CrossMark

ORIGINAL ARTICLE

Design of reciprocal recommendation systems for online dating

Peng Xia[1] · Shuangfei Zhai[2] · Benyuan Liu[1] · Yizhou Sun[3] · Cindy Chen[1]

Abstract Online dating sites have become popular platforms for people to look for potential romantic partners. Different from traditional user-item recommendations where the goal is to match items (e.g., books, videos) with a user's interests, a recommendation system for online dating aims to match people who are mutually interested in and likely to communicate with each other. We introduce similarity measures that capture the unique features and characteristics of the online dating network, for example, the interest similarity between two users if they send messages to same users, and attractiveness similarity if they receive messages from same users. A reciprocal score that measures the compatibility between a user and each potential dating candidate is computed, and the recommendation list is generated to include users with top scores. The performance of our proposed recommendation system is evaluated on a real-world dataset from a major online dating site in China. The results show that our recommendation algorithms significantly outperform previously proposed approaches, and the collaborative filtering-based algorithms achieve much better performance than content-based algorithms in both precision and recall. Our results also reveal interesting behavioral difference between male and female users when it comes to looking for potential dates. In particular, males tend to be focused on their own interest and oblivious toward their attractiveness to potential dates, while females are more conscientious to their own attractiveness to the other side of the line.

1 Introduction

Online dating sites have become popular platforms for people to look for potential romantic partners, offering an unprecedented level of access to possible dates that is otherwise not available through traditional means. According to a recent survey,[1] 40 million single people (out of 54 million) in the USA have signed up with various online dating sites such as Match.com, eHarmony and around 20 % of currently committed romantic relationships began online, which is more than through any means other than meeting through friends.

Many online dating sites provide suggestions on compatible partners based on their proprietary matching algorithms. Unlike in traditional user-item recommendation systems where the goal is typically to predict a user's opinion toward passive items (e.g., books, movies), when making recommendation of potential dates to a user (referred to as service user) on an online dating site, it is important that not only the recommended users match the user's dating interest, but also the recommended users are interested in the service user and thus likely to reciprocate when contacted. A successful online dating recommendation system should match users with mutual interest in each

✉ Peng Xia
  pxia@cs.uml.edu

[1] Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, USA

[2] Department of Computer Science, Binghamton Univeristy, State University of New York, Binghamton, NY, USA

[3] College of Computer and Information Science, Northeastern University, Boston, MA, USA

Springer

other and hence result in better chances of interactions between them and improved user satisfaction level.

In this paper, we study the reciprocal online dating recommendation system based on a large real-world dataset obtained from a major online dating site in China with a total number of 60 million registered users. In particular, *given a user, we seek to identify a set of users who are most likely to be contacted by the service user when recommended and reciprocate when contacted.*

The characteristics of our online dating network present unique opportunities and challenges to the reciprocal recommendation problem. First, there is a rich set of user attributes available in our dataset that can be used in the prediction models. These include a user's age, gender, height, weight, education level, income level, house ownership, geographic location, occupation, interests/hobbies, number of photos, etc. In addition, there are a variety of online dating specific information including a user's preference in potential dates (age range, height range, education level, income range, geography location, etc), and his/her dating and marriage plan (when to get married, whether to live with parents and have child after marriage, marriage ceremony style, etc). Moreover, our dataset contains the communication traces between users, i.e., who sent or replied to messages to whom and the associated timestamps. As shown in our earlier paper (Xia et al. 2014b), the communication trace of a user reflects his/her actual dating preference, which may significantly deviate from his/her stated preference and thus should play an important role in the design of the recommendation system.

Due to the heterogeneous dating nature in our recommendation problem (dating is restricted to users of opposite genders on the online dating site in our study), previous approaches designed for friend recommendation in conventional online social networks such as Facebook and LinkedIn cannot be directly applied. For example, the number of common neighbors is often used for friend recommendation for conventional social networks, i.e., the more common friends two users share, the more likely they will become friends and thus should be recommended to each other. On a heterosexual online dating site, however, a user is only interested in contacting other users of opposite gender, resulting in a bipartite network between males and females. There are no common neighbors between a service user and recommended users since they are of different genders. To this end, we will need to devise appropriate mechanisms that account for the special characteristics of the online dating network.

The contributions of this paper are summarized as follows.

- We present a recommendation system that aims to match users who are most likely to communicate with

each other. We introduce similarity measures that capture the unique features and characteristics of the heterogeneous online dating network. In particular, we build a preference model for each service user based on the attributes of users who have been contacted by the service user. We also characterize the interest similarity between two users if they send messages to same users, and attractiveness similarity if they receive messages from same users. Based on these similarity measures, we compute the compatibility between a service user and potential dating candidates and the recommendation list is generated to include candidates with top scores. Using a combination of similarity measures, we construct a set of content-based and collaborative filtering-based algorithms with different measures of compatibility between users.

- We evaluate the performance of our proposed algorithms on the real-world online dating dataset that consists of 200,000 users and around two million messages over a period of two months. Our results show that both content-based and collaborative filtering-based recommendation algorithms presented in our paper significantly outperform previously proposed approaches. Also, compared to the content-based algorithms, our collaborative filtering-based algorithms achieve much better performance in both precision and recall. Moreover, most of our proposed recommendation algorithms place the relevant recommendations (users who have been actually contacted by and replied to the service user) in the top 30 to 50 % positions of the recommendation list. This is an important performance measure as users tend to look at the list from top to bottom.

- Our results also reveal interesting behavioral difference between male and female users when it comes to looking for potential dates. Among the collaborative filtering-based algorithms, the best performance for male users is achieved when the recommender captures the attractiveness of recommended users to the service user and interest from the service user in recommended users. On the contrary, the best performance for females is achieved when recommended users are interested in the service user and the service user is attractive to recommended users. The results show that when looking for potential dates, males tend to be focused on their own interest and oblivious toward their attractiveness to potential dates, while females are more conscientious to their own attractiveness to and interest from the other side on the line.

The rest of the paper is organized as follows. Section 2 describes the related work on reciprocal recommendation as well as reciprocal relation prediction in online social

networks. Section 3 presents the reciprocal recommendation algorithms we proposed. Description and characteristics of our dataset are provided in Sect. 4. Section 5 presents the performance of our proposed algorithms. Finally, we conclude our paper in Sect. 6.

2 Related work

A few studies on the analysis of user behavior of online dating sites have provided valuable guidelines to design recommendation system for online dating. In Xia et al. (2014b), the authors analyze how user's sending and replying behavior correlate with several important user attributes, such as age, income, education level, and number of photos, and how much a user's actual preference deviates from his/her stated preference. The findings also correspond to the result of Pizzato et al. (2010a) that the recommendation system built on user's implicit preference outperforms that built on user's explicit preference.

There exists research that aims to predict user reciprocity in various online social networks. In Mori et al. (2012), a machine learning-based approach is proposed to find plausible candidates of business partners using firm profiles and transactional relations between them. The authors of Hopcroft et al. (2011) propose a Triad Factor Graph model to deal with two-way relation prediction in online social networks. In Xia et al. (2014a), both user-based and graph-based features are applied in a machine learning framework to predict user replying behavior in online dating networks. A new collaborative filtering approach called *CLiMF* (Shi et al. 2012) is proposed to learn latent factors of users and items and improves the performance of top-k recommendations for recommending friends or trustees on Epinions and Tuenti. Further, they improve the algorithm by optimizing Expected Reciprocal Rank, an evaluation metric that generalizes reciprocal rank in order to incorporate user feedback with multiple levels of relevance in Shi et al. (2013).

There have been recently several studies on the people-to-people recommendation for online social networks. Both content-based algorithm and collaborative filtering method are applied to recommend users to follow in Twitter (Hannon et al. 2010). A LDA-based method is employed in Zhao et al. (2013) to discover communities in Twitter-style online social networks, and matrix factorization is applied on each community to provide recommendations.

For online dating recommendations, the authors in Li and Li (2012) and Pizzato et al. (2010b) analyzed the characteristics of reciprocal recommendations in detail. In particular, Li and Li (2012) considers both local utility (users' mutual preference) and global utility (overall bipartite network) and proposes a generalized reciprocal

recommendation framework for both online dating sites and online recruiting sites. A content-based reciprocal algorithm (RECON) proposed in Pizzato et al. (2010b) learns the preference of both sides of users and defines a new evaluation metric (success rate) to evaluate the performance of their algorithm. In their following work Pizzato et al. (2011) and Pizzato and Silvestrini (2011), RECON is extended to consider both positive and negative preference, and collaborative filtering is applied with stochastic matching algorithm. In Zhao et al. (2014), a hybrid collaborative filtering-based algorithm that takes reciprocal links into account is proposed and shown to have good performance in recommending both initial and reciprocal contacts. A recent study of Akehurst et al. (2011) finds that users with similar profiles like and dislike similar people and are liked and disliked by similar people. This hypothesis is used to build a content-collaborative reciprocal recommender, which is evaluated on a popular online dating dataset. In Cai et al. (2010), collaborative filtering algorithms are used to capture the reciprocity in people-to-people recommendation. The authors in Tu et al. (2014) propose a two-side matching framework for online dating recommendations and design a latent Dirichlet allocation (LDA) model to learn the user preferences from the observed user messaging behavior and user profile features.

