Available online at www.sciencedirect.com
ScienceDirect
Int. J. Human-Computer Studies 69 (2011) 587–601
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Sequence-based trust in collaborative filtering for document recommendation

Duen-Ren Liu[a], Chin-Hui Lai[b,*], Hsuan Chiu[a]

[a]Institute of Information Management, National Chiao Tung University, Hsinchu, Taiwan
[b]Department of Information Management, Chung Yuan Christian University, Chungli, Taoyuan County, Taiwan

Received 14 December 2009; received in revised form 26 May 2011; accepted 1 June 2011
Communicated by E. Motta
Available online 12 June 2011

Abstract

Collaborative filtering (CF) recommender systems have emerged in various applications to support item recommendation, which solve the information-overload problem by suggesting items of interest to users. Recently, trust-based recommender systems have incorporated the trustworthiness of users into CF techniques to improve the quality of recommendation. They propose trust computation models to derive the trust values based on users' past ratings on items. A user is more trustworthy if s/he has contributed more accurate predictions than other users. Nevertheless, conventional trust-based CF methods do not address the issue of deriving the trust values based on users' various information needs on items over time. In knowledge-intensive environments, users usually have various information needs in accessing required documents over time, which forms a sequence of documents ordered according to their access time. We propose a sequence-based trust model to derive the trust values based on users' sequences of ratings on documents. The model considers two factors – time factor and document similarity – in computing the trustworthiness of users. The proposed model enhanced with the similarity of user profiles is incorporated into a standard collaborative filtering method to discover trustworthy neighbors for making predictions. The experiment result shows that the proposed model can improve the prediction accuracy of CF method in comparison with other trust-based recommender systems.

Keywords: Collaborative filtering; Recommender system; Sequence-based trust; Document recommendation

1. Introduction

Recommender systems have emerged in various applications to support item recommendation (Resnick and Varian, 1997; Schafer et al., 2007), solving the information-overload problem by suggesting items of interest to users. Various recommendation methods have been proposed. The collaborative filtering (CF) method (Resnick et al., 1994) has been successfully used in various applications. It predicts user preferences for items in a word-of-mouth manner. User preferences are predicted by considering the opinions (in the form of preference ratings) of other "like-minded" users.

Recently, trust-based recommender systems (Lathia et al., 2008; O'Donovan and Smyth, 2005) have incorporated the trustworthiness of users into the CF techniques to improve the quality of recommendation. Trust-based recommender systems can be classified in two categories: reputation trust and relationship trust (Kwan and Ramachandran, 2009). Reputation trust is a quantitative assessment that allocates a trust score to a specific person (or object) by accumulating a given user's accurate predictions made to other users (Cho et al., 2007; Kim et al., 2008; O'Donovan and Smyth, 2005). On the other hand, relationship trust is the trust between two users. One user trusts another based on past interactions or explicitly specified relationships (Hess et al., 2006; Hwang and Chen, 2007; Kopel and Kazienko, 2007; Lathia et al., 2008). Massa and Avesani (2004, 2007a,b) and Massa and Bhattacharjee (2004) propose a relationship-trust recommender system based on a user's web of trust, which explicitly specifies the friends s/he trusts. For instance, in Epinions.com, users are allowed to assign their

*Corresponding author. Tel.: +886 32655413; fax: +886 3 2655499.
E-mail address: chlai@cycu.edu.tw (C.-H. Lai).

1071-5819/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2011.06.001

personal trust value to the review writers. Through trust propagation from the web of trust, the trust value between two users can be predicted even though there is no direct trust value specified (connection) between them. Their work, however, relies on the user's explicit assignment of trust value that is not easy to collect and may create a heavy burden on users.

Some researchers (Hwang and Chen, 2007; Kim et al., 2008; O'Donovan and Smyth, 2005) have proposed trust computation models to derive the trust values based on users' past ratings of items. O'Donovan and Smyth (2005) suggest that if a user has usually delivered accurate predictions in the past, s/he merits being called reliable and trustworthy. The accuracy of a prediction indicates the degree of approximation between the predicted ratings and the real ratings given by users. A prediction on an item contributed from a given user (producer) is accurate to a target user (consumer) if the difference between their ratings on the item is within a predefined error bound. Generally, a user is more trustworthy if s/he has contributed more accurate predictions than other users. Their proposed trust metric is a reputation trust, which basically accumulates the given user's accurate predictions made to other users or a group of users. Their trust model includes the item level and profile level. The item-level/profile-level trust metric of a given user is derived by computing the ratio of accurate predictions that s/he has made to other users over a particular item/all items that s/he has rated in the past. In addition, Hwang and Chen (2007) propose a relationship trust metric to derive the trust value between two users by calculating the ratio of accurate predictions over all co-rated items, i.e., those items that have been rated by both of them. The proposed relationship trust metric is more personalized than the reputation trust metric. Their proposed trust metrics are combined with the standard CF technique to improve prediction quality for a MovieLens data set. In addition, there may be some irregular rating data due to malicious attacks (Mobasher et al., 2005, 2007; O'Donovan and Smyth, 2006; O'Mahony et al., 2004), which inject attacking profiles with biased rating data to affect the recommendation. Such irregular rating data may degrade the effectiveness of trust models and the accuracy of recommendations. O'Donovan and Smyth (2006) employed a time-based approach to reduce the effect of malicious attacks by selecting older user rating profiles and filtering out recent attacking profiles in order to build trust. Their time-based approach in building trust focused on robustness under malicious attacks rather than the aspect of users' various information needs over time.

Nevertheless, conventional trust-based CF methods do not address the computation of trust values based on users' various information needs for items over time. In general, a user only has one rating score on an item and there is no ordering relationship between the items (movies) in a user's rating history. That is, it does not matter whether a user saw a horror movie first and then a comedy movie, or the contrary. In knowledge-intensive environments, users usually have various information needs in accessing required documents over time, producing a sequence of documents ordered according to their access time. For such environments, the ordering of documents required by a user may be important. For example, a user may need to access documents with prerequisite and basic knowledge first, followed by documents with advanced knowledge.

In this work, we focus on the aspect of recommending documents in order to fulfill users' information needs in knowledge-intensive environments. We propose a sequence-based trust model to derive trust value based on users' sequences of document ratings. The proposed model considers time factor, document similarity and user's profile in computing the trustworthiness of users. Generally, the latest documents accessed by a given user more precisely reflect their current information needs. Thus, an accurate prediction made in the recent past contributes more trustworthiness than the one made earlier. By this rationale we give greater weight to the most recent documents in trust computation. Moreover, conventional trust computational models use the ratings on the same item to derive the accuracy of prediction and compute the trust value, and they do not address the computation of trust values based on similar documents.

In knowledge-intensive environments, users often have the information needs to access documents with similar contents. A user's rating of a document generally reflects the user's perception of the relevance of the document content to his/her information needs. Thus, the ratings on different documents with similar contents should also help to derive the trustworthiness of users. Accordingly, we consider the time factor and the ratings on similar documents to derive a sequence-based trust computation model. In addition, the recommended item is a text-based document, thus, content analysis is useful to select neighbors based on the similarity of user profiles which reveal users' interest on document content. The proposed model enhanced with the similarity of user profiles is incorporated into the standard CF method to effectively discover trustworthy neighbors for making predictions. For sequence-based data set in a knowledge-intensive environment, our experiment result shows that the proposed model can improve the prediction accuracy of the CF method in comparison with other trust-based recommender systems.

The paper is organized as follows. We present the related work in Section 2. An overview of our proposed sequence-based trust for recommendation is presented in Section 3. Section 4 describes our proposed trust computation models and the recommendation methods based on these models. The experiment results and evaluations are presented in Section 5. Conclusions and future work are finally made in Section 6.

2. Related work

2.1. Recommender systems

Recommender systems (RS) can be classified in two categories, content-based recommender systems (CB) (Pazzani and Billsus, 2007) and collaborative filtering systems (CF) (Konstan et al., 1997; Resnick et al., 1994). The former identifies items of special interest through analyzing item descriptions, while the latter filters or evaluates items by users' opinions. Details are described below.

A number of recommender systems apply a content-based technique to various domains, such as web pages (Pazzani and Billsus, 1997), news articles (Watters and Wang, 2000) and TV programs (Ali and Stam, 2004). Content-based recommender systems recommend interesting items to the user by analyzing their content descriptions. The content features are used to establish a characteristic profile. Most content-based recommender systems adopt information retrieval to analyze item content and build a profile for an item or a user. The content-based approach recommends items with similar attribute features to the customer profiles according to their preference in the past.

The collaborative filtering (CF) method predicts users' preferences by considering the opinions (in the form of preference ratings) of other "liked-minded" users (Konstan et al., 1997; Resnick et al., 1994). In general, CF methods can be roughly classified as user-based and item-based CF methods. User-based CF exploits historical data expressing preferences to form user neighbors and make recommendations based on those similar users' opinions. On the other hand, item-based CF determines recommendations by relying on items' associations, which are based on user's ratings among items. Sarwar et al. (2001) built a user-item matrix to identify relationships between different items and then find other similar products that users might like.

To provide useful recommendations, the user-based CF approach involves two steps: neighborhood selection and target user's rating prediction on items. The purpose of neighborhood selection is to select users who have similar interests to the target user. Several metrics have been proposed for similarity computing, e.g., Pearson correlation coefficient (Resnick et al., 1994). Eq. (1) is used to evaluate the Pearson correlation between target user c and recommender p

$$w_{c,p}^{Pearson} = \frac{\sum_{j\in(S_c^I\cap S_p^I)}(r_{c,j}-\overline{r}_c)(r_{p,j}-\overline{r}_p)}{\sqrt{\sum_{j\in(S_c^I\cap S_p^I)}(r_{c,j}-\overline{r}_c)^2}\sqrt{\sum_{j\in(S_c^I\cap S_p^I)}(r_{p,j}-\overline{r}_p)^2}}, \quad (1)$$

where S_c^I and S_p^I represent a document set rated by user c and p, respectively; $r_{c,j}$ is target user c's rating of item j; $\overline{r}_c$ is user c's average rating of items in the set $(S_c^I \cap S_p^I)$.

