

# Survey on Recommending Learning Path of Student using Machine Learning

Priti B. Jadhav<sup>1</sup>, Dr. R. B. Ingle<sup>2</sup>

M.E Student, Computer Engineering, Pune Institute of Computer Technology, Pune, India<sup>1</sup>

HOD, Computer Engineering, Pune Institute of Computer Technology, Pune, India<sup>2</sup>

**Abstract:** Many times student do not understand what is the field they are good at, what is the area they can improve to be better in. This paper introduce the adaptive questionnaire strategies, which changes the question based on student previous answer. This can help student to find out the depth of knowledge they have in that subject. Some of them are to check the students learning style. In that which student work better in what kind of courses. This paper gives idea about the different methods to set the questionnaire and the different classification algorithm to help us to understand the nature of the student like for what course they are suitable or they can find their correct career path with different classes the paper gives the techniques to make the classification of the students.

**Keywords:** Business Intelligence, Data Mining, Machine Learning, Web Application.

## I. INTRODUCTION

There are many courses and different fields are emerging every student have their different skills and many times they don't know where to go which path to select or depth of the subject they are good. To assess student's knowledge from various perspectives and various skills effectively, it is very important that question paper should be made by considering various constraints such as proper coverage of units from syllabus, coverage of difficulty levels, coverage of cognitive levels. There are many classification algorithm we can have to use for this type of system we will see in brief some algorithm for the recommendation to the student for the respective field of their expertise. The quality questions generation is very important in order to check the student knowledge in the subject. The adaptive nature will test student's ability to answer the questions which can be from different subject and of different level.

## II. BACKGROUND AND RELATED WORK

This section is literature review. In this section we will see the research papers related to the adaptive mechanism and the recommendations.

Lakshmi Sreenivasa Reddy. D et al. [4] learning styles vs suitable courses the has considered the learning styles of students and one questionnaire is made. This questionnaire is divided in to four dimensions global / sequential, visual/auditor, sensing/intuitive. The algorithms are used in this for ILS Attribute value frequency algorithm and CRT(classification and regression algorithm) algorithm for generating rules. The CRT tree got the 76%accuracy.

Vaibhav M. Kale et al.[9] in this we can see the question paper generation is for coverage the total syllabus as well as topics are covered. The algorithm presented in this paper attempts to solve the problem of dividing unit wise

questions, marks allocated to that and the format of paper by dividing the task of generation of question. way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

Sabine Graf et al. in [1] the paper Learning Management System is build in which the automatic, dynamic, and global student modeling is done adaptive mechanism aims at being easy to use for teachers by being generic and adaptable for teachers, allowing them to adjust the mechanism to their course structure and preferences. The Adaptive modeling is made based on student learning style.

Mohamed Amine Chatti t al. [3] implemented 16 different Tag based collaborative filtering recommendation algorithms, memory based as well as model based, and compare them in terms of accuracy and user satisfaction. The results of the conducted offline and user evaluations reveal that the quality of user experience does not correlate with high-recommendation accuracy. In this paper for recommendation they have used collaborative filtering, content based recommendation and hybrid recommendation. The main technique is the collaborative filtering recommendation. Memory based CF algorithm is used for rating prediction and recommendation.

Zhaoli Wang et al. [2] the resume scanning and then checking in the fields of programming skills and company's requirements we have to match and after matching those we can say the particular candidate is eligible for so and so company by matching it with company profile. The algorithm used in this paper genetic algorithm is used to meet the needs. The algorithm is used for training the data set and it build three kind of resume databases: positive feedback resume database (PosF), negative feedback resume database (NegF), non

deterministic resume database (NonF). But the problem here is the student can lie in the resume.

### III. CLASSIFICATION ALGORITHMS

There are three types of machine learning algorithm depending on its learning style viz. Supervised, Unsupervised and Semi supervised. In supervised learning the algorithm input data called training set which knows the label of the classes prior. The model is prepared through the working on the training set until it reach to the desired model we get. The problems which fall under this are classification and regression. Algorithms in this category are logistic regression propagation, back propagation neural networks.

Unsupervised learning, training data does not have known results. A model is prepared by deducing structures present in the input data. This may extract rules or mathematical rules to organise the data of similar type. This includes problems on the clustering, association rule mining and dimensionality reduction. Algorithms which comes under this are Apriori and K-means.

