Recommendation System

Manini Patel (1644004)
MTech(CSE), 2nd sem, SEAS, AU
Subject : Algorithms And Optimization For Big Data

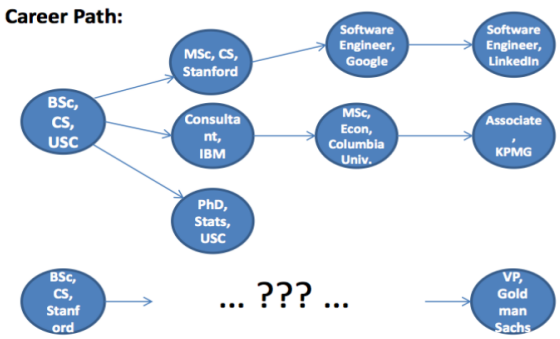
***Abstract*—Until recently job seeking has been a tricky, tedious and time consuming process, because people looking for a new position had to collect information from many different sources. Recommender system technology aims to help users in finding items that match their preferences; it has a successful usage in a wide-range of applications to deal with problems related to information overload efficiently. Good match between persons and jobs needs to consider both, the preferences of the recruiter and the preferences of the candidate. Here, we are suggesting a career path of a person to acquire his/her goal by making use of the data from LinkedIn users' public profiles. Two ways are presented here for suggesting a career path, one in which we read person's profile and suggest a path and in other one user enters his/her career goal and based on this we suggest a career path.**

Keywords: Recommendation system, Collaborative filtering, Skill recommendation, Career path suggestion

I. INTRODUCTION

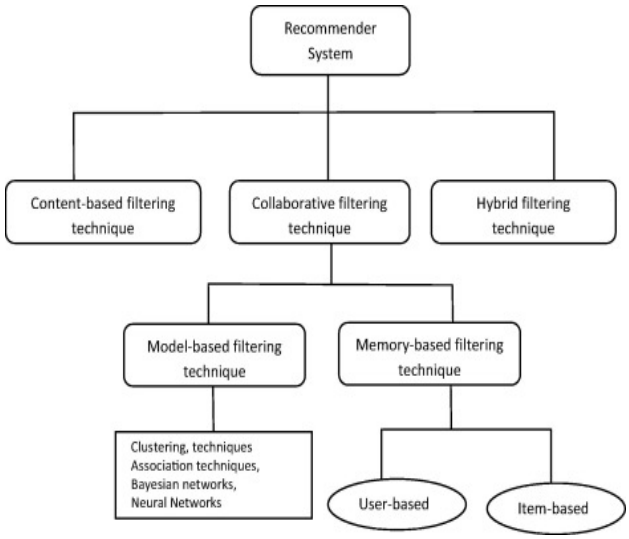
An online community consists of members sharing common interests and purposes administered through guidelines and policies within a computer system. Online community life has increasingly become a significant part of our social lives and has become a new channel through which organisations can connect with stakeholders, including job candidates. As increasing numbers of employers utilise these platforms to screen job candidates, job candidates are increasingly presenting themselves in online communities to impress employers. Membership in the LinkedIn (www.linkedin.com) online community has grown exponentially. The University of Massachusetts at Dartmouth released a study finding that 81% of Inc. 500 companies use LinkedIn for talent acquisition. LinkedIn is perhaps the most successful and widely used social networking site (SNS) for recruiters and job seekers and is the world's largest professional network on the Internet.

With the increasing volume of information available online, recommender systems have become a daily tool for Internet users, providing them with desirable help in finding information. The recommender systems which is a subclass of information filtering system, used to determine the interested items for a specific user by employing a variety of information resources that related to users and items.They are widely used by e-commerce sites, like LinkedIn that gives recommendation to the users for possible connections that they can add to their circles.Following figure gives a graphical representation of the modeling of peoples career paths.



II. RECOMMENDATION FILTERING TECHNIQUES

The use of efficient and accurate recommendation techniques is very important for a system that will provide good and useful recommendation to its individual users. Following shows the anatomy of different recommendation filtering techniques.



