# Summary of Applying Data Mining Techniques in Job Recommender System for Considering Candidate Job Preferences

# -Gupta, Anika and Dr. Garg, Deepak

## Summary

Recommender systems are an important aspect of ecommerce. The recommendation system worked on in this paper is about recommending the best jobs based on salary and candidate’s skills. Candidates can have similar looking profiles, but their choices may be different based on what position to choose, and what can they trade-off with – distance, salary, company rating. This recommender system takes Candidate information and applies classification rules along with their job preferences to produce recommended jobs list. The authors of this paper have specified “…the efforts have been put to judge the gap between the candidate’s choices, belonging to different groups, regarding the selection of different offered jobs. In this paper, it is tried to foresee the customer preferences regarding the jobs on four basic parameters of company preferences, position offered, pay-scale offered and job location…”[Page 2].

The authors have assumed that all the text categorization is in place and discrete with well labeled values, understandable by the system, and that the candidate or the system automatically updates the information dynamically. They have considered company group level, position offered, pay-scale offered and job location for recommending the best jobs for the candidates. They have categorized the company into groups instead of using continuous numerical rating. Decision Trees are used for determining the criteria for attribute splitting. A threshold taken for the sample size determined if the value was selected or not, and that was used to fill the job category with 1 or 0, depending on if the rule exists or not.

The entire process had multiple steps. First the jobs was shortlisted based on candidates eligibility, then using the decision tree inductive rule the job was initially categorized and normalized the weights. The final weights were generated for the jobs that were shortlisted based on the weight. The results are presented to he user in sorted order. The second part of the research was based on the candidate specialized job recommendation, where it takes the preference of the candidate into consideration. The same processes are repeated but a preference matrix is added.

They used a content based recommender system, Collaborative Filtering Based Recommender, and Knowledge Based Recommender. If two companies of the same category had same offers, but differences in salary, then these lead to a mismatch and unevenness of the normal trend. Thus there were a few trade-offs based on the certain aspects, like – position vs salary, company rating vs salary and position. According to the author, there were limitations in the recommendation when the knowledge base couldn’t form a preference matrix based on the candidate’s information, the entire model was reduced to only a Content Based Recommender. But if the candidate applies for a minimum of 10 jobs, then this limitation is rectified to some extent and the job preferences and efficiency was regained.

## Assumptions

There were a few assumptions made by the authors in this model. They are as follows:

* It is assumed that all the text categorization is already in place and we have discrete and well labelled values that are easily understandable by the system.
* It is assumed that either the candidate or the system itself updates the age, experience, education and skills field. Either the candidate can be alerted periodically for updating of these fields else automatic updating can be done with the help of Dynamic modification and extraction method used in. It uses TF-IDF value for feature extraction and information gain with threshold value for feature addition.
* Location preference matrix is not considered as it is assumed that the group has applied equally in all 4 regions and hence adding these will not result into any new information.

## Solution Approach

**Data Collection:** The data has been taken from the e-commerce websites which list the jobs and the requirement of the companies.

**Feature Selection:** Two major categories were chosen, Candidate and Job. For the candidate behaviour – age, gender, marital status, education, grade, major, experience, skills, current location, current employment status – were taken into consideration as features, and for the Job – required qualification and experience, skills required, employer, Field of industry, position offered, pay scale, location – were taken as features.

**Data Categorization:** The candidate data and the job data was categorized into different groups for finding the candidate’s behaviour belonging to a particular group for selecting a particular job on the basis of 4 different parameters – Company group, position offered, pay-scale, and location of the job.

**Model Selection:** The author’s proposed system uses CBR(Content Based Recommender), Model based CFR(Collaborative Filtering-based Recommender), and KBR(Knowledge Based Recommender).

**Mining of Decision Tree Induction Rules:** The evaluation criterion taken for determining the strength of the rules were lift, confidence and sample size. Lift determined the importance of the rule and confidence is for the reliability of the rule. The authors said “…A threshold value for the sample size was taken. If there exists a rule which has high lift and confidence but its sample size is less than the threshold value, then it is not taken into consideration, however strong the rule may be. All the rules that crossed the selection criteria were enlisted and checked for redundancy. After that, a common matrix representing all the job categories preferences, for a particular rule was made. In this matrix, corresponding to a particular rule, if the rule exists for a particular job category, then the corresponding field of the job category is made 1 else 0…” Then four preference matrices were generated each preserving the generated rules and the weights were normalized.

**Recommendation Generation:** The following steps were followed by the author in recommendation generation:

* Shortlisting of jobs for which the candidate is currently eligible for based on Major, Experience and Qualification, and remove any redundant values
* Calculating content based similarity for the short listed jobs for the candidate. Cosine similarity was used for calculating the similarity. A preference matrix vector is created with the weights assigned based on the skill requirement.
* Decision Tree Induction Rule is applied to the candidate’s category. First a basic categorization is done, after which a preference based categorization is done. Proper weights are assigned to the jobs by judging the general behaviour of each candidate and then normalize the weights.
* The final weights are generated by summing of the cosine similarities and the weight.
* Sorting the jobs in descending order of the score it obtained. This is the final recommendation of the jobs for the candidate.

## Relevance to our Project

The way the data was categorized and the algorithms used will be useful in our project. This paper was recommending jobs, and categorized continuous variables into discrete categories which can reduce the number of values by a large extent and thus improve model performance in speed and accuracy. The use of Decision Trees can be used as a classifier on certain aspects of the problem we are trying to solve. The categorization of companies based on their requirement can help us in categorizing the prediction of salary. The working of a recommender system proposed provides some insights on the initial work to be put into the dataset before producing the model.

**-Afzal Mukhtar – PES2201800675**