In the prediction phase, Eq. (2), Resnick's prediction formula (Resnick et al., 1994), is used to make predications. The predicted rating score is derived from the target user's average rating and his/her neighbors' relative opinions on the common rated items, as shown below

$$\hat{p}_{c,j} = \overline{r}_c + \frac{\sum_{p\in NS} w_{c,p}^{Pearson}(r_{p,j}-\overline{r}_p)}{\sum_{p\in NS}\left|w_{c,p}^{Pearson}\right|}, \quad (2)$$

where $\hat{p}_{c,j}$ represents the predicted rating that target user c may provide for item j; $\overline{r}_p$ is his/her average rating; $w_{c,p}^{Pearson}$ is the user similarity score between target user c and his/her neighbor p and NS is the set of neighbors that have been selected to provide their relative interests.

There are some famous commercial recommendation systems such as *Amazon.com* (Linden et al., 2003), *Ringo* (Shardanand and Maes, 1995) and *Last.fm*[1] (Schafer et al., 2007). Amazon.com uses an item-based CF to recommend similar items for customers. Ringo adopts a user-based CF for recommending music albums and artists. Last.fm also uses a collaborative filtering approach to recommend music based on users with similar musical tastes. These systems do not address the trust issues and do not consider users' information needs over time when making recommendations. In this work, we focus on the issue of recommending documents in order to fulfill users' information needs in knowledge-intensive environments. In such environments, users usually have various information needs in accessing required documents over time. The recommendation systems mentioned above do not address such issue. Accordingly, we consider the time factor and the ratings on similar documents to derive a sequence-based trust computation model to improve the recommendation quality. The details will be discussed in Section 3.

2.2. Information retrieval and filtering

Information retrieval (IR) transforms textual documents into a list of features and filters out non-relevant ones by three phases: stop-word removing, stemming and term weighting phases (Baeza-Yates and Ribeiro-Neto, 1999). Each document is described by a term vector, which consists of representative terms and their term weights. We employed the well-known *tf–idf* approach (Salton and Buckley, 1988) to calculate term weights.

According to Salton and Buckley (1988), any codified knowledge item d (e.g., documents, reports, etc.) can be represented as a feature vector of weighted terms in a n-dimensional space. The feature vector of document d is represented as $\rightharpoonup d = \langle w_{1,d}, w_{2,d}, \ldots, w_{n,d} \rangle$. The weight of term i in document d is $w_{i,d}$ is derived using Eq. (3)

$$w_{i,d} = \left(0.5 + \frac{0.5 \times tf_{i,d}}{\max_j tf_{j,d}}\right) \times \left(\log\frac{N}{df_i} + 1\right), \quad (3)$$

where $tf_{i,d}$ indicates the term frequency of term i in document d, df_i means the number of documents which contain the specific term i and N is the total number of documents.

The similarity between documents is usually measured by the cosine measure (Baeza-Yates and Ribeiro-Neto, 1999), which computes the cosine of the angle between their corresponding feature vectors. Two documents are considered similar if the cosine similarity value is high. The cosine similarity of two documents, X and Y, is

[1]http://www.last.fm/.

$simcos(X, Y) = \rightharpoonup X \cdot \rightharpoonup Y / (\| \rightharpoonup X \| \| \rightharpoonup Y \|)$, where $\rightharpoonup X$ and $\rightharpoonup Y$ are the feature vectors of X and Y, respectively.

2.3. Relationship trust and reputation trust

The purpose of designing a trust metric is to quantify the degree of trust (Weng et al., 2006). The trust value can be classified into direct and indirect trust, depending on whether a user actively indicates trust or not (O'Donovan, 2009). The meaning of direct trust is that a user expresses their opinion in value or opinion format to another person during their interaction. The ''friend'' list in *Epinions*, *Facebook* or the feedback from eBay exemplifies direct trust (Massa and Avesani, 2004, 2007a; Massa and Bhattacharjee, 2004). Conversely, indirect trust is derived through computation. Marsh (1994) claimed that trust can be viewed as a function of reputation, which can be computed over historical data.

With regard to the procedure of deriving the trust metric, two dimensions of trust metrics are defined: relationship and reputation (Kwan and Ramachandran, 2009). Relationship trust relies on qualitative measurements depending upon social network connections. A user decides his/her trust of another based on some private knowledge which was gained through past interactions, or explicitly specified relationships. Some researchers have named it personal trust or local trust (Golbeck and Kuter, 2009; Hwang and Chen, 2007; Massa and Bhattacharjee, 2004), whose value is limited between two users and diversified with different user pairs. Several examples such as Facebook and Epinions by which the user includes a friend in his/her list are this type. If the relationship trust is not explicitly indicated, it can be inferred from the rating data or other indirect information (Hwang and Chen, 2007; Lathia et al., 2008; Weng et al., 2006). On the other hand, reputation trust is a more quantitative assessment that allocates a score to a specific object or person within a particular context. An individual's reputation trust is collected from the members in the community. A famous example is eBay, on which each seller attains a trust value through several buyers' comments. Some researchers call it global trust or expert degree with similar concept (Cho et al., 2007; Kim et al., 2008; O'Donovan and Smyth, 2005).

As trust is applied in social networks, it provides more functions for the expansion of Internet intelligence. For example, users enjoy sharing documents with their friends or reading articles written by a credible writer. Those behaviors on the Internet form a so-called web of trust (WoT). The main concept of WoT is that even though two users were unknown to each other before, their friendship is still able to be inferred through other trust relationships which are known and related to the two users (Golbeck and Kuter, 2009; Kopel and Kazienko, 2007). People are linked through this relationship and then a social network is constructed. The trust relationships among people have attracted more and more attention (Riegelsberger et al., 2005). Lately, several social network applications (Yang and Chen, 2008) on the Web have become mature such as *MySpace* and *Facebook* (Lampe et al., 2007). Numbers of active users on such social network applications are growing rapidly. Thus, the trust issue is becoming increasingly important in terms of social networking.

2.4. Trust-based CF recommender systems

According to the trust characteristic presented above, trust-based recommender systems can be classified in two categories: reputation trust and relationship trust.

2.4.1. Reputation trust-based recommender system

Several researchers propose reputation trust as an auxiliary factor in the recommendation phase. Reputation trust is referred to as ''expert'' or ''professional degree'' by some researchers (Cho and Kwon, 2008; Cho et al., 2007, 2009; Kim et al., 2008). Cho et al. (2007) and Kim et al. (2008) judge whether someone is qualified as an expert by adopting Riggs's model (Riggs and Wilensky, 2001), which assigns scores to reviewers based on how close their ratings are to the average ratings. For example, Kim et al. (2008) use *Epinion.com* data to derive the degree of trust-based on users' expertise in categories, which are derived based on the quality of reviews and reputations of review raters/writers. O'Donovan and Smyth (2005) claim that accurate recommendation in the past is important and reliable, and they propose profile-level trust and item-level trust derived from user rating data. They use a simple version of Resnick's prediction formula (Resnick et al., 1994) to calculate a target user c's predicted rating on an item i_k from a recommender p's rating on i_k, as defined in Eq. (4)

$$\hat{P}^{p}_{c,i_k} = \bar{r}_c + (r_{p,i_k} - \bar{r}_p), \tag{4}$$

where $\hat{P}^{p}_{c,i_k}$ is a predicted rating of the target user c on item i_k by a recommender p; $\bar{r}_c$ and $\bar{r}_p$ refer to the mean ratings of target user c and recommender p, respectively; r_{p,i_k} is p's rating on i_k. The rating prediction, by recommender p for target user c, is *correct* if the predicted rating is within an error bound of c's actual rating, as shown in Eq. (5)

$$correct(i_k, p, c) \Leftrightarrow \left| \hat{P}^{p}_{c,i_k} - r_{c,i_k} \right| < \varepsilon, \tag{5}$$

where r_{p,i_k} is the actual rating of the item i_k given by the target user c, and ε is an error bound measuring the closeness.

According to this equation, recommender p is regarded as trustworthy if his/her prediction on item i_k in target user c's view is close to c's actual rating. A user is viewed as trustworthy if s/he always contributes precise predictions. Let (p, c_x, i_k) denote a recommendation that both user c_x and recommender p have rated item i_k and can be used to derive p's trustworthiness. Let U be the set of users and I be the set of items. All the recommendations that p has involved form a set called p's *RecSet*, as shown in Eq. (6). For each recommendation in *RecSet*(p), the trustworthiness of p on a specific item i_k for user c_x is measured as in Eq. (5). $CorrSet(p)$ defined in Eq. (7) stores all recommendations that recommender p has made approximate predictions on some item i_k for some user c_x

$$RecSet(p) = \{(p, c_x, i_k) | c_x \in U, i_k \in I; \text{ both } p \text{ and } c_x \text{ have rated } i_k\}, \tag{6}$$

$$CorrSet(p) = \{(p, c_x, i_k) | (p, c_x, i_k) \in RecSet(p) \text{ and } Correct(i_k, p, c_x)\}. \tag{7}$$

The profile-level trust, $Trust^{PL}(p)$, is calculated as the percentage of *correct* predictions that the recommender p has made; while the concept of item-level trust, $Trust^{IL}(p, i)$, is similar but focuses on a specific item i, as defined in Eq. (8).

$$Trust^{PL}(p) = \frac{|CorrSet(p)|}{|RecSet(p)|};$$
$$Trust^{IL}(p, i) = \frac{|\{(p, c_x, i) | (p, c_x, i) \in CorrSet(p)\}|}{|\{(p, c_x, i) | (p, c_x, i) \in RecSet(p)\}|}. \tag{8}$$

Both profile-level trust and item-level trust can be used in the recommendation phase. The neighbors of target users are selected by filtering out users whose profile-level trust values are lower than a specified *threshold*. The weight between user p and the target user c is derived by combining the value of profile-level trust with user *pearson similarity* (Eq. (1)) in a harmonic mean. Then, these user weights are applied in a modified version of Resnick's formula for prediction. The item-level trust can be applied similarly in the recommendation phase. Please refer to O'Donovan and Smyth (2005) for details.

