This is the author's version published as:


Copyright 2010 IEEE
A Social Matching System for an Online Dating Network: A Preliminary Study

Richi Nayak  
Computer Science Discipline  
Queensland University of Technology  
Brisbane, Australia  
r.nayak@qut.edu.au

Meng Zhang  
Computer Science Discipline  
Queensland University of Technology  
Brisbane, Australia

Lin Chen  
Computer Science Discipline  
Queensland University of Technology  
Brisbane, Australia

Abstract—Due to the change in attitudes and lifestyles, people expect to find new partners and friends via various ways nowadays. Online dating networks create a network for people to meet each other and allow making contact with different objectives of developing a personal, romantic or sexual relationship. Due to the higher expectation of users, online matching companies are trying to adopt recommender systems. However, the existing recommendation techniques such as content-based, collaborative filtering or hybrid techniques focus on users explicit contact behaviors but ignore the implicit relationship among users in the network. This paper proposes a social matching system that uses past relations and user similarities in finding potential matches. The proposed system is evaluated on the dataset collected from an online dating network. Empirical analysis shows that the recommendation success rate has increased to 31% as compared to the baseline success rate of 19%.

Keywords- Social Network Analysis, Recommender Systems, Social Matching, Clustering, online dating

I. INTRODUCTION

The current revolution of the World Wide Web is pushing the end users to exploit the web as a collaborative platform. According to a report, uploads account for up to 44% of Internet traffic in year 2007 as compared with year 1997 when only 10-20% of Internet traffic was uploads [1]. Users participate on the Web via Web blogs, forums, wiki, YouTube etc. This is resulting into stronger and more frequent online relations between people. As the popularity of Web grows, numerous online social networks have been introduced and the number is soaring.

Match making is no different. Match making has grown out of the old fashioned agency arrangement for the partner search to online dating networks. Online dating is fast becoming a high growth industry with the US expenditure on dating networks rising from US$32 million to US$56 million over the past three years [2]. The current booming in the use of online social networking by people from various different demographics, has increased the customer base of online dating companies and has made using these services more socially acceptable [2]. It is reported that there are around 8 million singles in Australia [3] and 54.32% of them are using online dating services. Users of online dating services are overwhelmed by the information retrieved by using these services. The process of selecting the right person among a vast amount of candidates becomes tedious and nearly ineffective if the automatic selection process is not available. There is an increasing demand for dating networks to deploy a solution that dynamically assists in the match making process by suggesting potential matches. Therefore, a social matching system, utilizing data mining to predict behaviors and attributes that could lead to future successful matches, which allows personalized matches be suggested to the users becomes a necessity.

Recommendation systems have existed for a long time to suggest users a product according to their web visit histories or based on the product selections of other similar users [4] [5]. One of the most reputed recommendation systems is Amazon (www.amazon.com). In most cases, the recommendation is an item recommendation which is inanimate. On the contrary, the recommendation in dating networks is made about people who are animate. Different from item recommendation, people recommendation is a form of two way matching. Where a person can refuse an invitation, products cannot refuse to be sold [6]. The goal of an e-commerce recommendation system is to find products most likely to interest a user, whereas, the goal of a social network recommendation system is to find the user who is mostly likely to interest the user and respond favorable to them. Current recommendation systems cannot handle this well [6].

There are few published examples of recommendation systems applied explicitly to online dating. Authors in [7] treat the online dating recommendation as any other recommendation system. Traditional recommendation algorithms, including user-user algorithms and item-item algorithms are used. In the user-user algorithm, the rating prediction of user X to user Y comes from ratings of those users who are similar to user X and have rated user Y. The Item-Item algorithm collects all the ratings of user X. When the prediction of rating X on user Y is needed, the prediction utilizes the rating of user X on all the other users who are similar to user Y. The problem with this method is that match making is different from item recommendation in that item cannot choose the buyer but dating service users can choose the dating candidate. In this method, the rating is the only parameter which affects the match making algorithm. But it is not the case in reality. Many factors, such as age,
job, ethnicity, education etc play important roles in the match making process. Authors in [6] proposed a theoretical generic recommendation algorithm for social networks that can easily be applied to an online dating context. Their system is based on a concept of social capital which combines direct similarity from static attributes, complementary relationship(s), general activity and the strength of relationship(s). However, this work is at a theoretical level and there have been no experiments carried out to prove the effectiveness of this theory. There are many weight factors in the proposed algorithm which may negatively influence it being an effective algorithm. Efficiency is another problem for this pairwise algorithm with a very high computation complexity.

