A Job Recommendation Method Optimized by Position Descriptions and Resume Information

Peng Yi[1] ,Cheng Yang[1] ,Chen Li[1] ,Yingya Zhang[1]
1. Science and Technology Department, Communication University of China
Beijing, China
Yipeng_cuc@163.com

***Abstract*—With the development of Internet technology, online job-hunting plays an increasingly important role in job-searching. It is difficult for job hunters to solely rely on key-words retrieving to find positions which meet their needs.To solve this issue, we adopted item-based collaborative filtering algorithm for job recommendations. In this paper, we optimized the algorithm by combining position descriptions and resume information. Specifically,job preference prediction formula is optimized by historical delivery weight calculated by position descriptions and similar user weight calculated by resume information. The experiments tested on real data set have shown that our methods have a significant improvement on job recommendation results.**

***Keywords*—Job Recommendation ; Item Based Collaborative Filtering ; Similarity Measure ; Historical Delivery Weight;Similar User Weight**

I. Introduction

With the wide spread of the Internet, online recruitment get much more concern. Job hunters often have to spend a lot of time and efforts in searching and browsing in order to find the suitable vacancy from the large numbers of recruitments information in major recruitment website. Job recommendation algorithms which utilize recommendation method to filter positions that do not meet the requirements and recommend the proper positions for job hunters are play an important role in the recruitment websites[1~4].

Based on the analysis of real recruitment data and the comparison of the existing recommendation methods, item-based collaborative filtering algorithm has been used as the basic algorithm for job recommendation[5].In this paper, we proposed an optimization algorithm to improve the accuracy of job recommendations. Historical delivery weight calculated by position descriptions and similar user weight calculated by resume information were added as two influencing factors in the preference prediction. The experiments tested on real recruitment data have shown that the optimization algorithm have greatly improved the final recommendation result..

II. Basic Job Recommendation Algorithm

A. Recommendation Algorithm

Recommendation algorithm is an prediction method which utilizes some mathematical theory to analyze user behavior and predict the user's preference of items. The basic recommendation algorithm is as follows:

(1)Content Based Recommendation

Content based recommendation generates recommendation lists based on item characteristics description and user historical interests[6]. More simply, content based recommendation algorithm recommend items which are similar to user's historical interest.

(2)Collaborative Filtering Recommendation

Collaborative filtering recommendation generates recommendation by utilizing the user-item rating matrix. The basic idea of the collaborating filtering algorithm is that users do selections usually referred to the like-minded friends[7]. At present, collaborative filtering algorithm can be divided into user-based collaborative filtering algorithm and item-based collaborative filtering algorithm.

(3)Knowledge Based Recommendation

Knowledge based recommendation does not rely on the user's and item's historical data, but utilize the previous knowledge and generate recommendations by rules[8]. More specifically, user specify their requirements, and then knowledge based algorithms give the recommendation that meet their needs.

(4)Hybrid Recommendation

In many practical recommendation system, single algorithm is often difficult to meet all of the demands. Researchers often integrate several recommendation algorithms and get a better recommendation result..

B. Basic Job Recommendation Algorithm

Compared with the recommendation algorithms above, content based recommendation results lack diversity, which means it will miss some suitable positions if user haven't deliver the similar one, the knowledge based recommendation is often applied to the purchase of low frequency items such as houses or cars, hybrid recommendation is too complicated and not suitable for online job recommendations. Collaborative filtering recommendation algorithm seems to be the right choice. In our previous researching[5], we compared the two different collaborative filtering recommendation methods, and item-based collaborative filtering got an better result in real worldwide job recommendation occasions. In this paper, item-based collaborative filtering has been chosen as our basic job recommendation algorithm.The steps of the algorithm are as follows:

Firstly, find out the positions that co-delivery users who applied for at least one common position with job hunter U_i

978-1-4673-9613-4/16/$31.00 ©2016 IEEE

have delivered and add this positions into candidate position set $Item^i : \{Item_1^i, Item_2^i, ..., Item_p^i\}$.

Secondly, for each position $Item_j^i$ in candidate position set, calculate the preference of user U_i by referring to the function below:

$$pref_o(U_i, Item_j^i) = \sum_{l=1}^{K} sim(Item_l^{U_i}, Item_j^i)$$

$Sim(\bullet)$ represents the similarity measure between two items. $Item_j^{U_i}$ represents the position l that user U_i has delivered and K represents the total number of positions that user U_i have delivered.

