Collaborative Filtering Algorithm for Employment Based on Employment Behavior

Jie Niu

Suzhou University, Suzhou234000, Anhui, China

Abstract

There are relatively less studies on the issues of the prediction of the recruitment companies that the job seekers may prefer in view of the network utilization records. This paper proposes a modeling method for the employment behavior of the job seekers, and builds a structure relation between the job seekers with the employment information, it can be found that the set of the neighbors is the most influential on the job seekers, and such similarity can be integrated into the collaborative filtering algorithm for employment that is based on the probabilistic matrix factorization. The experimental results show that compared with the traditionally applied social network of information and the tag information recommendation algorithm, the proposed CFAE recommendation algorithm can predict the actual demand of the job seekers more effectively and improve the recommendation accuracy.

Key words: Collaborative Filtering, Employment Behavior, Probabilistic Matrix Factorization.

1. INTRODUCTION

With the rapid development of the Internet technology, the data volume on the network has increased sharply, and job seekers gradually fall into the vast expanse of the sea of employment information, hence it has become more and more urgent for us to quickly and efficiently acquire the employment information that we need from such vast data. Under this background, recommendation system hence come with the tide, through the recommendation system to collect and analyze various kinds of information of users to learn their interest and behavior patterns, according to the analysis of the user interest and behavior patterns, recommend to the user the services that he may need.

In the numerous recommendation methods, the collaborative filtering is the most applied algorithm at present. Although the collaborative filtering algorithm for employment has gained huge success, it still has many problems, among which a very important one is that the traditional collaborative filtering algorithm often ignores the structural relations between job seekers. The effective use of the connection between the job seekers can enrich the information of the individual job seekers, so as to identify the personal interest of the job seekers more accurately. The existing solutions mainly include two kinds: One kind is to use an explicit social network relation (Choi, Ko and Han, 2012; Pagare and Patil, 2013; Rosa, Rodrigues and Basso, 2013); the other is by implicit tag information to calculate the degree of similarity between the job seekers (Su, Wang and Lochovsky, et al., 2012; Walun and Sadafale, 2013; Liu, Li and Tang, et al., 2014), so as to obtain the relation between the job seekers. In the actual data, however, it is fairly difficult to get enough social relations or tag information. One of the core steps of the collaborative filtering algorithm that is based on the mining of relation is to obtain the relation of the job seekers. At present, the ways of obtaining the relation of the job seekers are mainly divided into two categories, explicit and implicit category (Zhang, Li and Feng, et al., 2015; Bruyère, Cooper and Pelletier, et al., 2014; Salehi and Kamalabadi, 2013). Literature (Lee, Kaoli and Huang, 2014) proposes through the explicit social network relation to obtain the connection between the job seekers, and add the social relation matrix on the basis of the original score matrix increase, which has greatly improved the effect of the algorithm. Literature (Font, Serra and Serra, 2013) proposes using a random walk model that makes use of the trust relation network of the job seekers. Through conduct random walk on the social network to find similar items, it can increase the forecast rating sources to the target item, so as to reduce the impact caused by data sparsity. As in the actual system, it is difficult to obtain enough network relation. And literatures (Ahn, Kim and Choi, et al., 2014; Liu, Wu and Liu, 2013) propose the application of tag information to acquire implicit relation matrix, and then give the corresponding recommendations. Literatures (Kim and El Saddik, 2013; Hunag, Yuan and Zhong, et al., 2015) make use of tag information to calculate the near neighbors, and assumes that the near neighbors will directly affect the feature vector of the job seekers, so as to add a priori based on the near neighbor relation for the feature vector of the job seekers; On this basis, through learning the feature vector by gradient descent method, the near labor relation based matrix factorization model is further perfected.

Based on the above problem, this paper proposed a modeling method of the nearest neighbor based on the time series consumption behavior, by building a network of consumption that is based on time sequence, to obtain the mutual influence relations of the job seekers. As this modeling method only requires the consumption time of the job seekers, and do not need complicated information of the job seeker, such as the employment
labels, social relations and so on, and the influence of its computation is directed, which can find out the interational relation between the job seekers, therefore, it can accurately identify the collection of the neighbors that have the greatest influence on the current job seeker, and its applicable field is also very broad. On this basis, a recommendation algorithm called CFAE is designed, which successfully applies the neighbor set to the collaborative filtering algorithm for employment that is based on probabilistic matrix factorization. Finally, this paper carries out corresponding experiment on real job.zhaopin.com website recommendation data set, the results show that the CFAE (Collaborative Filtering Algorithm for Employment) can predict the actual score of the job seekers more effectively than the traditional application of social network information and tag information recommendation algorithm, thus improve the recommendation accuracy.

