by integrating the merits of both methods. Our experiment results show that the proposed method can enhance the quality of recommendations made by traditional methods.

1 INTRODUCTION

Because of the rapid development of information technologies in recent years, it is now relatively easy to access knowledge resources. In knowledge-intensive environments, knowledge workers need to access task-related codified knowledge (documents) to perform tasks. However, the huge volumes of documents that exist in various knowledge domains often lead to information overload. Thus, there is a need for document recommendation methods that support knowledge workers as they perform tasks by recommending appropriate documents to suit their information needs, i.e., task needs.

Workers may have various information needs when executing tasks. Because each worker's information needs may change over time, we model a worker's document referencing behaviour for a specific task as a knowledge flow (KF) to represent the evolution of his/her information needs (Lai and Liu, 2009). From the personal perspective, a worker's KF is derived from his/her past referencing behaviour to represent his/her personal needs. The topics and documents included in the KF are related to the worker's specific personal needs. From the group perspective, the information needs of the majority of the group's members are more important than those of individual members. A group's needs may partially reflect the needs of an individual worker that cannot be inferred from his/her past referencing behaviour. In other words, the group's knowledge complements that of the individual worker.

Recommender systems (Konstan et al., 1997, Balabanovic and Shoham, 1997) can alleviate the information overload problem and help workers identify and retrieve needed documents based on their preferences or information needs. However, the referencing behaviour of knowledge workers may vary over time, but most recommendation methods do not consider workers' KFs. Because traditional recommendation methods focus on personalized recommendations and have some limitations, several group-based recommendation methods have been proposed (Jameson, 2004, McCarthy and Anagnost, 1998, O'Connor et al., 2001). Existing group

Lai C., Liu D. and Chen Y..
RECOMMENDING DOCUMENTS VIA KNOWLEDGE FLOW-BASED GROUP RECOMMENDATION.
DOI: 10.5220/0003486903410349
In *Proceedings of the 6th International Conference on Software and Database Technologies* (ICSOFT-2011), pages 341-349
ISBN: 978-989-8425-77-5

recommendation schemes satisfy the information needs of most workers in a group, but they often neglect individual workers' preferences and do not consider recommendations in the context of a KF environment.

In this work, we propose a hybrid recommendation method that combines a KF-based group recommendation (KFGR) method with traditional collaborative filtering method. The traditional recommendation method focuses on the personal perspective rather than the group perspective; however, the group's information needs may be important because they partially reflect an individual's needs. In other words, the group's knowledge may complement that of the individual worker. Therefore, we take the group perspective into consideration to offset the drawback of the personal perspective. The KFGR method is a novel recommendation method which takes workers' KFs and their personal preferences into account to recommend documents for a group of workers with similar KFs. The drawback of the group perspective is that it may not satisfy the information needs of some individuals, since it focuses on the needs of the majority of group members. To resolve the problem, we combine the KFGR method with traditional recommendation method, i.e., collaborative filtering, to enhance the quality of recommendations. The proposed hybrid method achieves a trade-off between the group-based and personalized methods by combining the merits of both methods. The experiment results show that the proposed method can improve on the quality of recommendations provided by traditional recommendation methods.

The remainder of this paper is organized as follows. Section 2 contains a review of related works. In Section 3, we describe the KF model and the proposed hybrid recommendation method. In Section 4, we detail the experiment results and discuss their implications. Section 5 contains some concluding remarks.

2 RELATED WORK

2.1 Knowledge Flow

Knowledge flows among people and processes facilitate knowledge sharing and reuse. The concept of knowledge flows has been applied in various domains, e.g., scientific research, communities of practice, teamwork environments, industry, and organizations (Zhuge, 2006). KM enhances the effectiveness of teamwork by accumulating and disseminating knowledge among team members to facilitate peer-to-peer knowledge sharing (Zhuge, 2002). Luo et al. (2008) introduced the concept of textual knowledge flows based on the management of knowledge maps. In an organization, knowledge workers normally have various information needs over time when performing tasks. Thus, we define a knowledge flow from the perspective of a worker's information needs to represent the evolution of referencing behaviour and the knowledge accumulated for a specific task (Lai and Liu, 2009). Then, the KF-based recommendation methods are proposed for recommending task-related codified knowledge.