Finally, rank those position with their prediction preference and generate the recommendations by selecting the top N highest degree of preference.

III. OPTIMIZATION METHOD

Item-based collaborative filtering algorithm can recommend positions for job hunters. But when we launched an in-depth analysis on basic algorithm, it has shown that this method calculate items similarities among positions only depends on job hunters' delivery behavior and ignore the descriptions of positions or job hunters. In this paper, historical delivery weight calculated by position descriptions and similar user weight calculated by resume information were added as two influencing factors in the preference prediction formula. The specific optimization method is as follows.

A. Optimized by Position Descriptions

(1).Positions Similarity

The description of position mainly include the field, academic requirements, location, and other information. In this paper, five key factors has been considered in position similarity calculation. The specific formula is as follows:

$$simJ(Item_x, Item_y) = \prod_{l=1}^{2} f(J_l^{Item_x}, J_l^{Item_y}) * \frac{\sum_{l=3}^{5} f(J_l^{Item_x}, J_l^{Item_y})}{3}$$

$$f(J_l^{Item_x}, J_l^{Item_y}) = \begin{cases} 0 \cdots\cdots J_l^{Item_x} \neq J_l^{Item_y} \\ 1 \cdots\cdots J_l^{Item_x} = J_l^{Item_y} \end{cases}$$

$J_1^{Item_j}, J_2^{Item_j}, J_3^{Item_j}, J_4^{Item_j}, J_5^{Item_j}$ represent the position's field、academic requirements 、 title 、 location 、 category, respectively.

(2)Historical Delivery Weight

It is obvious that historical delivery data can reflect user's job searching intention. Based on the similarity calculated above, we increased the weight of candidate positions that similar to the historical delivery positions. The specific formula is as follows:

$$\omega_H(Item_j^i) = 1 + \max(simJ(Item_j^i, Item_m^{U_i})), Item_m^{U_i} \in Item^{U_i}$$

$\max(\bullet)$ Represents the maximum value in the bracket, $simJ(Item_j^i, Item_m^{U_i})$ represents the positions similarity between candidate position $Item_j^i$ and $Item_m^{U_i}$, $Item^{U_i}$ represents the historical delivery positions of user U_i .

B. Optimized by Resume Information

(1)User Similarity

The description of user mainly refers to user's resume information which consists of include his major and other basic information. As we all know, a job hunter's major will largely determine the successful rate in job application. In this paper the user description similarity has been divided into major similarity and basic information similarity.

(a)Major similarity

Due to the limited data, it is difficult to calculate the similarity of major just by the name of major. But, it is always true that the major similarity can reach very high among job hunters who apply for the same position. More specifically, we analyzed the major distribution of job applicants applied for the same position, calculated the support and confidence between majors, and obtained the final major similarity.Take major m_a, m_b as example.The support formula is as follows：

$$S(m_a, m_b) = \frac{\left|\{I_n | m_a \in M^{I_n}, m_b \in M^{I_n}, I_n \in I\}\right|}{|I|}$$

The confidence formula for both m_a to m_b and m_b to m_a as follows：

$$C(m_a \to m_b) = \frac{\left|\{I_n | m_a \in M^{I_n}, m_b \in M^{I_n}, I_n \in I\}\right|}{\left|\{I_m | m_a \in M^{I_n}, I_m \in I\}\right|}$$

$$C(m_b \to m_a) = \frac{\left|\{I_n | m_a \in M^{I_n}, m_b \in M^{I_n}, I_n \in I\}\right|}{\left|\{I_m | m_b \in M^{I_n}, I_m \in I\}\right|}$$

$|\bullet|$ represents the number of objects contained in the set。 I and I_n represent the collection of all positions and a particular position, respectively. M^{I_n} represents the major set of applicants who have applied position I_n .

The final major similarity measure can be calculated as follows：

$$simM(m_a, m_b) = \begin{cases} 0 & \cdots S(m_a, m_b) < \varepsilon \\ \max(C(m_a \to m_b), C(m_b \to m_a)) & \cdots S(m_a, m_b) \geq \varepsilon \end{cases}$$

$\max(\bullet)$ Represents the maximum value in the bracket, ε represents the threshold。

(b)Basic Information Similarity

User basic information mainly refer to the information of gender, educational background, location and political outlook. The basic information similarity among user U_i and U_o can be calculated is as follows:

$$simB(U_i,U_o)=\frac{\sum_{l=1}^{4}\omega_l * f(S_l^{U_i},S_l^{U_o})}{\sum_{l=1}^{4}\omega_l}$$

$$f(S_l^{U_i},S_l^{U_o})=\begin{cases}0\cdots\cdots S_l^{U_i}\neq S_l^{U_o}\\1\cdots\cdots S_l^{U_i}=S_l^{U_o}\end{cases}$$

$S_1^{U_i},S_2^{U_i},S_3^{U_i},S_4^{U_i}$ represent gender, educational background, location and political outlook respectively. ω_l represents the calculation weight.