This paper proposes a social matching system that combines the social network knowledge with content-based and collaborative filtering techniques of recommendation by utilizing users' past relations and user similarities to improve recommendation quality. It follows the premise that if two users are in a successful relationship (as shown by a positive message exchange) and a previous partner of each user who had exchanged a positive message is similar to the respective user, then the two previous partners have a high probability to be the best match to each other. This system includes a nearest neighbour algorithm which provides the system an add-on layer to group similar users [4][5]. It also includes a relationship-based user similarity prediction algorithm which is applied to calculate similarity scores and generate recommendations.

The proposed system is evaluated on the dataset collected from a popular Australian dating network. Empirical analysis shows that the proposed system is able to recommend the top-N users with high accuracy. The recommendation success rate has increased to 31% as compared to the baseline success rate of 19%. The baseline recall of the underlying dating network is also increased from 0.3% to 9.2% respectively.

The outline of this paper is as follows: Section 2 details the proposed social matching system; Section 3 presents the experimental results and analysis. Finally, we conclude in Section 4.

II. THE PROPOSED METHOD

A. Preliminaries

Data required by a dating network for recommending potential partners can be divided into the following features: (1) Personal profile for each user which includes self details on demographic, fixed-choice responses on Physical, Identity, Lifestyle, Career and Education, Politics and Religion and other attributes, free-text responses to various interests such as sport, music etc, and optionally, one or more photographs; (2) Ideal partner profile for each user which includes information about what user prefers in Ideal partner, usually the multiple choices on the attributes discussed before; (3) User activities on the network such as viewing the profiles of other members, sending pre-typed messages to other users; sending emails or chat invitations; and (4) Measure of relationships with other users such as willingness to initialize relationships and responding to invitations, and frequency and intensity with which all relationships are maintained. A relationship can be called successful for the purpose of match making when a user initiates a pre-typed message 1 (or “kiss”) as a token of interest and the target user sends back a positive kiss reply.

Let \( U \) be the set of users in the network. Let \( X \) be a user personal profile that includes a list of personal profile attributes, \( X = \{x_1, \ldots, x_n\} \) where each attribute \( x_i \) is an item such as body type, dietary preferences, political persuasion and so on. Consider the list of user’s ideal partner profile attributes as a set \( Y = \{y_1, \ldots, y_n\} \) where each attribute \( y_i \) is an item such as body type, dietary preferences, political persuasion and so on. For a user \( u \), value of \( x_i \) is unary, however, the values of \( y_i \) can be multiple. Let \( P = X + Y \) denote a user profile containing both the personal profile attributes and partner preference attributes. The profile vector of a user is shown as \( P(u) \).

There can be many types of user activities in a network that can be used in the matching process. Some of the main activities are “viewing profiles”, “initiating and/or responding kisses”, “sending and/or receiving emails” and “buying stamps”. The profile viewing is a one-sided interaction from the viewer perspective; therefore it is hard to define the viewers’ interests. The “Kiss interactions” are more promising to be considered as an effective way to show the distinct interests between two potential matches. A user is able to show his/her interest by sending a “kiss”. The receiver is able to ignore the “kiss” received, give a predefined negative reply or return a predefined positive reply. When a receiver replies a kiss with positive predefined message it is considered as a “successful” kiss or a “positive” kiss. Otherwise, it is judged as an “unsuccessful” kiss or a “negative” kiss.

B. Generation of Social Networks

An online dating network or part of it can be represented as graph that shows people and their relationships. It includes set of actors, actor roles, ties (also called as relationships that link the actors) and information flows. Previous research on computer-mediated communication has shown that the relationship built through computer networks are fundamentally equal to social networks set up in face-to-face process [8][9][10]. In social networks, people build up the trust on others through the information how these people have behaved in the past. Therefore, it will form higher percentage of successful communication among people when in online social spaces both actors and their interaction history are recorded and utilized in recommendation [8]. Based on this philosophy, a social network which describes the past relations between users and their previous contact users, is derived as shown in Figure 1 and used in match making.

Let user \( u_i \) be the user who has successfully interacted with more than a certain number of previous partners for a particular period. Let \( GrA \) be the set of users, \( GrA \subseteq U \), who user \( u_i \) has positively interacted. Let user \( u_j \) be the user who

---

1 We call a pre-defined message as “kiss” in this paper.
user \( u_x \) has positively interacted last, \( u_y \in GrA \). Let \( GrB \) be the set of users who are ex-partners of user \( u_x, u_y \in GrB \). Note: \( gender(u_x) = gender(GrB) \) and \( gender(u_y) = gender(GrA) \). Users \( u_x \) and \( u_y \) are called seed users as they provide us the network. Users in \( GrA \) and \( GrB \) are called as relationship-based users. The relationship between the user \( u_x \) and a user in \( GrB \) reflect the personal profile similarity between the two users. Similarly, the relationship between users \( u_x \) and a user in \( GrA \) reflect the personal profile similarity between the two users. This similarity value is evaluated by using an instance-based learning algorithm.

