

ONLINE COURSE RECOMMENDATION SYSTEM

Supreeth S Avadhani¹, Siddharth Somani², Vaibhav Nayak³, Sudhanva.BS⁴

^{1,2,3,4}BE, Department of CSE, NIE Mysore, Karnataka, India

Abstract - A recommender system or a recommendation system is a subclass of information filtering system that seeks to predict the rating or preference, an user would give to an item. Our Application uses K-nearest neighbour (KNN), K-Means and Collaborative filtering which are algorithms under Machine Learning.

Key Words: Machine Learning, K-Nearest Neighbour (KNN), Collaborative Filtering, K-Means Clustering

1. INTRODUCTION

With the increase in number of students and the variety of courses that colleges offer, the structure of education system became complex. This situation has made it hard for students to find the courses they want effectively. The main idea behind the recommendation systems for students is to allow them to make decisions to select the most appropriate course for their career path. A recommendation system includes user model, the counseled model and recommendation algorithmic program. Course recommendation is to take advantage of provided information and suggestions, to help students make better decisions regarding their courses. So, with the development of Education system, it's harder for students to find the course they want, and thus recommendation systems are applied more widely.

1.1 Prediction System

In this paper the focus on recommending courses to students based on the factors like student's marks, teacher rating, attendance, lacking skills (personal or related to cognitive domain). To achieve this machine learning algorithms like KNN, K-Means and collaborative filtering have been made use.

1.2 K-nearest neighbor (KNN)

It is an algorithm in pattern recognition which does not make any assumptions about the underlying data distributions (i.e non-parametric) used for classification or regression. In both cases, the input consists of the k closest training examples in the feature space.

- In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors and the object is assigned to the class most common among its k nearest neighbors.
- In k-NN regression, the output is the property value for the object. This value is the average of the values of k nearest neighbors.

1.3 K-Means

K-means clustering is a method that is popular for cluster analysis in data mining. The aim of k-means is to allocate every data point to one of the k clusters i.e to the one whose mean/centroid is nearest to the data point.

1.4 Collaborative Filtering

Collaborative filtering (CF) is a technique used by recommender systems. Collaborative filtering is a method used for automatically predicting (filtering) the interests of a user by collecting preferences or interests from many users (collaborating). The underlying assumption of the collaborative filtering approach is that if a person A has the same opinion as a person B on an issue, A is more likely to have B 's opinion on a different issue than that of a randomly chosen person.

2. RELATED WORK

Imran[1] et al. proposed a learning management system that is one of the technologies to be used in personalized recommendation system. The traditional algorithms focus only on user ratings and do not consider the changes of user interest and the credibility of ratings data, which affected the quality of the system's recommendation. Hence the paper presents an improved algorithm to solve this problem. The main idea is based on the assumption that similar users have similar preferences. By computing users similarity based on the user ratings to find the neighbors who have the similar interest with the active user. Then the active user's preference for an item can be predicted by combining

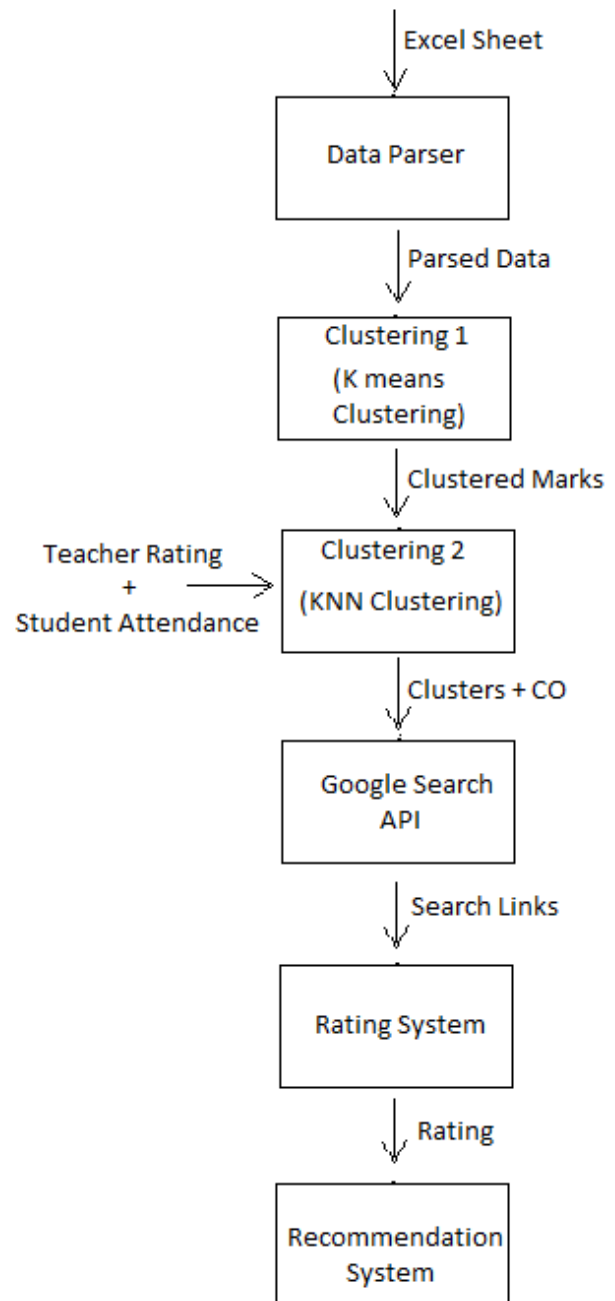
the neighbor's scores for the same item. Finally, top-N courses which the active user will most probably like are provided. But the similarity measurement in collaborative filtering algorithm only pays attention to the similarity score rather than the user interest.