Linear combination was utilized to integrate the two similarity. The final user similarity formula is as follows:

$$simU(U_i,U_\bullet)=0.6*simM(m_{U_i},m_{U_o})+0.4*simB(U_i,U_\bullet)$$

(2)Similar User Weight

It is obvious that these positions which similar user have applied are much more likely to be the candidate positions. In this paper, we increased the weight of the candidate positions whose certified applicants are more similar to target user based on the user similarity above. The specific formula is as follows:

$$\omega_U(Item_j^i)=1+\max(simU(U_i,U_\bullet)),U_\bullet\in U^{Item_j^i}$$

$\max(\bullet)$ represents the maximum value in the bracket, $U^{Item_j^i}$ represents a collection of users who have applied for candidate position $Item_j^i$.

C. Optimized Prediction Formula

Based the analysis above, we optimized the preference formula in basic recommendation algorithm by multiply the historical delivery weight and similar user weight.The specific formula is as follows:

$$pref(U_i,Item_j^i)=pref_\bullet(U_i,Item_j^i)*\omega_H(Item_j^i)*\omega_U(Item_j^i)$$

With the optimization formula above, we computed the preference of each position in candidate set and selected the top N positions with the highest preference as the recommendation positions for job hunter U_i. The process framework of our optimization method is presented as follows:

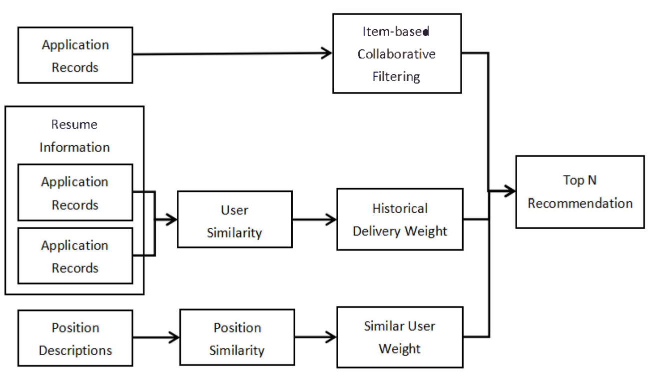

Fig. 1. Process Framework of Our Optimization Method.

IV. EXPERIMENTS AND ANALYSIS

A. Test Data

The test data was collected by a recruitment website which contains 737 positions information, 100 job-hunters' resume information and 1072 deliver records in the year of 2013.

TABLE I. CONTENTS OF TEST DATE

APPLICATION RECORDS	Application ID; Student ID; Position ID; Application Date;
POSITION DISCRIPTIONS	Position ID ;Title; Location;Type; Filed; Category
RESUME INFORMATION	Student ID;Major;Gender; Educational Background; Location; Political Outlook

B. Evaluation

In order to better verify the effectiveness of the recommended optimization, the precision and the recall rate are selected as the test index in the experiment.Specific introduction is as follows:

(1)Precision

Precision refers to the average correct rate of the recommended list. The proportion of recommended positions which have actually be applied by the recommended job hunter. The formula for precision is as follows:

$$precision=\frac{|R(u)\cap P(u)|}{|R(u)|}$$

R(u) represents a collection of recommended positions for job hunter U, I(u) represents positions that job hunter U has applied. $|\bullet|$ represents the number of the elements in a collection.

(2)Recall

Recall rate refers to the average coverage rate of the recommended list. The proportion of the applied positions which have been recommended.The formula for recall is as follows:

$$precision=\frac{|R(u)\cap P(u)|}{|P(u)|}$$

R(u) represents a collection of recommended positions for job hunter U, I(u) represents positions that job hunter U has applied. $|\bullet|$ represents the number of the elements in a collection.