2.2 Information Retrieval and Task-based Knowledge Support

A knowledge worker may acquire knowledge from a large number of documents. Since the documents can reveal the information needs of the knowledge worker, we need to filter the documents by using information retrieval (IR) techniques, which enable us to access specific items of information (Baeza-Yates and Ribeiro-Neto, 1999).

Information filtering with a similarity-based approach is often used to locate knowledge items relevant to the task-at-hand. The discriminating terms of a task are usually extracted from a knowledge item/task to form a task profile, which is used to model a worker's information needs. For example, Holz et al. (2005) proposed a similarity-based approach to organize desktop documents and proactively deliver task-specific information; while Liu et al. (2005) presented a *K*-Support system to provide effective task support for a task-based working environment.

2.3 Recommendation

2.3.1 Collaborative Filtering

Collaborative filtering (CF) is widely used in recommender systems. CF recommends various items, such as products, movies, and documents, based on the preferences of people who have the same or similar interests to those of the target user. The approach involves two steps: neighbourhood formation and prediction. The neighbourhood of a target user is selected according to his/her similarity to other users, and is computed by Pearson's correlation coefficient or the cosine similarity measure. Either the *k*-NN (nearest neighbours) approach or a threshold-based approach is used to choose n users that are most similar to the target user. We use a threshold-based approach in this paper.

2.3.2 Group-based Recommendation

Group recommender systems are used in various application domains, such as those that recommend music, movies, TV programs and tourist attractions. Generally, such systems can be classified as (1) those that aggregate individual users' profiles/preferences to form a group's profile/preferences (McCarthy and Anagnost, 1998); and (2) those that merge individual recommendation lists into a group recommendation list (O'Connor et al., 2001, McCarthy and Anagnost, 1998, Kim et al., 2010). Under the first approach, there is a high probability of discovering valuable recommendations that will satisfy the majority of the group's members. The second approach gives users more information when they need to make decisions and the recommendation results are relatively easy to explain. However, it is not easy to identify unexpected items, and it is very time-consuming if the group is large. Therefore, we follow the first approach and aggregate workers' topic domains based on their knowledge flows to generate profiles for a group.

3 HYBRID PERSONALIZED AND GROUP-BASED METHOD

3.1 Overview

In a knowledge intensive environment, a high degree of knowledge sharing can have a significant effect on the workers' efficiency. Each worker accumulates knowledge when he/she executes a task, and that knowledge can be shared with and reused by other team members with similar information needs. In this paper, we propose a personalized group-based recommendation method, i.e. KFGR-UCF, to facilitate knowledge sharing among a group of workers. The method combines the KF-based group recommendation method (KFGR) and user-based collaborative filtering method (UCF) to enhance the quality of document recommendation.

The rationale behind the proposed model is that a group's information needs may partially reflect an individual member's information needs that cannot be inferred from his/her past document referencing behaviour. In other words, the group's knowledge can be used to satisfy the individual member's needs. Thus, the group-based method can complement the personalized method. However, the group perspective may neglect the specific information needs of an individual, because it focuses on the information needs of the majority of the group's members. To resolve this problem, our hybrid recommendation method combines the merits of the two approaches to improve the recommendation quality. The group-based method recommends documents from the perspective of the majority's information needs, while the personalized methods recommend documents according to the specific needs of an individual.

The proposed recommendation method is comprised of three phases: 1) compiling individual knowledge flows (codified-level KFs and topic-level KFs); 2) grouping knowledge workers and generating group profiles; and 3) recommending documents to workers.

The first phase involves three steps: document profiling, document clustering, and KF generation. To accomplish tasks, knowledge workers may need to access various documents, and those documents can reflect the workers' preferences or requirements in different periods. We align the documents in a sequence, called a codified-level KF. Each document in the sequence is represented as an n-dimensional vector comprised of key terms in the document and their weights. Next, we cluster the documents into several topics based on their cosine similarity scores. To observe the evolution of information needs, we generate a topic-level KF (TKF) as a topic sequence by mapping the documents in the codified-level KF into corresponding clusters (topics).

In the second phase, we group similar knowledge workers into groups by using a KF similarity measure derived from the alignment similarity and aggregate profile similarity (Lai and Liu, 2009). The KF similarity score indicates whether the referencing behaviour of two workers is similar. After grouping the workers, each group's important codified knowledge can be elicited from the topics accessed by the group members. We compile group profiles to represent each group's important knowledge.

