University Recommendation System for MS

Mirgane Sneha1 Jambhale Priya2 Bitake Shubhangi3 Ivare Priyanka4
1,2,3,4SVPM's COE Malegaon(BK), Baramati

Abstract— Recommendation System is a subclass of information filtering system which takes input from students and provides Students with the most suitable output to fulfill his re-quirements. This system is used to present a new college admission system using data mining and machine learning techniques for tackling college admission prediction problems. This System uses content-based filtering and collaborative Filtering to provide aspiring students (Master of Science) with the most appropriate choices of colleges based on different parameters. The system analyses the student academics, merit, background, student records and the college admission criteria. Then it predicts the likelihood of colleges the student may enter. In addition to the high prediction accuracy is an advantage, as the system can predict suitable colleges that match the student's profiles and the suitable track channels through which the students are advised to enter.

Key words: Collaborative-based Filtering (HOSVD), Content-based Filtering (LSH), SOP, Recommendation System

I. INTRODUCTION

Every day an individual spends about half of their time searching for relevant information. Recommendation systems plays a major role in the information filtering systems dealing with how best it can recommend items or information relevant to the user. Recommendation system or recommender system can be applied to a variety of applications. E-Commerce site being widely dependent on it. Its task is to automatically personalize the efforts of the user by recommending items from a large collection of items to what the user needs. It provides hint or suggestion to the user according to his/her preferences or interests. There is tremendous measure of information available and the recommendation system is proven to be useful to prune the redundant information and making useful recommendation to the users. In E-Commerce, recommendation system plays an important role. RS increases deals by recommending items and it permits users to make decisions for example which item to buy. Items are in term of movies and books, etc. such as Amazon.com utilize the recommendation system for recommending books to the users. Most of recommendation systems focused on the preference given by user[6]. User’s preference for an item is called as rating. The two tasks are performed to provide the recommendation to the users.

1) Rating prediction: From the repository of information available, the ratings of unrated items are predicted [5].
2) Ranking: To rank the users rated items and then recommended to users to maximize the user’s utility. There are different recommendation techniques like collaborative filtering, content based, demographic, utility based, knowledge based, Graph Based and hybrid.

Every year number of student face the task in selecting university for Master of Science (MS) degree. This system helps to recommend to the student, university which is best for them. In propsed system we used two effective algorithm i.e. Locality Sensitive Hashing(LSH) which work on similarity between student's profile and university's profile and Higher Order Singular Value Decomposition which work on rating given to the university by past students.

Content-based filtering (CB) [1][4] [7] method suggests and recommends objects and information which are comparable in content to objects that the students have interested previously, or compared and matched to the users’ characteristics.

Multi-Criteria Collaborative filtering (MC-CF) [2] [3] [5] is a popular recommendation algorithm that bases its predictions and recommendations on the ratings or behavior of other users in the system. The fundamental assumption behind this method is that other users’ opinions can be selected and aggregated in such a way as to provide a reasonable prediction of the active user’s preference. Collaborative filtering (CF) algorithms usually separated into two parts: 1) Memory-based algorithm, 2) Model-based algorithm. In our system we used two effective algorithm i.e. Locality Sensitive Hashing(LSH) which work on similarity between student's profile and university's profile and Higher Order Singular Value Decomposition which work on rating given to the university by past students.

II. RELATED WORK

Many authors presented different recommendation systems and techniques to improve the suggestions or Recommendations for users. This section comprises a brief mention about some methods proposed by different authors. Suja Panikar, Prayag Mane, Chetan Pakhale, Shubham Fulzele.Akshay Rathi [1] discussed the “Online Grocery Recommendation System” the failures and success of slope-one and min hash algorithms.


Che-Rung Lee, Ya-Fang Chang [10] discussed the"Enhancing Accuracy and Performance of Collaborative Filtering Algorithm byStochastic SVD and Its MapReduce Implementation”, in this MapReduce, which allows distributed processing of massive data stored in a distributed file system by using stochastic SVD (SSVD)

Alper Bilge, Cihan Kaleli [11] work on ”A Multi-Criteria Item-based Collaborative Filtering Framework”. In this, we propose an item-based multi-criteria collaborative
filtering framework in order to determine appropriate neighbor selection method, compare traditional correlation approaches with multi-dimensional distance metrics.

### III. SYSTEM ARCHITECTURE

To the best of our knowledge, there are numerous Data Mining technologies available, but not meant for generating recommendations in education domain i.e. for students by predicting MS university recommendation for them. The proposed architecture of University Recommendation System is shown in Fig. 1.

The University Recommendation System works by collecting student’s feedback in the form of ratings for Colleges over multiple criteria’s like Infrastructure, Facility, Teachers, Placements, Admission Difficulty and Campus Life. The overall ratings of universities is considered by the past ratings on the basis of different criteria. For that University Recommendation System plays an important role.

A. Proposed System Has Two Main Components

Registration phase of student and Recommendation System which provides an easy to use interface for the students to interact with the system.

B. Registration Phase

In Registration phase, student have to login the system and complete the verification phase. The Students database contains student information.

C. Recommendation System

Recommendation system consists of three techniques:

1) Content based Filtering Technique
2) Multi-Criteria Collaborative based Filtering Technique
3) Sentiment Analysis

D. Content Based Filtering Technique

It is a technique which considered students Information such that his Score (GRE , TOFEL) , specialization, Financial budget, Interest. And also considered the University information such that cutoff of particular exams, Fee structure, coarse etc. This information are stored into the particular database. This technique used to compare datasets of University and dataset of student to each other and by using LSH (Locality Sensitive Hashing ) find the similarity between them. This is works on the basis of students interest. Then finding the jaccard’s coefficient and set one threshold value. whenever the jaccard’s coefficient value is greater than the threshold value then it considers the relevant recommendations[1].