In a recent study (Xia et al. 2015), we proposed different similarity measures that capture the unique features and characteristics of the online dating network, and our algorithms significantly outperform previously proposed approaches. In this paper, we further extend our previous work by studying the projection network characteristics, comparing the performance of our algorithms with state-of-the-art matrix factorization algorithm, analyzing the ranking effectiveness of collaborative filtering-based algorithms and investigating the reason on the poor performance given by content-based approaches.

The studies most relevant to our online dating recommendation problem are Pizzato et al. (2010b) and Zhao et al. (2014), in which a content-based algorithm (RECON) and a collaborative filtering-based algorithm (HCF) were presented and shown to outperform other approaches. In this paper, we will compare the performance of our proposed algorithms with these two algorithms.

3 Recommendation system

In this section, we propose a generalized reciprocal recommendation system that aims to match people with mutual interest in each other. We also introduce two previously proposed approaches, namely a content-based algorithm RECON (Pizzato et al. 2010b), a hybrid collaborative filtering algorithm (HCF) (Zhao et al. 2014), as

well as the matrix factorization (MF) algorithm (Koren et al. 2009) for comparison.

3.1 System design

The success of a reciprocal recommendation system lies in its ability to recommend users with whom the service user has mutual interest and thus they are likely to communicate with each other. The interaction records between a pair of users are a good indicator of actual interest and attractiveness between the sender and receiver. If a recommended user matches the service user's interest, the service user will be more likely to approach the recommended user. Also, if the service user's attractiveness is compatible with the recommended user's interest, the recommended user will be more likely to reply to the service user when contacted.

Figure 1 depicts an example of an online dating network. Based on user attributes and their communication traces, our goal is to match users with mutual interest in each other, for example, M3 and F3, M4 and F4. In the following, we will describe how to measure the similarity between a pair of users in terms of their dating interest and attractiveness, and how to construct various recommendation algorithms based on these similarity measures.

Our reciprocal recommendation system is comprised of the following four major components.

Extracting User-Based Features When a user registers with the online dating site, he/she needs to provide a variety of profile information including the user's gender, age, current location (city level), home town location, height, weight, body type, blood type, occupation, income range, education level, religion, astrological sign, marriage and children status, photographs, home ownership, car ownership, interests, smoking and drinking behavior. Most of these attributes are categorical features, except age, height, weight and number of photographs.
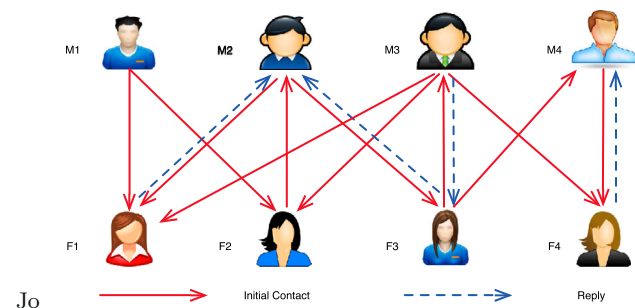

Fig. 1 Example of an online dating recommendation problem

Algorithm 1 Reciprocal Score(x, y)

Input: service user x, and recommended user y
Output: reciprocal score
begin
    /* initialize compatible scores s*/
    $s(x, y) = 0.0$
    $s(y, x) = 0.0$
    /* compute compatible scores for both x and y */
    foreach u in $Neighbor_1(y)$:
        $s(x, y) = s(x, y) + Similarity_1(x, u)$
    end for
    foreach v in $Neighbor_2(x)$:
        $s(y, x) = s(y, x) + Similarity_2(y, v)$
    end for
    /* normalize compatible scores for both x and y */
    if $|Neighbor_1(y)| > 0$
        $s(x, y) = \frac{s(x, y)}{|Neighbor_1(y)|}$
    if $|Neighbor_2(x)| > 0$
        $s(y, x) = \frac{s(y, x)}{|Neighbor_2(x)|}$
    /* compute reciprocal score */
    if $s(x, y) > 0$ and $s(y, x) > 0$
        return $\frac{2}{(s(x,y))^{-1} + (s(y,x))^{-1}}$
    else
        return 0.0
end

Extracting Graph-Based Features The online dating site we study is for heterosexual dating and only allows communications between users of opposite sex. The communication traces between users can be modeled as a bipartite network between males and females. A collection of graph-based similarity features are extracted from the bipartite graph that represent a user's active level in dating and attractiveness. The detailed definitions of these graph-based features are provided in Sect. 3.2.

Compute Reciprocal Scores Based on both user-based and graph-based features, we use the reciprocal score to measure the mutual interest and attractiveness between a pair of potential dates as described in Algorithm 1.

In Algorithm 1, $Neighbor_1()$ and $Neighbor_2()$ represent the neighbor set of a user with different directions in the bipartite network, and their formal definitions are given in Eqs. (6) and (7). $Similarity_1(,)$ and $Similarity_2(,)$ are customizable functions measuring the similarity between two users. We will discuss these functions in the following subsection. Given $Neighbor_1$, $Neighbor_2$, $Similarity_1$ and $Similarity_2$, $s(x, y)$ measures the similarity between user x and user y's neighbors, while $s(y, x)$ measures the similarity between user y and user x's neighbors. After computing $s(x, y)$ and $s(y, x)$, the reciprocal score is computed as the harmonic mean of these two similarity scores.

Generate Recommended User List For a given service user, a recommendation list will be generated by ranking these reciprocal scores. We will present the top-K users in the list to the service user. Note that the reciprocal score may not be symmetric if $Neighbor_1$ and $Neighbor_2$ are set as

different functions or $Similarity_1$ and $Similarity_2$ are set as different functions. This is different from the case in RECON where there is a unique reciprocal score for any pair of users.

3.2 Similarity functions

3.2.1 Content-based similarity functions

In content-based recommendation system, every recommended user can be represented by a feature vector or an attribute profile. These features can be numeric or nominal values representing some aspect of the user, such as age, height, income. Let A_x denote the set of known attributes (age, height, income, education level, etc) of user x, i.e.,

$$A_x = \{a : a \text{ is a known attribute of user } x\}$$

We denote the set of user-based attributes of user x as

$$U_x = \{v_a^x : \text{ for } a \in A_x\}, \qquad (1)$$

where v_a^x is the value of attribute a of user x.

The first similarity measure based on user attributes follows the work of RECON (Pizzato et al. 2010b), where the values of the numeric attributes (e.g., age, height) are grouped into ranged nominal values. For each service user, RECON builds his/her preference model by constructing the distribution of the receivers' user attributes. How much a service user is interested in a recommended user is then measured by comparing the attributes of the recommended user with the preference model of the service user. This is equivalent to computing the similarity between two users as follows:

$$\text{content-similarity}^a(x, y) = \frac{\sum_{a \in A_x \cap A_y} P_a(x, y)}{|A_x \cap A_y|} \qquad (2)$$

where

$$P_a(x, y) = \begin{cases} 1, & \text{if } v_a^x = v_a^y \\ 0, & \text{otherwise.} \end{cases} \qquad (3)$$

In RECON, numerical values are converted into categorical values, for example, users are divided into age groups 20–24, 25–29, 30–34, etc. Note that this method does not capture the numerical attribute information at the boundaries of continuous ranges. For example, a user of age 25 will not be considered to be similar to a user of age 24 as they fall into different age ranges. To avoid the loss of information, we modify the similarity measure defined in Eq. (2) as follows and use it as our second similarity measure, i.e.,

$$\text{content-similarity}^b(x, y) = \frac{\sum_{a \in A_x \cap A_y} Q_a(x, y)}{|A_x \cap A_y|}, \qquad (4)$$

where $Q_a(x, y) = P_a(x, y)$ for the nominal attributes as before; for numeric attributes, we define

$$Q_a(x, y) = \frac{v_a^* - |v_a^x - v_a^y|}{v_a^*}, \qquad (5)$$

where $v_a^* = \max_{i \neq j} |v_a^i - v_a^j|$ represents the maximum absolute difference for attribute a among all users. This new similarity measure results in a value between 0 when the attributes of the two users have the maximum difference (i.e., $|v_a^i - v_a^j| = v_a^*$) and 1 when the attributes of the two users are the same (i.e., $v_a^i = v_a^j$). For the above example, the age similarity between a 24-year-old user and a 25-year-old user will be 48/49 (very close to 1), where 49 is the max difference in user ages. It is clear that this new measure can better capture the similarity of numerical attributes between two users than the measure defined in Eq. (2), and as will be shown in Sect. 5, results in better performance in generating the recommendation list.

The above two similarity measures are based on user attributes and can be used to construct variations of content-based recommendation algorithms.