Bakhshinategh[2] et al. proposed that for many reasons our understanding of these relationships is still at a very rudimentary stage. In general, research in the disciplinary traditions that study business organizations has been fundamentally unbalanced: researchers interested in characterizing the environment have typically been content with very simple models of the firm while researchers interested in the internal dynamics of firms have usually been content with very simple models of the environment. The paper has developed a general framework integrating resources and competitive environment as sources of firm competitive advantage and performance, with a will to consider market power as well as efficiency rents. The authors have, for instance, highlighted the strategic importance of considering external resource acquisition as a strategic option, of developing absorptive capacity and of investing in isolating mechanisms.

Based on the above research papers, the are based recommendations solely on the marks obtained by the student. In reality however there are many more factors that will influence the future courses a student undertakes. This is why we are looking to use numerous attributes such as grades, attendance and areas of interest as a vital factors in basing the recommendations particular to the student.

3. PROPOSED APPROACH

3.1 Flow Chart



Step 1: Extracting Data from Input Excel Sheet-

- The input is an excel sheet containing the marks of every student in the class for each CO. The data of marks needs to be extracted from this sheet.
- The starting coordinates of each test are known to us and are given in the function getCoordinates. Using these coordinates, we iterate through the

sheet and extract the marks into a dictionary named allMarks, which maps the marks to the cognitive domain of the question that it represents.

- The marks of all tests are obtained from the sheet simultaneously using the concept of multithreading. There are a total of 6 threads, which will extract the data for the corresponding CO number.

Step 2: Classifying Students Based on Marks-

- The raw data is now fed into the K-Means clustering algorithm which outputs the threshold value for good students. There will be a threshold marks for each question in the test. We then can classify the students as “good” or “bad” based on the threshold marks obtained.
- The K-means clustering algorithm selects two initial centroids and clusters’ all the marks (considered as points) into two groups. The mean of each group is considered as the new centroid. The algorithm repeats the above process until previous and current centroids are same.
- To write to the excel sheet, we use the module openpyxl in python. The coordinates of where to write to are known to us, and are given in the function getWriteCoordinates. Using these coordinates, we thus write to the given cells by iterating over the rows.

Step 3: Clustering the students to identify lacking skill-

The KNN Clustering function opens the classifier.csv file which contains:

- Teacher rating* – Rating of the teacher for a particular CO.
- Attendance* – Attendance of the student for the CO.
- Conclusion One* – What skill the student is lacking.

| Value on the dataset | Corresponding Lacking |
|----------------------|-----------------------|
| 1 | Hard Work |
| 2 | Listening skills |
| 3 | Concentration |
| 4 | Self-Learning |

- Conclusion Two* – The corresponding Cognitive domain the student is lacking.

| Value on the dataset | Cognitive Domain |
|----------------------|------------------|
| 0 | L1 |
| 1 | L2 |
| 2 | L3 |
| 3 | L4 |
| 4 | L5 |
| 5 | L6 |

- The Teacher Rating and attendance is stored in matrix X and Conclusion One is stored in vector y and it is used to train the clustering model which uses the K Nearest Neighbours(KNN) algorithm. This is returned by the KNN Clustering function.
- getPrediction() is the main driver function which predicts the Conclusion One. x_test contains the input numpy array which has the teacher rating and attendance.

Step 4: Generating Course Specific Links-

The chapter names in each of the CO is given as an input to the Google search function which returns top results matching the study material. These links are then provided to the students in different clusters.