2. DESCRIPTION OF COLLABORATIVE FILTERING ALGORITHM FOR EMPLOYMENT

This section will describe in detail the CFAE recommendation algorithm designed in this paper. Firstly, introduce how to use the sequential information of the time consumption to explore the mutual influence relation between the job seekers, so as to determine the neighbor set that has the greatest impact on the job seeker; Then introduce to how to blend the neighbor set successfully in the collaborative filtering algorithm that is based on probabilistic matrix factorization; And then, briefly analyze the time complexity of CFAE; Finally, make analysis on the features of the algorithm.

2.1. Employment Behavior Modeling Based Nearest Neighbor Selection

In the relation based matrix factorization model, one of the core steps is the acquisition of to the relations of the job seekers. All the traditional collaborative filtering algorithms ignore the consumption time information of the job seekers or the items, however, the time sequence information of the items consumption of the job seekers may have part of the hidden rules, and making use of these rules can explore the relationship between the job seekers and the items to a certain extent. For example, job seeker A saw a part-time recruiter, and job seeker B saw the same part-time recruiter in a short period of time after A saw it, if this scenario happened for many times, it was likely that A had the potential influence relation to B. In order to find out this potential relation, we firstly introduce the time sequence based job seekers consumption network diagram, as shown in Figure 1.

In the consumption network diagram, $U$ is the set of job seekers, $E$ is the set of edges, $W$ represents the weight of the edge, the content in the "()" shows the number of items consumed by the job seekers. If within the set period of time (for example, 1 day), $U_i$ and $U_j$ have consumed the same item subsequently, then the weight of the edge $W_{i\rightarrow j}$ increased 1. After traversing all the items, the number of the items that are in line with such condition is the weight $W_{i\rightarrow j}$ of the directed edge from $U_i$ to $U_j$. According to this network diagram, we define that the weight of influence relations between the job seekers is as the following:

$$ T_{i\rightarrow j} = \frac{W_{i\rightarrow j}}{f(U_i, U_j)} $$

Where, $f(U_i, U_j)$ is the union of the sets of items that are consumed by job seeker $U_i$ and $U_j$, and $T_{i\rightarrow j}$ is the influence of i to j. In Figure 1, for example, suppose the number of union of the sets of items that are consumed by job seeker $U_i$, and job seeker $U_j$ is 100, then the influence of $U_i$ to $U_j$, $T_{i\rightarrow j} = 5/100 = 0.05$, while the influence of $U_j$ to $U_i$, $T_{j\rightarrow i} = 25/100 = 0.25$. Thus it can be seen that, the influence between two of the job seekers is directed, compared with the existing undirected computation methods, this method has better rationality.

In the same way, we can set up the consumption network diagram for the items (please refer to Figure 2), based on the similarity of the item consumption network diagram to Figure 1, the nodes stand for the items, the digit in the "()" demonstrates the number of job seekers who have consumed the item, and the weight of the edge stands for how many job seekers have consumed the two items at the endpoints subsequently. After the corresponding network diagram is set up, formula (5) can be applied to make the computation for the influence relation accordingly.

$$ S_{i\rightarrow j} = \frac{W_{i\rightarrow j}}{f(V_i, V_j)} $$
The experimental results show that, the method of the computation of influence that is proposed in this paper is simple and effective, by the strength of this influence relation, it is possible to seek for the set of neighbors that has the most significant influence on a specific job seeker (such as Top - 20), further incorporate these nearest neighbors into the collaborative filtering that is based on matrix factorization, so as to improve the accuracy of the recommendation system.