2.4.2. Relationship trust-based recommender system

Relationship trust metrics consider the truster's subjective opinions when predicting the trust value which s/he placed on the trustee. Epinions.com allows users to express their trust opinions by adding a reviewer into their web of trust list or block list according to whether the reviewer's reviews are valuable. Massa and Avesani (2007b) consider such kind of trust opinion as local trust (relationship trust), and take advantage of web of trust in *Epinions.com* to balance the collaborative recommender system's defects (Massa and Avesani, 2004, 2007a; Massa and Bhattacharjee, 2004).

Even though relationship trust presents an improvement on traditional CF recommender systems, the direct relationship trust data has some defects. This kind of data is not usual in most recommender systems and it is hard to collect. Besides, the quality of a reviewer's review cannot always maintain consistency and the relationship trust may vary according to the reviewer's quality and the user's interest. Hwang and Chen (2007) consider the truster's subjective opinions to obtain more personalization effects when predicting the trust value which s/he placed on the trustee. They calculate the personal (local) trust value of target user c with respect to recommender p, as shown in Eq. (9)

$$T_{c \to p} = \frac{1}{|(I_c \cap I_p)|} \sum_{i_k \in (I_c \cap I_p)} \left(1 - \frac{|\hat{p}^p_{c,i_k} - r_{c,i_k}|}{M}\right). \tag{9}$$

Recommender p predicting item i_k in target user c's view is denoted as $\hat{p}^p_{c,i_k}$. Instead of filtering with an error bound, they use all items that are co-rated by p and c to compute the personal trust. M is the range of the rating score between maximum and minimum rating scores. If two users have no co-rated items, which results in no direct trust relationships between them, trust propagation can be used to infer their trust value through indirect relationships in the web of trust. In addition, a user c's global trust can be derived as the average of the personal trust values given by neighbors directly connected to c in the web of trust. Resnick's prediction formula (Eq. (2)) is then used to make predictions by replacing the similarity score with the trust value as the weight to compute the weighted sum of the ratings given by neighbors. In their research, the experiment evaluation shows that the personal (local) trust-based CF method performs better than the global trust-based CF method.

3. An overview of sequence-based trust for recommendation

3.1. Concept of sequence-based trust

Most trust computation models consider accurate predictions derived from past rating records to infer the trust value. A prediction on an item contributed from a recommender (producer) is accurate for a target user (consumer) if the difference between their ratings on the item is within a predefined error bound. Generally, a user is more trustworthy if s/he has contributed more precise predictions than other users. From our point of view, the inference of trust value should not only depend on accurate predictions but also on the time when the rating was made. In knowledge-intensive environments, users normally have various information needs in accessing required documents over time, producing a sequence of documents ordered according to their access time. The sequence of required documents for a given user may also reveal the change in their information needs over time. Generally, the latest documents accessed by a given user more precisely reflect their current information needs. Similarly, an accurate prediction made in the recent past contributes more trustworthiness than one made some time ago.

Moreover, users often have the information needs to access documents with similar contents. A user's rating of a document generally reflects the user's perception of the relevance or usefulness of the document content to his/her information needs. Even though two users do not access the same documents, we can still infer that they may have similar information needs. It is possible that they may also have the same perspective on the usefulness of the document contents to their information needs, if the contents of their required documents are similar. Thus, the ratings of different documents with similar contents should also help to derive the trustworthiness of users. Accordingly, we consider the time factor and the ratings of similar documents to derive a sequence-based trust computation model.

In this work, we focus on the aspect of recommending documents to fulfill users' information needs in knowledge-intensive environments. We do not consider the issues of malicious attacks and the robustness of trust-based recommendation (Mobasher et al., 2005, 2007; O'Donovan and Smyth, 2006; O'Mahony et al., 2004). We assume that there is no malicious attack and users' rating behaviors are regular, i.e., the rating data truly reflects a user's opinion on how relevant and useful a document is to the user's information needs.

For the rest of the paper, a document sequence of a given user denotes the sequence of documents accessed by the user, while a user's rating sequence represents his ratings on the documents he had accessed over time. We use the term "target user" to describe the one who is recommended to and the term "recommender" to describe the one selected for recommending items to the target user.

Definition 1. A *document sequence* of a user u, denoted as S_u^D, is a sequence of documents that are accessed by u and are ordered by u's access time on documents. S_u^D is expressed as $\langle d_{k1,u}^{t_{u1}}, \ldots, d_{kj,u}^{t_{uj}}, \ldots, d_{kf,u}^{t_{uf}} \rangle$ and $t_{u1} < t_{u2} < \ldots < t_{uf}$, where $d_{kj,u}^{t_{uj}}$ denotes the document d_{kj} that the user u accessed at time t_{uj}; t_{u1} is the starting time index of the first document accessed in u's sequence; t_{uf} is the time index of the most recent document accessed in u's sequence.

Definition 2. A *rating sequence* of a user u, denoted as S_u^R, is a sequence of ratings on the documents that are accessed by u and are ordered by u's access time on documents. S_u^R is expressed as $\langle r_{u,d_{k1}}^{t_{u1}}, \ldots, r_{u,d_{kj}}^{t_{uj}}, \ldots, r_{u,d_{kf}}^{t_{uf}} \rangle$ and $t_{u1} < t_{u2} < \ldots < t_{uf}$, where $r_{u,d_{kj}}^{t_{uj}}$ denotes the rating of document d_{kj} that the user u accessed at time t_{uj}.

3.2. Proposed framework of sequence-based trust for recommendation

Fig. 1 shows the framework of our proposed sequence-based trust model and the CF recommendation methods based on the proposed model. First, documents are preprocessed by the *tf–idf* approach (Salton and Buckley, 1988) to generate document profiles describing the key contents of documents. In addition, the system records the user's accessing behavior, including the accessing time of documents and ratings of documents. Because each user has various information needs at different times, his/her documents are arranged as a document sequence ordered by their access time. Then, the similarities among document profiles are derived in the similarity computation step. Next, these document similarities and document ratings in users' document sequences are incorporated into

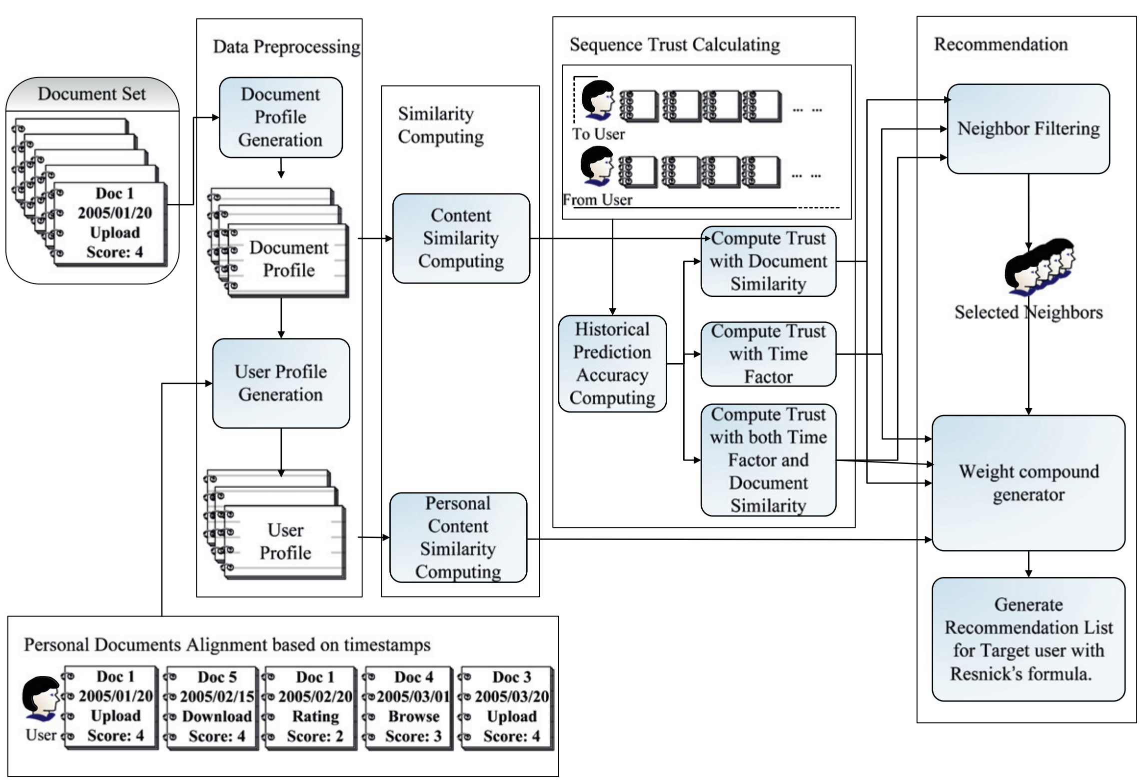

Fig. 1. Proposed framework of sequence-based trust for recommendation.

our trust model to obtain the sequenced-based trust values which denote the trustworthiness among users. We propose three kinds of trust models which consider time factor, document similarity and both time factor and document similarity, respectively. These trust values of users are used to discover highly trustworthy users as neighbors for a target user. Notably, the neighbors selected using different trust computation models for the same target user may vary. The proposed CF methods derive the predictions of document ratings for the target user based on the trust values and the document ratings of neighbors. Moreover, the proposed methods further improve the prediction process by considering the neighbors' similarity to the target user based on user profile similarity.