(3)F1-measure

F-1measure is comprehensive method which integrated precision and recall and could be effective in balancing the precision and recall.The formula for F-measure is as follows:

$$F1=\frac{2*precision*recall}{precision+recall}$$

C.Results and analysis

In our experiments, we chose Loglikelihood similarity as the similarity measure method in basic item-based collaborative filtering algorithm. For each target user, 1-9 positions have been recommended. The final comparison of between the original algorithm and the optimized algorithm is as follow:

TABLE II. PRECISION RESULTS

N	Basic Algorithm	Optimized Algorithm
1	44.74%	53.03%**(+18.5%)**
2	37.90%	44.55%**(+17.5%)**
3	37.78%	44.87%**(+18.8%)**
4	36.84%	44.61%**(+21.1%)**
5	33.33%	37.6%**(+12.8%)**
6	30.77%	34.72%**(+12.8%)**
7	27.82%	29.37%**(+5.6%)**
8	27.68%	27.81%**(+0.5%)**
9	19.03%	19.21%**(+1.0%)**

TABLE III. RECALL RESULTS

N	Basic Algorithm	Optimized Algorithm
1	38.20%	39.33%**(+3.0%)**
2	32.86%	34.29%**(+4.4%)**
3	32.67%	33.33%**(+2.0%)**
4	32.56%	33.14%**(+1.8%)**
5	28.13%	28.13%
6	26.67%	26.67%
7	25.17%	25.17%
8	25.83%	25.83%
9	17.17%	17.17%

TABLE IV. F1-MEASURE RESULTS

N	Basic Algorithm	Optimized Algorithm
1	41.21%	45.16%**(+9.6%)**
2	35.20%	38.75%**(+10.%)**
3	35.04%	38.25%**(+9.2%)**
4	34.57%	38.03%**(+10.0%)**
5	30.51%	32.18%**(+5.5%)**
6	28.57%	30.17%**(+5.6%)**
7	26.43%	27.11%**(+2.6%)**
8	26.72%	26.78%**(+0.2%)**
9	18.05%	18.13%**(+0.4%)**

It can be clearly seen from the results that optimization method have largely improved the precision and recall rate when 1-7 positions have been recommended. In real world job recommendation system, recommend 3-6 positions each time can meet the need of job hunters. That means our optimization method might be a promising candidate algorithm for job recommendations.

V. SUMMARY AND PROSPECT

The optimization algorithm proposed in this paper aim to improve the accuracy of job recommendation. Specifically, historical delivery weight calculated by position descriptions and similar user weight calculated by resume information were added as two influence factors in preference prediction. When test with real world data set, our optimization method largely outperformed the original algorithm. Our future work may verify the feasibility of our optimization method on more data set.

ACKNOWLEDGMENT

The work is supported by grants from National Science and Technology Supporting Program of China (2014BAH10F00), New Century Excellent Talents in University of Ministry of Education of China (NCET-12-0944), University Research Program of Communication University of China (3132015XNG1522). We thank the reviewers and editors for their helpful comments.

REFERENCES

[1] Rafter R, Bradley K, Smyth B. Personalised Retrieval for Online Recruitment Services[C]// 22nd Annual Colloquium on Information Retrieval (IRSG2000). 2000:382.

[2] Lee D H, Brusilovsky P. Fighting Information Overflow with Personalized Comprehensive Information Access: A Proactive Job Recommender[C]// International Conference on Autonomic and Autonomous Systems. 2007:21-21.

[3] Malherbe E, Diaby M, Cataldi M, et al. Field selection for job categorization and recommendation to social network users[C]// Ieee/acm International Conference on Advances in Social Networks Analysis and Mining. 2014:479-87..

[4] Almalis N D, Tsihrintzis G A, Karagiannis N, et al. FoDRA ¿¿¿ A new content-based job recommendation algorithm for job seeking and recruiting[C]// International Conference on Information, Intelligence, Systems and Applications. IEEE, 2015..

[5] Zhang Y, Yang C, Niu Z. A Research of Job Recommendation System Based on Collaborative Filtering[C]// Seventh International Symposium on Computational Intelligence and Design. IEEE, 2014:533-538.

[6] Pazzani M J, Billsus D. Content-Based Recommendation Systems[M]// The adaptive web. Springer-Verlag, 2007:325--341..

[7] Adomavicius G, Tuzhilin A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions[J]. IEEE Transactions on Knowledge & Data Engineering, 2005, 17(6):734-749..

[8] Jannach D, Zanker M, Felfernig A, et al. Recommender systems : an introduction[M]. Cambridge University Press, 2011..