3.2.2 Graph-based similarity functions

Based on the message traces between users, we define the following two graph-based measures to capture the user's active level and attractiveness:

$$Se(x) = \{y : x \text{ has sent a messages to } y\} \qquad (6)$$

$$Re(x) = \{y : x \text{ has received a message from } y\} \qquad (7)$$

where $Se(x)$ is defined as the set of out-neighbors of x and its cardinality reflects the activeness of user x; $Re(x)$ is defined as the set of in-neighbors of x and its cardinality reflects the attractiveness of user x.

Based on the graph-based measures defined in Eqs. (6) and (7), we introduce the following two similarity functions:

- *Interest similarity* Consider two users of the same gender, x and y. If they both contact a same user, it shows that they share common interest in looking for potential dates. The fraction of users who received messages from both x and y among all users who received messages from either x or y serves as a measure of the similarity between the dating interest of the two users, i.e.,

$$\text{interest-similarity}(x,y) = \frac{|Se(x) \cap Se(y)|}{|Se(x) \cup Se(y)|}, \tag{8}$$

Note that we adopt the Jaccard coefficient in the interest similarity measure as the number of shared receivers from a pair of users is typically far outnumbered by the total number of receivers from the two users. This will become clear based on the degree distribution of users shown in Fig. 4 and weight distribution of the projection network shown in Fig. 6 from the dataset description in Sect. 4. For the example shown in Fig. 1, together M1 and M2 contacted three different females among which F2 received messages from both of them. The interest similarity between M1 and M2 is thus 1/3.

- *Attractiveness similarity* Consider two users of the same gender, x and y. If they both receive messages from a same user, it shows that they share common attractiveness to potential dates. The fraction of users who sent messages to both x and y among all users who sent messages to either x or y serves as a measure of the similarity between the attractiveness of the two users, i.e.,

$$\text{attractiveness-similarity}(x,y) = \frac{|Re(x) \cap Re(y)|}{|Re(x) \cup Re(y)|}. \tag{9}$$

By the same token we adopt the Jaccard coefficient in the attractiveness similarity measure. For the example shown in Fig. 1, both F1 and F2 received messages from M1 and M3 while F1 also received messages from M2. The attractiveness similarity between F1 and F2 is thus 2/3.

3.3 Recommendation algorithms

Based on these four similarity functions, we construct the following two content-based algorithms and four collaborative filtering algorithms.

The content-based algorithms are constructed based on the similarity of user attributes, including:

- **CB1 (RECON)** Both $Neighbor_1$ and $Neighbor_2$ in Algorithm 1 are set as out-neighbors $Se()$, and both $Similarity_1$ and $Similarity_2$ are computed using content similarity defined in Eq. (2). This algorithm is the same as RECON (Pizzato et al. 2010b).
- **CB2** Both $Neighbor_1$ and $Neighbor_2$ are set as out-neighbors $Se()$, and both $Similarity_1$ and $Similarity_2$ are computed using content similarity function defined in Eq. (4), where we do not convert numeric attributes into nominal attributes.

Collaborative filtering-based algorithms make use of the communication history of the service user as well as the

decisions made by users with similar interest or attractiveness to help make recommendations. Based on different combinations of users' dating interest and attractiveness, we construct the following four collaborative filtering-based recommendation algorithms:

- **CF1** Both $Neighbor_1$ and $Neighbor_2$ are set as out-neighbors $Se()$, and both $Similarity_1$ and $Similarity_2$ are computed using attractiveness similarity defined in Eq. (9). Therefore, for Algorithm 1, we have

$$s(x,y) = \frac{1}{|Se(y)|} \sum_{k \in Se(y)} \text{attractiveness\_similarity}(x,k)$$

$$s(y,x) = \frac{1}{|Se(x)|} \sum_{k \in Se(x)} \text{attractiveness\_similarity}(k,y)$$

In this case, $s(x,y)$ sums up the attractiveness similarity (i.e., contacted by same users) between service user x and users who have been contacted by user y, capturing the attractiveness of the service user x to user y. Similarly, $s(y,x)$ captures the attractiveness of user y to service user x. Putting these two factors together, this algorithm captures the mutual attractiveness between the service user and recommended users. An example of CF1 algorithm is shown in Fig. 2a.
- **CF2** Both $Neighbor_1$ and $Neighbor_2$ are set as in-neighbors $Re()$. Both $Similarity_1$ and $Similarity_2$ are computed using interest similarity defined in Eq. (8). In this case, $s(x,y)$ sums up the interest similarity between service user x and users who have contacted user y, capturing the interest from user x to y. Similarly, $s(y,x)$ captures the interest from user y to x. Together, this algorithm captures the mutual interest between the service user and recommended users. An example of CF2 algorithm is shown in Fig. 2b.
- **CF3** $Neighbor_1$ is set as out-neighbors $Se()$, while $Neighbor_2$ is set as out-neighbors $Re()$. $Similarity_1$ is computed using attractiveness similarity defined in Eq. (9), and $Similarity_2$ is computed using interest similarity defined in Eq. (8). This algorithm captures the interest from recommended users in the service user and the attractiveness of the service user to recommended users. An example of CF3 algorithm is shown in Fig. 3a.
- **CF4** $Neighbor_1$ is set as in-neighbors $Re()$, while $Neighbor_2$ is set as out-neighbors $Se()$. $Similarity_1$ is computed using interest similarity defined in Eq. (8), and $Similarity_2$ is computed using attractiveness similarity defined in Eq. (9). This algorithm captures the attractiveness of recommended users to the service user and the interest of the service user in recommended users. An example of CF4 algorithm is shown in Fig. 3b.

Fig. 2 Example of **a** CF1 and **b** CF2 algorithms

(a)   (b)
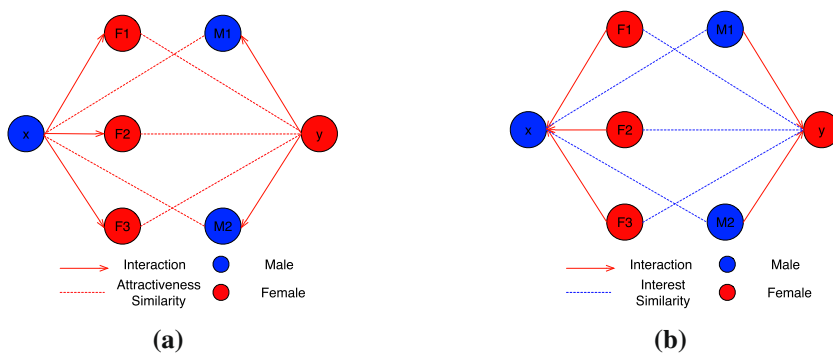
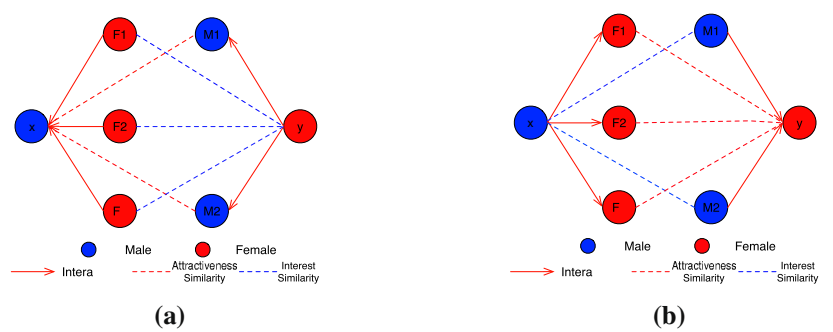**Fig. 3** Example of **a** CF3 and **b** CF4 algorithms

(a)   (b)

Figure 2 illustrates the mechanism of CF1 and CF2 algorithms. In the example of CF1 shown in Fig. 2a, service user x sent messages to users $F1$, $F2$ and $F3$. These out-neighbors of service user x share similar attractiveness with user y, i.e., user y and $F1$, $F2$, $F3$ received messages from at least one common user. Also, user y sent messages to users $M1$ and $M2$ who share similar attractiveness of user x. Therefore, CF1 captures the mutual attractiveness between the service user and recommended users. If the reciprocate score between x and y ranks in the top-K position, user y will be included in the recommendation list for service user x. Figure 2b shows an example of CF2, which captures the mutual interest between the service user x and user y, i.e., user x's interest in user y and user y's interest in service user x. Examples of CF3 and CF4 illustrated in Fig. 3 can be interpreted in a similar way.

In addition to the above algorithms, we also implement the content-based algorithm (RECON) proposed in Pizzato et al. (2010b) and the hybrid collaborative filtering algorithm (HCF) proposed in Zhao et al. (2014). In particular, RECON corresponds to our CB1 algorithm, where the $Neighbor_1$ and $Neighbor_2$ are set as $Se()$ function, and $Similarity_1$ and $Similarity_2$ are computed based on Eq. (2). HCF extends the baseline collaborative filtering approach by considering both initial and reciprocal contacts to compute the similarity between two users, where reciprocal links are given higher weight than single-direction contacts. These two algorithms are most related to our study

and have been shown to outperform many other approaches. Matrix factorization algorithm has been widely adopted in recommendation systems and link prediction problems. For traditional user-item recommendations, the rating matrix R can be factorized into two matrices P and Q, where each row of P represents the latent factor of a user and each column of Q denotes the latent representation of an item. To the best of our knowledge, matrix factorization algorithm has not been applied to recommendations in prior online dating recommendation research. In this paper, we will compare the performance of our proposed algorithms with three algorithms, namely RECON, HCF and MF, in Sect. 5.