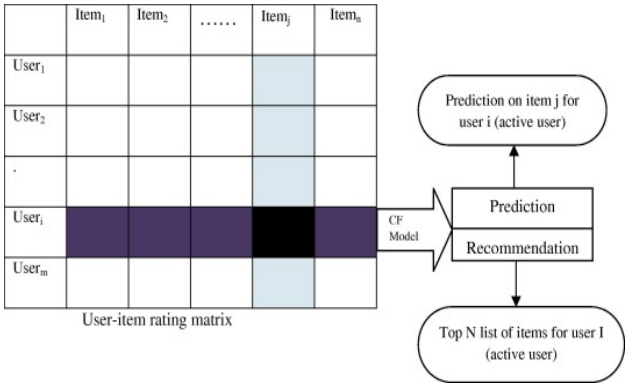
A. *Content-based Filtering*

Content-based technique is a domain-dependent algorithm and it emphasizes more on the analysis of the attributes of items in order to generate predictions. When documents such as web pages, publications and news are to be recommended, content-based filtering technique is the most successful. In content-based filtering technique, recommendation is made based on the user profiles using features extracted from the content of the items the user has evaluated in the past. Items that are mostly related to the positively rated items

are recommended to the user. CBF uses different types of models to find similarity between documents in order to generate meaningful recommendations.These techniques make recommendations by learning the underlying model with either statistical analysis or machine learning techniques. Content-based filtering technique does not need the profile of other users since they do not influence recommendation.[2][7]

B. Collaborative Filtering

Collaborative filtering is a domain-independent prediction technique for content that cannot easily and adequately be described by metadata such as movies and music. Collaborative filtering technique works by building a database (user-item matrix) of preferences for items by users. It then matches users with relevant interest and preferences by calculating similarities between their profiles to make recommendations. Such users build a group called neighborhood. An user gets recommendations to those items that he has not rated before but that were already positively rated by users in his neighborhood. Recommendations that are produced by CF can be of either prediction or recommendation. Prediction is a numerical value, R_{ij}, expressing the predicted score of item j for the user i, while Recommendation is a list of top N items that the user will like the most as shown in below Figure.[2][7][8]



C. Hybrid filtering

Hybrid filtering technique combines different recommendation techniques in order to gain better system optimization to avoid some limitations and problems of pure recommendation systems. The idea behind hybrid techniques is that a combination of algorithms will provide more accurate and effective recommendations than a single algorithm as the disadvantages of one algorithm can be overcome by another algorithm. Using multiple recommendation techniques can suppress the weaknesses of an individual technique in a combined model. The combination of approaches can be done in any of the following ways: separate implementation of algorithms and combining the result, utilizing some content-based filtering in collaborative approach, utilizing some collaborative filtering in content-based approach, creating a unified recommendation system that brings together both approaches.[2][7]

III. DATA PRE-PROCESSING

The raw data from LinkedIn is represented in natural language format and it was collected in JSON format, and this added to the task of data cleaning an additional level of processing requirement. JSON parsing is done over data and then they are converted into csv files that can be read by excel. During pre-processing,the data contained many non-ASCII characters like bullet points, right arrow and other garbage value so that data was removed.

IV. APPROACH FOR RECOMMENDATION SYSTEM

There are basically two approaches for Collaborative filtering:

1. Frequency approach : This first sees what career goal is right for the user or what the user wants then makes an exhaustive list of all the skills that users have in that profession after that make a matrix S of each user and their skills for that profession. The matrix S is then summed along the column to find how many users have a particular skill and then the top 3 frequencies are taken. All skills having the above frequency are shown to the user.

2. Jaccard approach : This first sees what career goal is right for the user or what the user wants. It then makes an exhaustive list of all the skills that users have in that profession. After that make a matrix S of each user and their skills for that profession and then make a matrix T of pairs of skills and how many time they occur together. After this the Jaccard Index of all pairs of skills is calculated and stored in a matrix J.[9]

V. IMPLEMENTATIONS AND RESULTS

In module-1 career is selected randomly and the user to whom career path is to be suggested is too. In modele-2 user is asked to select a Career from the list of options and then the suggested skill set is shown to the user.