4. Sequence-based trust models and CF methods

We will describe the proposed sequenced-based trust models in Sections 4.1–4.3 respectively. Then, in Section 4.4, we will introduce the recommendation methods based on the proposed models.

4.1. Sequence-based trust with time factor

This section presents the trust computation model considering the time factor. Each user has a document sequence and corresponding rating sequence ordered by time index. The documents/ratings of users are aligned according to their relative time index in corresponding sequences. Note that the document rating, which is given by a worker, on a scale of 1–5, indicates whether a document is perceived as useful and relevant to the worker's task.

The conventional trust model calculates the ratio of accurate predictions made according to past ratings without considering the time factor. Our proposed trust model derives the trust value of a given user not only based on the ratio of accurate predictions but also on the time that the accurate predictions were made.

In conventional trust models (Hwang and Chen, 2007; O'Donovan and Smyth, 2005), each document prediction provides equal weight when counting how much the target user may trust the recommender. More recent predictions should, however, provide more confidence, because people normally pay more attention to recent events. Thus, in order to show the time effect on trust relationship, we present a sequence-based trust model.

Similarly to the conventional trust computation models (Hwang and Chen, 2007; O'Donovan and Smyth, 2005), we also use a simple version of Resnick's prediction formula (Resnick et al., 1994) to calculate a target user c's predicted rating of a document d_k, $\hat{P}^{p}_{c,d_k}$, which is derived from a recommender p's rating of d_k, as defined in Eq. (10)

$$\hat{P}^{p}_{c,d_k} = \bar{r}_c + (r_{p,d_k} - \bar{r}_p), \tag{10}$$

where $\bar{r}_c$ and $\bar{r}_p$ refer to the mean ratings of target user c and recommender p; r_{p,d_k} is p's rating of document d_k. If $\hat{P}^{p}_{c,d_k}$ is close to the real rating score of user c on d_k, i.e., r_{c,d_k}, we conclude that both the target user c and the recommender p have a similar perspective on document d_k. The more similar the perspective, the more trust they have, as illustrated in Eq. (11)

$$T^{pure}_{c,p,d_k} = 1 - \frac{\left|\hat{P}^{p}_{c,d_k} - r_{c,d_k}\right|}{M}, \tag{11}$$

where T^{pure}_{c,p,d_k} is the pure trust value between target user c and recommender p pertaining to document d_k that is derived from the rating data without considering the time factor; M is the range of the rating score, which equals the difference of the maximum and minimum rating scores.

Generally, the latest documents accessed by a given user more precisely reflect his/her current information needs. Similarly, an accurate prediction made in the recent past contributes more trustworthiness than the one made some time ago.

A document sequence of a user c is a time-ordered sequence of the documents arranged by their access times as defined in Definition 1. Let S^D_c and S^R_c be the document sequence and rating sequence of target user c respectively, as defined in Definition 1. Let $d^{t_{cj}}_{k,c}$ denote the document d_k that the user c accessed at time t_{cj}. $r^{t_{cj}}_{c,d_k}$ denotes c's rating of document d_k; t_{c1} is the starting time index of the first document accessed in S^D_c; t_{cf} is the time index of the most recent document accessed in S^D_c. Let S^D_p and S^R_p be the document sequence and rating sequence of a recommender p respectively. Assume that a document d_k is accessed by user c at time t_{cj} and accessed by recommender p at time t_{pi}. The time factor $TF^{p,t_{pi}}_{c,t_{cj}}$ defined in Eq. (12) is the harmonic mean of the time weights of user c's rating $r^{t_{cj}}_{c,d_k}$ and user p's rating $r^{t_{pi}}_{p,d_k}$

$$TF^{p,t_{pi}}_{c,t_{cj}} = \frac{2 \times tw^{cj}_c \times tw^{pi}_p}{tw^{cj}_c + tw^{pi}_p}. \tag{12}$$

The two time weights tw^{cj}_c and tw^{pi}_p are calculated from the time index t_{cj} of user c's sequence and the time index t_{pi} of user p's sequence respectively. Higher time weights are given to ratings with more recent time indices. The time weight of a rating made at time t_{pi} by user p is defined as $tw^{pi}_p = (t_{pi} - t_{ps})/(t_{pf} - t_{ps})$, where t_{ps}/t_{pf} is the starting/latest time index in user p's sequence and t_{ps} is set to zero. The time weight of a rating made at time t_{cj} by user c is defined similarly. The time factor of a prediction will be high if both the time weights of the ratings are high, i.e., both the ratings are made in more recent time.

Eq. (11) derives the pure trust value of a prediction without considering the time factor. We further use the time factor of a prediction to denote the importance (weight) of the prediction contributing to the trustworthiness. The trust value of user c with respect to recommender p is then derived by taking the weighted average of the pure trust values of predictions made on co-rated documents between them. Consequently, $T^{TF}_{c,p}$, the sequence-based trust metric considering time factor is defined as in Eq. (13)

$$T^{TF}_{c,p} = \frac{\sum_{d^{t_{cj}}_{k,c} \text{ in } S^D_c} \sum_{d^{t_{pi}}_{k,p} \text{ in } S^D_p} (1-(|\hat{P}^{p,t_{pi}}_{c,d_k} - r^{t_{cj}}_{c,d_k}|/M)) \times TF^{p,t_{pi}}_{c,t_{cj}}}{\sum_{d^{t_{cj}}_{k,c} \text{ in } S^D_c} \sum_{d^{t_{pi}}_{k,p} \text{ in } S^D_p} TF^{p,t_{pi}}_{c,t_{cj}}}, \tag{13}$$

where $\hat{P}^{p,t_{pi}}_{c,d_k}$ is the target user c's predicted rating on a document d_k, which is derived from a recommender p's rating on d_k at time t_{pi} according to Eq. (10); $r^{t_{cj}}_{c,d_k}$ is the target user c's actual rating on document d_k accessed at time t_{cj}; S^D_c and S^D_p are the document sequences of the target user c and recommender p respectively; M is the range of the rating score, which equals the difference of the maximum and minimum rating scores.

In addition, any one document may appear in the user's document sequence several times. A user may give different ratings to the same document accessed at different time, because his/her information demand may vary over time. Fig. 2 shows an example for deriving the time factor. U_p is a recommender and U_c is a target user. Both of them have average rating with a score of three. Note that U_p is trustworthy if s/he has a similar view (ratings) to U_c on identical documents at recent time index of their document sequences. Doc1, Doc4 and Doc5 exist in both sequences. We use U_p's opinion to predict U_c's score. According to Eq. (13), we compute the weighted average on all co-rated items, and then we obtain the trust value 0.8217 considering the time factor.

4.2. Sequence-based trust with document similarity

In this section, we consider the ratings of similar documents to derive a sequence-based trust computation model. Even though two users do not access the same documents, their ratings of different documents with similar contents should also help to derive the trustworthiness of users.

There are two possible approaches to derive item (document) similarity: content-based, and rating-based approaches. The content-based approach uses cosine

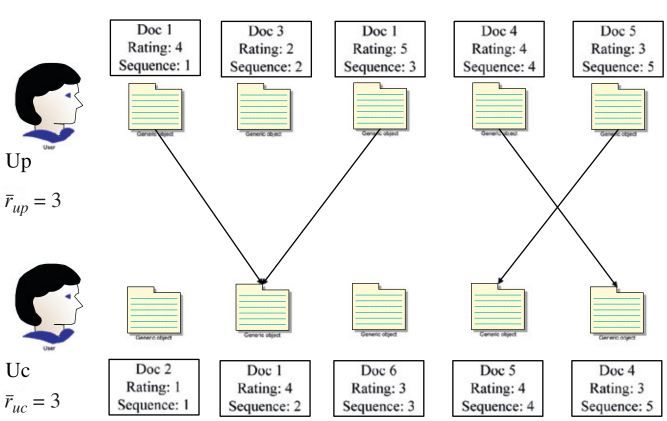

Fig. 2. Illustration of sequence-based trust with time factor.

similarity to derive the similarity of documents based on their document profiles, which are represented as term vectors by the *tf–idf* approach (Baeza-Yates and Ribeiro-Neto, 1999). The rating-based approach derives item similarity based on item ratings. The adjusted cosine similarity can be used to determine the similarity between two items, i and j, according to their ratings of common users, i.e., those users who have rated both items i and j (Sarwar et al., 2001). Two items are considered similar if their common users generally have similar tastes (ratings) concerning them.

In this work, we use a content-based approach to derive item (document) similarity since we focus on recommending documents to fulfill users' various information needs over time. For items such as documents with rich content descriptions, it is intuitive to use the content-based approach to derive the item similarity. Previous researches have also shown the effectiveness of using content-based approaches to boost the recommendation quality of conventional CF methods (Melville et al., 2002). Thus, we adopt the content-based approach to derive the document similarity for building our sequence-based trust models. The rationale for using content-based similarity is that the trust still exists if users have similar views on documents with similar contents. We note that conventional trust-based CF methods did not consider item similarity in computing their trust models. In this work, we have used content-based similarity to enhance the effectiveness of trust models by computing the trust degrees based not only on the *same* items, but on *similar* items as well.

