Indian Institute of Information Technology, Vadodara

January 4, 2018

Design Project Report
on
Elective Course Recommendation System
Submitted by

Mehak Piplani 201551072
Sakshee Jain 201551074
Sonam Kumari 201551079
Siddharth Jain 201551093

Under the supervision of **Prof. Pratik Shah**

1 *Abstract*

Choosing a right course in formative years is very important decision as student's future depends on this one decision. Student by himself is not mature enough to take right decision in his early life. Selecting wrong courses means mismatch between student's aptitude, capability and personal interest. Since there is no other reliable source generally available that can guide the student towards the most suitable direction, recommendation system has been evolved to provide him/her guidance in selecting a right course. Thus our project idea is to develop a system for helping the students to choose an elective course which would be best suited for him/her based on features like previous performance, skill set, interest, ability to learn, etc. To execute this, we planned to take few inputs from the user and with the help of an algorithm which would combine all the features and in result output a ranked list of courses with the expected grades in each subject which he/she should proceed with. We have used Collaborative based filtering, content based filtering, and singular value thresholding algorithm for kernel matrix completion for predicting grades of elective courses.

2 Introduction

There have been a number of researches conducted in the field of educational data mining to predict the performance of student. In this sequence, our project discusses the problem of elective subject selection and proposes a solution through a recommendation system which may help student to make the right choice which best suites his/her.

We are trying to provide a solution to this problem using the latest technologies and researches on recommendation system. Recommendation system enhance the teaching and learning, recommend good solution, analyze data and offer data to modify activity plan. In the last few years recommendation system have provided valuable solution in opting one choice out of available many choices by focusing on logical relationship. So that people behave intelligently and make the right choice.

Student assessments are traditional method to predict student performance such as failing or passing or forecasting successful completion of the course, in this continuation predicting the classification of degree or achievement. Data mining in last few years extracted a lot of information from student data. This information‘s helped in handling of student and increase students expertise level.

In this paper we have proposed To develop a recommendation system which will recommend a ranked list of elective courses along with a predicted grade in all the courses based on student assessment done in the past course, his/her interest and what he/she wants to do in future which would be given as input to the system. The parameters that are used, would provide the best suited course list to the student for their better academic performances.

3 Literature Survey

Through the various research papers first the concept of recommendation system and its applications was understood. I got to know the various kinds of recommendation systems content-based, collaborative and hybrid and all these in the case of Book recommendation gave same results. Then I moved on to course recommendation and learnt about various techniques like ACO(Ant Colony Optimization)[9], ML-based approaches like regression,SVM,clustering ,etc. The best motivation was provided by the paper[3] which provided us with 4 algorithms to predict the course a student should opt for.

After experiment and result analysis, it is found that among 4 association rule algorithm 'Apriori Approach' performs the best with better accuracy in course recommender system[1]. Collaborative Approach provides student an accurate prediction of the grade they may get if they choose a particular course[7]. In study about course recommender system for University College Dublin's online enrolment application, it is found that collaborative algorithm when evaluated using large historical enrolment data, it provides very encouraging performance in terms of both recall and coverage[6]. Item-based techniques hold the promise of allowing CF based algorithms to scale to large data sets and at the same time produce high-quality recommendations. It overcomes challenges to user-based technique i.e scalability , sparsity etc[8]. Some other alogs that can be used are Clustering Technology, Feed-forward back propagation probabilistic neural network and Classification using Fuzzy Logic and Rough Set.

Firstly, after reading various research papers on recommendation system, I understood the idea, types, and working of it. In those papers various methods were discussed and experiments were conducted with real datasets to assess the overall performance of the proposed approach. They have used collaborative recommendation system that employed association rules algorithm to recommend university elective courses to a target student based on what other similar students have taken[2]. Also they have calculated the faculty expertise because faculty also plays a great role as they teach those subjects.

Tadej and Matek proposed a recommendation system for github.com which recommends projects to users. System works on network analysis of graph constructed on github data using link prediction [5]. Feng Xie in his paper[10] proposed a link prediction approach for item recommendation with complex number. Complex number can model relations e.g. like vs. dislike similar vs. dissimilar which where not captured in previous link prediction methords. Ilham Esslimani in his paper[4] new method to densify the network using social network technic which will improve the performance of standard CF algorithm.

