International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 07, July-2021

Upgrading Students Via Probational System

The Course Recomendation

Sourabh Kadam
Dept. of Computer Science
SOE, MIT-ADTU
Pune, Maharashtra, India

Mayuresh Naidu
Dept. of Computer Science
SOE, MIT-ADTU
Pune, Maharashtra, India

Moumita Baidya
Dept. of Computer Science
SOE, MIT-ADTU
Pune, Maharashtra, India

Aditi Jaiswal
Dept. of Computer Science
SOE, MIT-ADTU
Pune, Maharashtra, India

Suresh Kapare
Dept. of Computer Science
SOE, MIT-ADTU
Pune, Maharashtra, India

***Abstract*— This paper is focused on the course recommendation process. On the Internet, where a lot of choices for different courses are massive, there is need to filter and should be prioritize. Recommendation system solves this problem by searching across massive amount of information to provide users with personalized course content. This paper explores the different characteristics and potentials of different prediction techniques in recommendation systems in order to recommend proper course to user which can also be useful for research or practice in the field of recommendation systems.**

Keywords*— Data preparation; data inspection; data preprocessing; approaches; model building; content based filtering; collaborative filtering; hybrid filtering .

I. INTRODUCTION

Internet has revolutionized the world of consumerism and unleashed a greater demand of courses by providing atrouble-free buying experience and delivery to the user. Knowledge is also an important descriptor in identifying humans. Our problem is defined as follows: whenever someone searches particular courses then our recommendation system finds the most similar questions to be asked for that user. For these purpose we will take available in use datasets of such websites like courser, udemy, edx etc. and then we will apply our algorithms on it to see what results our recommendation system returns. The proposed system is designed for both customers and sellers as it is beneficial for each of them.

The categorization of a course to recommend was, until now, manually performed by teams, since it requires both domain expertise and a comprehensive knowledge over the range of products. Manual classification can give multiple errors which will cause incorrect results, misleading analysis and also requires too much time. So we have developed such a recommended system which will give more accurate results so for this purpose we have used two techniques first one is RNN & LSTM along with some unsupervised algorithm and second is Word2Vec. In this project, we propose a method to recommend courses to the learners based on their knowledge. To build recommendation system to suggest courses this field has its applications in social media, e-commerce, and criminal law. It can is used for other ecommerce website to improve their recommendation system. This helps us in providing high-level features for courses and it can be used for matching user level preferences. Recommending videos is extremely challenging. Some of existing recommendation algorithms proven to work well on small problems but fail to operate on greater scale. Highly specialized distributed learning algorithms and efficient serving systems are required for handling massive user base.

II. OBJECTIVE

- To improve existing recommended systems.
- Generation of more accurate results using various techniques.
- To increase speed of the system as compared to others.

It can also be used for other E-commerce sites with just some modifications using related datasets.

III. LITRETURE SURVEY

Kunal Shah et al [1] This paper provides an overview of the complimentary schemes and describes the current generation of complimentary methods. Recommendation systems or recommendation systems (RS) are data filtering systems and software tools and strategies that provide suggestions to the user according to his or her needs. Many popular Ecommerce sites make extensive use of RS to promote news, music, research articles, books and product items. Recommendation programs use personal, anonymous and local information from the Internet. This paper attempts to explain the various limitations of the complimentary methods and their merits.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 07, July-2021

Goran Antolić et al [2] A recommendation program is a software program that aims to make recommendations. To do so, the recommendation system includes a number of features, such as data collection and processing, a recommendation model, a late processing recommendation and a user-friendly interface. Recommendation programs use one or a combination of several recommendation techniques. In this paper we present a recommendation program designed to give users recommendations based on their interests in various domains. We take the interests of the user into account for his or her activities and posts on the social network. A social network used as a source of information for the user (Facebook) provides an Open API that allows access to information about a user collected on a social network. Thanks to this data we overcome a problem called "cold start" and create a user profile. The recommendation program is often associated with only one domain, while the recommendation program described in this paper can generate recommendations from a variety of areas (movies and music). In addition the recommendations related to a specific area, our program is able to recommend web articles (random text), which are appropriate for a user who may be in more than one type of interest.

Yuri Stekh et al [3] Advanced and researched methods of data mining to solve problems predicting recommendations: group analysis, organizational rules, ambiguous set view. These methods are used in the collaborative filtering process.

Marwa Hussein Mohamed et al [4] Today's Recommender program is a new field of research in machine learning. The main idea of the recommendation program is to build relationships between products, users and to make the decision to choose the most appropriate product for a particular user. There are four main ways in which systems recommend generating a list of user filtering recommendations - content, interaction, census and hybrid. In content-based filtering the model uses object specification to recommend additional objects with close properties. Collaborative filtering uses user preferences such as items that the user has previously viewed or purchased, in summary of any measurements the user has provided those rating items and similar conclusions made in the list of other users' items. Guessing things that the user can find interesting. Filters look at user profile data such as age, gender, education and location to find similarities with other profiles to find new recommendation lists. Hybrid filtering includes all three filtering methods. This paper presents research on recommendation programs, strategies, challenges in facial recommendation programs and writes other research papers that address these challenges.

