Journal of Information & Computational Science 8: 16 (2011) 4061–4068
Available at http://www.joics.com

Reciprocal Recommendation Algorithm for the Field of Recruitment ⋆

Hongtao YU*, Chaoran LIU, Fuzhi ZHANG

School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China

Abstract

In order to solve the problem that the traditional recommendation algorithm can not carry out the two-way recommendation between users and items, by introducing the concept of reciprocal recommendation to extend the traditional recommendation algorithm, a reciprocal recommendation algorithm based on bi-directional matching is proposed. First of all, depending on the specific basis of the resume delivery on recruitment system and mining the potential preference information of users, an implicit preference function calculation method is proposed. Then, based on the consideration that different effects for the users explicit and implicit preference information in computing similarity, a similarity calculation method based on the integration of these two preference information is proposed. At last, the reciprocal recommendation is achieved according to the reciprocal value from high to low. The experimental results show that the proposed algorithm can not only solve the problem of two-way recommendation between job seekers and jobs, but also improve the accuracy of recommendation.

Keywords: Reciprocal Recommendation; Field of Recruitment; Explicit Preference; Implicit Preference

1 Introduction

Reciprocal recommendation is a kind of important method of the personalized recommendation, Compared with the traditional recommendation algorithm, both parties in the recommendation are all the objects with the right to choose freely.

In recent years, some scholars have paid attention to the research of reciprocal recommender systems. Luiz Pizzato et al.[1] introduced the concept of reciprocal recommendation for the first time, and put forward a reciprocal recommendation algorithm based on hybrid strategy. This algorithm divided recommendation process into two stages and adopted collaborative filtering and content-based algorithm for the recommendation of users. But, the algorithm has the problem of lacking information and does not solve the cold-start problem. After that, In [2][3], they proposed a content-based reciprocal recommendation algorithm RECON for the typical reciprocal recommendation instanceonline dating [9][10], this algorithm provides an effective recommendation for

⋆Project supported by the Natural Science Foundation of Hebei Province, China (F2011203219) and the Special Fund for Fast Sharing of Science Paper in Net Era by CSTD of Ministry of Education, China (20101333110013).

*Corresponding author.
Email address: yu5771@163.com (Hongtao YU).

December 2011

lots of users based on the communion information between them. But, this algorithm only applies to the dating field that the information of recommended targets is symmetric, so it has certain limitations. Malinowski et al.[4] proposed a bidirectional recommendation algorithm based on collaborative filtering, they regard the matching between users as a two-way recommendation process and search candidates for both sides by using different recommendation technologies. Although the algorithm has realized bidirectional recommendation, it only adopts the binary representation with Yes and No when expresses users preferences, and it can not measure the degree of users preferences for each index well, so the quality of recommendation is not high.

Richi et al.[5] proposed a social matching system based on the online dating site, it combines the content-based social network knowledge with the technology of collaborative filtering, groups similar users by using nearest neighbor algorithm, and realizes the similarity computation and recommendation by using relationship-based similarity prediction algorithm so as to satisfy the higher requirements and expectations of users. But, it is only in the primary study stage, the method of similarity computation for reciprocal value, which is not perfect, has affected the quality of recommendation. S.Bull et al.[6] proposed I-help system based on user model, this system is built on an agent architecture. Based on the analysis of requirements for help, it searches and arranges candidate helper from different personal agents. As a learning tool, it has been used widely. J.Greer et al.[7] also introduce a prototype system PHelpS. Based on the idea of reciprocal recommendation, it seeks companion that provides help for the person who encounters difficulties in the work. Both systems are all designed to search for potential candidates for users, but they depend on a lot of user models and the amount of calculation is very large.