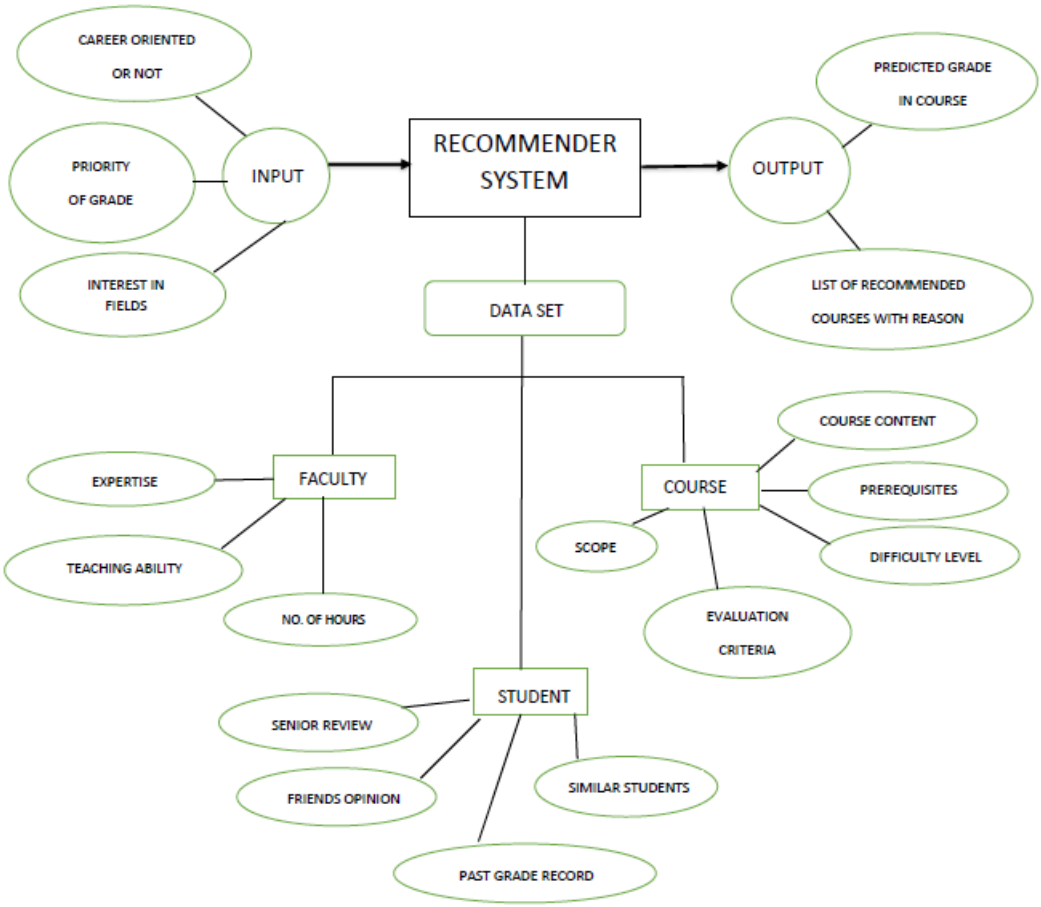
4 System Design

The Recommendation system we have proposed takes input from the students regarding their interests and future goals. The system also takes the information about the past grades, prerequisites and evaluation criteria of courses from the

data set provided to the system.

The system tries to predict the grades of students in the elective courses offered, using various techniques like user-based collaborative filtering, item-based collaborative filtering, kernel based threshold mechanism and content based approach. Then it compares the results and uses the method giving least error.

The output of the system is a ranked list of courses according to the input given by the student along with the predicted grades by the above mentioned method.



