International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 05 | May -2017 www.irjet.net p-ISSN: 2395-0072

Implementation of an Automated Job Recommendation System Based on Candidate Profiles

Vinay Desai, Dheeraj Bahl, Shreekumar Vibhandik,Isra Fatma

Prof. Vijay Kolekar , Dept. of Computer Engineering, K.J. College of Engg & Mgt. Research, Pune, India

---

Abstract - *Dealing with the enormous amount of recruiting information on the Internet, a job seeker always spends hours to find useful ones. To reduce this laborious work, we design and implement a recommendation system for online job-hunting. In this paper, we contrast user-based and item-based collaborative filtering algorithm to choose a better performed one. We also take background information including students' resumes and details of recruiting information into consideration, bring weights of co-apply users (the users who had applied the candidate jobs) and weights of student used-liked jobs into their commendation algorithm. At last, the model we proposed is verified through experiments study which is using actual data. The recommended results can achieve higher score of precision and recall, and they are more relevant with users' preferences before.*

***Key Words*- recommendation system; item-based collaborative filtering; content-based filtering; Vector Space Model(VSM); Mahout**

1. INTRODUCTION

The increasing usage of Internet has heightened the need for online job hunting. According to *Jobsite's* report 2014, 68% of online jobseekers are college graduates or post graduates. The key problem is that most of job-hunting websites just display recruitment information to website viewers. Websites just display recruitment information to website viewers. Students have to retrieve among all the information to find jobs they want to apply. The whole procedure is tedious and inefficient. By creating an easy job recommendation system where everyone will have a fair and square chance. This saves a lot of potential time and money both on the industrial as well as the job seeker's side. Moreover, as the candidate gets a fair chance to prove his talent in the real world it is a lot more efficient system. The basic agenda of every algorithm used in today's world be it a traditional algorithm or a hybrid algorithm is to provide a suitable job that the user actually seeks and wishes for.

1.1 LITERATURE SURVEY

There are endless algorithms to help a seeker find the right job, some are the traditional algorithms while some are newly found and there are a large number of hybrid algorithms which are a combination of many algorithms.
All these algorithms have only goal to seek a righteous job for the candidate.

CF is a popular recommendation algorithm that bases its predictions and recommendations on the ratings or behaviour of other users of the system.
It also uses a profound technique called as Information Retrieval (IR). Information Retrieval is a new and advanced technique used for achieving the most accurate and desired result without compromising on the efficiency of the result.

1.2 RELATED WORK

A. Recommendation Algorithms

1) Content-based filtering (CBF):
In Content-based methods, features of items are abstract and compared with a profile of the user's preference.

In other words, this algorithm tries to recommend items that are similar to those that a user liked in the past. It is widely applied in information retrieval(IR). However it performs badly in multimedia field such as music or movie recommendation because it is hard to extract items attributes and obtains user's preference sometimes.

2) Collaborative Filtering (CF):
CF is a popular recommendation algorithm that bases its predictions and recommendations on the ratings or behavior of other users in the system. There are two basic types:

User-based CF and Item-based CF.

- **User-based CF**: find other users whose past rating behavior is similar to that of the current user and use their ratings on other items to predict what the current user will like. The working of User –based CF is a quite simplified technique; all it does is examine the past interests of the user and based on the past results the system makes an accurate result of the candidate who has applied for a job.

Item-based CF: Rather than using similarities between users' rating behavior to predict preferences, item–based CF uses similarities

- Between the rating patterns of items. Since finding similar items is easier than finding similar users, and attributes of items are more stable than users' preference, item-based methods are suitable for off-line computing. The preferred outcome of any of the two methods is to provide a suitable job. There are some drawbacks in

Collaborative Filtering as a whole aspect. Collaborative Filtering approaches often suffer from three problems: cold start, scalability and sparsely. These three drawbacks are very problematic at times and crucial opportunities can be missed because of this.

B. Methods of Similarity Calculation

1) Cosine Similarity

Cosine similarity uses two N-dimensional vector's cosine Value to indicate the degree of similarity between them. It is widely used in information retrieval (IR).