In the last phase, we propose a hybrid of KF-based group recommendation and user-based CF (KFGR-UCF), which considers both the group and personal perspectives, to recommend suitable documents to knowledge workers. The group-based approach derives a group-based score (preference) of a group, k, for a target document based on the topic-level KFs of the group's members. Note that similar documents are grouped into clusters (topics), so topic-level KFs should provide a larger number of related documents to satisfy workers' task needs than codified-level KFs. Thus, the group-based approach employs the topic-level KF to predict a group's ratings on documents.

3.2 Knowledge Flow Model

A worker's knowledge flow (KF) represents the evolution of his/her information needs and preferences during a task's execution (Lai and Liu, 2009). Workers' KFs are identified by analyzing their knowledge referencing behaviour based on their historical work logs, which contain information about previously executed tasks, task-related documents and the accessed time of documents.

A KF comprises two levels: a codified level and a topic level. The knowledge in the codified-level indicates the knowledge flow between documents based on the access time. In most situations, the knowledge obtained from one document prompts a knowledge worker to access the next relevant document (codified knowledge). Hence, the task-related documents are sorted in order of the times they were accessed to obtain a document sequence as the codified-level KF.

Documents with similar concepts and access times are grouped together automatically to form a topic-level abstraction of the task knowledge. Note that each topic may contain several task-related documents. The codified-level KF is abstracted to form a topic-level KF, which represents the transitions between various topics. Since the task knowledge in the topic level may flow between topics, it could prompt the worker(s) to retrieve knowledge from the next related topic.

3.3 Document Profile Generation

Two profiles, a document profile and a topic profile, are used to represent a worker's KF. A document profile can be represented as an *n*-dimensional vector comprised of the key terms in the document and their respective weights derived by the normalized *tf-idf* approach. Based on the term weights, terms with higher values are selected as discriminative terms to describe the characteristics of the document. The document profile d_j is comprised of these discriminative terms. Let the document profile be DP_j=<$dt_{1j}:dtw_{1j},dt_{2j}:dtw_{2j},\cdots, dt_{nj}:dtw_{nj}$>, where dt_{ij} is a term i in d_j and dtw_{ij} is the degree of importance of the term i to the document d_j, which is derived by the normalized *tf-idf* approach. The document profiles are used to measure the similarity of the documents

3.4 Knowledge Flow Mining and Extraction

When performing a task in a knowledge-intensive and task-based environment, a worker usually requires a large amount of task-related knowledge to accomplish the task. By analyzing a worker's referencing behaviour for a specific task, the corresponding knowledge flow of the task is derived by a knowledge flow extraction method. For a specific task, the method derives two kinds of KFs, a *codified-level KF* and a *topic-level KF*, to represent the worker's information needs. Each worker has his/her own codified-level KF, which represents his/her accumulated knowledge for a specific task at the codified level.

The topic-level KF, which is derived by clustering documents with similar content and access times in the codified-level KF, is represented by a topic sequence. Based on the order of documents in each worker's codified-level KF, documents with similar content are grouped into clusters by using a hierarchical agglomerative clustering method with a time variant (HACT) algorithm. When clustering a series of time-ordered documents, i.e., the codified-level KF, the algorithm considers the documents' contents as well as the times the documents were accessed.

We adopt the average linkage hierarchical clustering method (Jain et al., 1999) to group documents that have similar profiles and are within the same time window into clusters by using the cosine measure to calculate the similarity between the profiles of two documents. Then, the clustering result with the best quality is selected to derive the topic-level KF. Note that a cluster represents a topic set and has a topic profile (derived from the document cluster), which describes the features of the topic.

Topic Profile Generation

Documents in the same cluster contain similar content and form a topic set. The key features of the cluster are described by a topic profile derived from the profiles of documents in the cluster. Let $TPf_x =< tt_{1x} : ttw_{1x}, tt_{2x} : ttw_{2x}, \cdots, tt_{nx} : dtw_{nx} >$ be the profile of a topic (cluster) x, where tt_{ix} is a topic term and ttw_{ix} is the weight of the topic term.