2.2. Matrix Factorization Model

After the completion of the mining of the influence relation between the corresponding job seekers and the set of the nearest neighbors is found, apply it to the matrix factorization model. At this point, the feature vector of the job seekers should be influenced by its nearest neighbor job seekers, namely similar job seekers shall have similar feature vector.

\[
\bar{U}_u = \sum_{i \in N_u} T_{u\rightarrow i} U_i, \quad \bar{V}_v = \sum_{j \in N_v} S_{v\rightarrow j} V_j
\]  

(3)

Wherein, \( U \) and \( V \) represent the approximate feature vectors, \( U_u, N_u \) represent the set of the neighbors of job seeker \( u \) and product \( i \) respectively. During the learning of the feature vectors, this paper takes into comprehensive consideration of the dual influence of the characteristics of the job seekers themselves and the influence related to the characteristics of the job seekers, and the feature vector of each job seeker not only obeys the Gaussian prior mean value of 0 so as to prevent over fitting, but also shall be similar to the feature vector of the relation job seekers. In addition, as we have considered the information of the time sequence, its influence relation is more explicit and more in line with the actual situation:

\[
p(U \mid T, \sigma_u^2, \sigma^2) = \prod_{u=1}^{N_u} p(U_u \mid 0, \sigma_u^2 I) \times \prod_{v=1}^{N_v} \left[ \sum_{i \in N_u} T_{u\rightarrow i} U_i, \sigma^2 I \right] 
\]

(4)

\[
p(V \mid S, \sigma_v^2, \sigma^2) = \prod_{i=1}^{M} N(V_i \mid 0, \sigma_v^2 I) \times \prod_{j=1}^{M} \left[ \sum_{j \in N_v} S_{j\rightarrow v} V_j, \sigma^2 I \right] 
\]

(5)

Through the Bayesian inference, its posterior probability is as the following:
\[ p(U,V|R,T,S,\sigma^2,\sigma^2_v,\sigma^2_v) \propto p(R|U,V,\sigma^2_R)p(U|T,\sigma^2,\sigma^2_v)p(V|S,\sigma^2,\sigma^2_v) \]
\[ = \prod_{i=1}^{N} \prod_{j=1}^{M} \mathcal{N}(R_{ij}|g(U_i^T V_j),\sigma^2_R) \times \prod_{i=1}^{N} \left( \sum_{T_{ij}} U_i \right) \times \prod_{j=1}^{M} \left( \sum_{S_{ij}} V_j \right) \]
\[ = \prod_{i=1}^{N} N(U_i|0,\sigma^2_u I) \times \prod_{j=1}^{M} N(V_j|0,\sigma^2_v I) \]

In order to facilitate the solving, conduct logarithmic processing to the posteriori probability obtained by formula (6) as the following:
\[ \ln p(U,V|R,T,S,\sigma^2,\sigma^2_v,\sigma^2_v) = -\frac{1}{2\sigma^2_R} \sum_{i=1}^{N} \sum_{j=1}^{M} I_{ij}^2 (R_{ij} - g(U_i^T V_j))^2 - \frac{1}{2\sigma^2_u} \sum_{i=1}^{N} U_i^2 U_i - \frac{1}{2\sigma^2_v} \sum_{j=1}^{M} V_j^2 V_j \]
\[ - \frac{1}{2\sigma^2_R} \sum_{i=1}^{N} \sum_{j=1}^{M} T_{ij}^2 (U_i - \sum_{i=1}^{N} T_{ij} U_i)^2 \]
\[ - \frac{1}{2\sigma^2_u} \sum_{i=1}^{N} \sum_{j=1}^{M} S_{ij}^2 (V_j - \sum_{j=1}^{M} S_{ij} V_j)^2 \]
\[ - \frac{1}{2}(N \times K) \ln \sigma^2_R + (M \times K) \ln \sigma^2_u + (N \times K) \ln \sigma^2_v + (M \times K) \ln \sigma^2_v + C \]

To maximize the posteriori probability, it is equivalent to minimize the objective function as the following:
\[ L(R,T,S,U,V) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{M} I_{ij}^2 (R_{ij} - g(U_i^T V_j))^2 \]
\[ + \frac{\lambda_v}{2} \sum_{i=1}^{N} U_i^2 U_i + \frac{\lambda_u}{2} \sum_{i=1}^{N} V_i^2 V_i + \frac{\lambda_T}{2} \sum_{i=1}^{N} \left( U_i - \sum_{i=1}^{N} T_{ij} U_i \right)^2 \left( U_i - \sum_{i=1}^{N} T_{ij} U_i \right) \]
\[ + \frac{\lambda_S}{2} \sum_{j=1}^{M} \left( V_j - \sum_{j=1}^{M} S_{ij} V_j \right)^2 \left( V_j - \sum_{j=1}^{M} S_{ij} V_j \right) \]
\[ \lambda_v = \sigma^2_u / \sigma^2_u, \lambda_u = \sigma^2_v / \sigma^2_v, \lambda_T = \sigma^2_R / \sigma^2_R, \lambda_S = \sigma^2_v / \sigma^2_v \] (9)