2) Tanimoto Coefficient: Tanimoto coefficient, also known as the Jaccard index,measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets.

X Y

Jaccard (X, Y) =X Y(2)

3) Log Likelihood

Similar to Tanimoto coefficient, the Log likelihood method calculate similarity based on the common preference two users shared. Given the total number of items and the number of each user rated items, the final result is the impossibility of that the two users have such common preference.

4) The City Block Distance

The city block distance is the sum of the lengths of the projections of the line segment between the points onto the coordinate axes.

(,) | | 0 −*n*

i i i D x y x y (3)

2. RECOMMENDATION SYSTEM OF STUDENTS JOB HUNTING(SJH)

A. Procedure of SJH Recommendation

There are four steps in our system as *Fig. 1* shows:

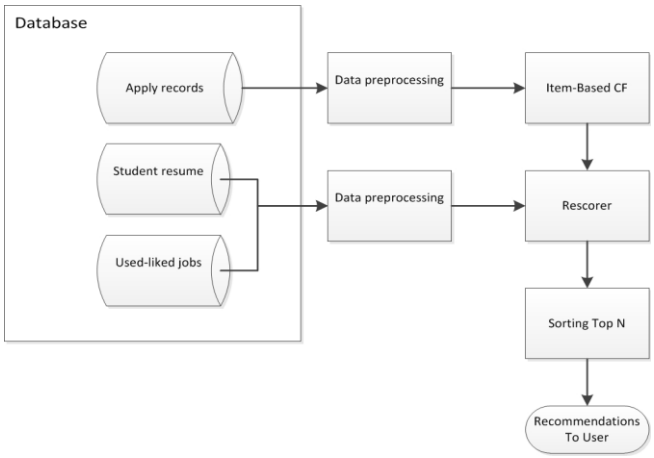


Figure. 1. Procedure of SJH recommendation

- Data Preprocessing: In this step, we clean the raw data to filter useless data including inactive users and expired recruiting information.

B. *Item-based CF Deals with Boolean Data*:

Rather than using similarities between users' rating behavior to predict preferences, item– based CF uses similarities between the rating patterns of items. Since finding similar items is easier than finding similar users, and attributes of items are more stable than users' preference, item-based methods are suitable for off-line computing. The preferred outcome of any of the two methods is to provide a suitable job.

The procedure is presented below:

for each jobi useri applied{
for each co-applied userj who applied jobi{
find out jobs that userj applied;
add these jobs to candidate set;
}
delete jobi from candidate set;
}

IV. EXPERIMENTS

In this section, user-based and item-based CF algorithms are tested on our data set respectively. Then item-based CF,the better performed one, to be applied on the Student JobHunting recommendation system. At last, we evaluate the performance of improved recommender that using used-liked job and co-apply users weights based on item-based algorithm. The implementation of our experiments is based on ApacheMahout.

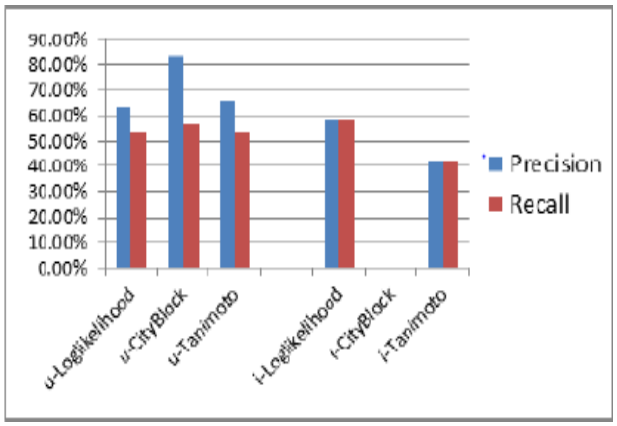


Fig. 2 shows that all three user-based algorithm and itembasedalgorithm which using Log likelihood similarityreached higher precision and recall than other two algorithms.