4 Dataset description

4.1 Dataset overview

The dataset used in our study is obtained through a collaboration with baihe.com, one of the major online dating sites in China. Our dataset includes the profile information of 200,000 users uniformly sampled from users registered in November of 2011. Of the 200,000 sampled users, 139,482 are males and 60,518 are females, constituting 69.7 and 30.3 % of the total number of sampled users, respectively. For each user, we have his/her message sending and receiving traces (who contacted whom at what

time) in the online dating site and the profile information of the users that he or she has communicated with from the date that the account was created until the end of January 2012. Note that the site is for heterosexual dating and only allows communications between users of opposite sex.

After a user creates an account on the online dating site, he/she can search for potential dates based on information within the profiles provided by the other users including user location, age, etc. Once a potential date has been discovered, the user then sends a message to him/her, which may or may not be replied by the recipient. In this paper, we focus on the prediction of whether a user will reply to initial messages sent by other users. Subsequent interactions between the same pair of users do not represent a new sender–receiver pair and can not be used as the only indicator for continuing relationship as users may choose to go off-line from the site and communicate via other channels (e.g., email, phone or meet in person).

Since we only have eight full weeks' worth of online dating interaction records for our sample users, we will consider the activities of each user during the first eight weeks of his/her membership. Table 1 describes the characteristics of the dataset. More detailed description and analysis of the dataset can be found in our recent work (Xia et al. 2013, 2014b).

For both males and females, we obtain the distribution of the number of messages sent by each user per week given that a user sends at least one message during the week, and

Table 1 Dataset description

Type	Initial contact links	Reciprocal links (reply rate)
Male to sample female	1,586,059	150,917 (9.5 %)
Female to sample male	328,645	58,946 (17.9 %)

plot its complementary cumulative density function (CCDF) in Fig. 4a. We observe that the distributions exhibit heavy tails. Most users only sent out a small number of messages: 94.6 % of males and 96.5 % of females sent out less than 100 messages during the first eight weeks of their membership. On the other hand, there are small fractions of users that sent out a large number of messages. According to the online dating site, most of these highly active users are likely to be fake identities created by spammers and their accounts have been quickly removed from the site.

The distribution of number of messages received by a user is plotted in Fig. 4b. A female is likely to receive more messages than a male. Most users only received a small number of messages: 99.3 % of males and 90.1 % of females received less than 100 messages during the first eight weeks of their membership. On average, a male received 7 messages, while a female received 35 messages during the first eight weeks.

4.2 Projection network characteristics

Based on user communication traces, we construct several projection networks for each gender and direction of communications (sending or receiving). The sending projection network is constructed by adding an edge between two users who have sent messages to at least one common receiver, while the receiving projection network is constructed by adding an edge between two users who received messages from at least one common sender. The weight of each edge in a sending or receiving projection network denotes the number of common receivers or senders between the two nodes, respectively. Figure 5 illustrates the sending projection network of male users and receiving projection network of female users corresponding to the example shown in Fig. 1.
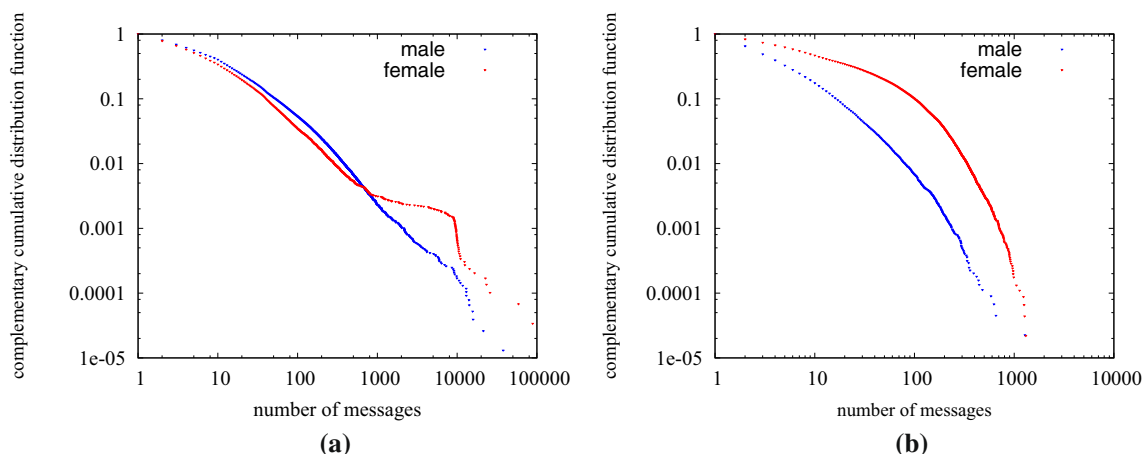
Fig. 4 a CCDF of the number of messages a user sent out during the first eight weeks of his/her membership. **b** CCDF of the number of messages a user received during the first eight weeks of his/her membership

Table 2 describes several important network measurements of the sending and receiving projection networks for male and female users.

Figure 6 shows the node degree and edge weight distributions for the sending projection network of both male and female users. The node degree in a sending projection network represents the number of other users with whom the user shares some degree of similar interest. We observe that a male shares similar interest with a larger number of peers than a female. The median degrees for male and female users are 89 and 33, respectively. We also observe that most edges in the sending projection networks have a low weight (i.e., most pairs of users contact very few common receivers), in particular, 73.1 and 76.8 % of the edges in the male and female sending projection networks have a weight of 1. Also, the CCDF of edge weight

distribution of females lies above that of males, indicating that females tend to share more common receivers than males.

Figure 7 plots the distributions of node degree and edge weight of the receiving projection network for both male and female users. The median degrees for male and female users are 283 and 551, respectively. Most of the edges have a low weight (i.e., most pairs of users are contacted by very few common senders), in particular, 85.5 and 70.6 % of the edges in the male and female receiving projection networks have a weight of 1.

5 Evaluation

For a given service users in our test set, we rank the recommended users by comparing their reciprocal scores and recommend the top-K users in the list. We evaluate the performance of each algorithm by comparing the top-K users in the recommended list with the receivers contacted by the service user in test set.

5.1 Experiment setup

Most of the active users in our dataset are newly registered users. They are usually very active in looking for a potential date in the first one or two weeks after registration
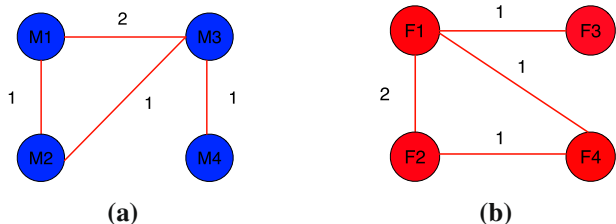
Fig. 5 a Male sending projection network and **b** female receiving projection network corresponding to the example shown in Fig. 1

Table 2 Number of nodes and edges in projection networks

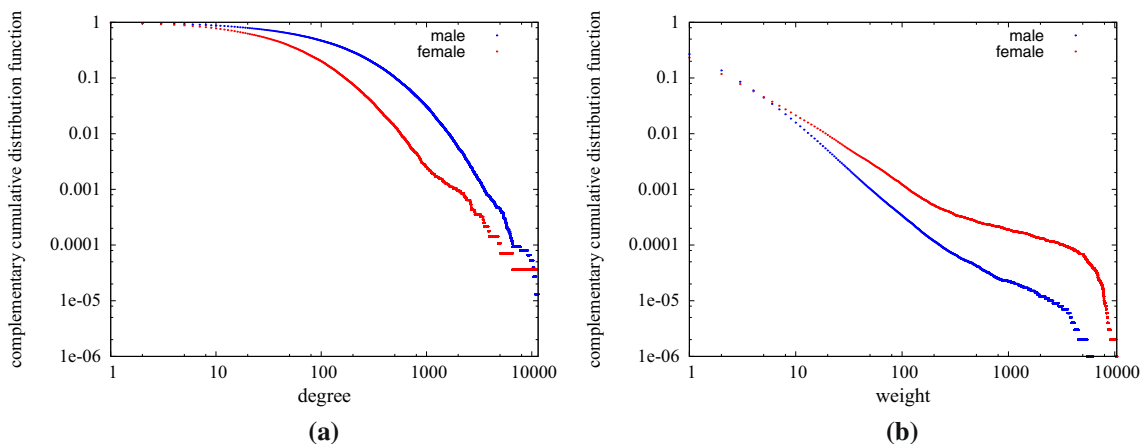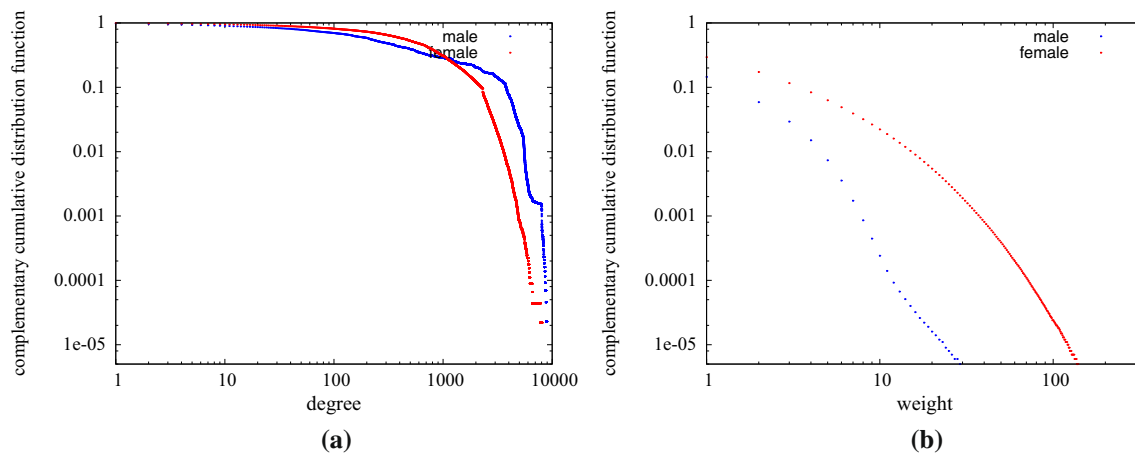| Network type	Gender	Nodes	Edges
Sending projection network	Male to female	75,379	7,716,078
	Female to male	28,550	1,025,738
Receiving projection network	Male from female	43,420	22,603,491
	Female from male	45,214	18,858,211