3.5 Grouping Knowledge Workers and Generating Group Profiles

To find a target worker's neighbours, we compare his/her topic-level KF with those of other workers to compute the similarity of their KFs. Such similarity measurement is used to indicate whether the KF referencing behaviour of two workers is similar. Since each KF is a sequence, the sequence alignment method (Oguducu and Ozsu, 2006), which computes the cost of aligning two sequences, can be used to measure the similarity of two KF sequences. Based on this concept, we use a hybrid similarity measure, comprised of the KF alignment similarity and the aggregated profile similarity, to evaluate the similarity of two workers' KFs (Lai and Liu, 2009).

3.5.1 Building Group Profiles

The members of a group have similar KFs because their information needs are similar; and they usually need to refer to related documents for a specific topic. Thus, the group-based approach derives the group-based score (preference) of a group k for a target document based on the topic-level KFs (TKFs) of the group's members. Since similar documents are grouped into clusters (topics), a larger number of related documents that may satisfy workers' task needs can be recommended by considering topic-level KFs rather than codified-level KFs. We identify the important topics that the members accessed and compute their weights based on each member's KF (Eq. (1)). Let $GTR_{k,x}$ be group k's accumulated rating for topic x, which indicates the weight of topic x in group k. In addition, let T_u be the set of topics in the topic-level KF of user u, and let U_k be the set of users in group k. $GTS_k = \bigcup_{u \in U_k} T_u$ is the set of topics accessed by members of group k.

$$GTR_{k,x} = \frac{\sum_{u \in U_k} PTR_{u,x}}{|U_k|} \quad (1)$$

where $|U_k|$ is the number of workers in the group. $PTR_{u,x}$ is the personal rating of worker u for topic x , indicating the importance of topic x to worker u. The rating is derived by Eq. (2) based on u's topic-level knowledge flow, assuming that topic y_t is the topic accessed by u at time index t.

$$PTR_{u,x} = \frac{\sum_{t=1}^{t_{now}} \overline{TR}_{u,y_t} \times tw_{t,t_{now}}^{u,y_t} \times csim(TPf_x, TPf_{y_t})}{\sum_{t=1}^{t_{now}} tw_{t,t_{now}}^{u,y_t} \times csim(TPf_x, TPf_{y_t})} \quad (2)$$

where $\overline{TR}_{u,y_t}$ is the average rating of worker u for topic y_t; $\overline{TR}_{u,y_t}$ is derived by averaging the ratings of worker u for documents belonging to topic y_t. TPf_x / TPf_y is the topic profile of topic x / topic y_t described in Section 3.4; and $csim(TPf_x, TPf_y)$ is the profile similarity between topic x and topic y_t measured by the cosine formula. In addition, $tw_{t,t_{now}}^{u,y_t}$ is the time weight of topic y_t accessed by worker u at time t. It is defined as $tw_{t,t_{now}}^{u,y_t} = \frac{t - St}{t_{now} - St}$, where St is the start time of the worker's KF and t_{now} is the time the worker accessed the most recent topic in his/her KF.

Based on Eq. (1), we can derive the group's ratings for topics based on the members' personal ratings for those topics. A higher $GTR_{k,x}$ score means that the topic x is more important to group k.

3.6 Recommendation Phase

This phase combines the KF-based group recommendation method (KFGR) with the personalized methods to generate recommendation lists for workers. In the following sub-sections, we discuss KFGR and the hybrid method, i.e., the KFGR-UCF method.

3.6.1 The KFGR Method

Some topics may be of interest or important to the majority of the group's members. Since documents related to those topics will probably satisfy the workers' information needs, the proposed group-based approach considers the importance of the topics accessed by group members. Let $\overline{Gr}_{k,i}$ be the group rating based on the document ratings in knowledge flows of group members, as shown in Eq. (3).

$$\overline{Gr}_{k,i} = \frac{\sum_{u=1}^{M} (r_{u,i} \times tw_{t,t_{now}}^{u,i})}{\sum_{u=1}^{M} tw_{t,t_{now}}^{u,i}} \quad (3)$$

where $r_{u,i}$ is worker u's rating for document i, and $tw_{t,t_{now}}^{u,i}$ is the time weight of document i that worker u gives it rating at time t. The value of $\overline{Gr}_{k,i}$ is derived from the personal ratings of group k's members for document i. It is a weighted average group rating of group k for document i derived by considering its document ratings given by group members and its time factors in members' knowledge flows.