Aim at the deficiencies of existing reciprocal recommendation algorithm in accuracy and combine with the characteristics of recruitment, in this paper we propose a preference function calculation method based on users interaction history; a new user similarity measurement method is presented by considering users explicit preferences and implicit preferences comprehensively; On the basis of these, a novel reciprocal recommendation algorithm to solve the problem of bidirection matching is proposed. Compared with the existing reciprocal recommendation algorithm, the proposed algorithm has improved the recommendation accuracy and it can be applied to the asymmetric reciprocal recommendation domain.

2 Background

2.1 Reciprocal recommendation

For a reciprocal recommendation problem between job-seeker and job, the recommendation process can be divided into two parts: Job recommendation and job-seeker recommendation. The design idea of these two parts is the same roughly.

For a job-seeker, the job with higher matching degree should be recommended to him. When given a user u, v is the job candidates set of u, then the degree of preference $P_1(u, v)$ between job-seeker u and every job v in V is larger than the degree of preference $P_1(u, v_0)$ between the same job-seeker u and every job v_0 not in V, this is shown as:

$$R_1(u) = \{v : P_1(u, v) > P_1(u, v_0), \forall v \in V, \forall v_0 \notin V\} \tag{1}$$

Similarly, for a job, the job-seeker with higher matching degree should be recommended to it. When given a job v, U is the job-seeker candidates set of v , then the degree of preference $P_2(v, u)$

between v and every job-seeker u in U is larger than the degree of preference $P_2(v, u_0)$ between the same job v and every job-seeker u_0 not in U, this is shown as:

$$R_2(v) = \{u : P_2(v, u) > P_2(v, u_0), \forall u \in U, \forall u_0 \notin U\} \tag{2}$$

For the reciprocal recommendation between job-seeker u and job v, the recommendations for both sides should be the objects which are the most consistent with their preferences. Therefore, the reciprocal recommendation for job-seeker u is shown as:

$$RR(u) = \{v : v \in R_1(u) \ and \ u \in R_2(v)\} \tag{3}$$

The reciprocal recommendation needs to consider the preference of job-seeker u and job v. To get a single recommendation set, we need to combine R_1 and R_2. This is shown as:

$$RR(u, v) = \omega_1 P_1(u, v) + \omega_2 P_2(v, u) \tag{4}$$

Where ω_1 and ω_2 are weight parameters, which should be chosen according to the specific situation.

2.2 Introduction of vector space model

In VSM(Vector Space Model), the text is composed of a group of terms, each term is given a different weight according to its importance. We can map the text into a feature vector $V(d)$=(t_1,$\omega_1(d)$;$\cdots$;t_n,$\omega_n(d)$) by using VSM. Where t_i(i=1,2$\cdots$,n) is a term which is different from each other, $\omega_i(d)$ is the weight of t_i in d and be defined the frequency of occurrence in d-$tf_i(d)$ usually. Generally, we use TF-IDF algorithm to calculate the weight of term, TF-IDF [8] is shown as:

$$\omega_i(d) = \frac{tf_i(d)log(\frac{N}{n} + 0.1)}{\sqrt{(\sum_{i=1}^{n}(tf_i(d))^2 \times log^2(\frac{N}{n_i} + 0.1)}} \tag{5}$$

Where N is the number of all text,n_i is the number of text which contains term t_i. The higher the frequency of a term in the document is, the bigger its weight is. The relevance between text d_1 and text d_2 is expressed by the cosine value of the angle between two vectors, this is shown as:

$$Sim(d_1, d_2) = cos\theta = \frac{\sum_{i=1}^{n}\omega_i(d_1) * \omega_i(d_2)}{\sqrt{(\sum_{i=1}^{n}\omega_i^2(d_1))(\sum_{i=1}^{n}\omega_i^2(d_2))}} \tag{6}$$

3 User Preferences and Similarity Calculation

3.1 Explicit preference and implicit preference

Definition 1 *Given a job-seeker u, the feature vector P_0 and the preference vector P_0' are established respectively according to his/her personal information and preference to the job v, the vector $P_E(u)$={$P(u)$|(P_0,P_0')} is called the explicit preference vector of job-seeker u.*