In the above formula, \( \lambda_v = \sigma^2_u / \sigma^2_u \), \( \lambda_u = \sigma^2_v / \sigma^2_v \), \( \lambda_T = \sigma^2_R / \sigma^2_R \), \( \lambda_S = \sigma^2_v / \sigma^2_v \). Through the method of gradient descent, the feature vector of each job seeker (item) can be obtained, and the computation method of its gradient is as the following:
\[ \frac{\partial L}{\partial U_i} = \sum_{j=1}^{M} I_{ij}^2 V_j g(U_i^T V_j) (g(U_i^T V_j) - R_{ij}) + \lambda_v U_i + \lambda_T \left( U_i - \sum_{j=1}^{M} T_{ij} U_i \right) - \lambda_u \left( U_i - \sum_{j=1}^{M} T_{ij} U_i \right)^2 \]
\[ \frac{\partial L}{\partial V_j} = \sum_{i=1}^{N} I_{ij}^2 U_i g(U_i^T V_j) (g(U_i^T V_j) - R_{ij}) + \lambda_v V_j + \lambda_S \left( V_j - \sum_{j=1}^{M} S_{ij} V_j \right) - \lambda_u \left( V_j - \sum_{j=1}^{M} S_{ij} V_j \right)^2 \]

Wherein, \( g'(x) \) is the derivative of \( g(x) \), in this paper, \( g'(x) = \frac{e^{-x}}{1+e^{-x}} \).

3. EXPERIMENTAL RESULT AND ANALYSIS

In this section, firstly, the data set used in the experiment is introduced, then the evaluation standard is illustrated and comparison is conducted on the algorithms, and finally, the comparison experimental result of CFAE model and other methods is presented, and the corresponding analysis is carried out on the experimental result.

3.1 Experimental Data Set

In order to compare the influence of different information (relations) to the recommendation result, the experimental data set should contain the rating information, tag information and the social relations of the job seekers. For this purpose, this paper makes use of the data fetched from the job.zhaopin.com website as the experimental data set. Job.zhaopin.com is a well-known talent recruitment website, which provides score rating, discussion and recommendation services to the job seekers. It owns the biggest recruitment information database in China at present, and is one of the largest online communities in China. In the website, each job seeker can rate the
recruiters within the scope of [1, 5). In addition, job.zhaopin.com website allows the job seekers to find their friends via E-mail. To sum up, job.zhaopin.com data set is suitable for the experimental research in this paper.

In this paper, two groups of data sets are extracted from the job.zhaopin.com website: one set is the rating information of the job seekers to full time recruitment units, and the social relation and tag information of the job seekers; the other set is the rating information of the job seekers to part time recruitment units and other corresponding information.

3.2. Evaluation Standard

In this paper, RMSE is adopted as the evaluation standard for the experiment in this paper. RMSE measures the accuracy of the prediction through the computation of the deviation between the predicted job seeker’s rating and the actual job seeker’s rating. RMSE provides an intuitive method for the recommendation quality, and is the most commonly used recommendation quality measurement methods. The smaller the overall RMSE of the recommendation algorithm is, the higher the recommendation quality will be, so to speak.

Assuming that rating vector of the algorithm for the prediction of \( C \) items is represented as \( \{ p_1, p_2, \ldots, p_C \} \), the corresponding actual rating vector is \( \{ r_1, r_2, \ldots, r_C \} \), then the RMSE of the algorithm is expressed as

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{C} (p_i - r_i)^2}{C}}
\]  

(12)

3.3. Experimental Result and Analysis

In this paper, four kinds of methods are selected as comparison algorithms, the probabilistic matrix factorization (PMF), time sequence information based tensor factorization (BPTF), social relation based matrix factorization (Social MF), and tag information based matrix factorization (Tag MF). In the experiments, in order to reduce the complexity of the model computation, this paper sets \( \lambda_u = \lambda_v = 0.001 \), and assumes \( \lambda_p = \lambda_q = 0 \). As Social MF has no item relations, thus set \( \lambda_s = 0 \), where the initial value of the feature vector \( U \) and \( V \) is obtained through random extraction of normal distribution with the mean value of 0. After that, in the operation of a single iteration, the feature vector \( U \) and \( V \) have iterative update on the basis of the value of the previous iteration, until it achieves convergence. For Tag MF and CFAE, in this paper, Top-20 most similar or most influential job seeker is selected as the neighbor of the target job seeker. And the number of the neighbors of the job seeker in Social MF is acquired based on the data of social relations, and the number of the neighbors of its item is 0.