Fig. 6 CCDF for **a** degree and **b** weight distribution for the sending projection networks

Fig. 7 CCDF for **a** degree and **b** weight distribution for the receiving projection networks

Table 3 Experiment dataset

Male	Female	# of messages in training set	# of messages in test set
24,602	8250	730,110	270,294

(Xia et al. 2013). We select the user interactions within 10 days from user registration time as the training data, and interactions in the remaining time as the test set.

We filter the service users by selecting users who have sent or replied at least 5 messages in the training period, and use the interactions between these users in the test period as the test set. For the training set, we count all the interactions initiated, received or replied to by the selected users in the training period, and use these interactions to train our recommendation system. Table 3 summarizes the experiment dataset.

For RECON (CB1) and CB2, we manually pick 20 features over all 39 features. These selected features include age, height, weight, city, education level, income, house status, marriage status, children status, physical looking, car status, number of photographs, smoking habit, drinking habit, marriage status, parents status, children plan, dating method and wedding plan. Among these features, age, height, weight and number of photographs are treated as numeric values. For HCF, we performed several experiments to get the optimal weight parameter s in the computation of the success score between two users (Zhao et al. 2014). For matrix factorization algorithm, we treat the single-direction communications between two users as rating 1 and reciprocal communications between two users

as rating 2. Different experiments are conducted to find the optimal number of latent factors that minimizes the objective function.

5.2 Evaluation metrics

For a given service user, we define three set of users: T as the set of users we have recommended to the service user, I as the collections of users who have been contacted by the service user and R as the set of users who have been contacted by the service user and replied to the service user in the test set. We define the following two different evaluation metrics:

$$I\text{-}Precision = \frac{|I \cap T|}{|T|}, \quad I\text{-}Recall = \frac{|I \cap T|}{|I|}, \qquad (10)$$

and

$$R\text{-}Precision = \frac{|R \cap T|}{|T|}, \quad R\text{-}Recall = \frac{|R \cap T|}{|R|}, \qquad (11)$$

where I-Precision and R-Precision measure the ratio of users in the recommendation list who have been contacted by or exchanged messages with the service user, respectively. I-Recall and R-Recall measure the ratio of users who have been contacted by or exchanged messages with the service user in the list of recommended users. From another perspective, I-Precision and I-Recall measure an algorithm's performance in recommending users that the service user is interested in and thus likely to contact, and R-Precision and R-Recall measure an algorithm's performance in recommending users who have mutual interest with the service user and are thus likely to reciprocate when contacted.

5.3 Evaluation results

In this subsection, we apply our recommendation algorithms on the experiment datasets and compare the performance with RECON (Pizzato et al. 2010b), HCF (Zhao et al. 2014) and MF (Koren et al. 2009). We first report the performance of recommending users that the service user will contact and then evaluate the performance of recommending users who will reciprocate when contacted by the service user.

5.3.1 I-Precision and I-Recall

We first examine the performance of these algorithms in recommending users whom the service user is interested in and thus likely to contact.

Figures 8 and 9 show the performance of the two content-based algorithms, namely CB1(RECON) and CB2. We observe that by preserving the values of numeric attributes in the similarity measure, CB2 significantly outperforms CB1 in both precision and recall. The improvement is more pronounced for females than for males.

Figures 10 and 11 show the *I-Precision* and *I-Recall* of the collaborative filtering-based recommendation algorithms for male and female users, respectively. We observe that for both male and female users, the four algorithms proposed in this paper, CF1 to CF4, significantly outperform previously proposed HCF algorithm. For male users, while there are some difference in the performance of CF1, CF2 and CF3, the difference is rather small when compared with CF4, which is much more effective in attracting the service user to contact the recommended users. However, for female users, the four algorithms (CF1 to CF4) all perform similarly. There is no algorithm significantly outperforming others. Note that the CF4 algorithm captures the interest of the service user in recommended users as

well as the attractiveness of the recommended users to the service user. The results indicate that when it comes to looking for potential dates, males tend to be more focused on their own interest and oblivious toward their attractiveness to potential dates, while females do not show such behavior. We observe that the matrix factorization algorithm outperforms CF1, CF2 and CF3 algorithms for male users, but not for female users. The reason is that female-to-male matrix (0.02 % Sparseness) is more sparse than the male-to-female matrix (0.05 % sparseness). Our algorithms, especially CF4, yield better performance than the matrix factorization algorithm because the latter only considers the direct communications between users while our algorithm can utilize the communications between users who share similar interest or attractiveness.

5.3.2 R-Precision and R-Recall

We now examine the performance of these algorithms in recommending users who will be contacted by and exchange messages with the service user.

Figures 12 and 13 show the *R-Precision* and *R-Recall* of the two content-based algorithms for male and female users, respectively. Similar to *I-Precision* and *I-Recall*, CB2 performs much better than CB1(RECON), as CB2 does not convert numeric attributes into nominal attributes and thus does not loss information of these numeric attributes.

The performance of the collaborative filtering-based algorithms for male and female users is shown in Figs. 14 and 15, respectively. The algorithms proposed in our paper (CF1–CF4) still achieve much better results than HCF. The performance of the matrix factorization algorithm is similar to the case for *I-Precision* and *I-Recall*. It outperforms CF1, CF2, CF3 for male users, but not for female users. For male users, while CF4 still outperforms the other
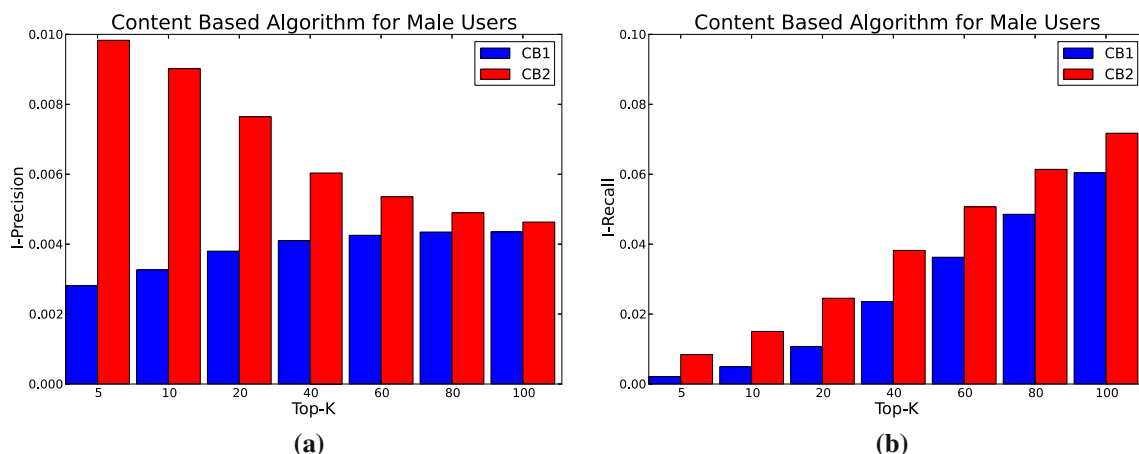