Under this circumstance, we continue to evaluate these four methods via some other variables, for example the number of neighborhood.

C. Evaluation Of User-based CF

1) Contrast of different similarities

Since user's preference on jobs values (0, 1) in the job apply records, and Log likelihood, City Block and Tanimotoare three methods of similarity calculation that are suitable for Boolean data, in this experiment we recommended three items to test precision and recall with these three methods. We chose different neighborhood numbers to reduce its influence.

TABLE I. PRECISION OF USER-BASED CF WITH DIFFERENTSIMILARITY METHODS

Precision	Log likelihood	City Block	Tanimoto
n=5	75.00%	94.44%	77.78%
n=10	62.82%	83.33%	65.38%
n=20	56.41%	71.79%	56.41%
n=40	56.41%	71.79%	56.41%

TABLE II. RECALL OF USER-BASED CF WITH DIFFERENT SIMILARITY

Recall	Log likelihood	City Block	Tanimoto
n=5	53.85%	56.41%	53.85%
n=10	53.85%	56.41%	53.85%
n=20	56.41%	58.97%	56.41%
n=40	56.41%	64.10%	56.41%

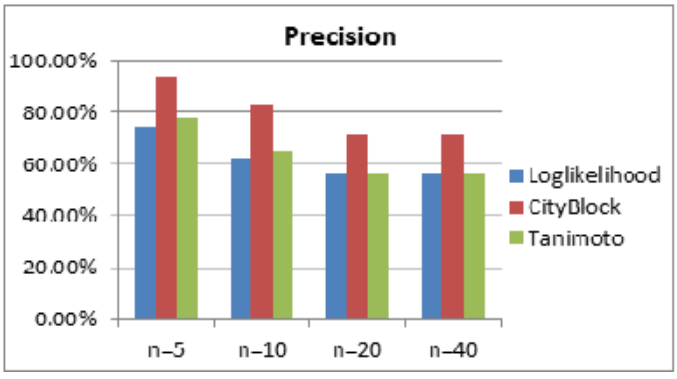


Figure. 3. Precision of user-based CF with different similarity methods

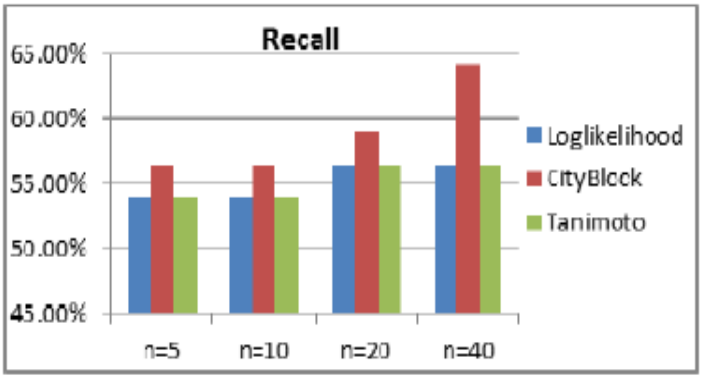


Figure. 4. Recall of user-based CF with different similarity methods

It should be noted that the *Generic Recommender IR Stats Evaluator* that Mahout offered just focuses on users who have recommended items, while automatically ignores users that cannot be recommended. So the result of evaluation seemed very good. In next section we will test the recommendation results artificially.

D. Evaluation of Item-based CF

According to the result of section IV.C, when using Log likelihood similarity method, item-based algorithm performed well. So we decided to use item-based CF and selected Log likelihood method to compute candidate items' similarities in the Student Job Hunting recommendation system.