**Personal Profile Similarity:** An instance based learning algorithm is developed to calculate similarity between the seed user and the relationship-based users. Attributes in personal profile are of categorical domain. So an overlap function is used that determines how close the two users are in terms of attribute \( x_i \).

\[
S_{x_i}(u_1, u_2) = \begin{cases} 1, & x_i(u_1) = x_i(u_2) \\ 0, & \text{Otherwise} \end{cases}
\]

where \( u_1 \) is a seed user and \( u_2 \in GrB \) or \( u_2 \in GrA \). This matching process is conducted between a seed user \( u_x \) and \( GrB \) users, as well as the corresponding partner seed user \( u_y \) and \( GrA \) users. Their personal profiles can be compared as these set of users belong to same gender.

All attributes are not equally important when selecting a potential match [8][10]. For example, analysis of dataset of a popular dating network\(^2\) shows that attributes such as height, body type, and have children are specified more frequently than attributes such as nationality, industry and have pets in user personal profiles. Therefore, for each attribute score, one particular weight is assigned when combined them all together. The weight is set according to the percentage that all members have indicated that attribute in their personal profiles existing in the network. Inclusion of the weight values according to the network statistic allows us to reflect the common user interest in the network.

\[
SimScore(u_1, u_2) = \sum_{i=1}^{n} S_{x_i}(u_1, u_2) \times \text{weight}_{x_i}
\]

**Solving cold-start problem in the network:** A recommendation system can suffer from cold-start problem [4] when the number of relationship-based users is very low in a network or new users are to be included in the matching process. This research utilises the \( k \)-means clustering algorithm that helps to increase the size of \( GrA \) and \( GrB \) by finding similar users according to the seed users \( u_x \) and \( u_y \), respectively. Users in the network for a specified duration are grouped according to their personal profiles. Let \( C^m = \{C_1, ..., C_m \} \) be the cluster of male members of the network where \( c_k \) is the centroid vector of cluster. Let \( C^f = \{C_1, ..., C_f \} \) be the cluster of female members of the network.

---

\(^2\) Due to privacy reason the detail of this network is not given.
where $c_k$ is the centroid vector of cluster. The user personal profile and preference attributes $P = X + Y$ where $X = \{x_1, ..., x_n\}$ and $Y = \{y_1, ..., y_n\}$ are used in the clustering process.

Appropriate clusters corresponding to gender of the seed users, $u_s$ and $u_e$ are used to test which cluster matches best with the seed user,

$$(\max_{c_k \in C} \langle S^k (P(u_s),c_k) \rangle)$$

where $S^k$ shows the maximum similarity between a centroid vector and a user profile vector. Cosine similarity is employed in the process. Members of the matched cluster are used in extending the size of $GrA$ or $GrB$.

**Putting it all together:** Once the similarity scores, identifying profile similarity between the seed user and a potential match, $SimScore(u_s,GrB_j)$ and identifying profile similarity between the seed partner and a recommendation object, $SimScore(u_y,GrA_i)$ are obtained, these scores are combined using a weighted linear strategy.

$$Match(GrA_i,GrB_j)$$

$$= w_1 \times SimScore(u_s,GrB_j) + w_2 \times SimScore(u_y,GrA_i)$$

These weights have been determined with empirical analysis. Experiments show better results when $w_1$ is given higher weight than $w_2$. Therefore $w_1$ and $w_2$ are set as 0.6 and 0.4 respectively. For each recommendation object in $GrA$, matching partners are ranked according to their $Match(GrA_i,GrB_j)$ score, and top-$n$ partners from $GrB$ become the potential match of $GrA_i$.

### III. EXPERIMENTS AND DISCUSSIONS

This section presents the empirical analysis of the proposed method.

**A. Dataset: The online dating network**

The proposed method is tested with the dataset collected from a real life online dating network. There were about 2 million users in the network. We used the three months of data to generate and test networks of relationship-based users and recommendations.

The activity and measure of relationship between two users in this research is “kiss”. This relationship measure is an effective way to show the distinct interests between two potential matches. Analysis of the dataset shows that positive kiss replies have a strong correlation with stamp purchase behaviour on the network. Thus, positive kiss reply indicates not only the member’s interest but also a good sign for the network revenue. Therefore the number of positive kisses is used in testing the proposed social matching system.