Moreover, group members may access and rate the target documents, so we also take the members' ratings into account to obtain the predicted rating of a document in a group. Let $GDR_{k,i}$ be the predicted group rating of group k for a target document i, as shown in Eq. (4). The value of $GDR_{k,i}$ is derived from linearly combing two parts: group rating based on the document ratings of group members and group rating based on the topic-level KF (TKF). The group rating based on the document ratings of group members is obtained by the group members' ratings for document i. The group rating based on the TKF is the weighted sum of group k's ratings on topics by using the similarity measures of the topics to the target document as the weights.

$$GDR_{k,i} = Aw_{k,i} \times \overline{Gr}_{k,i} + (1 - Aw_{k,i}) \times \frac{\sum_{x \in GTS_k} csim(TPf_x, DPf_i) \times GTR_{k,x}}{\sum_{x \in GTS_k} csim(TPf_x, DPf_i)} \quad (4)$$

where $GTR_{k,x}$ is the predicted group rating of group k for topic x measured by Eq. (1); TPf_x is the profile (term vector) of topic x; DPf_i is the profile (term vector) of document i; GTS_k is the topic set of group k; and $\overline{Gr}_{k,i}$ is the weighted average group rating of group k for document i derived by considering the time factor, as shown in Eq. (3).

$Aw_{k,i}$ is the activity weighting of group k for document i, and is defined as Eq. (5).

$$Aw_{k,i} = \begin{cases} \beta + (1-\beta) \times min(\frac{2 \times |M_{k,i}|}{|Gr_k|}, 1), \; if \; |M_{k,i}| > 0 \\ 0, \; if \; |M_{k,i}| = 0 \end{cases} \quad (5)$$

where $|M_{k,i}|$ is the number of group members that rated the target document i; $|Gr_k|$ is the number of members in group k; and β is an adjusting weight determined by the experimental analysis.

The value of $Aw_{k,i}$ is in the range of 0 to 1. It will be high if most of group members rate the document i, implying that $\overline{Gr}_{k,i}$ is reliable for representing group k's rating on document i. That is, the group rating based on the document ratings of group members (i.e., $\overline{Gr}_{k,i}$) will contribute more to the predicted group rating, i.e., $GDR_{k,i}$. On the contrary, if a few group members rate the document i, the value of $Aw_{k,i}$ will be small. Thus, the group rating based on TKFs will contribute more to the predicted group rating.

Here, we consider the ratings of group members who have rated the target document and the predicted group rating for the document. The latter is derived as the weighted sum of group k's ratings for topics in GTS_k by using the cosine similarity between the profiles of the target document and topics as the weights.

3.6.2 The Hybrid KFGR-UCF Method

In this section, we linearly combine the KFGR method with user-based CF (UCF) to recommend documents to a target worker. The recommendation list is generated by combining the predicted ratings of KFGR and UCF. As mentioned earlier, KFGR uses the group's information needs based on the members' KFs to make recommendations. It recommends a group's preferred documents to a target worker, and considers the group members' preferences (i.e. ratings on target documents) as well as the group's accumulated ratings on topics. Meanwhile, the UCF method recommends documents to a target worker based the ratings of workers with similar information needs. The similarity between workers is determined by calculating Pearson's correlation coefficient based on the workers' ratings for documents. Thus, the predicted rating of a document is obtained from neighbours who have similar preferences to the target worker and whose similarity scores are higher than a threshold θ. To improve the performance of the KFGR and UCF recommendation methods, we combine them linearly. Based on the hybrid method, the predicted rating of worker a for document i, $PDR_{a,i}$, is derived by Eq. (6).

$GDR_{k,i}$ is the predicted rating of group k for document i based on Eq. (4); $Psim(R_a, R_u)$ is Pearson's correlation coefficient between user a and user u measured by their rating vectors R_a and R_u; $\overline{r}_a$ and $\overline{r}_u$ are the average ratings of worker a and worker u respectively; $r_{u,i}$ is the rating given by worker u for document i; and $\alpha_{KFGR\text{-}UCF}$ is a parameter used to adjust the weight between group-based prediction and user-based CF prediction.

$$PDR_{a,i} = \alpha_{KFGR-UCF} \times GDR_{k,i} + (1 - \alpha_{KFGR-UCF}) \times \left(\overline{r}_a + \frac{\sum_{u \in Neighbor(a)} Psim(R_a, R_u) \times (r_{u,i} - \overline{r}_u)}{\sum_{u \in Neighbor(a)} Psim(R_a, R_u)} \right) \quad (6)$$