Definition 2 *Given a job v , the feature vector P_1 and the preference vectorP_1' are established respectively according to its attribute information and preference to the job-seeker u, the vector $P_E(v)$={$P(v)$|(P_1,P_1')}is called the explicit preference vector of job v.*

The collection of implicit information mainly depends on the exchanges between users. When a job-seeker u is interested in a job v, she/he will send to a resume which includes her/his attributes and interests. The employer will decide to hire her/him or not after reading the resume. If the employer gives some positive replies (e.g. take an interview), the matching will be successful. If the replies are negative or nothing, the matching is failed. The information that got by the exchanges indirectly is called users implicit preference.

Definition 3 *Given a attribute set A of job v, a is a attribute element in A, v_a is one of the value of a, if $P_I(u)=\{P(u)|(n,v_a);a\in A\}$, then $P_I(u)$ is called the implicit preference of job-seeker u. Where n is the times that v_a appears in the resume which is sent by u.*

Definition 4 *Given a attribute set B of job-seeker u, b is a attribute element in B, v_b is one of the value of b, if $P_I(v)=\{P(v)|(m,v_b);b\in B\}$, then $P_I(v)$ is called the implicit preference of job v. Where m is the number of resume which is positive among the resumes with attribute value v_b.*

We record each attributes usage of all users and get degree of preference by computing the proportion of each one in all attributes, and intercept data record of different time to track the change of $user's$ interest.

3.2 Preference function integration-based similarity calculation

In view of the shortage of traditional similarity calculation method, we measure the similarity between users by integrating the user's explicit and implicit preference. The similarity calculation of explicit information expressed by text has been introduced in section 2.2. The calculation formula of similarity between job-seeker u and job v is shown as:

$$Sim1(u,v) = cos\theta = \frac{\sum_{i=1}^{n}\omega_i(u)\omega_i(v)}{\sqrt{(\sum_{i=1}^{n}\omega_i^2(u))(\sum_{i=1}^{n}\omega_i^2(v))}} \tag{7}$$

Where $\omega_i(u)$ is the weight of its preference attribute of job-seeker u, $\omega_i(v)$ is the weight of its feature attribute of job v, n is the number of all the attribute.

For the implicit preference, the similarity between job-seekers preference p_u and job is calculated by the exchange of resumes, the formula is as follows:

$$Sim2(p_u,v) = \frac{1}{n}\cdot\frac{\sum_{i=1}^{n}V_{a,i}}{N} \quad (V_{a,i}\in P_u) \tag{8}$$

Where $V_{a,i}$ is the value of its attribute of job v, n is number of all the attribute, N is the number of resumes that sent by every job-seeker.

As $user's$ preference is composed of explicit information and implicit information, the comprehensive similarity is calculated by Equation (9):

$$Sim(u,v) = \alpha Sim1(u,v) + \beta Sim2(p_u,v) \tag{9}$$

Where α and β are weight parameters, and $\alpha+\beta=1$. In our experiment, we set α and β to 0.4 and 0.6 respectively.

$Sim(u,v)$ satisfies the preference of job-seeker, but the reciprocal recommendation should satisfy both sides. According to Equation (9), we can get the similarity between preference of job v and job-seeker u. The reciprocal value can be computed by Equation (10):

$$Recip-sim = \frac{2sim(u,v)\cdot sim(v,u)}{sim(u,v)+sim(v,u)} \tag{10}$$

4 Reciprocal Recommendation Algorithm

The idea of reciprocal recommendation algorithm is as follows. First of all, the useful information is extracted from users resumes. Then we locate the explicit preferences of users and get the implicit preferences indirectly according to the condition of sending and receiving resumes, similarity of different preferences is calculated using different methods. Finally, the comprehensive similarity is calculated and recommendation is generated. The main steps are as follows.