Sung-Woo Byun et al [5] The need for technology to recommend content is increasing as the number of media content on TV and the Internet increases. According to the specific recommendation, the research conducted in the screening recommendation program is increasing. But these researchers have problems with the onset and onset of colds. To resolve these issues, this paper proposes a recommendation system based on user usage in terms of viewing time patterns and interests. With this, we bring out the best in 259 media contents with 157 users and calculate your preference scores..

Silvana Aciar et al [6] Consumer reviews, ideas and shared experiences in product use are a powerful source of information about consumer preferences that can be used in recommendation programs. Apart from the importance and significance of such information, there is no comprehensive way to legitimize the process of selecting and retrieving ideas and the use of returned comments. In this paper, a new recommendation system is proposed based on consumer product reviews. The priority approach is tailored to the system. The proposed approach has been demonstrated using case studies of the digital camera recommendation system.

François Fournier [7] The recommendation system is a filter system to deliver information such as movies, music, books, news, photos, web pages, tools to the user. This information is filtered to interest the user. The purpose of the recommendation program is usually "to help consumers learn about new products and desirable items among thousands of options" [1] [2]. Data filtering programs, in general, aim to remove unwanted or unwanted information from the database. They aim to introduce relevant information and reduce overload while improving signal-to-audio level at the semantic level. According to a review by Jijjin [3] in some texts in 2001, "the definition of 'recommendation' seems to differ from one author to another. He added that "some view the 'recommendation system' as a general adjective representing various forms of recommendation / prediction including collective, social and content filtering, Business networks and organizational rules. [6]". This looks like the current thinking these days in the field and is the definition chosen by Herlocker et al.

IV. METHODOLOGY

In this, we will discuss which algorithms and techniques we used.

A. Text pre-processing

Data pre-processing is important in deep learning, so following pre-processing is done on data.

1) Spliting the dataset variable

We took the column course difficulty as the target column to classify the courses according to the difficulty label. It has three types of beginner, intermediate, mixed. We applied Label Encoder to encode the categorical merit. All other required columns in the dataset were taken as the dependent variables to classify courses on the basis of difficulty label to recommend users.

2) Lower ()

We lowercase the text to reduce the size of our vocabulary of the dataset. Thus it reduces the sources to be used for.

3) Remove or convert numbers

We can either remove numbers or convert the numbers using python APIs. To avoid the loss of Information we prefer more converting the numbers into words.

4) Remove punctuation

Punctuations like !@$ etc. makes the model confused. Hence the a word for example, 10000 dollar!, 10000USD$ model will take those two values as two different values

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 07, July-2021

because of that punctuations. Hence it is best to remove punctuations.

5) Remove whitespaces

In some textual data there are more than one whitespaces due to some inconveniences from the person putting the data. So removing whitespaces reduces the size required to store the datasets easily with less cost and resources.

6) Removing stopwords

Stop words do not contribute to the meaning of a sentence. Hence, they can safely be removed. The NLTK library has a set of stop words and we can use these to remove stop words from our text and return a list of word tokens.

7) Stemming and lemmatization

Stemming finds root form of a word. Stem or root is the part to which inflectional affixes (-ed, -ize, -de, -s, etc.) are added. The stem of a word is generated by removing the prefix or suffix of a word. So, stemming word may not result in actual words. Lemmatization also changes a word to its root form. But the difference is that lemmatization ensures that the root word belongs to the language. For the purpose of valid words we use lemmatization. In NLTK use the WordNetLemmatizer to get the lemmas of words.

8) Histogram

Histograms group the data in bins and are the fastest way to get idea about the distribution of each attribute in dataset. It issues us a count of the number of observations in each bin created for visualization. Histograms also help us to see feasible outliers.

B. Vectorization

1) Bag-of-word model

With data we gathered data; we modeled them by a bag-of-word model. In this model, each word shown in the whole document corpora becomes an attribute. Then, each it is represented by a bit vector, indicating whether each word appears or not. This model is based on two different things; First, Word chances for one text position are unrestrained of the words that occur in other positions a Second, the probability of encountering a specific word is unrestrained of its position.

This assumption is incorrect, but it is known that this does not seriously affect classification task. We combined course title, key words, and abstract to construct a set of words representing a course.

2) TF-IDF

For automated text analysis we use TF-IDF and it is very useful for scoring words in machine learning algorithms for Natural Language Processing (NLP).

3) Word2vec

We used Word2vec to group the vectors of similar words together in vectorspace. It will help to detect similarities mathematically.