The value of $\alpha_{KFGR\text{-}UCF}$ is between 0 and 1. It is derived from conducting experiments by systematically adjusting its values in an increment of 0.1. When the value of $\alpha_{KFGR\text{-}UCF}$ is 1, $PDR_{a,j}$ is mainly derived by the KFGR method. That is, the recommendations are totally dominated by the group preferences. In contrast, when the value of $\alpha_{KFGR\text{-}UCF}$ is 0, $PDR_{a,i}$ is mainly derived by UCF method. This means that the recommendation is dominated by personal interests. Thus, the optimal value (i.e., the lowest MAE value) was chosen as the best setting. Based on the predicted ratings derived by Eq. (6), documents with high ratings are used to compile a recommendation list. Then, the top-*N* documents are recommended to the target worker.

4 EXPERIMENTS AND EVALUATIONS

A number of experiments were conducted to evaluate the proposed hybrid method. We discuss the experiment setup and the results in Sections 4.1 and 4.2 respectively.

4.1 Experiment Setup

We collected the data for the experiments from a laboratory in a research institute. The dataset is comprised of over 600 documents that had been accessed by about 60 workers. It also includes usage logs, which provide information about the workers' access behaviour, i.e., browsing, rating, downloading, and uploading documents. The log data is used to analyze the preferences of each user. In the laboratory environment, each worker has to complete a research task during a set time period; thus, he/she needs to access task-related documents (research papers). We can discover the workers' knowledge flows from their usage logs. The ratings given to documents on a scale of 1 to 5 indicate their relevance and usefulness to the worker's task. Then, we divide the data set into two parts: 70% for training and 30% for testing.

To measure the recommendation quality of the methods, we use the Mean Absolute Error (MAE) which is widely used in recommender systems (Breese et al., 1998, Herlocker et al., 2004). MAE measures the average absolute deviation of the predicted rating and the true rating. The lower the MAE score, the better the accuracy of the recommendation method. The MAE is derived by Eq. (7):

$$MAE = \frac{\sum_{i}^{N} \left| \hat{P}_i - r_i \right|}{N} \tag{7}$$

where N is the number of documents, $\hat{P}_i$ is the predicted rating of document *i*, and r_i is the real rating of document *i* given by the user.

4.2 Experiment Results

In the following sub-sections, we will discuss how to determine the parameters used in the experiments, and compare the performance of the proposed method and the traditional methods.

4.2.1 The Analysis of β

In this experiment, we will discuss how to determine the value of the activity weighting β (Eq. (5)) for the KFGR method. The KFGR method described in Section 3.6.1 is a hybrid method which linearly combines two parts of group ratings by using an activity weighting. One part is the group rating based on the TKFs, while the other part is the group rating based on the weighted average ratings of topics, as shown in Eq. (2) and Eq. (3) respectively. Because group members' information need may change over time, these two parts also takes the time factor into account. To combine these two parts, the activity weighting is derived from the majority opinion of group members on documents, i.e., Eq.(4), and is used to adjust the relative importance between these two parts.

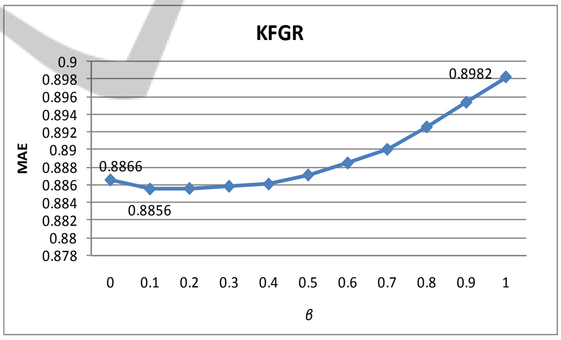

Figure 1: The MAE values under different β for KFGR.

For the KFGR, the activity weighting β is a decimal which ranges from 0 to 1. The other parameter (1-β) is the weight to adjust the activity weight by considering how many group members who have accessed the target document. To obtain the best MAE score, we systematically adjust the values of β in increments of 0.1 for the KFGR, as shown in Figure 1.