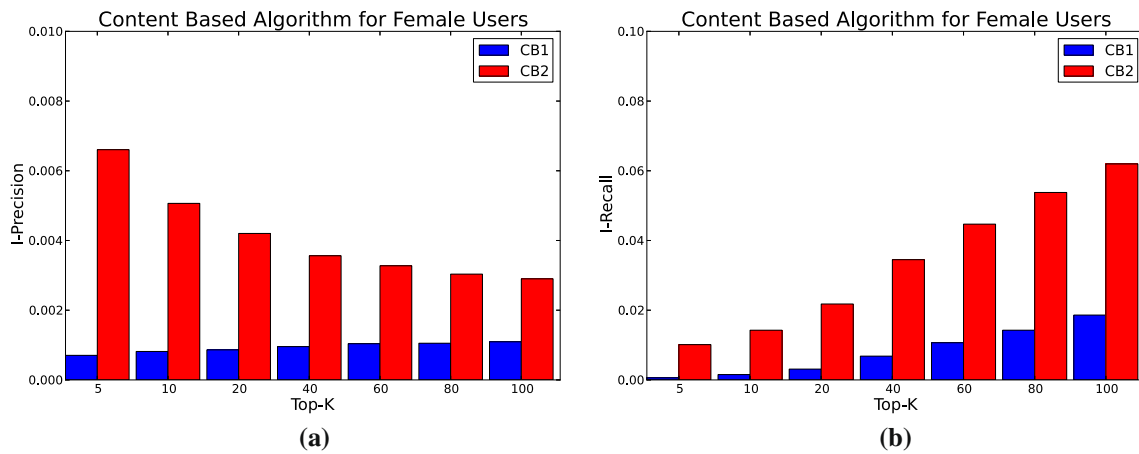**Fig. 8** *I-Precision* and *I-Recall* of content-based algorithms for male users

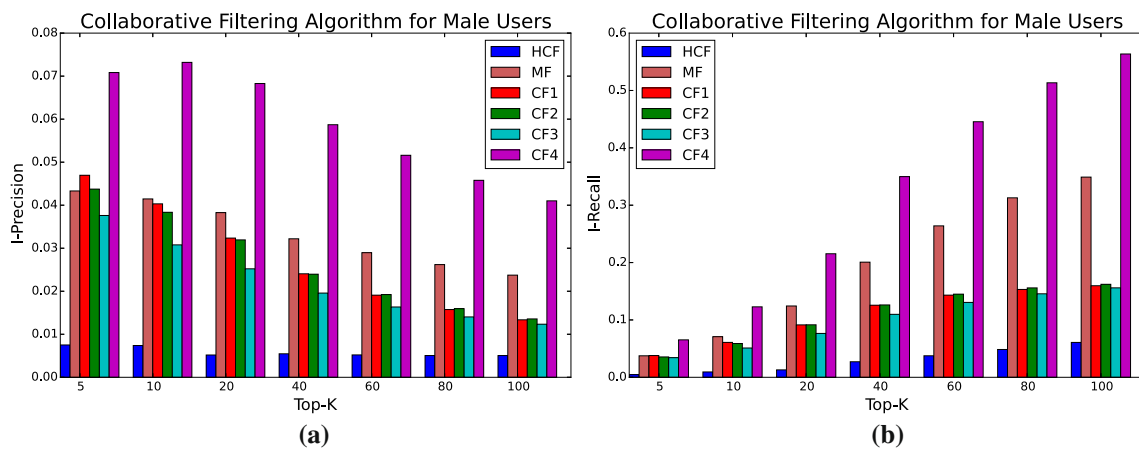**Fig. 9** *I-Precision* and *I-Recall* of content-based algorithms for female users

Fig. 10 *I-Precision* and *I-Recall* of collaborative filtering algorithms for male users
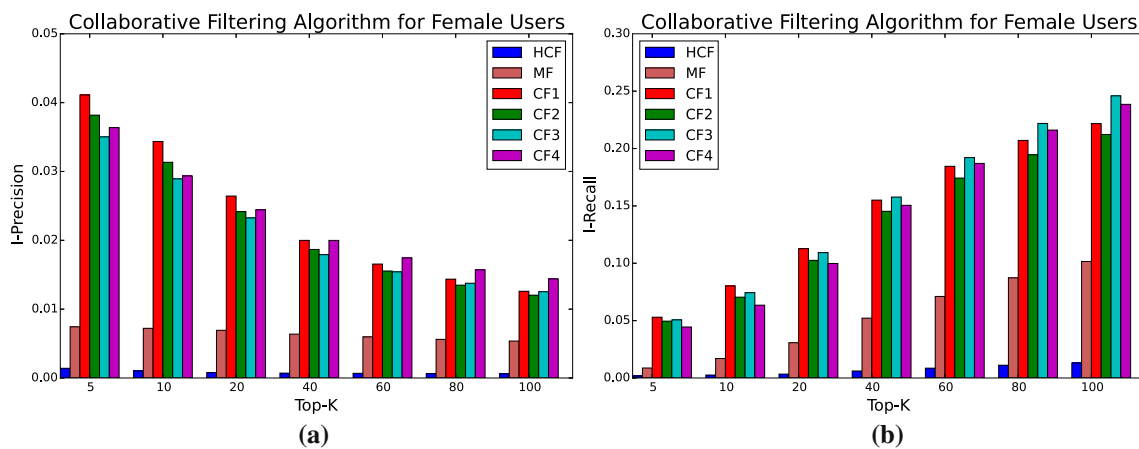

Fig. 11 *I-Precision* and *I-Recall* of collaborative filtering algorithms for female users

algorithms, the difference is not as pronounced as for *I-Precision* and *I-Recall* measures. For female users, CF1 and CF2 show very similar behavior. Recall that CF1/CF2 captures the mutual interest/attractiveness between the

service user and recommended users. This shows that learning the mutual interest and mutual attractiveness between two users has similar effects for recommending potential dates for females. Unlike for male users, CF4
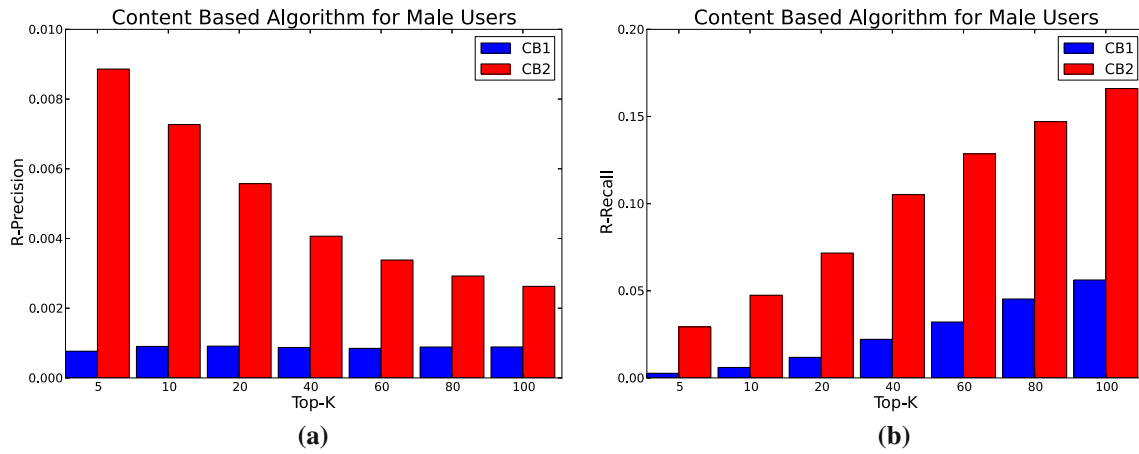
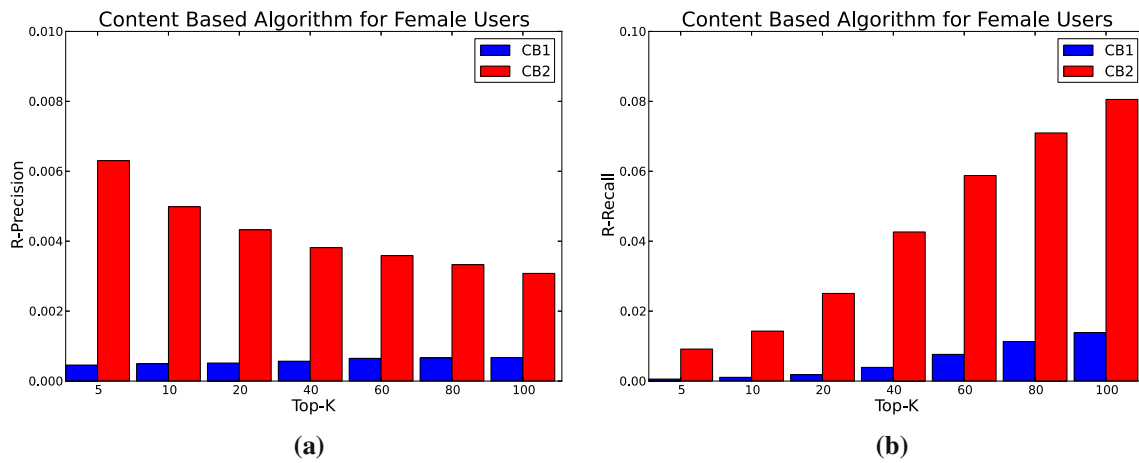**Fig. 12** *R-Precision* and *R-Recall* of content-based algorithms for male users

