Career Path Recommendation System

Rajvi Prajapati - 1401033
School Of Engineering & Applied Sciences, Ahmedabad University.

***Abstract*—In the recent years the Web has undergone a tremendous growth in terms of both content and number of job-seekers. This has led to a serious problem of information overloading in which it is becoming difficult for the job-seekers to locate the authentic information from different sources in the given time. Recommender systems have been developed to address this problem,by predicting a user's preference or his/her similarity with other group of users.Our aim through this exam is to find out relationships between jobs and people skills making use of data from LinkedIn user's public profiles.**

***Keywords*|– Quadtree, Spatital Autocorrelation, Geary's Index, Collaborative Filtering, Pearson's co-efficient**

I. INTRODUCTION

In job recommendation systems, when a user logs onto the platform, the platform reads user's profile which contains various parameters like CandidateID, Skills, Work-Experience, Preferred job location, Education etc. In my simple and intuitive model, I have considered two parameters: Skills and Location. Based on these details, each user will get relevent suggestions for set of skills to be acquired and job recommendations.

The motivation behind selecting these criteria is that 25% of all searches for jobs indeed specify only a location and not keywords. Also,there are many job seekers who aren't sure about which additional skills would be required in their jobs.

II. PROPOSED ALGORITHMS

A. Module-1

Here, Our primary focus will be on the preferred job location of the user. We can use latitude and longitude in order to identify any candidate's location. A data-structure called Quadtrees are used to partition a two dimensional space by recursively subdividing it into four quadrants or regions. We will provide recommendation to users in those regions.

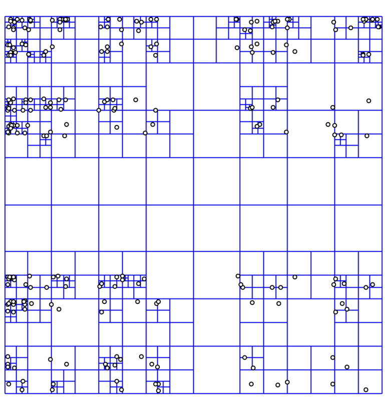

Figure 1. :Recursively sub-division into four sub-regions with point data

Now, we want to make clusters of candidates with similar skills in same neighbourhood. For this, Spatial Autocorrelation is a statistical measure which examines the spatial ordering of geographic data. It will deal with both skills and locations of spatial features. Spatial autocorrelation will determine whether adjacent or neighbouring values in the geographic data vary together, and if so, how. So, Geary's index is applied to measure correlation among the candidates in the regions resulting from Quadtree decomposition.

The index measures the similarity of i^{th} candidate and j^{th} candidate's skills, c_{ij}, which can be calculated as follows:

$$c_{ij} = (z_i – z_j)^2$$

where z_i and z_j are the values of the skills for candidate i and j. A locational similarity w_{ij} was used by Geary, and w_{ij}= 1 if i and j candidate shares a common boundary and w_{ij}= 0 if not. Geary's index is expressed as follows:

$$c = \frac{\sum_i \sum_j w_{ij} C_{ij}}{2 \sum_i \sum_j w_{ij} \sigma^2}$$

where σ^2 is the variance of the attribute z values.

- If the value of c = 1, the skills of candidates are distributed independently of location.
- If c <1, similar skills coincide with similar preferred job locations.
- If c >1, skills and corresponding job locations are dissimilar.

Thus, we will try to find the presense of correlation in the region and then recommend set of skills with a view that suggested skills will be useful to our user.

Algorithm 1 Module-1

Represent location of each user as coordinates (longitude-latitude).
for *Each region* **do**
  Find the spatial correlation of the entire region.
  if *Correlation lies within a specified range* **then**
    Users share similar preferences(in terms of skills).
  else
    Split the region into four regions depending upon the latitude and longitude.
  end
end

The output of the Quadtree Decomposition algorithm is that every user is assigned a particular region based on the correlation.i.e. every users will have suggested jobs based on preferable locations in terms of skillset.

B. An approach to Module-2

Here, we want to build a recommendation system in which we give a career goal as an input and get related skills as an output. Algorithm for the same is given below:

Algorithm 2 Module-2

Step1: Select a career goal for recommendation.
Step2: Identify all the locations (longitude and latitude) with the jobs relevent to career goal.
Step3: Map the user in the exact region (node) of the Quadtree according to his/her job locations.
Step4: Find a subset of users in the region who share similar skills with the active user. Select top-10 similar users.
Step5: Select top 5 relevent skills from each of these top-10 users to form a top set of 50 skills.
Step6: Recommend top-10 skills from the top set with highest correlation in the region.