Eq. (11) derives the pure trust value of a prediction for an identical document without considering the document similarity. Eq. (14) is used to predict a trust value based on documents with similar contents. Let $T^{DS}_{c,p}$ be the sequence-based trust metric considering document similarity. The target user c accessed document d_k at time t_{cj} and recommender p accessed document d_l at time t_{pi}, where documents d_k and d_l have similar contents. DS^{p,d_l}_{c,d_k} denotes the document similarity between documents d_k and d_l that is derived by use of the cosine similarity

$$T^{DS}_{c,p} = \frac{\sum_{d^{t_{cj}}_{k,c} \text{ in } S^D_c} \sum_{d^{t_{pi}}_{l,p} \text{ in } S^D_p \text{ and } DS^{p,d_l}_{c,d_k} \geq \theta} (1-(|\hat{P}^{p,t_{pi},d_l}_{c,d_k} - r^{t_{cj}}_{c,d_k}|)/M) \times DS^{p,d_l}_{c,d_k}}{\sum_{d^{t_{cj}}_{k,c} \text{ in } S^D_c} \sum_{d^{t_{pi}}_{l,p} \text{ in } S^D_p \text{ and } DS^{p,d_l}_{c,d_k} \geq \theta} DS^{p,d_l}_{c,d_k}}, \tag{14}$$

where $\hat{P}^{p,t_{pi},d_l}_{c,d_k}$ is the target user c's predicted rating of a document d_k, which is derived from a recommender p's rating of a similar document d_l at time t_{pi}, as defined similarly in Eq. (10). Note that predictions are conducted for those documents with similarity higher than a predefined *threshold*, θ. The document similarity is regarded as a weight of the prediction contributing to the trustworthiness. The trust value of target user c on recommender p is then derived by taking the weighted average of the predicted trust values based on the similarity of documents. The trust value with document similarity solves the problem whereby both users have no item in common.

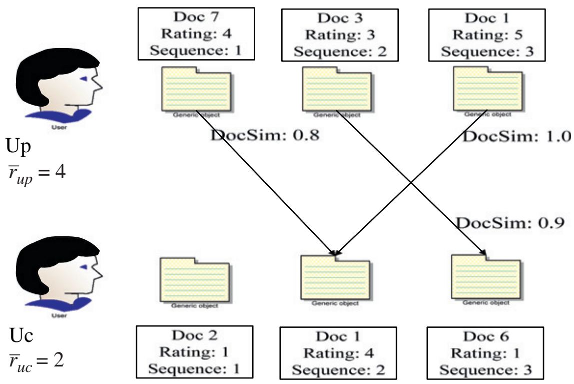

Fig. 3. Illustration of sequence-based trust with document similarity.

Fig. 3 shows an example for deriving the trust value considering document similarity. U_p is a recommender with mean rating 4 and U_c is a target user with mean rating 2. Note that U_p is trustworthy if s/he has a similar opinion to U_c on similar documents. Doc7 is similar to Doc1 and Doc3 is similar to Doc6. Therefore, U_p's opinion on these documents is exploited to predict U_c's score. Based on Eq. (14), the predicted trust value between user U_p and user U_c is 0.8074.

4.3. *Sequence-based trust with both time factor and document similarity*

In order to gain the advantage of both time factor and document similarity, we combine them to derive a sequence-based trust metric, $T_{c,p}^{H}$, as defined in Eq. (15). The trust metric is a hybrid of time factor and document similarity. The trust value of target user c on recommender p will be higher if p has contributed more recent and accurate predictions on documents more similar to user c's documents neighbors for a target user. Let NS be a neighbor set; $T_{c,p}^{Factor}$ be the sequence-based trust value of target user c on a recommender p; *Factor* may be *TF*, *DS* or *H* which represents one of our proposed trust models. *TF*, which denotes the sequence-based trust model with time factor, utilizes users' time-ordered sequences arranged according to the access times of the documents to derive trust values (Section 4.1). *DS*, which denotes the sequence-based trust model with document similarity, obtains the trust value of a prediction on users' different documents with similar content (Section 4.2). *H*, which denotes the sequence-based trust model with both time factor and document similarity, derives the trust value by combing the effects of time factor and document similarity (Section 4.3). The trustworthy users are selected as the neighbors for a target user according to Eq. (16). *Factor* may be *TF*, *DS* or *H* which represents one of our proposed trust models

$$NS = \{T_{c,p}^{Factor} > threshold\}. \tag{16}$$

4.4.2. *Recommendation with sequence-based trust weighting*

This section illustrates the proposed recommendation methods based on our sequence-based trust models. Such methods utilize the sequence-based trust values as weightings and the document ratings of the selected neighbors to make recommendations. The predicted rating of a document d for a target user c, $\hat{P}_{c,d}$, is calculated by Eq. (17)

$$\hat{P}_{c,d} = \overline{r}_c + \frac{\sum_{p \in NS} T_{c,p}^{Factor} (r_{p,d} - \overline{r}_p)}{\sum_{p \in NS} \left| T_{c,p}^{Factor} \right|}, \tag{17}$$

where NS is a neighbor set of the target user c; p is a neighbor of user c; $\overline{r}_c / \overline{r}_p$ is the average rating of documents given by the target user c/ user p; $r_{p,d}$ is the rating of a document d given by user p; $T_{c,p}^{Factor}$ is the sequence-based trust value between a target user c and user p; *Factor* may be *TF*, *DS* or *H* which represents one of our proposed trust models, as described in Sections 4.1–4.3. According to the Eq. (17), documents with high predicted ratings are used to compile a recommendation list, from which the top-N documents are chosen and recommended to the target user.

$$T_{c,p}^{H} = \frac{\sum_{d_{k,c}^{t_{cj}} \text{ in } S_c^D} \sum_{d_{l,p}^{t_{pi}} \text{ in } S_p^D \text{ and } DS_{c,d_k}^{p,d_l} \geq \theta} \left(1 - \left(\left| \hat{P}_{c,d_k}^{p,t_{pi},d_l} - r_{c,d_k}^{t_{cj}} \right| / M \right)\right) \times TF_{c,t_{cj}}^{p,t_{pi}} \times DS_{c,d_k}^{p,d_l}}{\sum_{d_{k,c}^{t_{cj}} \text{ in } S_c^D} \sum_{d_{l,p}^{t_{pi}} \text{ in } S_p^D \text{ and } DS_{c,d_k}^{p,d_l} \geq \theta} TF_{c,t_{cj}}^{p,t_{pi}} \times DS_{c,d_k}^{p,d_l}}. \tag{15}$$

4.4. *Sequence-based trust recommendation*

In the recommendation phase, the trust value is used as a filtering mechanism to select neighbors with high trust values for a target user. Such trust values and the item ratings of neighbors are incorporated into our recommendation methods to make document predictions for the target user. In addition, the recommended item is a text-based document, thus, content analysis is useful to select neighbors based on the similarity of user profiles which reveal users' interest on document content. The following sections describe the details.

4.4.1. *Recommendation with sequence-based trust filtering*

According to the trust relationship between users as illustrated in Sections 4.1–4.3, users whose trust values are higher than a pre-specified *threshold* are selected as

4.4.3. *Recommendation considering time factor, document similarity and profile similarity*

In this section, we introduce profile similarity and employ it to improve the predicting process. Since the recommended item is a text-based document, content analysis is useful to provide other preference information which cannot be discovered through rating data only. A user profile contains useful information which reveals a user's interest on document content in his/her history of document access. Therefore, when we judge whether a recommender is qualified to be a target user's neighbor, the user profile similarity offers another dimension to be explored.

A user profile expressed as a vector of terms represents a user's interests in document contents. The profile of a user is derived by aggregating the profiles of documents in his document sequence considering time factor and document ratings. Each user's documents are arranged in sequence according to their usage log file. In aggregating the document profiles, a document's term weight is multiplied by the time weight, which is determined according to the time index of the document, and the normalized rating of the document, as defined in Eq. (18)

$$\overrightarrow{UP_c} = \sum_{d_{k,c}^{t_{cj}} \text{ in } S_c^D} tw_c^{cj} \times \frac{r_{c,d_k}^{t_{cj}}}{M} \times \overrightarrow{DP_{d_k}}. \tag{18}$$

In Eq. (18), $\overrightarrow{UP_c}$ is a term vector of target user c's aggregated profile. Note that only top-N terms will be selected. $\overrightarrow{DP_{d_k}}$ is a vector of term weights for a document d_k accessed by user c at time t_{cj}. At this time point, user c gives rating $r_{c,d_k}^{t_{cj}}$ on document d_k and the rating score range is M. Notably, the time weight of a document accessed at time t_{cj} by user c is defined as $tw_c^{cj} = (t_{cj} - t_{cs})/(t_{cf} - t_{cs})$, where t_{cs}/t_{cf} is the starting/latest time index in user c's sequence and t_{cs} is set to zero. The user profile of recommender p, $\overrightarrow{UP_p}$, is derived similarly. The user profile similarity $UPS(c, p)$ can then be calculated using the cosine formula $cosine(\overrightarrow{UP_c}, \overrightarrow{UP_p})$.

In the recommendation step, the trust value considering time factor and document similarity is used as a filtering strategy according to Eq. (16). Nevertheless, some minor modification is made in the prediction of the target user c's rating on document d (see Eq. (19))

$$\hat{P}_{c,d} = \bar{r}_c + \frac{\sum_{p \in NS} H(UPS(c,p), T_{c,p}^H)(r_{p,d} - \bar{r}_p)}{\sum_{p \in NS} \left| H(UPS(c,p), T_{c,p}^H) \right|}, \tag{19}$$

where $H(UPS(c,p), Trust_{c,p}^H)$ is the harmonic mean of user profile similarity $UPS(c,p)$ and the trust value considering time factor and document similarity $Trust_{c,p}^H$

$$H(UPS(c,p), Trust_{c,p}^H) = \frac{2(UPS(c,p))(Trust_{c,p}^H)}{UPS(c,p) + Trust_{c,p}^H}. \tag{20}$$

The harmonic mean of a list of numbers is sensitive to the smallest elements of the list, Compared to the arithmetic mean, the harmonic mean tends to decrease the impact of large outliers and increase the impact of small elements. The advantage of using a harmonic mean is to balance the tradeoff between $Trust_{c,p}^H$ and $UPS(c,p)$. The harmonic mean is high when both $Trust_{c,p}^H$ and $UPS(c,p)$ are high. The harmonic mean tends to be small if one of the values is abnormally small.