(1)Extract users preference: according to the content of users resume, the attribute information is extracted and converted to vector form to get users explicit preference P_E; then we track the case of sending and receiving of users resumes and record the number of resumes which include different attributes, and locate the implicit preference P_I according to the proportion of each attribute.

(2)Similarity calculation: calculate the similarity between users using Equation (7)(8) and (9) in turnthen calculate the reciprocal score by Equation (10).

(3)Generate recommendation: the reciprocal scores are ranked, then according to the size of every reciprocal score to recommend them one by one following the top-n recommendations.

Based on the above steps, the similarity calculation algorithm and reciprocal recommendation algorithm are described as follows.

Algorithm1 Similarity calculation

Input: Job-seeker u, Job v, Parameter α,β
Output: Sim(u,v)
01:Extract u's explicit preference Pu
02:Extract job v's attribute information
03: Get $\vec{u}$, $\vec{v}$
04: $Sim_1(u,v) \leftarrow \cos(\vec{u},\vec{v})$
05:**For each attribute** $a \in A$ **do**
06: calculate the value n of attribute a in V
07: **If** $n = 0$ **then**
08: $Sim_2(P_u, v) \leftarrow 0$
09: **else**
10: $Pa \leftarrow$ n/N
11: Get $Sim2(P_u, v)$
12: **Endif**
13:**Endfor**
14:$Sim(u,v) \leftarrow \alpha Sim_1(u,v) + \beta Sim_2(P_u, v)$
15:**return** $Sim(u,v)$

The algorithm is composed of three parts. The first part, from line 1 to 4, is to extract the explicit vector information of both sides and calculate similarity of explicit preference based on vector space model. The second part, from line 5 to 13, is to calculate similarity of implicit preference. The last part, from line 14 to 15, is to calculate the comprehensive similarity by combining the explicit similarity and implicit similarity according to Equation (9).

Algorithm2 Reciprocal recommendation

Input: Job-seeker u, N number of jobs to provide
Output: List of recommendations R
01:Find u's preference Pu
02:**For each** v **do**
03: $Su \leftarrow Sim(u,v)$
04: **If** $S_u > 0$ **then**
05: Find v's preference Pv
06: $Sv \leftarrow Sim(v,u)$
07: $Sv \leftarrow Recipe - sim(Su, Sv)$
08: **Endif**
09:**Endfor**
10:**For i=1 to N do**
11: **If** $S_{vi} \geqslant S_{vi+1}$ **then**
12: *Sort* $R\{v_1, v_2, \ldots v_n\}$
13: **Endif**
14:**Endfor**
15:**return** R

The algorithm realizes the reciprocal recommendation between users. Line 1-9 is to calculate the reciprocal score according to the similarity calculated by algorithm 1; line 10-15 is to create a list of reciprocal recommendations and recommend them to the user.

5 Experiments and Evaluation

5.1 Dataset

To evaluate the effectiveness of reciprocal recommendation algorithm, we conduct experiment on a group of users. The users of this experiment are students who are from different professionals and different educations in a university, the background of students covers several professionals which are very popular in the current colleges and universities (such as: Computer, Machinery, Finance et al). In addition, 80 real data have been collected from the recruitment website (http://www.zhaopin.com) which is well-known at home, the job data contains different domains to ensure the diversity. Then according to track the sending and receiving of resumes, 1000 data have been chosen to be the set of experiment data.

5.2 Evaluation metrics

We use precision and recall metrics to evaluate the performance of our algorithm. The precision and recall can be calculated by Equation (11) and (12), respectively.

$$Precision(u) = \frac{\{x|x \in Set_1, x \ replied \ positively \ by \ u\}}{\{y|y \in Set_1 \bigcap y \in Set_2\}} \tag{11}$$

$$Recall(u) = \frac{\{x|x \in Set_1, x \ replied \ positively \ by \ u\}}{\{z|z \in Set_2, z \ replied \ positively \ by \ u\}} \tag{12}$$

Where Set_1 is the list of recommendations, Set_2 is the set of resumes that user u has sent.