5 Method

5.1 Singular Value Thresholding Algorithm

The first method is a simple rst-order and easy-to-implement algorithm that is extremely efficient at addressing problems in which the optimal solution has low rank. The algorithm is iterative and produces a sequence of matrices X^k, Y^k and at each step, mainly performs a soft-thresholding operation on the singular values of the matrix Y^k.
The singular value thresholding algorithm for approximately solving the nuclear norm minimization problem and by extension, problems of the form

$$\text{minimize } ||X||_* \tag{1}$$

$$\text{subject to} A(X) - b \tag{2}$$

where A_i s a linear operator acting on the space of $n1 \times n2$ matrices and $b \in R_m$. This algorithm is a simple rst-order method, and is especially well suited for problems of very large sizes in which the solution has low rank. We sketch this algorithm in the special matrix completion setting and let P be the orthogonal projector onto the span of matrices vanishing outside of so that the $(i, j)^{th}$ component of P(X) is equal to X_{ij} if $(i, j) \in \Omega$ and zero otherwise.

5.2 Content Based Algorithm

A content-based recommendation system recommends an item to a user based upon a description of the item and a profile of the user's interests. Content-based recommendation systems may be used in a variety of domains ranging from recommending web pages, news articles, restaurants, television programs, and items for sale. Although the details of various systems differ, content-based recommendation systems share in common a means for describing the items that may be recommended, a means for creating a profile of the user that describes the types of items the user likes, and means of comparing items to the user profile to determine what to recommend. The profile is often created and updated automatically in response to feedback on the desirability of items that have been presented to the user.

5.3 Grade Prediction method

The method predicts student grade in elective course from grades obtained in prerequisites subjects and average performance of student in all courses taken.

The data set consist of four things i.e. Student, Elective, Subject (which constitute the elective subject) and weight(assigned to subject for a elective). The obtained marks in prerequisite subject will be used for the purpose of finding the score of student in respective elective.

Student elective score: To find the score of elective following formula will be used

$$ElectiveScore = \sum_{i=0}^{n} S_i * W \tag{3}$$

Where i indicate subject number Si indicate obtained score grade in the ith subject Wi indicate weight of ith subject in curriculum Using this formula score of each elective is calculated. After multiplying subject percentage score with respective weight value of each subject which constitute the elective curriculum, sum all of them. This data table is used to recommend the most suitable elective subject for the student.

5.4 Collaborative filtering:

5.4.1 User Based Approach

The steps followed in user-based CF to make a prediction for Student St are as follows:
Step 1: Similarity between the target student St and every other student is calculated.
Step 2: Based on their similarity value with student St , set of k students, most similar to target student St is then selected.
Step 3: Finally, prediction of grades for a student St in course It is generated by taking the weighted average of the grades scored by the k similar students to course it.
For step 1 there are many algorithms but Pearson -r correlation coefficient performs the best. Using this we calculated sim(u,v) i.e. similar value of user u with v.

$$Sim_{u,v} = \frac{\sum_{i \in I}(r_{u,i} - \bar{r_u})(r_{v,i} - \bar{r_v})}{\sqrt{\sum_{i \in I}(r_{u,i} - \bar{r_u})^2}\sqrt{\sum_{i \in I}(r_{v,i} - \bar{r_v})^2}} \tag{4}$$

Here $r_{u,i}$ denotes the rating of user u for item i, and r_u is the average rating given by user u calculated over all items rated by u. Similarly, r_v denotes the rating of user v for item i, and r_v is the average rating given by user v calculated over all items rated by v.
Now for last step, grades of target student is calculated using an adjusted weightage sum formula, to take into account the fact that different students have different grades distributions.

$$P_{u,j} = \bar{r_u} + \frac{\sum_{v \in V} Sim_{u,v}(r_{v,i} - \bar{r_v})}{\sum_{v \in V} |Sim_{u,v}|} \tag{5}$$

5.4.2 Item Based

The main difference between item-based CF and user-based CF is that item-based CF generates predictions based on a model of item-item similarity rather than user-user similarity. In item-based collaborative filtering, first, similarities between the various courses are computed. Then from the set of courses previously taken (score) by the target user, k courses most similar to the target course are selected. For computing the prediction for the target course, weighted average is taken of the target student's scores on the k similar courses earlier selected.
Let the set of students who have scores of both courses i and j be denoted by U, then similarity coefficient ($Sim_{i,j}$) between them is calculated as

$$Sim_{i,j} = \frac{\sum_{u \in U}(r_{u,i} - \bar{r_u})(r_{u,j} - \bar{r_u})}{\sqrt{\sum_{u \in U}(r_{u,i} - \bar{r_u})^2}\sqrt{\sum_{u \in U}(r_{v,i} - \bar{r_v})^2}} \tag{6}$$

