

Recommender System For Users

Jayveersinh Gadhavi - 1644003, Sem 2

MTech in CSE with specialisation in Data science and Analytics

School of Engineering and Applied Science-AU

Abstract—Websites such as LinkedIn are now a days in huge demand for the development of one's Professional Network. Most of the people are creating their accounts in such systems in which they can post their CVs which is always helpfull in getting jobs. Many people are lacking behind in terms of skills for getting good jobs, as they do not have proper direction that in which areas they can develop their skills and what will be the career goals after acquiring the skills. Can we help them out? Can we give them some suggestions? The answer is Recommender System. Thus basic idea of the report is to design such a Recommender System in which following two modules must be covered. 1- A module that reads users profile and suggest a career path in terms of skillset – to be acquired. 2- A module in which user enters a career goal and based on this career goal and other related information the platform suggest a career path. Algorithms and results will be shown.

I. INTRODUCTION

What is Recommender System: A recommender system or a recommendation system is a subclass of information filtering system that seeks to predict the "rating" or "preference" that a user would give to an item. This System is used in linkedin,facebook and there are also recommender systems for experts, collaborators, jokes, restaurants, garments, financial services, life insurance, romantic partners and Twitter pages(From Wikipedia).

Personalization system such as recommender system attracted the interest of many researcher and practitioners. Many techniques for suggestion of career path and job recommendation have been developed and applied. These includes the one of the possible approach for career path recommendation system.

a) *Algorithms:* Algorithms that are used for implementing the Recommender systems are collaborative-based filtering, content-based filtering and Hybrid Recommender system approach.

A. Collaborative-Based Algorithm

Collaborative filtering methods are based on collecting and analyzing a large amount of information on users behaviors, activities or preferences and predicting what users will like based on their similarity to other users. A key advantage of the collaborative filtering approach is that it does not rely on machine analyzable content and therefore it is capable of accurately recommending complex items such as movies without requiring an "understanding" of the item itself. Many algorithms have been used in measuring user similarity or item similarity in recommender systems. For example, the k-nearest

neighbor (k-NN) approach and the Pearson Correlation as first implemented by Allen. Collaborative filtering is based on the assumption that people who agreed in the past will agree in the future, and that they will like similar kinds of items as they liked in the past.(From Wikipedia)

B. Content-Based Algorithm

Content-based filtering methods are based on a description of the item and a profile of the users preference. In a content-based recommender system, keywords are used to describe the items and a user profile is built to indicate the type of item this user likes. In other words, these algorithms try to recommend items that are similar to those that a user liked in the past or is examining in the present. In particular, various candidate items are compared with items previously rated by the user and the best-matching items are recommended. This approach has its roots in information retrieval and information filtering research.(From Wikipedia)

C. Hybrid Recommender System

Hybrid approaches can be implemented in several ways by making content-based and collaborative-based predictions separately and then combining them; by adding content-based capabilities to a collaborative-based approach (and vice versa), or by unifying the approaches into one model. Several studies empirically compare the performance of the hybrid with the pure collaborative and content-based methods and demonstrate that the hybrid methods can provide more accurate recommendations than pure approaches. These methods can also be used to overcome some of the common problems in recommender systems such as cold start and the sparsity problem.(From Wikipedia)

II. PREPROCESSING OF DATA

Database provided is in JSON format, thus JSON parsing is done over data and then they are converted into a nice csv files that can be read by excel. Cleaning of data must be Done before processing of the Data. Data cleansing, data cleaning, or data scrubbing is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data. Cleaning of Data AND processed Data is already been uploaded in the GitHub repo.

III. NEW SOLUTIONS

There are plenty of ideas to come up with the solution for recommender system. Thus the basic idea for this report is as follows.

A. Matching and Categorising/clustering by using 1st,2nd and 3rd Degree connection

The terms 1st, 2nd and 3rd degree came from the LinkedIn Suggestions for Users. 1st Degree is direct connection, 2nd Degree is the connection of connection, and 3rd degree is Unknown connection.

a) *For Module 1:* Categories can be form by analysing Skills of all connections. For Example, Suppose some people are preffering data analytics then they will form a cluster. Thus the skills of users belong to such a category are scanned and thus Suggestions from the users having high skills and high ranked profile will be given to the users who are lacking in the skills. In this way Recommender can work by clustering method for suggesting the skillsets to be aquired.

b) *For Module 2:* Same procedure of categorising is carried out as above. Thus a list of all the positions or jobs will be listed by analysing the posts of all users in same category. Indexing will be done among them and top career goals will be listed infront of user. Thus the user can select the suggested career goals and can futher work on them.