Experiment 1. Algorithm result under different setting of dimensionality of the feature vector

Firstly, the experiment compares the results of various algorithms under different setting of dimensionality of the feature vector. In the experiments, we set the dimensions of feature vector \( K=5,10,20 \) respectively. The other parameters of the algorithm in the experiment are set as the value that enables the optimum performance of the algorithm respectively. And the result of the RMSE comparison of the algorithm under different setting of dimensionality of the feature vector is shown in Table 1, as can be seen from the experimental results:

1) With the increase of \( K \), the algorithm has certain improved accuracy. However, it should be pointed out that, to a certain extent the increase of \( K \) can increase the time complexity of the model;

2) Compared with PMF, BPTF, Tag MF, Social MF and CFAE have great improvement, which further illustrates the big role that the time sequence information and the relation information between the job seekers plays in the improvement of the accuracy of the traditional collaborative filtering algorithm;

3) The accuracy of BPTF has improved compared with that of the Social MF, which is caused mainly by the relative scarcity of the social relations as shown in the Social MF. Compared with Tag MF and CFAE, the result of BPTF is a little bit worse.

4) The accuracy of Social MF is not as high as that of Tag MF and CFAE, this is largely due to the fact that Social MF does not take the relations between the items into consideration, in addition, Social MF does not consider that the influence relation between good friends is directed.

5) CFAE has further improvement than Tag MF, which shows that the influence relation put forward in this paper can effectively improve the accuracy of the algorithm. In addition, as only simple time information is required by CFAE, from the acquisition of the information and its application scope, CFAE also has greater advantage.
Table 1. RMSE Comparison Result under Different Setting of Dimensionality K

<table>
<thead>
<tr>
<th>Model</th>
<th>K=5</th>
<th>K=10</th>
<th>K=20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full Time Recruitment Unit</td>
<td>Part Time Recruitment Unit</td>
<td>Full Time Recruitment Unit</td>
</tr>
<tr>
<td>PMF</td>
<td>0.7511</td>
<td>0.7358</td>
<td>0.7465</td>
</tr>
<tr>
<td>BPTF</td>
<td>0.7317</td>
<td>0.7279</td>
<td>0.7267</td>
</tr>
<tr>
<td>Social MF</td>
<td>0.7339</td>
<td>0.7309</td>
<td>0.7307</td>
</tr>
<tr>
<td>Tag MF</td>
<td>0.7298</td>
<td>0.7267</td>
<td>0.7240</td>
</tr>
<tr>
<td>CFAE</td>
<td>0.7294</td>
<td>0.7251</td>
<td>0.7238</td>
</tr>
</tbody>
</table>

Experimental results show that, compared with traditional PMF, CFAE has relatively great improvement, which fully illustrates the rationality and effectiveness of the influence relation proposed in this paper. Compared to Social MF, Tag MF and BPTF, although there is no significant improvement in the accuracy of the algorithm proposed in this paper, in the actual recommendation system, the social network relations or the tag information that is required by Social MF and Tag MF need Social network is often difficult to get or extremely sparse, whereas CFAE only needs the time information that is relatively easier to obtain, therefore, the practical application scenarios of the algorithm are more extensive. But for BPTF, as BPTF requires Markov Monte Carlo algorithm to carry out the parameter estimation, its time efficiency is greatly reduced compared with that of CFAE. In addition, CFAE has also provided a new train of thought for the acquisition of the relation behavior and how to extend the probabilistic matrix factorization model.