Here $r_{u,i}$ denote the scores of student u for course i, and ru is the average scores scored by user u calculated over all courses previously taken by u. Similarly, $r_{u,j}$ denotes the rating of student u for course j.
To compute the predicted rating for a target item i for target user u, we use the following formula.

$$P_{u,j} = \frac{\sum_{j \in I} Sim_{i,j} * r_{u,j}}{\sum_{j \in I} |Sim_{u,v}|} \tag{7}$$

6 Experiment

6.1 For Grade Prediction

The first experiment involves completion of the matrix which has enteries of students grades in particular courses. The above methods have been applied on the matrix to find the most efficient method through the process of cross validation.

- **Exp 1 -** We predicted the grades hiding sets of 6-10 enteries for 7 courses.Then we compared the predicted value with actual value and stored the maximum positive and negative deviations and average deviation.This process was continued 10 times and all the results were analyzed.

- **Exp 2 -** We predicted the grades hiding sets of 15-20 enteries for 7 courses.Then we compared the predicted value with actual value and stored the maximum positive and negative deviations and average deviation.This process was continued 10 times and all the results were analyzed.

- **Exp 3 -** For grades prediction of a course having prerequisites, courses required prerequisites are identified for VI semester. Since here prediction is based only on student's own performance, there is no need of hiding records of other student it will not affect the predicted result. We used the average performance of a student and performance in prerequisite courses values from data-set. Each prerequisite course is given a weight which range from 0 to 1 representing the importance of prerequisite course to take present course.

6.2 For Listing Courses

For rank list of courses based on evaluation criteria and student's preference of course structure,content based approach is used. The course score for each course is calculated, then the courses which are above the average score of all courses are recommended ranked by their scores. Profiles are created for both student and course representing student interest and course attribute.Here also prediction is based only on student's own preferences and course credit structure, there is no need of hiding records of other student it will not affect the predicted rank list of courses.

7 Result

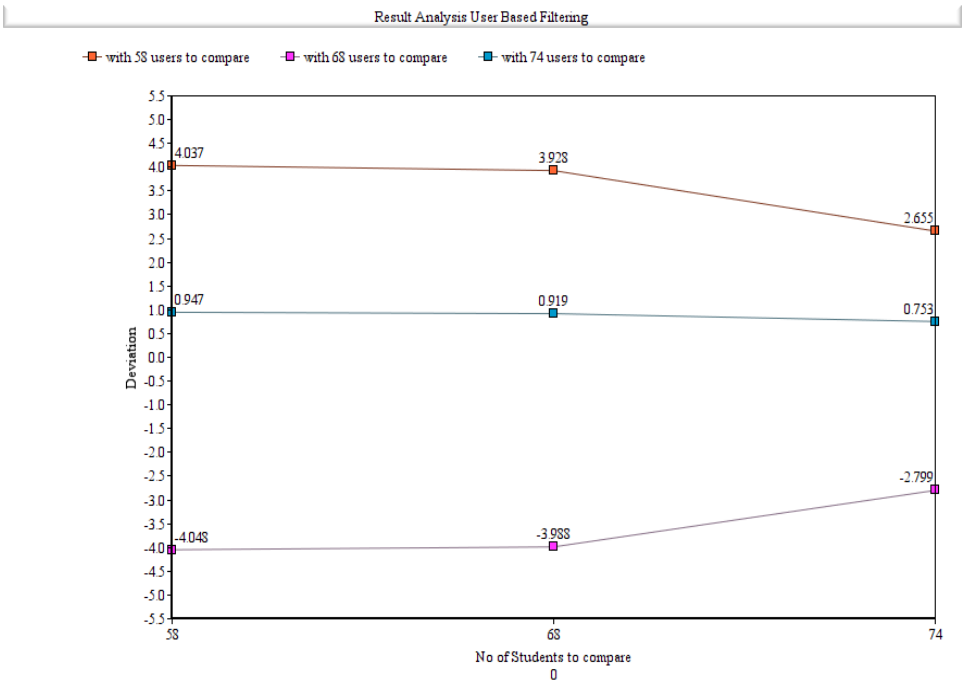
Singular Value Thresholding Algorithm - Since Data set for item is less i.e. 18 courses and 77 students so for kernel matrix completion between different courses comes to out in between (5,12). Using these course prediction of grades for student deviated maximum to 2.13 positive deviation and -2.14 negative deviation for few students, average deviation is 1.14 . Hence, It shows that this technique gives decent result for the given data set.