5. Experiments and evaluations

In this section, we conduct experiments to evaluate the recommendation quality for our proposed methods and compare them with other trust-based recommendation methods. We describe the experiment set-up in Section 5.1 and demonstrate the experimental results in Section 5.2.

5.1. Experiment setup

In our experiment, we collect a data set from a research institute laboratory. We build a knowledge management system (KMS) to collect documents related to knowledge workers' tasks. The data set contains users' access and rating behaviors concerning documents over time in conducting research tasks. Workers' tasks are research-based tasks and their research domains are recommender systems, data mining, information retrieval, workflow systems, knowledge management, etc. There are over 500 research-related documents and about 50 users in the data set. We extract knowledge from these documents to derive the document profiles. Generally, each document profile consists of 800 distinct terms after information extraction by the *tf–idf* approach described in Section 2.2. To share and retrieve the task-related knowledge, workers have four access behaviors in regard to documents, and such user behaviors, i.e., upload, download, browse or rate documents, are recorded in a log. For example, uploading behavior means that a worker "uploads" a document to the KMS to actively share the task-related knowledge, while downloading behavior means that a worker "downloads" a document relevant to his/her task from the KMS. Each user may access 45 documents on average, according to the log data. For the rating behavior, the user may give a document a rating on a scale of 1–5 to indicate whether the document is perceived as useful and relevant. In addition, if a user did not rate a document, a default score is assigned according to the user's access behavior. In this work, uploading and downloading behavior are regarded as more important than browsing behavior. Thus, a default score of 3 is given for browsing behavior, and 4 for uploading or downloading behavior. A high rating, i.e., 4 or 5, indicates that the document is perceived as useful and relevant; while a low rating, i.e., 1 or 2, suggests that the document is deemed to be not useful. The ratio of each behavior in the data set is 11% for browsing, 21% for uploading, 38% for downloading, and 30% for rating behavior. The rating distribution of documents is 7.2% for score 1, 6.1% for score 2, 19.1% for score 3, 65.3% for score 4, and 2.3% for score 5.

Users who conduct similar research tasks usually have similar information needs, i.e., similar interests in acquiring task-relevant documents. For example, two users, who execute two similar research tasks, "document recommender systems" and "blog recommender systems", may need similar task knowledge, such as recommendation techniques, data mining, and text mining. For our data set, each user may generally have 2–8 neighboring users with similar information needs. Thus, we select 2–8 neighbors for each target worker in our experiments. Since it is difficult to obtain such a data set, using the real application domain restricts the sample size of the data and the number of participants in the experiments. Due to the small size of the data set, the number of neighbors is also small.

In our experiment, the data set is divided into a training set and a testing set. We basically use a stratified sampling approach to select the target workers from each group of workers with approximately similar information needs. Some users accessed and rated less than the required *threshold* of 10 documents. These users were not used in the experimental analysis. Accordingly, 20% of the users in the data set were selected as the target workers. The data of non-target workers is included in the training set. The data on target workers is further divided as 70% for training and 30% for testing based on the access time of documents in those workers' document sequences. The training set is used to generate recommendation lists, while the test set is used to verify the quality of the recommendations. The training data is separated from the testing data. Accordingly, we evaluate the performances of our proposed methods and compare them with the traditional CF method and other trust-based recommendation methods.

5.1.1. Evaluation metrics

To measure the quality of recommendations, we use the Mean Absolute Error (*MAE*), which evaluates the average absolute deviation between a predicted rating and the user's true rating (Eq. 21). The lower the *MAE*, the more accurate the method will be

$$MAE = \frac{\sum_{k=1}^{N} \left| \hat{P}_{d_k} - r_{d_k} \right|}{N}. \tag{21}$$

Here N is the amount of testing data, $\hat{P}_{d_k}$ is the predicted rating of document d_k and r_{d_k} is the real rating of document d_k.

5.1.2. Methods compared in the experiment

In the trust-based recommendation methods, the trust value is obtained by the use of different trust computation models for selecting neighbors for a target user. Thus, we use different strategies based on these models to make recommendations and then analyze their recommendation performances. These recommendation strategies are defined as follows.

CF: the standard Resnick model in GroupLens (Resnick et al., 1994). The Pearson correlation coefficient (Eq. (1)) is used in filtering and making predictions.

Profile-TrustCF (Profile-TCF): the profile-level trust is used in filtering and the weight which combines both the profile-level trust and user similarity derived by Pearson correlation coefficient is used to make predictions (O'Donovan and Smyth, 2005), as described in Section 2.4.1.

Item-TrustCF (Item-TCF): the item-level trust is used in filtering and the weight which combines both the item-level trust with user similarity derived by Pearson correlation coefficient is used to make predictions (O'Donovan and Smyth, 2005), as described in Section 2.4.1.

Personal-TrustCF (Personal-TCF): personal trust between two users is calculated by averaging the prediction error of their co-rated items (Hwang and Chen, 2007), as presented in Section 2.4.2.

Time-SeqTrustCF (T-STCF): recommendation with sequence-based trust with time factor using Eqs. (13) and (17).

DocSim-SeqTrustCF (DS-STCF): recommendation with sequence-based trust with document similarity using Eqs. (14) and (17).

Time-DocSim-SeqTrustCF (T-DS-STCF): recommendation with sequence-based trust with both time factor and document similarity using Eqs. (15) and (17).

Time-DocSim-UserProfileSim-SeqTrustCF (T-DS-UPS-STCF): recommendation considering user profile similarity as well as the sequence-based trust with both time factor and document similarity using Eqs. (15) and (19). The weight in the prediction formula is derived using the harmonic mean of the trust value and user profile similarity.

5.2. Experimental results

In the experiments, we compare various recommendation methods from different aspects in Sections 5.2.1–5.2.4. There are 50 users in our data set. The data set shows that each user may generally have 2–8 neighboring users with similar information needs on documents. For most users in our data set, there are 4–6 neighboring users with similar information needs. Thus, we select 2–8 qualified users as target user's neighbors and we compare the *MAE* under different number of neighbors.

The DS-STCF, T-DS-STCF and T-DS-UPS-STCF methods use the ratings of similar documents to compute the trust degrees. We adopt the content-based approach to determine similar documents, which have cosine similarity greater than or equal to a *threshold* θ. The *threshold* of the document similarity, i.e. θ, ranges from 0 to 1. The approach of setting $\theta = 1$ is similar to traditional trust models that employ users' ratings on the same items to derive trust degrees. The setting of the *threshold* for the document similarity may affect the recommendation quality. A suitable *threshold* should be decided to select "similar enough" documents in the trust models. It is difficult to determine the optimal value for the *threshold*. In this work, we conduct several experiments by trying different *threshold* values that are chosen based on the content similarity of documents as judged by human experts. A suitable value for the *threshold* is then chosen according to the experiment result.

5.2.1. Comparison of sequence-based trust with vs. without time factor

In this section, we evaluate the effect of time factor in sequence-based trust model. NT-STCF is a recommendation method with sequence-based trust without considering the time factor, which is derived from Eq. (13) by setting the time factor as 1. Fig. 4 shows that T-STCF performs better than NT-STCF. The time factor indeed contributes to improve the sequence-based trust metric.

5.2.2. The effect of time factor and document similarity

In this experiment, we compare our proposed sequence-based trust methods, including T-STCF (time factor), DS-STCF (document similarity) and T-DS-STCF. Fig. 5 shows that T-STCF performs better than DS-STCF. The time factor contributes more than document similarity in sequence-based trust model. In general, the hybrid of time-factor and document similarity, T-DS-STCF, performs slightly better than T-STCF. Considering both the time factor and document similarity in deriving the trust values can have better improvement of the recommendation quality than the sequence-based trust method considering time factor only.

5.2.3. Comparison of the weighting methods in prediction

In this experiment, we compare the recommendation quality of applying different weighting methods in the recommendation methods to derive the predicted ratings for documents. T-DS-STCF uses the trust value derived from the hybrid of time-factor and document similarity, namely T-DS-trust, as the weighting (Eq. (17)); T-DS-PS-STCF uses the harmonic mean of T-DS-trust and Pearson similarity (PS) of users as the weighting for prediction, which is similarly derived as Eq. (19); T-DS-UPS-STCF uses the harmonic mean of T-DS-trust and user profile similarity (*UPS*) as the weighting for prediction (Eq. (19)). For these three methods, the sequence-based trust, T-DS-trust, is used as a filtering mechanism to select the qualified neighbors for a target user.

The *MAE* values under different number of neighbors are shown in Fig. 6. The result shows that the sequence-based trust combined with user profile similarity performs better than that combined with Pearson similarity. The prediction is more accurate when the weight is derived by combining the sequence-based trust with the user profile similarity than when the weight is obtained by only using the sequence-based trust. This also implies that the user profile similarity can improve the recommendation quality in most circumstances.

5.2.4. Comparison with other methods

We compare our proposed methods, i.e., T-STCF, DS-STCF, T-DS-STCF, T-DS-UP-STCF, with the CF method, and other traditional trust-based recommendation methods under various numbers of neighbors (k). We also perform a statistical hypothesis test, the pair-wise and one-tailed t-test to examine the differences in performance between the recommendation methods. The *MAE* values of each method and the t-test results are listed in Table 1. For the t-test, the confidence level is set as 95%. Thus, the p-value with * means that the method is statistically significant at 0.05 level.