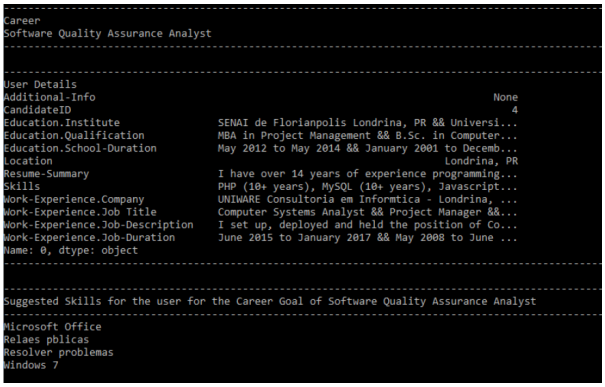


Fig : Suggesting Career Path using Frequency Approach

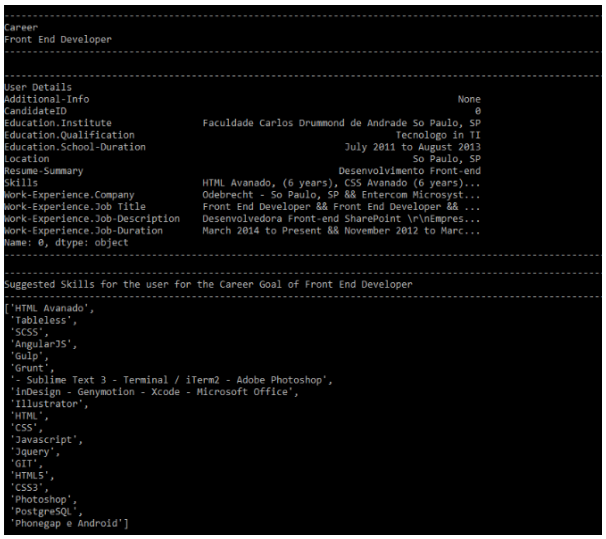


Fig : Suggesting Career Path using Jaccard Indices Approach

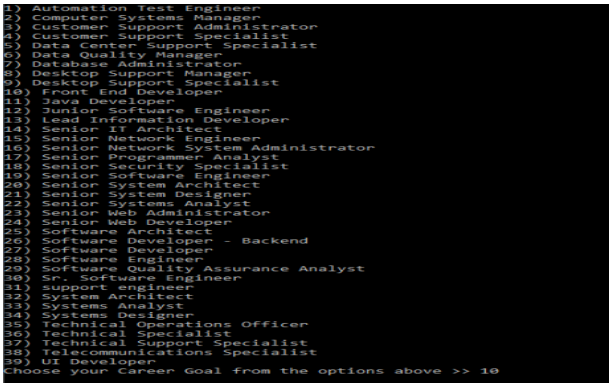


Fig : Giving options to the user for Career Path

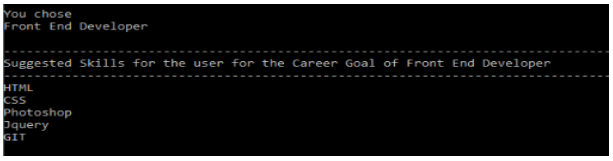


Fig : Suggesting Career Path using Frequency Approach

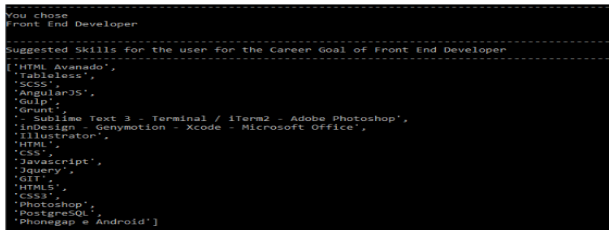


Fig : Suggesting Career Path using Jaccard Indices Approach

The Frequency Approach, although has a time complexity $O(n^2)$ it's suggestions may not be as beneficial as that of the Jaccard Indices Method, which has a time complexity of $O(mn^2)$ where m is the number of users and n is the number of skills. This is true because the second method takes into account as to which skills are most frequently seen together and hence, giving a better and more connected career path to the user.

CONCLUSION

The exponential increase in information (the so-called "Big Data paradigm") has caused difficulties in searching for desirable information and in addressing the increase in content. Therefore, the necessity of developing scalable recommender systems is on the rise. Memory-based collaborative recommendation engine algorithms can be a pretty powerful thing. The one we experimented with in python may be primitive, but its also simple to understand, and simple to build. It may be far from perfect, but robust implementations of recommendation engines, such as Recommendable, are built on similar fundamental ideas.

REFERENCES

[1] https://www.analytics.com/blog/2016/06/quick-guide-build-recommendation-engine-python
[2] http://www.sciencedirect.com/science/article/pii/S1110866515000341
[3] https://www.quora.com/Which-algorithms-are-used-in-recommender-systems
[4] https://www.dataquest.io/blog/pandas-python-tutorial/
[5] https://www.codementor.io/jadianes/build-data-products-django-machine-learning-clustering-user-preferences-du107s5mk
[6] http://www.inf.ed.ac.uk/teaching/courses/inf2b/learnnotes/inf2b-learn-note02-2up.pdf
[7] https://en.wikipedia.org/wiki/Recommender system
[8] https://en.wikipedia.org/wiki/Collaborative filtering
[9] https://en.wikipedia.org/wiki/Jaccard index