5.3 Experimental results and analysis

(1) Success rate

The comparison of three methods is shown in Table1. The results indicate that the comprehensive method is better then the rest. This shows that our algorithm has a high success rate. It also shows that $user's$ interest is depended on the predefined preference and interaction history, implicit preference can reflect the real interest of user more accurately than explicit preference, but the comprehensive consideration can make the recommendation more reliable.

Table 1: Comparison of success rate for three methods

	Explicit	Implicit	Comprehensive
Number of total recommendations	653	723	746
Number of successful recommendations	366	453	504
Success rate	56.1%	62.7%	67.6%

(2) The influence of the number of resumes on the performance of algorithm

Fig.1 shows the relationship between the number of generated recommendations and the number of resumes that are sent. For explicit method, with the increase of resumes, the number of generated recommendations for each user keeps the same, because this method does not depend on the number of resumes. But for implicit method, the number keeps rises. The reason is that we can not locate users preference when the number of resumes is very small in training set, it is hard to generate recommendations with accuracy. The curve of comprehensive method keeps stable, which shows that explicit method and implicit method can compensate each other weaknesses, and they tend to be stable and convergence when there are a large number of resumes.

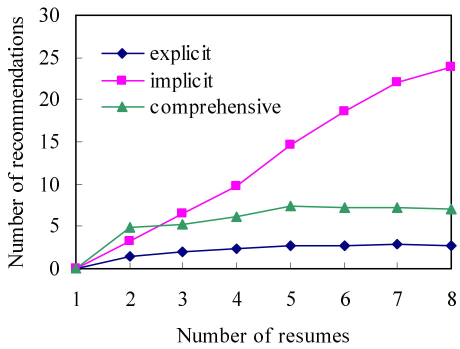

Fig. 1: The relationship of recommendation and resume

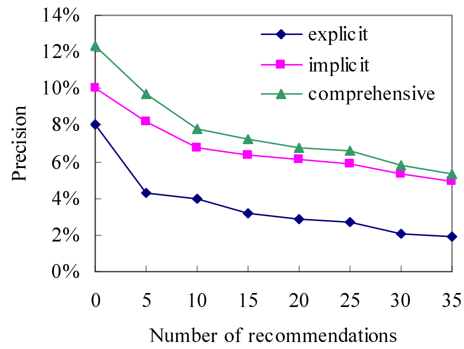

Fig. 2: The accuracy comparison of three methods

(3) Evaluating the algorithm by precision and recall

In Fig.2, With the number of recommendation increasing, the precision decreases. But, the effect of comprehensive method is the best, and the effect of implicit method is close to that of comprehensive method. The reason is that implicit preference can get much more abundant information than explicit preference and its decisive role is much bigger. Although the explicit preference is close to it when the number of recommendation is small, its overall effect is not ideal.

