Suggesting Career Path and Skills

Prerak Raja

School of Engineering and Applied Science
Ahmedabad University

***Abstract*—The user is willing to know skill set which will be required for him to go to his desired career path.In this paper, we discuss about suggesting a skill set for user to be acquired and there by also suggesting user his career path on the basis of his career goal and his other related information.We are given data set of some of the user profiles and then I am extracting some features based on the data available.K means clustering and shortest path algorithm have been applied for solving the problem.**

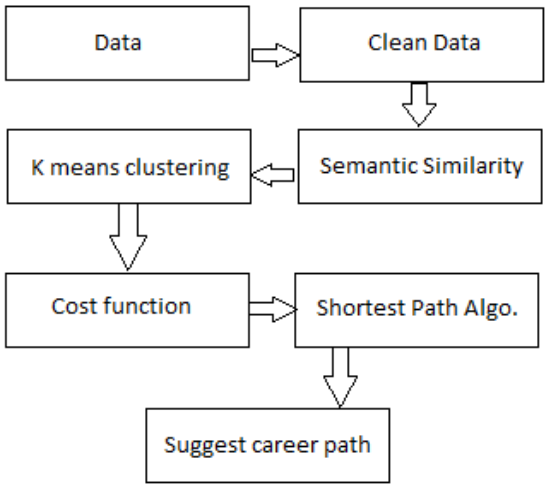
I. Introduction

It has been becoming difficult for college students to decide their career options upon graduation.Also,same situation has been arousing for current employees to decide on their own about their future goals.So,Mostly people try to take help from their colleague or their friends to suggest them their career path or they end up doing whatever everyone is doing.So,the best way to overcome this problem is by suggesting a career path by using our algorithms,so that one can achieve their goals easily.

First,persons career path is modeled according to his work experience in companies based on time.Each experience of paricular person is represented as node.There are multiple nodes presented in one particular user profile when we start reading his profile.Here each node is our feature.Now,when person has to achieve a certain goal we have many career paths available but we choose best path to reach his goal .We choose path based on the highest probability of achieving a goal.

II. Dataset

We are given dataset of several users based on the different positions.The data consist of:ID,Additional Info,skills,Work Experience,Education,Job Duration,Qualification,etc.The first thing we need to do is getting the information regarding work experience.For university degree hand rules are applied to represent them into standard format.For positions,people have habit of using their own names so for that I have used k means clustering.K means clustering is used mainly for solving the problem of positions that user writes in their profiles.For clustering, I have calculated distance between two positions by use of average semantic similarity.Hence, after clustering similar positions are grouped together so as to interpretation of different user can be made similar.



III. Approach

Data Clean: I have done some of the data cleaning manually and some of remaining part in python using inbuilt libraries.Data cleaning was required to easily fetch results from csv file and extract useful features.

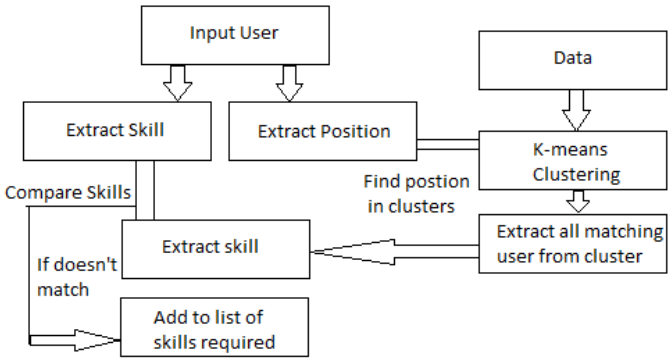
Semantic Similarity:This is one of the most useful part which tends to lessen the burden while generating output.User has tendency to write about their skills in their own way.If we take example,one user might be software developer but he may write his skill in profile as programmer,developer,software engineer.Hence,I took the similar words and counted them as same so that they can be categorized in one common position that is let say software developer.

K means Clustering: In my approach role of k means clustering was to separate different career positions into different clusters such as there are clusters made for software engineer,marketing,consultant,etc. The main benefit of this is if we know some skill set of the given user then we just have to check skill set required or career position required in that particular domain of interest(cluster).

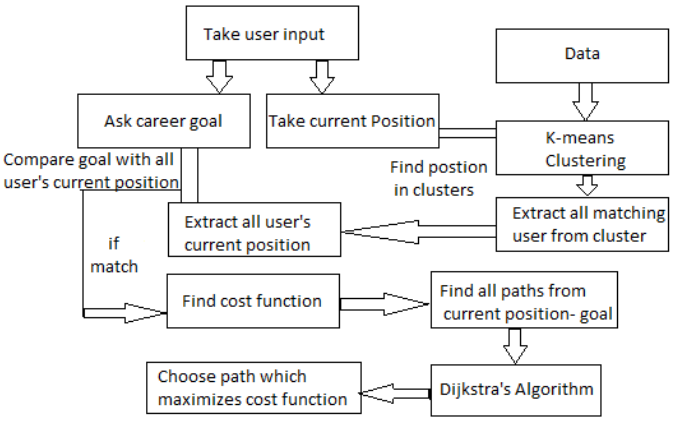
Cost function:We create graph of individual user in which node represent user's position(software developer,Sr. software developer,CEO) and edges represent transition from one position to another.For given user, we have different nodes(positions he has been to).For suggesting career path,We

ask user to input his goal(g).We already have user's current position(c) and now we have to reach to desired goal.For that,I have defined cost function which calculates probability of x/y.Here,x represents number of times transition occurs from c to j in our database and y represents summation over number of times transition occurs from c to all different position users can achieve.We will end up with different paths available for user to achieve his goal.We will choose one path that will maximize the cost function.

Shortest Path Algorithm:I have used Dijkstra's algorithm for finding shortest path between nodes in a graph.User has specified his career goal.We have already defined the cost function.This cost function will act as weights of edges of graph.As mentioned earlier,We will now implement our algorithm and find the path which leads to maximized cost function.In Dijkstra's algorithm,we mainly find path with minimum cost but here in our case,I am finding maximized cost path.The reason for that is the cost function that we defined earlier is calculating probability of number of times every user is opting career goal as mentioned by the user from the given current position.So it is clear that greater the probability ,greater are the chances for user to go and take that career path.



Problem 1: Suggest skills



Problem 2: Suggest career path

IV. Problems Faced

The data was given dirty.So how to clean data and then how to extract different features from descriptive data given in profile.I faced problems in extracting skills of particular user.Also,Position titles as written by particular user is different fro others so it was difficult to group them as same.Also,finding the estimated parameters and thereby suggesting a career path was difficult task and i am still not getting desired results for that.

V. Conclusion

Raw data given to us was natural language format so it was hard to preprocess data.Without preprocessing of data,it was difficult to extract some of the features.Clustering has been applied to group similar position as 1 cluster but it will contain some error for sure.Accuracy for clustering depends on how clean is the data and how well we have done semantic analysis.For generating career path for user the algorithm might get confused or may give results with errors if user inputs a goal which is not relevant to his acquired skills because we have clustered position so suggesting path for goal present in different cluster is not possible.

References

[1] https://www.3pillarglobal.com/insights/measures-of-semantic-similarity
[2] https://www.analyticsvidhya.com/blog/2014/09/data-munging-python-using-pandas-baby-steps-python/
[3] http://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html
[4] https://en.wikipedia.org/wiki/Dijkstra