Below is the block-diagram to execute above algorithm:

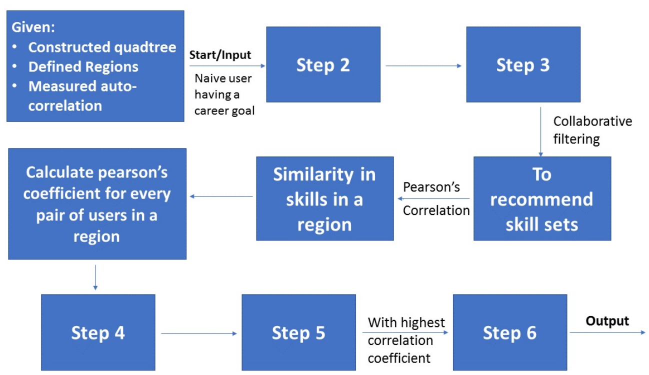

Figure 2. Block Diagram of algorithm for Module-2

Here, equation for pearson's correlation can be given as

$$C_{a,b} = \frac{\sum_{i=1}^{m} (r_{a,i} - \hat{r}_a)(r_{b,i} - \hat{r}_b)}{\sqrt{\sum_{i=1}^{m} (r_{a,i} - \hat{r}_a)^2 \sum_{i=1}^{m} (r_{b,i} - \hat{r}_b)^2}}$$

Figure 3. Equation for Pearson's Co-efficient

where $r_{a,i}$ and $r_{b,i}$ are frequency of job skills required for any career path.

III. Results and Interpretations

A. Complexity

For **data processing**, Let N be the size of the data, m be the total length of dictionary and U be the total number of users. we are mapping user's profile to our dictionary. So, over all space complexity is $\mathcal{O}(NmU)$

In **module-1**,we are partitioning a region with n users into k partitions with nearly equal sizes. Overall time required for collaborative filtering is proportional to $k.(\frac{n}{k})^2 = \frac{n^2}{k}$.

For **module-2**, let k be the number of skills with user and N be the additional skills to be acquired by user per career path. So,complexity becomes $\mathcal{O}(kN)$ in this case.

B. Proof of Correctness

As we saw above, spatial autocorrelation statistics allows us to measure interdependence for in spatial distribution and to use statistical methods to test hypotheses about spatial interdependence.

C. Output

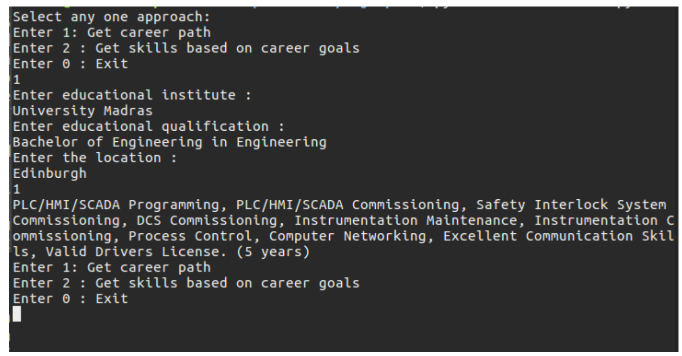

Figure 4. Output of Module-1

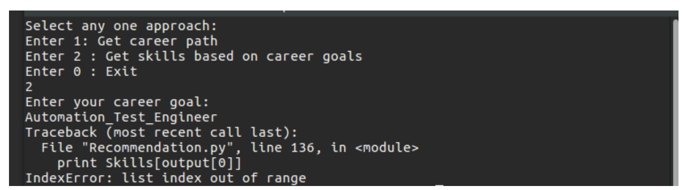

Figure 5. Output for Module-2

IV. Conclusion and Future work

From above experiment, we can summarize that user location has proved to be an important aspect. By integrating a mechanism work for adding location with skill set into the recommender system, it has been possible to produce recommendations that better suits the users preferences. Instead of using quadtree the usage of binary tree might help in improving the split in the future. This is because using a quadtree forces us to split into four equal parts while using a binary tree allows us to divide the region depending upon the data points.

References

[1] http://recommender-systems.org/collaborative-filtering/
[2] Musale, D. V., Nagpure, M. K., Patil, K. S., Sayyed, R. F. (2016). Job Recommendation System Using Profile Matching And Web-Crawling. INTERNATIONAL JOURNAL, 1(2)
[3] https://en.wikipedia.org/wiki/Spatial_analysis