B. Collaberative Filtering using Jaccardian indices Approach

First of all what career goal is right for the user or what the user wants is being analysed. It then makes a list of all the skills that users have in that profession. Then we form a matrix S of each user and their skills for that Career. Formation of Matrix T of pairs of skills and how many time they occur together will be done. After this the Jaccard Index of all pairs of skills is calculated and stored in a matrix J where $J(i, j)$ is the Jaccard Index of Skills i and j. The Jaccard Index of two values are calculated as follows:

$$J = \frac{AB}{A + B - AB} \quad (1)$$

where, A: Number of times that A occurs B: Number of times that B occurs AB: Number of times that AB occur together. Then the Skills having the top 3 Jaccard Indices are suggested to the user Thus this idea is almost similer to previously discussed idea for module 2 of first idea but this approach can be applied on both the Modules.

IV. IMPLEMENTATIONS AND RESULTS

The colloborative Approach Using Jaccard indices approach will give the following outcomes for our recommender System. The Jaccard Indices Method, which has a time complexity shown by:

$$O(mn^2)$$

Where m is the number of users and n is the numberof skills is having a very high efficiency with good accuracy and this is true because it is giving a better and more connected career path to the user.

```
(tensorflow) F:\IET\Sem-V\AOBD\Final Exam\python -W ignore readUser2.py

Career
Front End Developer

-----
User Details
Additional Info: None
CandidateID: 0
Education.Institute: Faculdade Carlos Drummond de Andrade So Paulo, SP
Education.Qualification: Tecnologo in TI
Education.School.Duration: July 2011 to August 2013
Location: So Paulo, SP
Resume.Summary: Desenvolvimento Front-end
Skills: HTML Avanado, (6 years), CSS Avanado (6 years)...
Work-Experience.Company: Odebrecht - So Paulo, SP & Entiercom Microsyst...
Work-Experience.Job.Title: Front End Developer && Front End Developer && ...
Work-Experience.Job.Description: Desenvolvedora Front-end SharePoint Vintapres...
Work-Experience.Job.Duration: March 2014 to Present && November 2012 to Marc...
Name: 0, dtype: object

-----
Suggested skills for the user for the Career Goal of Front End Developer
-----
['Tableless',
 'SCSS',
 'AngularJS',
 'Gulp',
 'Grunt',
 'Sublime Text 3 - Terminal / iTerm2 - Adobe Photoshop',
 'InDesign - Gensymotion - Xcode - Microsoft Office',
 'Illustrator',
 'HTML',
 'CSS',
 'Javascript',
 'jQuery',
 'GIT',
 'HTML5',
 'CSS3',
 'Photoshop',
 'PostgreSQL',
 'Phonegap e Android']

Time taken: 0.25881771861215463
```

Fig. 1. Showing Suggested skills for the User

```
You chose
Front End Developer

-----
Suggested Skills for the user for the Career Goal of Front End Developer
-----
['HTML Avanado',
 'Tableless',
 'SCSS',
 'AngularJS',
 'Gulp',
 'Grunt',
 'Sublime Text 3 - Terminal / iTerm2 - Adobe Photoshop',
 'InDesign - Gensymotion - Xcode - Microsoft Office',
 'Illustrator',
 'HTML',
 'CSS',
 'Javascript',
 'jQuery',
 'GIT',
 'HTML5',
 'CSS3',
 'Photoshop',
 'PostgreSQL',
 'Phonegap e Android']

Time taken: 0.39154865733425415
```

Fig. 2. Showing Suggested Career paths for User

a) *For Module 1:* Figure 1 is a Result of implementation of the Jaccard indices method for suggesting skills to the User.

b) *For Module 2:* Figure 2 is a Result of implementation of the Jaccard indices method for suggesting career goals to the user after choosing the career.

Github Code Details the required files such as Input data,processed data, data cleaning algorithms, Jaccard indices algorithm and results are beang uploaded in AOBD_Exam_Submission Folder in repository AOBD17_1644003. Following is the link: https://github.com/jv-wave/AOBD17_1644003

V. CONCLUSION

The Implementation of both the modules are done using Python language with the Jaccard Indexing approach. Thus the concept of Recommender System is in demand and it is a very Open and fundamental issue for the IT Society.

REFERENCES

- 1) <https://en.wikipedia.org/wiki/LinkedIn>
- 2) https://en.wikipedia.org/wiki/Data_cleansing
- 3) https://en.wikipedia.org/wiki/Recommender_system
- 4) https://en.wikipedia.org/wiki/Collaborative_filtering
- 5) https://en.wikipedia.org/wiki/Jaccard_Index
- 6) <https://www.quora.com/Which-algorithms-are-used-in-recommender-systems>
- 7) <https://www.analytics.com/blog/2016/06/quick-guide-build-recommendation-engine-python>
- 8) <http://www.inf.ed.ac.uk/teaching/courses/inf2b/learnnotes/inf2b-learn-note02-2up.pdf>