From Table 1, the traditional CF method performs worse than others no matter what the number of neighbors is. Profile-TCF also performs unsatisfactorily. Both of them show a wide gap from the other metrics. The performances of most methods under $k=4$ and 6 are better than their performances under $k=2$ and 8. This implies that the appropriate number of neighbors chosen for prediction in our data set is 4 or 6. The performance is not satisfactory if very few neighbors ($k=2$) are selected for prediction. In addition, the performance may be worse

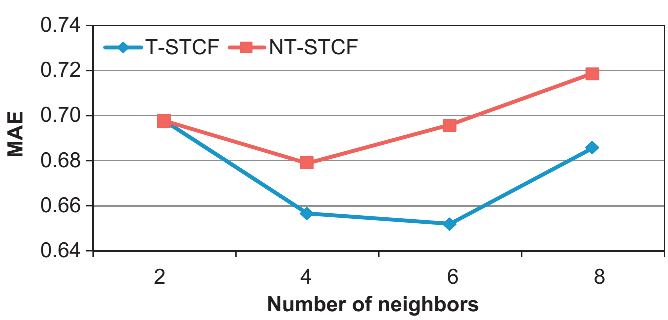

Fig. 4. The effect of the time factor.

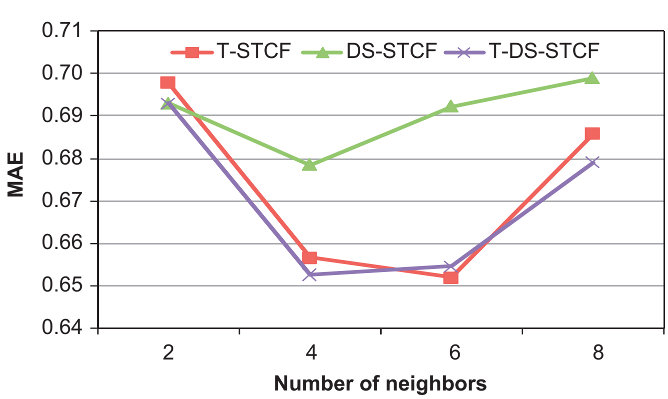

Fig. 5. Comparing the *MAE*s of sequence-based trust methods.

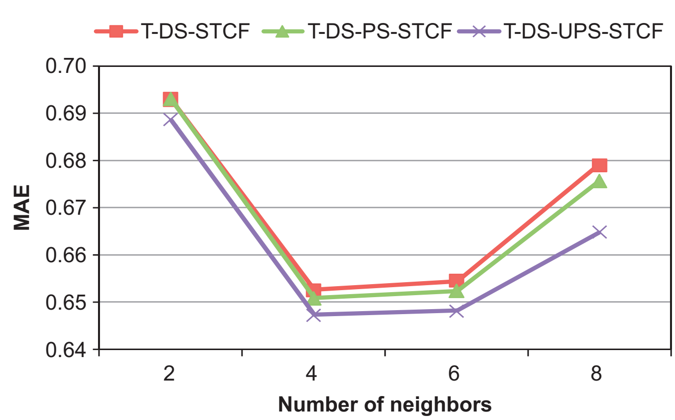

Fig. 6. Comparing the *MAE*s of T-DS-STCF, T-DS-PS-STCF and T-DS-UPS-STCF.

Table 1
Comparison of all methods under different k.

Methods	$k=2$		$k=4$		$k=6$		$k=8$	
	MAE	*p*-value	*MAE*	*p*-value	*MAE*	*p*-value	*MAE*	*p*-value
T-STCF	0.6978		0.6567		0.6520		0.6858	
CF	0.7716	0.0147*	0.7875	0.0003*	0.8447	0.0000*	0.8766	0.0000*
Profile-TCF	0.7703	0.0158*	0.7912	0.0004*	0.8060	0.0002*	0.8379	0.0005*
Item-TCF	0.7251	0.1722	0.7161	0.0347*	0.7187	0.0236*	0.7113	0.2545
Personal-TCF	0.7325	0.0099*	0.6996	0.0025*	0.7141	0.0018*	0.7159	0.1012
DS-STCF	0.6930		0.6785		0.6923		0.6991	
CF	0.7716	0.0097*	0.7875	0.0015*	0.8447	0.0003*	0.8766	0.0001*
Profile-TCF	0.7703	0.0104*	0.7912	0.0011*	0.8060	0.0027*	0.8379	0.0009*
Item-TCF	0.7251	0.1301	0.7161	0.1135	0.7187	0.2055	0.7113	0.3459
Personal-TCF	0.7325	0.0043*	0.6996	0.1261	0.7141	0.0579*	0.7159	0.2098
T-DS-STCF	0.6930		0.6526		0.6546		0.6790	
CF	0.7716	0.0097*	0.7875	0.0002*	0.8447	0.0000*	0.8766	0.0000*
Profile-TCF	0.7703	0.0104*	0.7912	0.0003*	0.8060	0.0003*	0.8379	0.0002*
Item-TCF	0.7251	0.1301	0.7161	0.0251*	0.7187	0.0252*	0.7113	0.1524
Personal-TCF	0.7325	0.0043*	0.6996	0.0012*	0.7141	0.0019*	0.7159	0.0423*
T-STCF	0.6978	0.0543	0.6567	0.1183	0.6520	0.3311	0.6858	0.3801
DS-STCF	0.6930	–	0.6785	0.0272*	0.6923	0.0189*	0.6991	0.0548
T-DS-UP-STCF	0.6886		0.6473		0.6482		0.6649	
CF	0.7716	0.0067*	0.7875	0.0001*	0.8447	0.0000*	0.8766	0.0000*
Profile-TCF	0.7703	0.0072*	0.7912	0.0001*	0.8060	0.0002*	0.8379	0.0001*
Item-TCF	0.7251	0.0997	0.7161	0.0168*	0.7187	0.0156*	0.7113	0.0797
Personal-TCF	0.7325	0.0018*	0.6996	0.0005*	0.7141	0.0008*	0.7159	0.0154*
T-STCF	0.6978	0.0053*	0.6567	0.0103*	0.6520	0.2794	0.6858	0.1449
DS-STCF	0.6930	0.0147*	0.6785	0.0108*	0.6923	0.0079*	0.6991	0.0168*
T-DS-STCF	0.6930	0.0147*	0.6526	0.0082*	0.6546	0.0026*	0.6790	0.0439*

*Significance marked using $p < 0.05$.

if dissimilar users are also selected as neighbors for prediction as the case of $k=8$.

The result shows that the *MAE* values of our proposed sequence-based trust methods are smaller than other methods. We note that our experiment data set contains users' access and rating behaviors on documents over time in conducting research tasks. Conventional trust-based CF methods do not address the computation of trust values and recommendations for such sequence-based data, in which documents and ratings are ordered according to their access time. Accordingly, our proposed sequence-based trust methods, T-STCF and DS-STCF, perform better than conventional trust-based CF methods, including Personal-TCF, Profile-TCF and Item-TCF. The *t*-test results of Table 1 show that the differences in performance between T-STCF and conventional trust-based CF methods (Profile-TCF, Item-TCF and Personal-TCF) are statistically significant under $k=4$ and 6, while the differences are not significant under $k=2$ and 8. This implies that T-STCF, which considers sequence-based trust and a time factor, can improve the recommendation quality. For our data set, the appropriate number of neighbors chosen for prediction is 4 or 6. Thus, T-STCF may not make a significant improvement to recommendation quality under 2 and 8 neighbors. Although DS-STCF performs better than conventional trust-based CF methods in terms of *MAE*, the differences in performance are generally not significant.

For sequence-based documents and ratings, the time factor and document similarity contribute to derive more reliable personal trust values and make more accurate recommendations for target users. Moreover, the hybrid of time-factor and document similarity, T-DS-STCF, performs better than conventional trust-based CF methods, and performs better than T-STCF and DS-STCF in most circumstances. The differences in performance between T-DS-STCF and conventional trust-based CF methods are generally statistically significant, except for the comparison with Item-TCF under $k=2$ and 8. The differences in performance between T-DS-STCF and DS-STCF methods are statistically significant under $k=4$ and 6, while the differences between T-DS-STCF and T-STCF are not significant.

In addition, T-DS-UPS-STCF, which uses a harmonic mean of user profile similarity and trust value considering both time factor and document similarity in the computation model, generally performs better than other methods. The differences in performance between T-DS-UPS-STCF and conventional trust-based CF methods are generally statistically significant, except for the comparison with Item-TCF under $k=2$ and 8. The differences in performance between T-DS-UPS-STCF and other sequence-based trust methods (T-STCF, DS-STCF and T-DS-STCF) are generally statistically significant, except for the comparison with T-STCF under $k=6$ and 8. This implies that considering user profile similarity and the sequence-based trust with both time factor and document similarity in a recommendation method indeed improves the recommendation performance.

For our data set, the appropriate number of neighbors chosen for prediction is 4 or 6. Thus, our proposed sequence-based trust methods can have significant improvement on recommendation quality under 4 and 6 neighbors. In comparisons with Item-TCF, our proposed methods perform better than Item-TCF, although the differences in performance are not significant under 2 and 8 neighbors. The result shows that adopting our proposed sequence-based trust model in CF methods for sequence-based and knowledge-intensive (document) environments can have better improvement on recommendation quality than conventional trust-based CF methods.

6. Conclusions and future work

In this research, we propose sequence-based trust recommendation methods to derive the degree of trust-based on users' sequences of ratings of documents. Conventional trust-based CF methods do not address the computation of trust values in the context of sequence-based and knowledge-intensive (document) environments. Our proposed methods consider time factor and document similarity in computing the trustworthiness of users. Moreover, the proposed methods are further enhanced by considering user profile similarity in the recommendation phase. The rationale behind using the time factor is that the predictions generated close to the current time provide more trustworthiness than those far away from the current time. In addition, the ratings of different documents with similar contents should also help to derive the trustworthiness of users. Accordingly, we employ the time factor and the ratings of similar documents to derive a sequence-based trust computation model. Moreover, we exploit user profile similarity to discover more user preference information which is not easy to observe though rating data. Finally, the proposed models are incorporated into the standard CF method to effectively discover trustworthy neighbors for making recommendations. From the experimental results, we discover that the prediction accuracy of recommendation is indeed improved using these two factors, and our trust metric performs satisfactorily when both factors are combined and incorporated with user's interest over time.