For the activity weight of the KFGR method, the lowest MAE occurs when β is 0.1. Thus, we set β=0.1 for the activity weighting of the KFGR method to predict document ratings. When β is 0, the activity weighting is totally derived from such majority ratio. However, when β is 1, the activity weight is also equal to 1 too. The predicted rating of

KFGR is totally derived from the group members' ratings based on TKFs, i.e. Eq. (3).

4.2.2 The Analysis of Time Factor and Activity Weighting

In this experiment, we compare KFGR, KFGR-NT, KFGR(AW=1) and KFGR-NT(AW=1) to analyze the effects of considering the time factor and the activity weighting in KFGR and KFGR-NT methods respectively, as shown in Figure 2. Both KFGR and KFGR(AW=1) methods take the time factor into account to obtain the group ratings. In the KFGR(AW=1) method, the activity weighting is set as 1 for the predicted ratings of documents. Similarly, both the KFGR-NT and KFGR-NT(AW=1) methods do not consider the time factor.

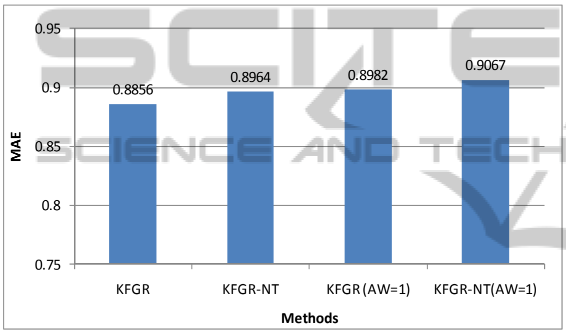

Figure 2: Comparison of KFGR and KFGR-NT.

From Figure 2, KFGR, which considers the time factor, outperforms KFGR-NT. Also, when setting the activity weighting as 1, the performance of KFGR (AW=1) is better than the KFGR-NT (AW=1). The KFGR method is more capable of satisfying users' information needs. In addition, the KFGR outperforms KFGR (AW=1), while KFGR-NT outperforms KFGR-NT (AW=1). Thus, considering the activity weighting based on the majority ratio is effective in improving the recommendation quality. The KFGR method has the best performance of recommendation. In the following experiments, we consider the time factor in KFGR, and assess the performance of the proposed hybrid methods.

4.2.3 Evaluation of the Hybrid KFGR-UCF Method

Here, we evaluate the performance of UCF and the hybrid KFGR-UCF. We first determine the value of the parameter $\alpha_{KFGR\text{-}UCF}$ for the hybrid KFGR-UCF method. The parameter is used to adjust the relative importance of KFGR and UCF, whose value ranges from 0 to 1. When $\alpha_{KFGR\text{-}UCF}$ is 0, the predicted rating is derived entirely by the UCF method; otherwise, when $\alpha_{KFGR\text{-}UCF}$ is 1, the predicted rating is derived entirely by the KFGR method.

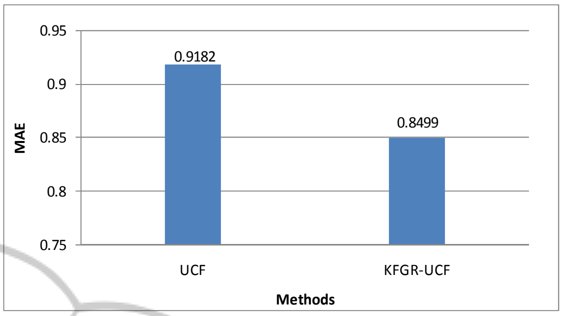

Figure 3: Comparison of UCF and KFGR-UCF.

To obtain the best MAE, we systematically adjust the value of $\alpha_{KFGR\text{-}UCF}$ in increments of 0.1. The optimal MAE value (0.8499) is generated by setting $\alpha_{KFGR\text{-}UCF}$ at 0.6. The importance weight of KFGR is 0.6, while that of UCF is 0.4. That is, the KFGR method is relatively more important than the UCF method in the hybrid of KFGR-UCF. The bar chart in Figure 3: compares the performance of UCF and KFGR-UCF. Since the KFGR-UCF clearly outperforms UCF, we conclude that the hybrid KFGR-UCF method improves the recommendation quality. More specifically, it is capable of predicting the information needs of individual users from a group's perspective.

