Course Recommendation System

Report submitted in partial fulfillment of the requirement for the degree of

B.Tech

in

Computer Science & Engineering

ANd9GcRcmIsloy3g35B_JsRNO8FT1UVoHApPujW-ZfaOSnKNCyg35wmufV3iEg

Under the supervision

of

Ms Charu Gupta

Designation & Deptt.

by

Achintya Sarkar 00720802715 , CSE-A

Himanshu Ratnani 03220802715 , CSE-A

Aayush Tiwari 00420802715 , CSE-A

Abhishek Jain 0062002715, CSE-A

Department of Computer Science & Engineering

Bhagwan Parshuram Institute of Technology

PSP-4, Sec-17, Rohini , Delhi-89

November 2018

**DECLARATION**

This is to certify that Report titled “Course Recommendation System”, is submitted by us in partial fulfillment of the requirement for the award of degree B.Tech. in Computer Science & Engineering to BPIT, GGSIP University, Dwarka, Delhi. It comprises of our original work. The due acknowledgement has been made in the report for using others work.

**Date: 1-11-2018** Achintya Sarkar 00720802715 , CSE-A

Himanshu Ratnani 03220802715 , CSE-A

Aayush Tiwari 00420802715 , CSE-A

Abhishek Jain 0062002715, CSE-A

**Certificate by Supervisor**

This is to certify that Report titled “Course Recommendation System” is submitted by Abhishek Jain 0062002715, Achintya Sarkar (00720802715) , Abhishek Jain 0062002715, CSE-A , Aayush Tiwari 00420802715 , in partial fulfillment of the requirement for the award of degree B.Tech in Computer Science & Engineering to BPIT, GGSIP University, Dwarka, Delhi. **It is a record of the candidates own work carried out by them under my supervision.** The matter embodied in this Report is original and has not been submitted for the award of any other degree.

  (Signature)

**Date: 1-11-18 Supervisor**

**Certificate by HOD**

This is to certify that Report titled “Course Recommendation System ” is submitted by Abhishek Jain 0062002715, Achintya Sarkar (00720802715) ,Himanshu Ratnani (03220802715), CSE-A , Aayush Tiwari 00420802715 , under the guidance of Ms Charu Gupta in partial fulfillment of the requirement for the award of degree B.Tech in Computer Science & Engineering to BPIT, GGSIP University, Dwarka, Delhi. The matter embodied in this Report is original and has been dully approved for the submission.

  (signature)

**Date: 1-11-18 Dr. Deepali Virmani**

ACKNOWLEDGEMENT

We would like to express a word of thanks to all those who have provided us with sincere advices and information during the course of the period we thought and worked for the creation of the product. It was indeed a great pleasure for us to work in a very cooperative, enthusiastic and learning atmosphere it was a very informative experience and we all have grown a lot during the course of creation of the final product and are much prepared to take on future projects and complete them before deadlines with utmost sincerity and delinquency .

we would like to express my sincere gratitude to our mentor, Ms. Charu Gupta for guiding us through this project. we thank her for the invaluable guidance and encouragement she provided throughout this project. The valuable suggestions, given by the m’aam is one of the main reasons for the completion of this project.

I wish to express my sincere to Ms.Deepali vermani for her strict supervision, constant encouragement, inspiration and guidance, which ensured the worthiness of my report.

(Signature of the students)

Table of Contents

Introduction

Nowadays, the amount of educational resources spread at Internet is huge and diverse . Massive Online Open Courses (MOOCs) such us Coursera, Udacity, EdX, to name a few, are gaining momentum[3] . It is possible to find courses from almost every knowledge domain[1]. This vast offer overwhelm any user willing to find courses according his/her background. This task can be tedious because it involves access to each platform, search available courses, select some courses, read carefully each course syllabus, and choose appropriate content. This process can be unmanageable if we extend our search beyond online courses to educational content.

In this work we propose a system for online courses recommendation, although MOOCs courses are primarily focused. To do so, we rely on course reviews , NLP [4] and we use a set of tools for using sentiments and classify courses based on them . Accordingly each course is assigned to a set of users who get a particular scores based on correlation[2] . Each document is a combination of topics and each topic is a probability distribution over words . Topic models are a type of graphical model based on Bayesian networks or SVM Classifiers [1]. The generative process described by a topic model does not make any assumptions about the order of words as they appear in documents. The only information relevant to the model is the number of times words are produced, this is known as the “bag-of- words” assumption [1].

Problem Statement

Nowadays, the amount of educational resources spread at Internet is huge and diverse . Massive Online Open Courses (MOOCs) such us Coursera , Udacity, EdX, to name a few, are gaining momentum. It is easy for a sophomore or even some senior year students to get lost in choosing the right course . The recommendation system we propose is trying to solve the issue where the course is recommended based on users general rating score based on their answers and academic record.

Kmeans

***k*-means clustering** is a method of [vector quantization](https://en.wikipedia.org/wiki/Vector_quantization), originally from [signal processing](https://en.wikipedia.org/wiki/Signal_processing), that is popular for [cluster analysis](https://en.wikipedia.org/wiki/Cluster_analysis) in [data mining](https://en.wikipedia.org/wiki/Data_mining). *k*-means clustering aims to [partition](https://en.wikipedia.org/wiki/Partition_of_a_set) *n* observations into *k* clusters in which each observation belongs to the [cluster](https://en.wikipedia.org/wiki/Cluster_(statistics)) with the nearest [mean](https://en.wikipedia.org/wiki/Mean), serving as a prototype of the cluster. This results in a partitioning of the data space into [Voronoi cells](https://en.wikipedia.org/wiki/Voronoi_cell).