Our experiments were conducted using a real application domain, i.e., research tasks in a research institute's laboratory. The real application domain restricted the sample size of the data and the number of participants in the experiments, since it is difficult to obtain a data set that contains sequences of documents and ratings. Because of this limitation, in our future work, we will evaluate the proposed approach on other application domains involving larger numbers of users and documents. In our future work, we will also investigate how to infer user's reputation with respect to sequence-based profile-level and item-level trust from our sequence-based concept. Moreover, our current study does not consider trust propagation from the web of trust. Our future study will investigate the integration of the sequence-based trust models with trust propagation in the web of trust.

In this work, we do not consider the issues of malicious attacks and the robustness of trust-based recommendation (Mobasher et al., 2005, 2007; O'Donovan and Smyth, 2006; O'Mahony et al., 2004). The irregular rating behaviors with biased rating data caused by malicious attacks may also degrade the effectiveness of our proposed models. Our current work assumes that there is no malicious attack and users' rating behaviors are regular. Nevertheless, malicious attacks and the robustness of trust-based recommendations are interesting issues worthy of discussion. In our future work, we will further investigate the robustness of our proposed sequence-based trust models under malicious attacks.

Moreover, our proposed sequence-based trust models use item similarity to boost the effectiveness of trust models by computing the trust degrees based on similar items. In this work, we have used the content-based approach to derive item similarity. However, the content-based approach cannot be applied if items do not have content descriptions. The rating-based approach derives item similarity based on item ratings, and can be applied to derive the similarity of items without content descriptions. Our proposed trust models can also incorporate the rating-based item similarity in computing the trust degrees. Integrating the rating-based approach into the trust models is an interesting issue worth further investigation. Our future work will compare the content-based approach with the rating-based approach in boosting the effectiveness of our sequence-based trust models.

Acknowledgments

This research was supported in part by the National Science Council of the Taiwan under the Grant NSC 99-2410-H-009-034-MY3.

References

Ali, K., Stam, W.v., 2004. TiVo: making show recommendations using a distributed collaborative filtering architecture. In: Proceedings of the Tenth ACM International Conference on Knowledge Discovery and Data Mining. Seattle, WA, USA, pp. 394–401.

Baeza-Yates, R., Ribeiro-Neto, B., 1999. Modern Information Retrieval. Addison Wesley Longman, New York.

Cho, J., Kwon, K., 2008. Source credibility model for neighbor selection in collaborative web content recommendation. In: Zhang, Y., Yu, G., Bertino, E., Xu, G. (Eds.), Progress in WWW Research and Development, LNCS 4976. Springer, Berlin/Heidelberg, pp. 68–80.

Cho, J., Kwon, K., Park, Y., 2007. Collaborative filtering using dual information sources. IEEE Intelligent Systems 22 (3), 30–38.

Cho, J., Kwon, K., Park, Y., 2009. Q-rater: a collaborative reputation system based on source credibility theory. Expert Systems with Applications 36, 3751–3760.

Golbeck, J., Kuter, U., 2009. The ripple effect: change in trust and its impact over a social network. In: Golbeck, J. (Ed.), Computing with Social Trust, Human-Computer Interaction Series. Springer, London, pp. 169–181.

Hess, C., Stein, K., Schlieder, C., 2006. Trust-enhanced visibility for personalized document recommendations. In: Proceedings of the 2006 ACM Symposium on Applied Computing. Dijon, France, pp. 1865–1869.

Hwang, C.-S., Chen, Y.-P., 2007. Using trust incollaborative filtering recommendation. In: Okuno, H., Ali, M. (Eds.), New Trends in Applied Artificial Intelligence, LNCS 4570. Springer, Berlin/Heidelberg, pp. 1052–1060.

Kim, Y.A., Le, M.-T., Lauw, H.W., Lim, E.P., Liu, H., Srivastava, J., 2008. Building a web of trust without explicit trust ratings. In: Proceedings of IEEE 24th International Conference on Data Engineering Workshop (ICDEW 2008), pp. 531–536.

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 (3), 77–87.

Kopel, M., Kazienko, P., 2007. Application of agent-based personal web of trust to local document ranking. In: Proceedings of the First KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications. Wroclaw, Poland, pp. 288–297.

Kwan, M., Ramachandran, D., 2009. Trust and online reputation systems. In: Golbeck, J. (Ed.), Computing with Social Trust, Human-Computer Interaction Series. Springer, London, pp. 287–311.

Lampe, C.A.C., Ellison, N., Steinfield, C., 2007. A familiar face(book): profile elements as signals in an online social network. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. San Jose, California, USA, pp. 435–444.

Lathia, N., Hailes, S., Capra, L., 2008. Trust-based collaborative filtering. In: Karabulut, Y., Mitchell, J., Herrmann, P., Jensen, C. (Eds.), Trust Management II, IFIP Advances in Information and Communication Technology 263. Springer, Boston, pp. 119–134.

Linden, G., Smith, B., York, J., 2003. Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Computing 7 (1), 76–80.

Marsh, S.P., 1994. Formalising Trust as a Computational Concept. Dissertation. University of Stirling.

Massa, P., Avesani, P., 2004. Trust-aware collaborative filtering for recommender systems. In: Meersman, R., Tari, Z. (Eds.), On the Move to Meaningful Internet Systems: CoopIS, DOA, and ODBASE, LNCS 3290. Springer, Berlin/Heidelberg, pp. 492–508.

Massa, P., Avesani, P., 2007a. Trust-aware recommender systems. In: Proceedings of the 2007 ACM Conference on Recommender Systems. Minneapolis, MN, USA, pp. 17–24.

Massa, P., Avesani, P., 2007b. Trust metrics on controversial users: balancing between tyranny of the majority. International Journal on Semantic Web and Information Systems 3 (1), 39–64.

Massa, P., Bhattacharjee, B., 2004. Using trust in recommender systems: an experimental analysis. In: Jensen, C., Poslad, S., Dimitrakos, T. (Eds.), Trust Management, LNCS 2995. Springer, Berlin/Heidelberg, pp. 221–235.

Melville, P., Mooney, R.J., Nagarajan, R., 2002. Content-boosted collaborative filtering for improved recommendations. In: Proceedings of the Eighteenth National Conference on Artificial Intelligence. Edmonton, Alberta, Canada, pp. 187–192.

Mobasher, B., Burke, R., Runa, B., Chad, W., 2005. Effective attack models for shilling item-based collaborative filtering systems. In: Proceedings of the WebKDD. Chicago, Illinois, USA.

Mobasher, B., Burke, R., Runa, B., Chad, W., 2007. Toward trustworthy recommender systems: an analysis of attack models and algorithm robustness. ACM Transactions on Internet Technology 7 (4), 23.

O'Donovan, J., Smyth, B., 2005. Trust in recommender systems. In: Proceedings of the Tenth International Conference on Intelligent User Interfaces. San Diego, California, USA, pp. 167–174.

O'Donovan, J., Smyth, B., 2006. Is trust robust? An analysis of trust-based recommendation. In: Proceedings of the 11th International Conference on Intelligent User Interfaces. Sydney, Australia, pp. 101–108.

O'Mahony, M., Hurley, N., Kushmerick, N., Silvestre, G., 2004. Collaborative recommendation: a robustness analysis. ACM Transactions on Internet Technology 4 (4), 344–377.

O'Donovan, J., 2009. Capturing trust in social web applications. In: Golbeck, J. (Ed.), Computing with Social Trust, Human-Computer Interaction Series. Springer, London, pp. 213–257.

Pazzani, M., Billsus, D., 1997. Learning and revising user profiles: the identification of interesting web sites. Machine Learning 27 (3), 313–331.

Pazzani, M., Billsus, D., 2007. Content-based recommendation systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (Eds.), The Adaptive Web, LNCS 4321. Springer, Berlin/Heidelberg, pp. 325–341.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J., 1994. GroupLens: an open architecture for collaborative filtering of netnews. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work. Chapel Hill, North Carolina, United States, pp. 175–186.

Resnick, P., Varian, H.R., 1997. Recommender systems. Communication of the ACM 40 (3), 56–58.

Riegelsberger, J., Sasse, M.A., McCarthy, J.D., 2005. The mechanics of trust: a framework for research and design. International Journal of Human–Computer Studies 62 (3), 381–422.

Riggs, T., Wilensky, R., 2001. An algorithm for automated rating of reviewers. In: Proceedings of the First ACM/IEEE–CS Joint Conference on Digital Libraries, Roanoke. Virginia, United States, pp. 381–387.

Salton, G., Buckley, C., 1988. Term weighting approaches in automatic text retrieval. Information Processing and Management 24 (5), 513–523.

Sarwar, B., Karypis, G., Konstan, J., Reidl, J., 2001. Item-based collaborative filtering recommendation algorithms. In: Proceedings of the Tenth International Conference on World Wide Web. Hong Kong, China, pp. 285–295.

Schafer, J., Frankowski, D., Herlocker, J., Sen, S., 2007. Collaborative filtering recommender systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (Eds.), The Adaptive Web, LNCS 4321. Springer, Berlin/Heidelberg, pp. 291–324.

Shardanand, U., Maes, P., 1995. Social information filtering: algorithms for automating ''word of mouth''. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95). Denver, Colorado, USA, pp. 210–217.

Watters, C., Wang, H., 2000. Rating news documents for similarity. Journal of the American Society for Information Science and Technology 51 (9), 793–804.

Weng, J., Miao, C., Goh, A., 2006. Improving collaborative filtering with trust-based metrics. In: Proceedings of the ACM Symposium on Applied Computing. Dijon, France, pp. 1860–1864.

Yang, S.J.H., Chen, I.Y.L., 2008. A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network. International Journal of Human–Computer Studies 66 (1), 36–50.