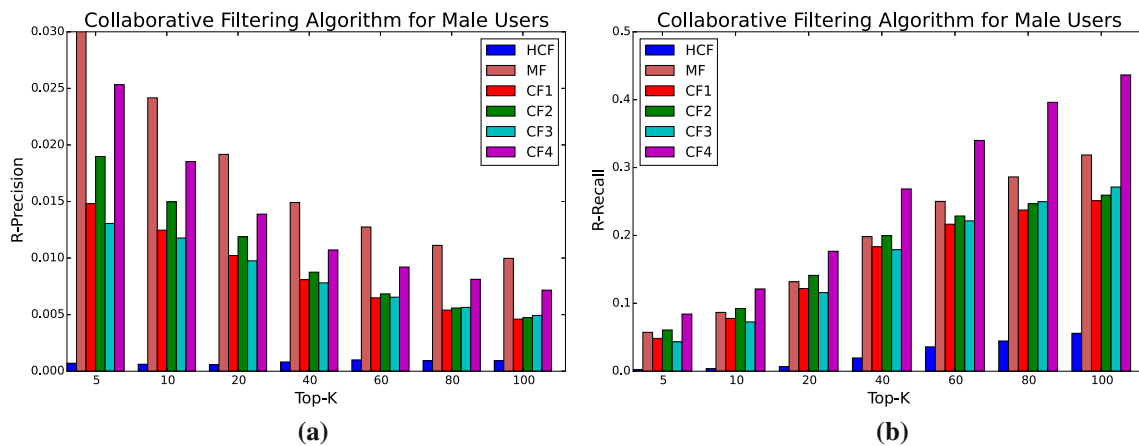**Fig. 13** *R-Precision* and *R-Recall* of content-based algorithms for female users

Fig. 14 *R-Precision* and *R-Recall* of collaborative filtering algorithms for male users

does not outperform other algorithms for female users. On the contrary, when the recommendation list (K) becomes large, CF3 starts to outperform the other algorithms. Recall

that CF3 is symmetric to CF4, capturing the interest from recommended users in the service user and the attractive of the service user to the recommended users. The results
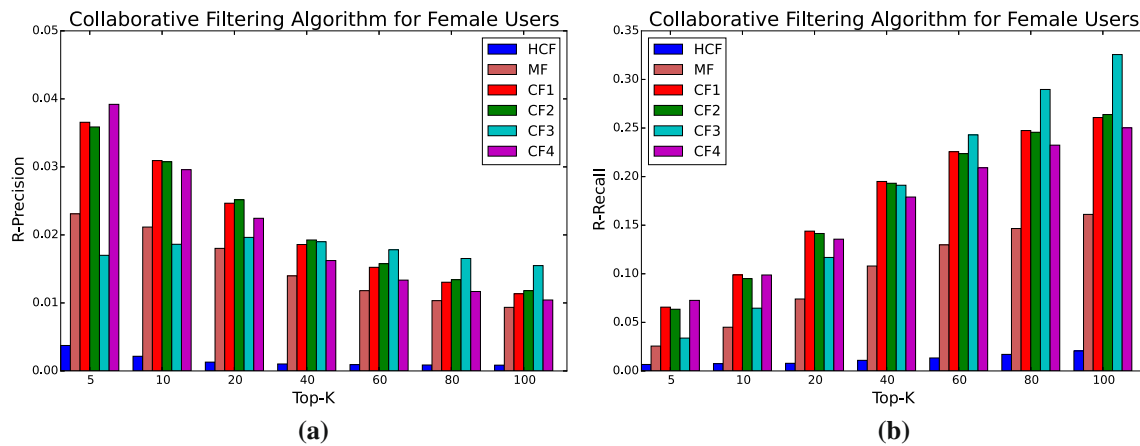
Fig. 15 *R-Precision* and *R-Recall* of collaborative filtering algorithms for female users
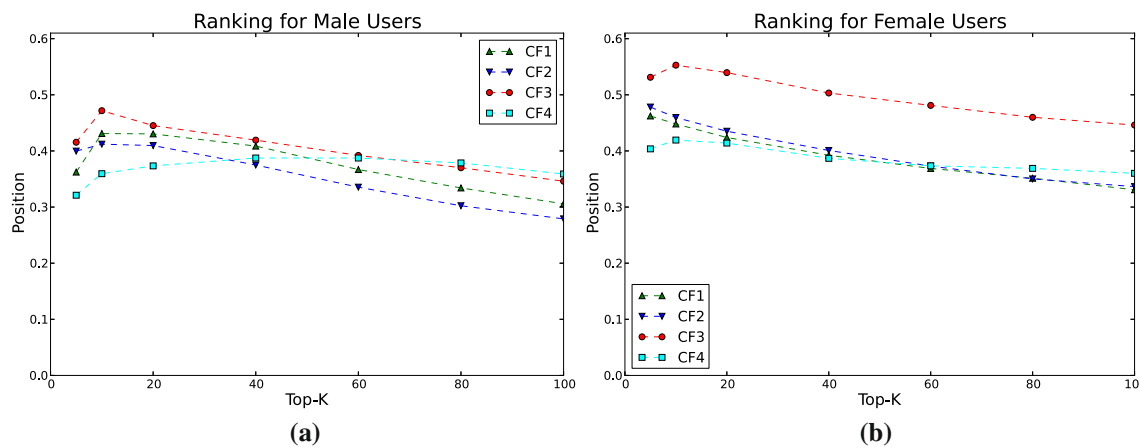

Fig. 16 Average effective recommendation position of proposed recommendation algorithms for **a** male users and **b** female users

indicate that when females look for potential dates, they are more conscientious to their own attractiveness to the other side of the line and the other sides' interest in them.

Note that in our experiments collaborative filtering-based algorithms (CF1–CF4) yield much higher precision and recall than content-based algorithms (CB1, CB2). One main drawback of collaborative filtering-based approach is that it can not address the cold start problem. On the other hand, content-based algorithm RECON (Pizzato et al. 2010b) provides a feasible solution to the cold start problem by recommending users who are interested to the new service user.

5.4 Ranking effectiveness

In addition to precision and recall, the relative positions of relevant recommendations are also an important measure for a recommendation system. Relevant recommendations are defined as users in the recommendation list who have actually exchanged messages with the service user, i.e., the

service user has followed the recommendation by contacting the recommended user who in turn has replied to the service user. Since a user usually looks at the recommendation list from top to bottom, a recommendation system ranking the relevant recommendations in top positions should be considered better than those with similar performance in precision and recall but ranking the relevant recommendations in lower positions.

Figure 16 plots the average positions of the relevant recommendations in the recommendation list (normalized by the size of recommendation list). All of these algorithms rank the effectively recommended users in the top 30 to 50 % of the recommendation list except for CF3 for female users which ranks the relevant recommendations around the halfway of the recommendation list.

5.5 Discussions

To illustrate the relatively poor performance of the content-based algorithms when compared with the collaborative
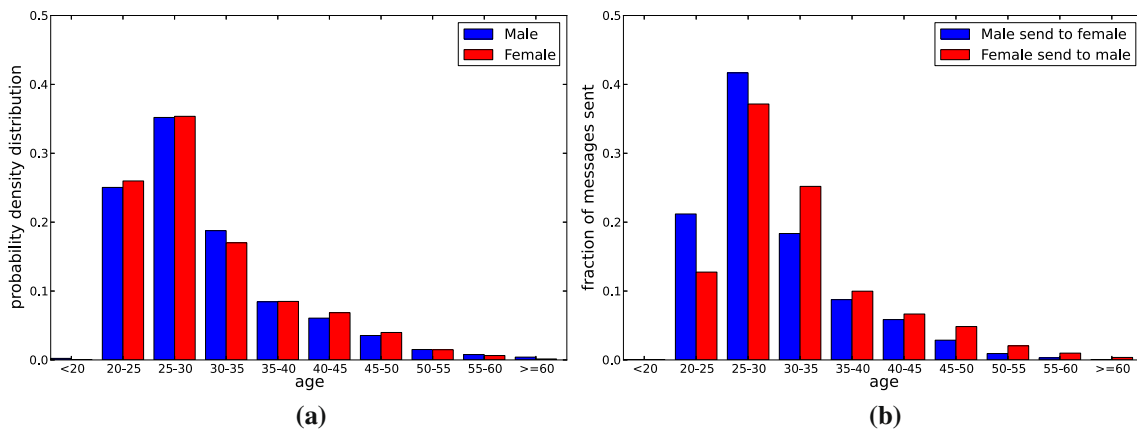
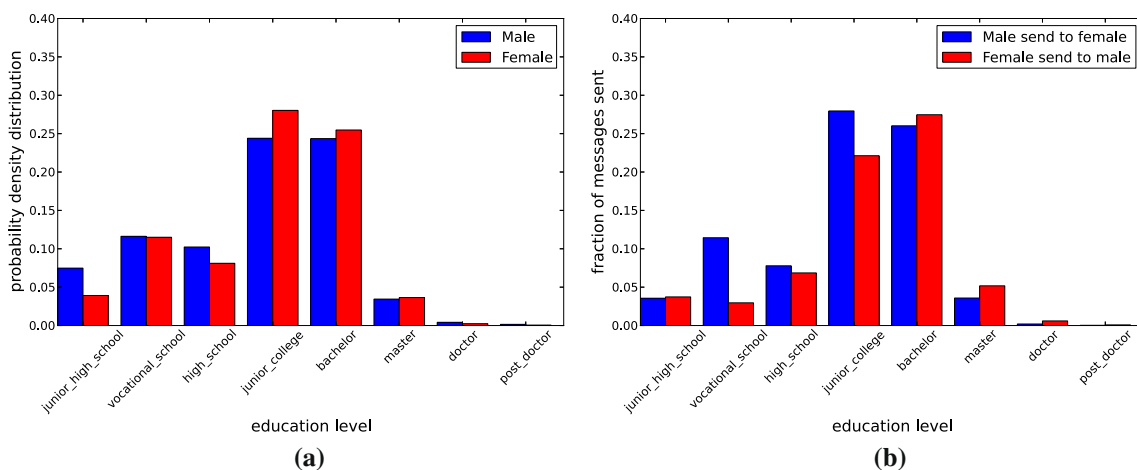**Fig. 17** **a** Age distribution of all users, **b** age distribution of messages sent