![Fig. 1: System Architecture](image)

E. Multi-Criteria Collaborative Filtering Technique

The ratings provided by students for items are the key input to Collaborative Filtering recommender systems. This technique based on the overall ratings given by the past Students on the basis of Multi-criteria of that particular university. We apply HOSVD for dimensionality reduction on the high-dimensional dataset of past students ratings. By using Dimensionality Reduction Technique Reduced the large no of data into the small data. Then similarity evaluation is find out. This is the best approach of recommended the Top-n universities of that particular student[3].

F. Sentiment Analysis

It is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine the students Statement of Purpose (SOP) Document it is positive, negative or neutral[8].

### IV. ALGORITHMS

1) Min Hash
2) HOSVD

A. Min Hash Algorithm

Min Hash Algorithm [1] is used to find the similarity between two sets with the help of Jaccard coefficient and Random Hash Value. The Jaccard similarity coefficient is a commonly used to indicate the similarity between two sets. Suppose for sets A and B, Jaccard coefficient is defined to be the ratio of the similar number of elements of their intersection and the number of elements of their union:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Where,

- Jaccard Co-efficient Range is from Zero to One. If the value is close to zero then it is irrelevant and if the value is close to one then we get the more relevant value.

With the help of this algorithm in our proposed system we will use to compare student profile and universities database. We will set a threshold value to get more accurate and feasible result as it will discard all results below that threshold value. Output will be in the form of tuples of universities having high similarities.
1) Steps to perform Min Hash Algorithm:
   Step 1: Taking set of student database and university database.
   Step 2: Then prepare a merge list that contains all details of student and university.
   Step 3: Now build a hash array which having size is equal to that size of merge list.
   Step 4: Build up a Min-Matrix.
   Step 5: Then Populate Min-Matrix from the required merge list.
   Step 6: Now we have to consider Student objects, merge list and find common objects.
   Step 7: Substitute value in Min-Matrix of that element from that hash array, repeat that for University objects also.
   Step 8: Count all the similar hash value elements from the Min-Matrix.
   Step 9: Now we will calculate the Jaccard Co-efficient:
   \[ J(A, B) = \frac{|A \cap B|}{|A \cup B|} \]
   Where, A is student objects and B is University objects.
   Step 10: Set threshold value.
   Step 11: Create a tuple which contains the items which have Jaccard value more than the threshold value.

2) Example

<table>
<thead>
<tr>
<th>STUDENT</th>
<th>INDEX</th>
<th>MERGE LIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score30</td>
<td>0</td>
<td>Score30</td>
</tr>
<tr>
<td>Branch:CS</td>
<td>1</td>
<td>Branch:CS</td>
</tr>
<tr>
<td>Name:USA</td>
<td>2</td>
<td>Name:USA</td>
</tr>
<tr>
<td>Name:USA</td>
<td>3</td>
<td>Name:USA</td>
</tr>
<tr>
<td>Branch:MECH</td>
<td>4</td>
<td>Branch:MECH</td>
</tr>
<tr>
<td>Cutoff</td>
<td>5</td>
<td>Score30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>UNIVERSITY</th>
<th>INDEX</th>
<th>MERGE LIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name:USA</td>
<td>0</td>
<td>Score30</td>
</tr>
<tr>
<td>Branch:MECH</td>
<td>1</td>
<td>Branch:MECH</td>
</tr>
</tbody>
</table>

   Fig. 2 :merging

<table>
<thead>
<tr>
<th>Hash value Index</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>University</td>
<td>10</td>
<td>NO</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
</tr>
</tbody>
</table>

Total no.of common element = 4
Total no.of elements = 6
So that, Jaccard Coefficient = 4/6 = 0.66
That means similarity between student and university = 66.66%.

B. HOSVD Algorithm

HOSVD is Higher Order Singular Value Decomposition [3] can be applied on three (or more) dimension called Tensor[3]. Steps for implementation is given below:
Step 1: HOSVD is applied on training data in 3-order tensor for dimensionality reduction to get the best approximation of rating information.
Step 2: The approximated data by HOSVD is used for clustering using cosine-based similarity.

\[
\text{Similarity} = \cos (A, B) = \frac{A^T B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{k} A_i B_i}{\sqrt{\sum_{i=1}^{k} A_i^2} \sqrt{\sum_{i=1}^{k} B_i^2}}
\]

Step 3: After overall rating prediction, recommender system forms the neighbours using cosine similarity presented for active user from corresponding cluster and makes predictions and Top-N recommendations.

Overall rating = \[ \frac{\sum_{i=1}^{n} \text{rating}(A_i, B_i)}{\sum_{i=1}^{n} \text{rating}(A_i, B_i)} \]

V. ADVANTAGES

1) URS Recommend the overall best Universities of MS to Students.
2) Application process to particular university through our website can be made.
3) Collaborative technique make system more efficient and scalable.
4) Analyze module will help to understand where you are lacking.
5) Web based application will make it more scalable.
6) Performs analysis of SOP for which students consult and pay.

VI. CONCLUSION

The Recommendation system examined how to incorporate information from metadata into recommendation algorithms. It evaluates the possibility to combine the various recommendations techniques. It describes the conventional Content, Collaborative Filtering recommendation approaches.

REFERENCES


Computational Intelligence and Data Mining (CIDM), 978-1-4673-5895-8/2013 IEEE