1) The performance of improved recommender: *We evaluate the capability of original item-based recommender and the improved recommender that takes co-apply users' weight and used-liked jobs' weight into account when recommending two or three jobs for each student. The numstands for number of recommended items. The results are recorded in Table V and Table VI.*

TABLE III. PERFORMANCE OF IMPROVED RECOMMENDER IN SJHSYSTEM(R_NUM=3)

r_num=3	Precision	Recall	Reach	F1-measure
original recommender	50.62%	47.13%	93.10%	48.81%
improved recommender	51.85%	48.28%	93.10%	50.00%

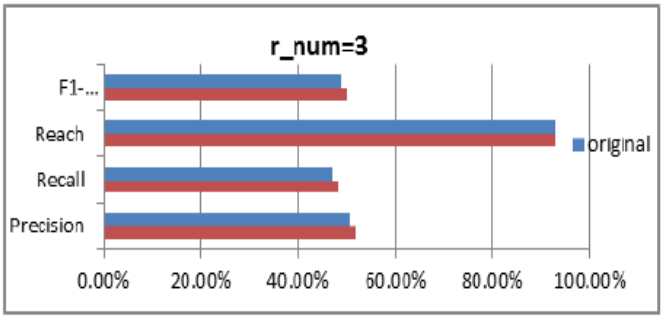
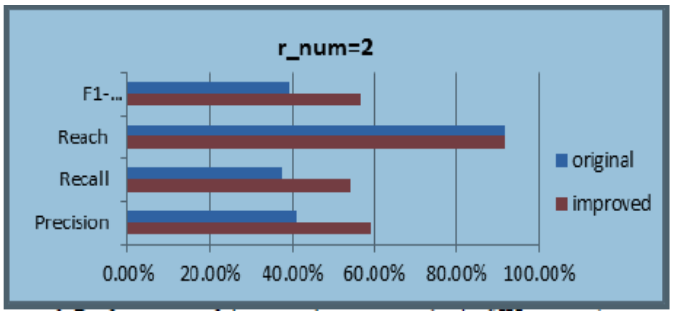


Figure. 5. Performance of improved recommender in SJH system(r_num=3)

TABLE VI. PERFORMANCE OF IMPROVED RECOMMENDER IN SJH SYSTEM (R_NUM=2)

r_num=2	Precision	Recall	Reach	F1-measure
original recommender	40.91%	37.50%	91.67%	39.13%
improved recommender	59.09%	54.17%	91.67%	56.52%



As *Fig.5* showed, when recommending three jobs for each student, the improved recommender had a little promotion at Precision, recall and F1 score. When the number of recommended item came to two, as *Fig.6* showed, all these three indicator score increased significantly. And the R*each* Rate remains as before. Because the sparseness of our apply records dataset, recommender can only offer 3 or 4recommended results for some students. To evaluate the overall recommender's performance, we considered the number of recommended items as three and two. If user U has three recommended jobs---(job 1, job 2, job 3), when recommending jobs for U, the improved recommender just re-ranking the three jobs recommended from the traditional recommender, so that it has no influence on the precision, recall and F1 score. However, when evaluating the Top 2 recommended jobs, the improved recommender changed the order of these three jobs, so that the top 2 jobs are different from former ones. The increased scores suggest that the improved recommender works well for the reason that jobs take precedence (1st and 2nd recommended jobs) are better than latter ones (3rd recommended job).

V. CONCLUSION

On the basis of this study and various techniques to research and after implementation of algorithms the CF based algorithm for its better performance and overall factors. Of course a lot of improvement and hybrid algorithms need to be implemented alongside CF algorithm. To further optimize the recommendation system, and integrate the system for better performance we keep in check the sparsity of user profile and use some methods of filling user's preference matrix can be utilized.

VI.REFERENCES

[1] Pazzani M J, Billsus D. Content-based recommendation systems [M]//The adaptive web. Springer Berlin Heidelberg, 2007: 325-341.

[2] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions [J]. Knowledge and Data Engineering, IEEE Transactions on, 2005, 17(6): 734-749.I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[3] Schafer J B, Frankowski D, Herlocker J, et al. Collaborative filtering recommender systems[M]//The adaptive web. Springer Berlin Heidelberg, 2007: 291-324.

[4] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C]//Proceedings of the 10th international conference on World Wide Web. ACM, 2001: 285-295.