Algorithms and Optimization of Big Data

End Term Project

Jainil Vachhani
School of Engineering and Applied Sciences
Ahmedabad University
Jainilvachhani@gmail.com

Abstract— **A recommender system is a subclass of information filtering system that seeks to predict the rating or preference that a user would give an item. Recommender systems have become increasingly popular in recent years and are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general. As such, they are a billion-dollar industry. This paper discusses a simple application of recommender system that includes finding a match between person and job based on preferences of recruiter and candidate. This paper also describes a system that can recommend a job based on candidate's profile and suggest a career path, in terms of the skillset to be acquired.**

Keywords— Recommender Systems, Data Analysis, Machine Learning

I. INTRODUCTION

A Recommender System

Recommender systems typically produce a list of recommendations in one of two ways through collaborative filtering or through content based filtering. Truly powerful recommendation system algorithms are based on Machine Learning. Simple recommendation algorithms do not perform well enough, because they do not learn and adapt to user's behavior. Netflix, Amazon, Facebook, LinkedIn, YouTube are some application where this system is widely used.

B Personalized Recommender System

Personalization is a means of meeting the customer's needs more effectively and efficiently, making interactions faster and easier and, consequently, increasing customer satisfaction and the likelihood of repeat visits.

C Non Personalized Recommender System

A non-personalized recommender system is one that makes the same recommendations for everyone. The simplest example is a retailer that shows the ten most popular products on their homepage.

D Collaborative Filtering

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.

E Content Based Filtering

Another common approach when designing recommender systems is content-based filtering. Content-based filtering methods are based on a description of the item and a profile of the user's preference. n 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.

II. CAREER RECOMMENDER SYSTEM AND DATA

In this system when a user logs in, based on his interest and skills information regarding skillset or relevant career path.

Module 1: Here, system recommends necessary skills required by a user for getting the particular job. This module can be implemented using personalized as well as non-personalized system.

Module 2: In this module, a career path is recommended to the user based on his current skills. This includes current job profile as well. Personalized recommendation is used in this case

The provided dataset contains 39 different job posts along with relevant candidate information. The data was converted in csv format for cleaning and to apply other functionality. The below figure shows that data cleaning:
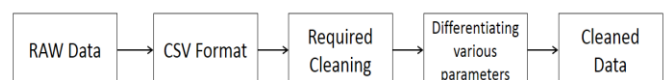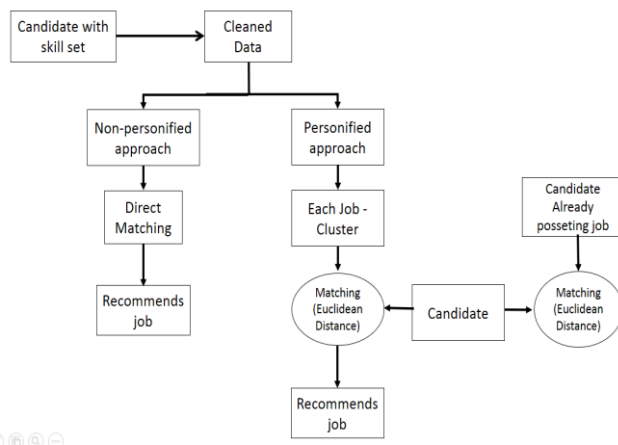
III. ALGORITHM

We use K Means clustering to provide a solution for both the modules and Euclidean distance is used to measure clustering distance.

The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. The two approaches are as follows:

- Matching of candidate to job with the candidate having the required skill set.

- The matching of candidate with other candidate with whom the given candidate has common skillset.

o The first approach let us know which job is appropriate for a given candidate.

o In the second approach, based on the profile match he will get a recommendation of another candidate.

o We have to calculate distance between a candidate and job for matching. There would be a distance assigned between each candidate and job in form of weight and candidate would get recommendation of job with minimum weight

The workflow of the algorithm is given below:

COMPARISON

- Content-based filtering:- Although this algorithm works for product recommended system it works badly in our module as it is hard to recommend job to new candidates and candidates who are willing to change their job. This algorithm would fail.

- Collaborative Filtering (CF):- Both user based and item based approach for this algorithm has three major problems that are cold start, scalability and sparsity.

CONCLUSION

The proposed algorithm works efficiently under given scenario and dataset. There is may or may not work efficiently under other parameters or different dataset. Also, it takes huge computational time to implement on large dataset.

[1] Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(3), 611-622

[2] Rowel Atienza, (2017, March 30) GAN by Example using Keras on Tensorflow Backend Retrieved from http://www.medium.com