Fig. 18 **a** Education distribution of all users, **b** education status distribution of messages sent
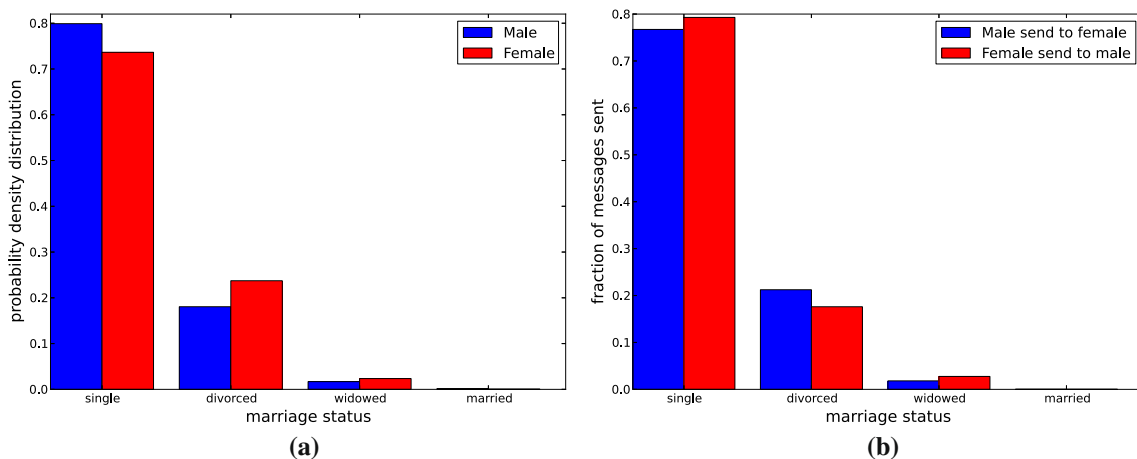

Fig. 19 **a** Marriage distribution of all users, **b** marriage status distribution of messages sent

filtering-based approaches, we examine the effectiveness of using a user's attributes to determine whether another user would send a message to him or her.

We plot several attribute distributions of all users as well as those of users who received messages for the age (Fig. 17), education level (Fig. 18) and marriage status

(Fig. 19) attributes. We observe that for these attributes, the distributions for message receivers are quite similar to those for all users with small Bhattacharyya distances, indicating that these attributes are not very effective in making recommendations. Specifically, the Bhattacharyya distance between the two age distributions is 0.122 for males and 0.064 for females. The Bhattacharyya distance between the two education distributions is 0.155 for males and 0.011 for females. For marriage status, the Bhattacharyya distance is below 0.032 for both males and females. We also examined other attributes, which show small Bhattacharyya distances too. For users in the age range of 20–30, junior college or bachelor degrees, or single marriage status, it is difficult to distinguish them as these users constitute the majority of the population.

6 Conclusions

Matching users with mutual interest in each other is an important task for online dating sites. In this paper, we propose a set of similarity-based reciprocal recommendation algorithms for online dating. We introduce several similarity measures that characterize the attractiveness and interest between two users and select most compatible users for recommendations. We evaluate the performance of our proposed algorithms on a large dataset obtained from a major online dating site in China. Our results show that the collaborative filtering-based algorithms achieve much better performance than content-based algorithms in both precision and recall, and both significantly outperform previously proposed approaches. Our results also show that male and female users behave differently when it comes to looking for potential dates. In particular, males tend to be focused on their own interest and oblivious toward their attractiveness to potential dates, while females are more conscientious to their own attractiveness to the other side on the line.

References

Akehurst J, Koprinska I, Yacef K, Pizzato L, Kay J, Rej T (2011) CCR—a content-collaborative reciprocal recommender for online dating. In: Proceedings of the twenty-second international joint conference on artificial intelligence

Cai X, Bain M, Krzywicki A, Wobcke W, Kim YS, Compton P, Mahidadia A (2010) Collaborative fileterfing for people-to-people recommendation in social networks. In: Austrlian joint conference on artifical intelligence

Hannon J, Bennett M, Smyth B (2010) Recommending Twitter users to follow using content and collaborative filtering approaches. In: Proceedings of the 2010 ACM conference on recommendation system

Hopcroft J, Lou T, Tang J (2011) Who will follow you back? Reciprocal relationship prediction. In: Proceedings of the 2011 international conference on information and knowledge management

Koren Y, Bell R, Volinsky C (2009) Matrix factorization techniques for recommender systems. IEEE Comput 8:30–38

Li L, Li T (2012) MEET: a generalized framework for reciprocal recommender systems. In: Proceedings of ACM international conference on information and knowledge management

Mori J, Kajikawa Y, Kashima H, Sakata I (2012) Machine learning approach for finding business partners and building reciprocal relationships. Expert Syst Appl Int J 39(12):10402–10407

Pizzato LA, Silvestrini C (2011) Stochastic matching and collaborative filtering to recommend people to people. In: Proceedings of the 2011 ACM conference on recommendation system

Pizzato L, Chung T, Rej T, Koprinska I, Yacef K, Kay J (2010a) Learning user preference in online dating. In: Proceedings of the preference learning (PL-10) tutorial and workshop. In: European conference on machine learning and principles and practice of knowledge discovery in databases

Pizzato L, Rej T, Chung T, Koprinska I, Kay J (2010b) RECON: a reciprocal recommender for online dating. In: Proceedings of the 2010 ACM conference on recommendation system

Pizzato L, Rej T, Yacef K, Koprinska I, Kay J (2011) Finding someone you will like and who won't reject you. In: Proceedings of the 19th international conference onuser modeling, adaptation, and personalization

Shi Y, Karatzoglou A, Baltrunas L, Larson M, Oliver N, Hanjalic A (2012) CLiMF: learning to maximize reciprocal rank with collaborative less-is-more filtering. In: Proceedings of the 2012 ACM conference on recommendation system

Shi Y, Karatzoglou A, Baltrunas L, Larson M, Hanjalic A (2013) xCLiMF: optimizing expected reciprocal rank for data with multiple levels of relevance. In: Proceedings of the 2013 ACM conference on recommendation system

Tu K, Ribeiro B, Jensen D, Towsley D, Liu B, Jiang H, Wang X (2014) Online dating recommendations: matching markets and learning preferences. In: Proceedings of 5th international workshop on social recommender systems, in conjunction with 23rd international world wide web conference

Xia P, Ribeiro B, Chen C, Liu B, Towsley D (2013) A study of user behaviors on an online dating site. In: Proceedings of the IEEE/ACM international conference on advances in social networks analysis and mining

Xia P, Jiang H, Wang X, Chen C, Liu B (2014a) Predicting user replying behavior on a large online dating site. In: Proceedings of 8th international AAAI conference on weblogs and social media

Xia P, Tu K, Ribeiro B, Jiang H, Wang X, Chen C, Liu B, Towsley D (2014b) Characterization of user online dating behavior and preference on a large online dating site. In: Missaoui R, Sarr I (eds) Social network analysis—community detection and evolution, Springer International Publishing, Switzerland, pp 193–217. http://link.springer.com/chapter/10.1007%2F978-3-319-12188-8_9

Xia P, Liu B, Sun Y, Chen C (2015) Reciprocal recommendation system for online dating. In: Proceedings of the IEEE/ACM international conference on advances in social networks analysis and mining

Zhao G, Lee ML, Hsu W, Chen W, Hu H (2013) Community-based user recommendation in uni-directional social networks. In: Proceedings of the 2013 international conference on information and knowledge management

Zhao K, Wang X, Yu M, Gao B (2014) User recommendation in reciprocal and bipartite social networks—a case study of online dating. In: IEEE intelligent systems