4) Recommenders

Using data model discussed so we now can implement personalized recommender for course. We can call recommendation system a system to fill out missing preference data. There can be lots of strategies to decide proper values for missing preference. Filling with the course-title's average or rating's average can be a simple baseline. Recommender systems can make use of either or both collaborative filtering and content-based filtering which also known as the personality-based approach, as well as other systems knowledge-based systems is one of them. Collaborative filtering approach consider user's past behavior to build model(purchase history and ratings etc.) as well as similar decisions made by other users. Then this model will also help to recommend or predict that the user may have an interest in. Content based filtering approach utilizes a series of discrete, pre-tagged characteristics of an item in order to recommend additional items with similar properties. Current recommender systems can be combination of one or more approaches which turns into a hybrid system which we are going to follow.

Each one of the system has its strengths and weaknesses. Some of them require a large amount of information about a user to make accurate recommendations. Whereas some needs very little information to start, it is far more limited in scope. Recommender systems are a useful for users discover items they might not have found otherwise. Recommender systems can often implemented using search engines indexing non-traditional data.

In this section, we discuss the implementation of recommender algorithms.

5) Content based filtering

This system will use item feature to recommend user. In this type of algorithm courses are first represented by their attributes. A user data is then built by taking user ratings on these attributes. The ratings can be taken from user actions like item read, bought, clicked. This model of user ratings on course attributes is then applied to any new item via their attributes to generate recommendations.
A particular variation of this is case based content filtering wherein certain cases are built around item attributes on which user input is taken and then recommendation is generated.
The only requirement of this recommender is data and does not require large set of users. This system will use item feature to recommend user. In this recommendation will be calculated based on cosine similarities.

6) Collaborative filtering

Collaborative filtering is based on two assumptions, 1. User tastes are individually stable or move in sync with each other over time and 2. The system is ranged within a domain of agreement. For the purpose of automatic prediction we going to use collaborative filtering so with the help of user's is interests and collecting his taste and preferences we were able to recommend course to him.

7) Hybrid filtering

Recommender systems can use a hybrid approach, combining multiple filtering like collaborative filtering, content-based filtering, and other approaches. To provide users more accurate courses we implemented content based filtering and collaborative filtering. Recommender systems can use a hybrid approach, combining multiple filtering like collaborative filtering, content-based filtering, and other approaches. To provide users more accurate courses we

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 07, July-2021

implemented content based filtering and collaborative filtering.

C. Classification techniques

With the help of classification techniques we were able to classify data into their respective category.

1) Long short-term memory

LSTM is an artificial RNN architecture for deep learning. For processing entire data sequence and it has feedback connections. It is well suited for classification, processing as well as making prediction.

2) Confusion matrix

For the purpose of performance evaluation of classifier confusion matrix is implemented. Confusion matrix is of n*n. Confusion matrix helps us to calculate model's accuracy.

3) Naive Bayes

Naive Bayes will help us to convert the results from your test into the real probability of the event. It will help to find conditional probability.

4) Recurrent Neural Networks

In this project we used RNN for sequence classification. RNN helped us, recognizing characteristics of sequential data we were able to predict next likely course.

V. FUTURE SCOPE

Despite many algorithms and current technology there is still room for improvement. Nowadays, there are many challenging problems faced by the ecommerce industry. One of them is the problem faced by sellers while uploading pictures of products on the platform for sale and the consequent manual tagging involved. Therefore, it gives rise to the misclassifications which leads to its absence from search results.

An image based search algorithm can be the true potential of ecommerce by enabling customers to click a picture of an object and search for related projects without the need for typing. So in future we can make such a system where by clicking picture of an object user can get their products so there is no need of typing.

Thus with some modification with this model using domain related dataset we can use this as recommendation system in many places.

VI. CONCLUSION

In this paper, we have presented a Personalized course Recommendation System, which recommends related course to user. This research paper will help other researchers who want to build their own recommendation system.

REFERENCES

[1] Kunal S., Akshaykumar S., Saurabh D. and Kisandas A. 2018 February. Recommender systems: An overview of different approaches to recommendations. 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). IEEE.

[2] Goran A. and Ljiljana B. 2017 July. Recommender system based on the analysis of publicly available data. 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE.

[3] Yuri S., Mykhoylo L., Vitalij A. and Vitalij C. 2015 November. Methods and tools for building recommender systems. In The Experience of Designing and Application of CAD Systems in Microelectronics. IEEE.

[4] Marwa Hussien M., Mohamed Helmy K. and Mohamed Hasan I. 2019 February. Recommender Systems Challenges and Solutions Survey. In 2019 International Conference on Innovative Trends in Computer Engineering (ITCE). IEEE.

[5] Sung-Woo B., So-Min L., Seok-Pil L., Kwang-Yong K. and Cho K. 2016 January. A recommendation system based on object of the interest. In 2016 18th International Conference on Advanced Communication Technology (ICACT). IEEE.

[6] Silvana A., Debbie Z., Simeon S. and John D. 2006 December. Recommender System Based on Consumer Product Reviews. In 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06). IEEE.

[7] François Fournier 2015 November. Recommender Systems: Technical Report and Literature Review.