The problem is computationally difficult ([NP-hard](https://en.wikipedia.org/wiki/NP-hardness)); however, there are efficient [heuristic algorithms](https://en.wikipedia.org/wiki/Heuristic_algorithm) that are commonly employed and converge quickly to a [local optimum](https://en.wikipedia.org/wiki/Local_optimum). These are usually similar to the [expectation-maximization algorithm](https://en.wikipedia.org/wiki/Expectation-maximization_algorithm) for [mixtures](https://en.wikipedia.org/wiki/Mixture_model) of [Gaussian distributions](https://en.wikipedia.org/wiki/Gaussian_distribution) via an iterative refinement approach employed by both *k-means* and *Gaussian mixture modeling*. Additionally, they both use cluster centers to model the data; however, *k*-means clustering tends to find clusters of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes.

The algorithm has a loose relationship to the [k-nearest neighbor classifier](https://en.wikipedia.org/wiki/K-nearest_neighbor), a popular [machine learning](https://en.wikipedia.org/wiki/Machine_learning) technique for classification that is often confused with *k*-means due to the *k* in the name. One can apply the 1-nearest neighbor classifier on the cluster centers obtained by *k*-means to classify new data into the existing clusters. This is known as [nearest centroid classifier](https://en.wikipedia.org/wiki/Nearest_centroid_classifier) or [Rocchio algorithm](https://en.wikipedia.org/wiki/Rocchio_algorithm).

Need For the project

The proposed project suggests an optimum course for each student separately . This gives each user the flexibility to try new course which they will be able to complete within stipulated time and gain knowledge without having to worry if the suggested course would be too easy or too difficult for them . This system also saves the time of user which could have been lost due to sheer number of online courses available on the internet .

Survey

Recommendation system [1] are of 3 types content based collaborative and hybrid recommendation system . These are governed by some rules which come under fuzzy logic [8] . The further analysis of review can be done

using sentiment analysis .In the proposed project the accuracy of recommendation can be adjusted so as to give probabilistic recommendation to a certain degree and rest based on user item matrix.

Requirement Specifications (h/w and s/w requirements)

The course recommender system would suggest best suitable courses to the new student after taking some inputs from him/her on some important attributes that will be needed to find the similar user to this new student .

This recommendation should be based on the attributes given by the student and should give the best suitable course .

The system should also be able to do semantic analysis of the user’s reviews and should bifurcate them into good or bad reviews.

The system should also be well aquinted with the new course added to various sites , whose users are not available and should identify the course to be good or bad according to the reviews present on the course.

The system should have a UI for the new user,where he or she can enter his or her details and search for the course needed and should also contain an UI which would show the courses recommended .

System analysis and design

The system recommends the user the best course from a large no. of courses available on different sites on the same field of interest. It helps the user to save his or her precious time to search over the internet about the course and read thousands of comments and reviews that may be discrete and analogous. Even then the user would not be able to search all the sites and all the courses available on a particular topic .

The system maintain a dataset of users and their reviews on different courses on different sites.

The system analyzes the best course taken by a similar group of users on the basis of the their reviews on the courses . Then the system takes some fields of inputs from new user to find the similar users existing in our dataset and then suggest the course best reviewedby the the similar group of users.

Work Done

The current progress has been the data collection and labelling of reviews from websites . The algorithms used to make a classifier for the reviews and get its sentiment has been able to differentiate and give scores to reviews with a classifier.score of 70% but it is a work in progress for multi class classification for differential scoring of reviews . The training of the classifier was done for only english reviews and nltk library was used to get the features from the classifier.

Implementation

The course recommender is one of the most powerful tool a student can get his hands on its one of the tools that is a game changer in terms of being effective and providing results to the student . One of the best thing a student can do in his life is to use his time effectively and given proper guidance any student can shoot to the moon . Course recommendation is a special project taken in consideration the problems a student goes throw his daily life by fixing it and can be used by any student with a keen interest to learn and grow . As a student we know how big the web really is and in this ocean of information finding the one we need is too tough going throw numerous pages to get what we need is not that effective approach so we tackled our this daily approach and created this product course recommendation system . So you want to learn somethings okay anyone can name to so many websites udemy ,coursera ,etc etc where you can search the course you want thats fine but how to know this one is the best you do’t know every website that is there you have not gone throw all the courses present at that website . Ratings can be deceiving and so many courses have so many ratings which among them is for you so we created this project who provides you with the best course for you in whole of the internet which will give you the best result there is .

Being a smart worker is always better than being a hard worker . Our algorithm also consider how well versed you are with the subject and also keeps in consideration your older courses taken to recommend the best course their is . The recommended courses can be paid or free.

Conclusion

Our 4 months of work bear fruits and our course recommending system is working correctly and our algorithm is being effective and providing good results to the queries provided their is still a lot of room for improvement . We are happy with the results provided by the system .performance of the algorithm can be improved due to our lack of knowledge and time we will leave this project for further improvement afterwards . Future work the project can be expanded by adding blogs and websites that can be added and can be helpful to the students we can also add books to the recommendation system newsletter can be added the one that student can subscribe to for being up to date to the current things and youtube channels that are helpful can also added with the system a forum or discussion feature can also be added to the product for making the product much better and being all round recommending system .

Future Work

Our Next step would be to implement a recommendation algorithm for the labeled courses and also to optimise the obtained features such that only the necessary information is extracted from each review and the training takes lesser time to implement . But before training appropriate training data (User , Item matrix etc. ) has to be extracted. Once goes as planned the recommendation system might be able to work in real word cases .