Item Based Algorithm - Since Data set for item is less i.e.18 courses and 77 students so for similarity comparison between different courses comes to out in between (5,12). Using these course prediction of grades for student deviated maximum to 4.1 positive deviation and -2.9 negative deviation for few students, average deviation is 1.137 . Hence, It shows that Item Based Collaborative approach performing poor in given dataset, because of less no of k-neighbours(similar courses).

User Based Algorithm - Since here we are calculating similarity between students which are large in numbers than courses, it is giving better accuracy in predicting Grades. Graph given below includes maximum positive deviation, maximum negative deviation and average deviation with different no of k-neighbours.

It shows that when data or no of k-neighbours(similar users) is reduces deviation from actual value is increases.

In Exp 3, Courses which do not have prerequisite have maximum deviation of +3 to -3 of actual grade obtained, this is because grade predicted is average score of student till now which may vary much for student. For other courses



maximum deviation of -2 to +2 of actual grade obtained, and average deviation is -1 to +1.

8 Conclusion and Future Work

The above experiments show that User Based Collaborative performs better than Item Based Collaborative Filter and Singular Value Thresholding Algorithm for grade prediction for available amount of data. The efficiency of the system can be further improved by increasing the data set to get better results like for comparison we can include all student's record so far studying in the particular Institute. For course's ranked list, system can be developed by including different modules like based on placements and faculty etc. As these parameters also play an important role in the decision of the course selection .This project can be further expanded to MOOC courses with few modifications in the system as technologies and development has advances so much that there is not much time when MOOC courses will also be part of the curriculum.

References

[1] Sunita B Aher and Lobo L.M.R.J. Article: A comparative study of association rule algorithms for course recommender system in e-learning. *International Journal of Computer Applications*, 39(1):48–52, February 2012. Full text available.

[2] Amer Al-Badarenah and Jamal Alsakran. An automated recommender system for course selection. 2016.

[3] Hana Bydzovská. Course enrollment recommender system. In Tiffany Barnes, Min Chi, and Mingyu Feng, editors, *EDM*, pages 312–317. International Educational Data Mining Society (IEDMS), 2016.

[4] Ilham Esslimani, Armelle Brun, and Anne Boyer. Densifying a behavioral recommender system by social networks link prediction methods. *Social Netw. Analys. Mining*, 1(3):159–172, 2011.

[5] Tadej Matek and Svit Timej Zebec. Github open source project recommendation system. *CoRR*, abs/1602.02594, 2016.

[6] Michael P. O'Mahony and Barry Smyth. A recommender system for online course enrolment: An initial study. In *Proceedings of the 2007 ACM Conference on Recommender Systems*, RecSys '07, pages 133–136, New York, NY, USA, 2007. ACM.

[7] Sanjog Ray and Anuj Sharma. *A Collaborative Filtering Based Approach for Recommending Elective Courses*, pages 330–339. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[8] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, WWW '01, pages 285–295, New York, NY, USA, 2001. ACM.

[9] Janusz Sobecki and Jakub M. Tomczak. *Student Courses Recommendation Using Ant Colony Optimization*, pages 124–133. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[10] Feng Xie, Zhen Chen, Jiaxing Shang, Xiaoping Feng, and Jun Li. A link prediction approach for item recommendation with complex number. *Knowledge-Based Systems*, 81(Supplement C):148 – 158, 2015.