5 CONCLUSIONS

We have proposed a hybrid KFGR-UCF method which combines the KF-based group recommendation method (KFGR) with the user-based collaborative filtering method (UCF) to enhance the quality of recommendations. Our method recommends documents from two perspectives, i.e., a group perspective and a personal perspective. From the personal perspective, some documents are only relevant to a worker's specific information needs, i.e., they are not related to the group's information needs. A member's personal information needs are derived from his/her previous referencing behaviour. From the group perspective, there are some documents that most group members consider relevant. The group's information needs may partially reflect an individual member's information needs that cannot be inferred from his/her past referencing behaviour; hence, the group's knowledge can complement the individual member's knowledge. In this work, we take the group perspective into consideration to offset the drawback of the personal perspective. However, the group perspective may neglect the information needs of an individual because it focuses on the needs of the majority of the group's members. Since the group-based method and the personalized method have distinct advantages, we combined them to exploit their respective merits. Our experiment results show that the hybrid method certainly improve the recommendation quality.

ACKNOWLEDGEMENTS

This research was supported by the National Science Council of the Taiwan under the grant NSC 96-2416-H-009-007-MY3 and NSC 99-2410-H-009-034-MY3.

REFERENCES

Baeza-Yates, R. & Ribeiro-Neto, B. 1999. *Modern Information Retrieval,* Boston, Addison-Wesley.

Balabanovic, M. & Shoham, Y. 1997. Fab: content-based, collaborative recommendation. *Communication of the ACM,* 40**,** 66-72.

Breese, J. S., Heckerman, D. & Kadie, C. 1998. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In: *Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence,* 43-52.

Herlocker, J. L., Konstan, J. A., Terveen, L. G. & Riedl, J. T. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS),* 22**,** 5-53.

Holz, H., Maus, H., Bernardi, A. & Rostanin, O. 2005. A lightweight approach for proactive, task-specific information delivery. In: *Proceedings of the 5th International Conference on Knowledge Management (I-Know)*, 101-127.

Jain, A. K., Murty, M. N. & Flynn, P. J. 1999. Data clustering: a review. *ACM Computing Surveys (CSUR),* 31**,** 264-323.

Jameson, A. 2004. More than the sum of its members: challenges for group recommender systems. In: *Proceedings of the working conference on Advanced visual interfaces,* Gallipoli, Italy, 48-54.

Kim, J. K., Kim, H. K., Oh, H. Y. & Ryu, Y. U. 2010. A group recommendation system for online communities. *International Journal of Information Management,* 30**,** 212-219.

Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R. & Riedl, J. 1997. GroupLens: applying collaborative filtering to Usenet news. *Communications of the ACM,* 40**,** 77-87.

Lai, C. H. & Liu, D. R. 2009. Integrating Knowledge Flow Mining and Collaborative Filtering to Support Document Recommendation. *Journal of Systems and Software,* 82**,** 2023-2037.

Liu, D. R., Wu, I. C. & Yang, K. S. 2005. Task-based K-Support system: disseminating and sharing task-relevant knowledge. *Expert Systems With Applications,* 29**,** 408-423.

Luo, X., Hu, Q., Xu, W. & Yu, Z. 2008. Discovery of Textual Knowledge Flow Based on the Management of Knowledge Maps. *Concurrency and Computation: Practice and Experience,* 20**,** 1791-1806.

Mccarthy, J. F. & Anagnost, T. D. 1998. MusicFX: an arbiter of group preferences for computer supported collaborative workouts. In: *Proceedings of the ACM conference on computer supported cooperative work (CSCW)*, Seattle, Washington, United States, 363-372.

O'connor, M., Cosley, D., Konstan, J. A. & Riedl, J. 2001. PolyLens: a recommender system for groups of users. In: *Proceedings of the seventh conference on European Conference on Computer Supported Cooperative Work*, Bonn, Germany, 199-218.

Oguducu, S. G. & Ozsu, M. T. 2006. Incremental click-stream tree model: Learning from new users for web page prediction. *Distributed and Parallel Databases,* 19**,** 5-27.

Zhuge, H. 2002. A Knowledge Flow Model for Peer-to-peer Team Knowledge Sharing and Management. *Expert Systems With Applications,* 23**,** 23-30.

Zhuge, H. 2006. Discovery of Knowledge Flow in Science. *Communications of the ACM,* 49**,** 101-107.