Experiment 2. Running time of the algorithm

In this experiment, we compared the various algorithms on the specific running time of each iterative update. The running environment of the experiment is as follows: Intel Core i3 CPU, primary frequency 2.67GHZ, Windows 7 system and 2G memory. In the experiment, set the feature vector dimension K = 5, and the results are shown in Table 2. As can be seen from the experimental results, the running time of the algorithm meets:

\[
\text{BPTF} > \text{CFAE} \approx \text{Tag MF} > \text{Social MF} > \text{PMF}. 
\]

The running time of BPTF is far higher than several of the other algorithms. This is mainly because it consumes too much time for BPTF to conduct the Markov Monte Carlo training, so the time efficiency of the algorithm is not high. In addition, it also can be shown from Table 2 that: The more relations that are taken into consideration, the higher its time complexity is. At the same time, as can be seen from the experiment: for PMF and BPTF, the running speed of full time recruitment units is faster than that of the part time recruitment units, which is the opposite for the rest of the algorithms on the other hand. This is mainly because PMF and BPTF do not consider the neighbor relation, therefore, their time complexity is only related to the rating data volume of the training set. The rating data of full time recruitment units is scarcer than that of the part time recruitment units, so the running time it requires is relatively less; However, for the rest of the algorithms, they have all considered the relation between the neighbors, as the full time recruitment unit has more job seekers and items than that of the part time recruitment unit data set, hence its complexity is higher.

Table 2. Comparison Result of the Running Time of a Single Iteration in Training

<table>
<thead>
<tr>
<th>Model</th>
<th>Full Time Recruitment Unit</th>
<th>Part Time Recruitment Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMF</td>
<td>1.7</td>
<td>1.9</td>
</tr>
<tr>
<td>BPTF</td>
<td>15.2</td>
<td>18.6</td>
</tr>
<tr>
<td>Social MF</td>
<td>2.9</td>
<td>2.7</td>
</tr>
<tr>
<td>Tag MF</td>
<td>4.2</td>
<td>3.9</td>
</tr>
<tr>
<td>CFAE</td>
<td>4.4</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Experiment 3. Comparison of Experimental Results under Different Rating Scarcity

In order to compare the effect of algorithm under different scarcity, this paper takes the number of ratings of the job seekers as the rating basis, and divides the job seekers into four groups, the number of rating is [0:10],[10:100],[100,500] and above 500 respectively. Figure 6 shows that in the two sets of data, the proportion of each group of job seekers.
After the corresponding division to the data, this paper makes use of the training set to study the corresponding model, then conduct the computation of RMSE in the test set to the four groups of job seekers respectively, and the results are shown in Figure 4 and Figure 5. It can be seen from the experiment of the two data sets that:

In the case of extremely sparse data (the number of rating is less than 10), the improvement effect of CFAE is not significant, which is worse than the Social MF and Tag MF with the introduction of additional information. This is mainly because the data sparse can cause the extreme scarcity of the network diagram that CFAE established, and thus influence its accuracy; while the parameter training method is introduced in the BPTF, so it is better than CFAE in the effect. However, CFAE is still better than the traditional PMF.

When the rating data of the job seekers increase, CFAE is better than Tag MF, Social MF and BPTF, thus further proves that the introduction of the influence relation in this paper can effectively improve the accuracy of the recommendation system.

From Figure 4, it also can be seen that: whether the CFAE proposed in this paper or the other algorithms, the RMSE result does not get smaller continuously with the increment of the number of ratings, when the number of ratings increases to above 500, the effect of the algorithm does not continue to get better but slightly declines. This is mainly because when the number of ratings is relatively small, the scarcity of the data influences the effect of the algorithm, at this point, over fitting phenomenon [8,25] is prone to occur in this model, that is, its accuracy in the training sample is good, while the accuracy in the test sample is very poor; with the increase in the number of ratings, the data becomes denser, thus alleviates the impact of the sparse data, and the effect of the algorithm is getting better gradually; After the training samples of the ratings are too many, the interest of the job seekers will spread, which will cause the result that the model cannot learn the characteristic preferences of the job seekers, which may affect the accuracy of the model. Therefore, it has a certain impact on the prediction result of the model to have good control on the sample size. As can be seen from Figure 4: For the purposes of the data sets in this paper, when the sample size of the job seekers is in the range of [100:500), the effect of the algorithm is optimal.

4. CONCLUSION

This paper makes use of the consumption time sequence of the job seekers to establish the consumption network diagram of the job seekers, through the network diagram, the mining of potential mutual influence relations between the job seekers is conducted and the nearest neighbors are looked for, then integrated into the collaborative filtering recommendation algorithm that is based on matrix factorization, so as to improve the accuracy of the rating prediction. As compared with social network information and tag information, the
consumption time information is easier to get, therefore, the collaborative filtering recommendation algorithm based on employment behavior has a wider application scope. Experiments on the real job.zhaopin.com recommendation data set show that this method has made certain improvement in the effect compared with the traditional recommendation algorithms.

REFERENCES