Step 5: Recommending a link based on collaborative filtering-

Package needed: Scikit-surprise

If using anaconda, use conda install -c conda-forge scikit.

| CO | User | Links | Ratings |
|----|------|--|---------|
| 2 | CO1 | 1http://leesei.github.io/learn-to-code/ | 2 |
| 3 | CO1 | 1https://github.com/washam/coding-interview-university | 0 |
| 4 | CO1 | 1https://www.quora.com/What-is-a-good-free-online-course-on-distributed-computing | 4 |
| 5 | CO1 | 1https://gregio.com/projects/coding_interview_university/ | 3 |
| 6 | CO1 | 1https://www.coursera.org/courses?query=distributed%20systems | 3 |
| 7 | CO1 | 2https://www.coursera.org/lecture/cloud-computing/2-2-what-is-a-distributed-system-nvMXE | 2 |
| 8 | CO1 | 2http://leesei.github.io/learn-to-code/ | 1 |
| 9 | CO1 | 2https://www.iare.ac.in/sites/default/files/TH20AUTONOMICUS%20%20REGULATION%20AND%20SYLLUBUS_6.pdf | 2 |
| 10 | CO1 | 2https://gregio.com/projects/coding_interview_university/ | 1 |
| 11 | CO1 | 2https://gregorium.com/replicating-a-bsc-in-computer-science-through-moocs/ | 3 |
| 12 | CO1 | 3https://www.quora.com/What-is-a-good-free-online-course-on-distributed-computing | 1 |
| 13 | CO1 | 3https://www.coursera.org/courses?query=distributed%20systems | 4 |
| 14 | CO1 | 3https://github.com/washam/coding-interview-university | 4 |
| 15 | CO1 | 3http://gregorium.com/replicating-a-bsc-in-computer-science-through-moocs/ | 0 |
| 16 | CO1 | 3https://gregio.com/projects/coding_interview_university/ | 0 |

Fig. Excel sheet used for training the clustering model (Input)

- The links generated in step 4 are rated by the users. This dataset is available in collabAllCo.csv file which contains the CO (based on which the link is generated), UserID, Links and Corresponding ratings.
- This data is used to fit the KNN algorithm. After the model is trained, we suggest the most rated

link for each user. The output is written to a file named collabOutput.csv.

Education - Volume 1: CSEDU, 347-2017, Porto, Portugal

| ID | User | Suggested Link |
|----|-------------------------|---|
| 1 | CO1 SACHIN M S | https://www.quora.com/What-is-a-good-free-online-course-on-distributed-computing |
| 2 | CO1 Shivani S Avadhani | https://www.quora.com/What-is-a-good-free-online-course-on-distributed-computing |
| 3 | CO1 T M ABHISHEK | https://www.coursera.org/courses?query=distributed%20systems |
| 4 | CO1 LIKHITHA R | https://www.quora.com/What-is-a-good-free-online-course-on-distributed-computing |
| 5 | CO1 PRATHEEK M P | https://www.quora.com/What-is-a-good-free-online-course-on-distributed-computing |
| 6 | CO1 R SHREESHA | https://www.quora.com/What-is-a-good-free-online-course-on-distributed-computing |
| 7 | CO1 RACHANA G | https://www.quora.com/What-is-a-good-free-online-course-on-distributed-computing |
| 8 | CO1 RAGHAV A VADAVADAGI | https://www.coursera.org/courses?query=distributed%20systems |
| 9 | CO1 RAGHUVARDHAN G | https://www.quora.com/What-is-a-good-free-online-course-on-distributed-computing |
| 10 | CO1 RAJ ABHISHEK | https://www.quora.com/What-is-a-good-free-online-course-on-distributed-computing |
| 11 | CO1 RASHMI G | https://www.quora.com/What-is-a-good-free-online-course-on-distributed-computing |
| 12 | CO1 REA FERNANDES | https://www.quora.com/What-is-a-good-free-online-course-on-distributed-computing |
| 13 | CO1 ROHIT KUMAR | http://bmsce.in/sites/default/files/syllabus_new_scheme_2014-2018.pdf |
| 14 | CO1 SAGAR S P | https://www.quora.com/What-is-a-good-free-online-course-on-distributed-computing |
| 15 | CO1 SAHANA M | https://www.quora.com/What-is-a-good-free-online-course-on-distributed-computing |

Fig. Predicted ratings and suggested link (output)

5. CONCLUSION

We hereby conclude that with the increase in student requirements, personalized recommendation has been paid more attention. As per the literature survey most of the recommendation systems use collaborative filtering which is not an effective method. Our project aims to provide a better solution by analyzing the strengths and weaknesses by taking into consideration marks, attendance and areas of interest.

ACKNOWLEDGEMENT

It gives us great pleasure in presenting this project report titled "Online Course Recommendation System" and we wish to express our immense gratitude to the people who provided invaluable knowledge and support in the completion of this project. Their guidance and motivation has helped in making this project a great success. We express our gratitude to our project guide Dr. Abhinandan S, who provided us with all the guidance and encouragement throughout the project development.

REFERENCES

- [1] PLORS: A personalized learning object recommender system. Hazra Imran¹ · Mohammad Belghis-Zadeh¹ · Ting-Wen Chang² · Kinshuk¹ · Sabine Graf. Vietnam journal of computer science. February 2016, volume 3, issue 1
- [2] A Course Recommender System Based on Graduating Attributes. Behdad Bakhshinategh¹, Gerasimos Spanakis², Osmar Zaiane¹ and Samira ElAtia¹. 9th International Conference on Computer Supported