Table I lists the details of the users and kisses in the network. A user who has logged on in the website during the chosen three months period is called as “active” user.

The seed users and relationship based users come from this set of users. The sample dataset shows that the gender distribution among network members is quite even.

A kiss sender is called “successful” when the target user sends back a positive kiss reply. There are about 50 predefined messages (short-text up to 150 characters) on the social network that a user can choose to show his/her interest to another member. These kiss messages are manually defined as positive or negative showing the user interest towards another member. A kiss communication between two members is defined as “successful kiss” when a target user sends back a positive kiss message reply after receiving an initial message from the sender. Likewise the “negative kiss” communication happens when a target user sends back a negative kiss message reply after receiving an initial message from the sender. There are a large number of kisses exist in the network that have never been replied by the target users and this type of kiss communication is called as “null kiss”.

Table I. User and Kiss Statistics for the three months chosen period

<table>
<thead>
<tr>
<th>3 Months Data</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td># of distinct active users</td>
<td>163,050</td>
</tr>
<tr>
<td># of female users</td>
<td>82,500</td>
</tr>
<tr>
<td># of male users</td>
<td>80,550</td>
</tr>
<tr>
<td># unique kiss senders</td>
<td>122,396</td>
</tr>
<tr>
<td># unique successful senders</td>
<td>91,487</td>
</tr>
<tr>
<td># unique kiss recipients in the network</td>
<td>198,293</td>
</tr>
<tr>
<td># unique kiss recipients who are active during the chosen period</td>
<td>83,865</td>
</tr>
<tr>
<td># unique kisses</td>
<td>886,396</td>
</tr>
<tr>
<td># unique successful kisses</td>
<td>171,158</td>
</tr>
<tr>
<td># unique negative kisses</td>
<td>346,193</td>
</tr>
<tr>
<td># unique null kisses</td>
<td>369,045</td>
</tr>
</tbody>
</table>

The number of unique recipients is larger than the active members in the network for the chosen 3-months period, as members regardless of them being inactive can receive kisses. It can be noted that for each kiss sender, there is about 4 kiss replies (including successful and negative both) on an average. It can also be seen that about 75% kiss senders have received at least one positive kiss reply. The amount of successful kisses is less than one fourth of the sum of negative and null kisses. A further kiss analysis shows a strong indication of Male members in the network for initiating the first activities such as sending kisses (78.9% vs 21.1%). They are defined as proactive behavior users in the paper. While female members who are reactive behaviour users usually wait for receiving kisses.

**B. Evaluation Criteria**

Rather than evaluating the performance of exact rating of matches in the recommendation list, we are interested in predicting whether a potential match would be amongst the user’s favorites. It will be tested by whether the user has made initial contact to the users in the recommendation list.
Our evaluating a match as being liked it in the top-n reflects our belief that knowing the actual rating of a potential match is not as important as knowing where the rating was relative to other ratings for a given user.

Let \( U \) be the set of network’s active users. Let \( GrA \) be the group of users who are going to receive potential partners’ recommendations and \( GrB \) be the group of users who become the potential partners. Let \( U = GrA \cup GrB \) where \( GrA \cap GrB \neq \emptyset \). The following metrics are used in evaluation:

\[
\text{Success Rate (SR)} = \frac{\text{Number of unique successful kisses \( GrA \) to \( GrB \)}}{\text{Number of unique kisses \( GrA \) to \( GrB \)}}
\]

\[
\text{Baseline Success Rate (BSR)} = \frac{\text{Number of unique successful kisses \( GrA \) to \( U \)}}{\text{Number of unique kisses \( GrA \) to \( U \)}}
\]

\[
\text{Success Rate Improvement (SRI)} = \frac{\text{Success Rate (SR)}}{\text{Baseline Success Rate (BSR)}}
\]

\[
\text{Recall} = \frac{\text{Number of (Kissed Partners} \cap \text{Recommended Partners)}}{\text{Number of (Kissed Partners)}}
\]

Success Rate (SR) is the proportion of kisses from \( GrA \) to \( GrB \) that are successful. Baseline Success Rate (BSR) indicates the current success rate of the underlying dating network without utilising the recommendation of potential partner matches. Success Rate Improvement (SRI) is the improvement gained using the recommendations. Ideally it should be greater than 1. Recall would estimate how many of all the potential matches in the user’s top-n were predicted correctly.