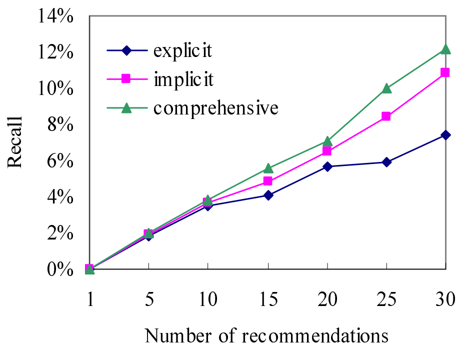

Fig. 3: Comparison of recall for three methods

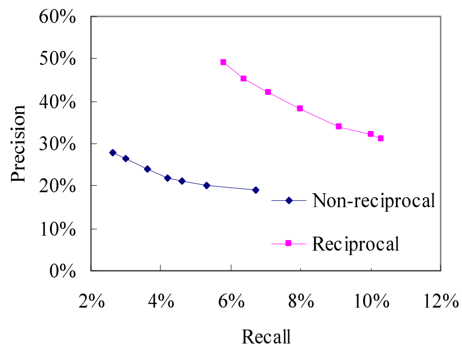

Fig. 4: Comparison of reciprocal and non-reciprocal algorithm

In Fig.3, Three is no obvious difference for three methods when the recommendation is small. With the number increasing, the recall rises. The main reason is that recall and precision restrict each other on the large-scale dataset. As the increase of recommendations, the precision will decline gradually and the recall will rise. At the same time, the advantage of comprehensive method will highlight, and the implicit method is worse than it and better than explicit method.

(4) Evaluating the algorithm by the index of reciprocal

The main goal of the reciprocal algorithm is to achieve mutual benefit of both sides. Fig.4 shows the comparison of the reciprocal algorithm and the traditional non-reciprocal recommendation. Every point of the line is top-5,10,15,20,30,40,50 from left to right. For the top-10, the precision of reciprocal recommendation is 45%, and the precision of non-reciprocal is 27%. The improvement of 18% dues to the reciprocity; Equality for the top-50, the recall of reciprocal is 10.3%, and the recall of non-reciprocal is 6.7%, the improvement of 3.6% also dues to the result of reciprocal.

The main reason of these figures is that job-seeker and employer are the main parts in recruitment system, and based on the premise of satisfying both parties, the two-way choice is a recommendation mechanism that both sides reach an agreement. Our recommendation algorithm change the object function that be optimized by original recommender system, it requires that the job-seeker and the job are in the front of the recommendation list of the opposite side at the same time, so the final recommendation can fit the individuation of recruitment website. From this we can see that compared with the non-reciprocal recommendation which only considers the preference of one side, reciprocal recommendation has a great improvement on precision and recall and improves the efficiency of recommender system.

6 Conclusion and Further Work

Reciprocal recommendation is a new research direction in the area of recommender systems. We have done some beneficial explorations in this area. In this paper we presented a reciprocal recommendation algorithm for the field of recruitment and gave the calculation method of users preference and the measurement method of similarity. The proposed algorithm combines the explicit preference and implicit preference of user and can locate the users characteristic and interest much more accurately. Based on the calculation of users comprehensive similarity, the success rate of recommendation has improved. In this paper, we only consider the reciprocal recommendation in the field of recruitment. How to improve its compatibility and apply it in a broader domain will be the next work.

References

[1] LuizPizzato, Reciprocal Recommenders, In 8th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems, ACM, USA, 2010: 53 – 64.

[2] LuizPizzato, RECON: A Reciprocal Recommender for Online Dating, ACM, Barcelona, 2010: 207 – 214.

[3] LuizPizzato, Reciprocal Recommender System for Online Dating, Proceeding of the 4th ACM conference on Recommender Systems, New York, 2010: 353 – 354.

[4] J.Malinowski, Matching people and jobs: A bilateral recommendation approach, In Proceedings of the 39th Annual Hawaii International Conference on System Sciences, 2006: 1 – 9.

[5] Richi Nayak, A Social Matching System for an Online Dating Network: A Preliminary Study, IEEE International Conference on Data Mining Workshops, 2011: 352 – 357.

[6] S. Bull. User modeling in i-help: What, why, when and how, In Modeling: Proceedings of the Eighth International Conference, Germany, 2001: 117 – 126.

[7] J.Greer, Supporting Peer Help and Collaboration in Distributed Workplace Environments, International Journal of Artifical Intelligence in Education, 1998, 9: 159 – 177.

[8] Zhongxiu He, The calculation method research of webpage content based on vector space model, Computer and modernization, 2010, 9 (14): 53 – 55.

[9] L.Brozovsky, Recommender system for online dating service, ACM, Ostrava, 2007: 1 – 12.

[10] L. Pizzato, Learning user preferences in online dating, Technical Report 656. Australia, 2010.