C. Results and Discussion

Choosing a good seed user: The first step in the proposed method is generation of networks of relationship based users. It is important to have a good seed in order to generate a network of relationship based users. Data analysis of the network shows that with randomly chosen 300 seed users, 81% of them have at least 50 unique previous partners as shown in Table II. Only 11% of them have more than 50 unique previous partners. Data analysis shows that many users in the network receive 30 to 50 positive (unique users’) kiss messages back. A social science research also has revealed that an average woman dates 24 men before finding her "Mr Right" [10]. Therefore, in this research if a seed user has sent at least 30 kisses and has received more than 30 positive kiss replies, this seed user is considered as a good seed user and will be chosen for continuous process.

Table II. Kiss Information for the selected 300 seed users

<table>
<thead>
<tr>
<th># of unique positive kisses (PK)</th>
<th># of unique seed users</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;50</td>
<td>189</td>
</tr>
<tr>
<td>50&lt;PK&gt;100</td>
<td>26</td>
</tr>
<tr>
<td>100&lt;PK&gt;300</td>
<td>13</td>
</tr>
<tr>
<td>&gt;300</td>
<td>3</td>
</tr>
</tbody>
</table>

In an online dating network, male users are normally proactive users while Female users are reactive users [10]. Thus, as a kiss initiating user who triggers the relationship-based user network, \( u_x \) is usually Male. A Male seed user \( u_x \) always generates a bigger size group which is far more than its seed partner, a Female seed user generates. To increase the relationship-based users in \( GrB \) particularly if gender(\( u_x \)) = Male, three seed partners \( u_x \) are used to generate ex-partners. Therefore the size of \( GrB \) increases three times than before.

Matching prediction performance: Figure 2 shows that the Success Rate (SR) decreases as the number of potential matches \( (GrB) \) is increased for a user in \( GrA \). This result confirms that higher the total score generated by the proposed matching system, \( \text{Match}(GrA, GrB) \), the more relevant and accurate matches are made. For example, users with higher total score in top-5 recommendation list received highest percentage of positive kiss reply.

There are a number of null kiss replies in the dataset. A null kiss reply can be transformed to positive kiss reply and negative kiss reply. If all the null kiss reply is able to transform to positive kiss reply, the success rate (SR) can be obtained as 66% for top-20 users.

The BSR of the underlying online dating network is 19%. This indicates that there is a 0.19 possibility that a kiss message would be responded positively by the receiver. SRI (Success Rate Improvement) stands for the comparison between recommending potential matches using the proposed system and using without the proposed system. It can be seen from Figure 2 that the proposed system is always better. This result describes that the potential matches offered by the system interest the user (as shown by figure 3) and also the receivers show high interests towards matches offered by the system interest the user (as shown by \( SR \) in figure 2). However, it can be seen that with the increased number of recommendations the value of SRI decreases as shown in figure 2.

It concludes that more matching recommendations will attract user attention and trigger more kisses to be sent. However, more recommendation will also lead to low quality recommendations. Users will miss their distinct targets when they counter on loads of matching recommendations.
Adoption of recommendation techniques in social networks for recommending people to people is currently gaining importance in the data mining community. This paper presents a personalized social matching system for matching network members by combining their past relationships in social networks and user similarities. Empirical analysis shows that the proposed system produces more successful matches than using the current database retrieval techniques. Results show that the success rate has been improved from the baseline results of 19% to 31% by using the proposed system. Due to the use of small networks of relationship-based users, the proposed system is able to generate recommendations in acceptable time frame – it takes about 2 hours to generate recommendations for 100,000 users excluding the offline activities such as clustering of users, creating views etc.

This is a preliminary study and a lot of work needs to be done to produce an improved recommendation system. One future work is to introduce another similarity measure that calculates the compatibility between the recommendation pairs. More advanced similarity measures will be suggested to measure similarity between users. Additionally, social network analysis techniques will be explored to include the implicit relationships between users.

IV. CONCLUSION

Adoption of recommendation techniques in social networks for recommending people to people is currently gaining importance in the data mining community. This paper presents a personalized social matching system for matching network members by combining their past relationships in social networks and user similarities. Empirical analysis shows that the proposed system produces more successful matches than using the current database retrieval techniques. Results show that the success rate has been improved from the baseline results of 19% to 31% by using the proposed system. Due to the use of small networks of relationship-based users, the proposed system is able to generate recommendations in acceptable time frame – it takes about 2 hours to generate recommendations for 100,000 users excluding the offline activities such as clustering of users, creating views etc.

This is a preliminary study and a lot of work needs to be done to produce an improved recommendation system. One future work is to introduce another similarity measure that calculates the compatibility between the recommendation pairs. More advanced similarity measures will be suggested to measure similarity between users. Additionally, social network analysis techniques will be explored to include the implicit relationships between users.

V REFERENCES