Semi supervised falls in between supervised and unsupervised. This includes graph based methods, density based methods

#### A. Decision Tree Algorithm

Decision tree algorithm is supervised algorithm. The algorithm like Decision tree, Random forest, popularly used for data science problems. It helps to split the data in which the class label is already decided in the training set. when the new data arrives we check the data falls in which categories and according to that we classify new data in the respective labels. The most popular decision tree algorithms are:

- Classification and Regression Tree (CART)
- Conditional Decision Trees
- C4.5 and C5.0
- Decision Stump
- M5
- Chi-squared Automatic Interaction Detection(CHAIID)

#### B. Naive Bayes

It is a classification technique based on Bayes theorem with an assumption of independence between predictors. In simple terms, a Naive Bayes classifier [1] first assume that the one particular feature is present in a class and that is unrelated to the presence of any other feature. Naive Bayesian model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to perform even more sophisticated classification methods. Algorithm also works for big amount of data or large data sets. It predict fast and multi class prediction can be

#### C. AVF

This algorithm is linear with respect to data size and it needs AVF [4] finds the outlier in categorical data this

method is is good to find the categorical data which discussed in [4]. In paper they have used the CRT for the rule generation to classify the data for the value  $k=40$  by using AVF algorithm.

#### D. SVM

This is supervised learning algorithm of classification. In this the total data set is divided into points in space. This space has the all data set mapped properly. The data is then divided by some function which properly chooses the maximum data falling into particular classes. Margin should be maximum in order to classify the data. The main issue here to select the proper hyper plane which divides the whole data.

#### E. CART

This is the type of decision tree models where the target variable can take a finite set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values are called regression trees.

#### F. Genetic Algorithm.

A genetic algorithm (GA) [2] is a method for solving both constrained and unconstrained optimization problems based on a natural selection process that mimics biological evolution. The algorithm repeatedly modifies a population of individual solutions.

It has the different steps of selection, genetic operation (crossover, mutation) and termination.

During each new generation the portion of existing population is taken for new breed. The way of doing it is by the fitness function. Then the operation performed on the selected data. Lastly the termination condition which varies on the different factors.

#### G. Fuzzy Classifier

This algorithm uses the fuzzy logic for the classification. It is rule based fuzzy classifier. It works with the probability and does not follow the strict boundaries like Decision tree, Naive Bayes or Random Forest. It uses the fuzzy logic and instead of saying the conclusion is completely wrong it might say it is 0.0001percent right. Example: If one student score 64 marks in data mining subject and other student have 65 which we decide it as the excellent score out of 80. And if the class labels are excellent in between 65 to 80 and other are 50 to 65 medium. So the thing is fuzzy classifier will not directly say the student is dull or medium which exactly should not happen.

### IV. ALGORITHMS FOR QUESTIONNAIRE

In this section we will discuss algorithm for the question generation. Generation of question papers through a question bank is an important activity in learning management systems and educational institutions.



### A. Randomized Algorithm.

There is Randomized algorithm for that purpose but system should not be completely randomized it should have some other characteristics. In GRE test the algorithm The Revised GRE does not adapt within a section, only between sections. Each question is weighted the same. Difficult questions and easy questions are randomly mixed throughout the section

The randomized algorithm will only pick the question randomly which can be good for the test but the depth of knowledge can not derived from that.

### B. CAT

CAT (Computerized Adaptive Testing) is use to pick the questions based on the student ability. CAT successively selects the question according to the student ability seeing what his/her performance is in the previous question. Depending on the student ability the test will be tailored and the questions will be fetched.

Working: It selects the questions from the pool.

Student can answer it right or wrong. Depending on that the system will fetch the next question. It continues these steps until the termination condition occur.

### C. Algorithm used in GRE

GRE exam is based on Computer Adaptive Testing. The test changes according to the question level have been answer. If you gave the answer of previous question the next question gets harder. It has five different section two verbal reasoning, two quantitative reasoning and either the experimental or research section. Five sections may appear in any order. GRE 2011 next section depends on the previous section performance. In [10] author used GRE to prepare adaptive questionnaire. With questions organized around a set of testable topics, the adaptive engine estimates a student's proficiency with maximum likelihood techniques and draws questions from conceptual areas where the student is weakest.

There is different parameters used in [9] author has stated different steps to generate the question paper. Unit-wise distribution of marks: It shows marks allotted for each unit in subject. Difficulty level-wise distribution of marks: It contains three levels easy, medium and hard. It shows marks allotted for each level. Distribution across cognitive levels of Bloom's taxonomy: Cognitive level of Bloom's taxonomy contains six levels such as knowledge, comprehensive, application, analysis, synthesis and evaluation. Question paper format: It is the format of question paper of the examinations and it shows the distribution of the marks across questions of question paper.

## V. CONCLUSION

In this paper we have discussed multiple methods for classification/prediction and question set generation. The system we required for the student classification based on

their performance in the test we require some sort of unbiased classifier which will not be strict. The classifier we identified is the fuzzy classifier which gives the classification based on the fuzzy logic which is not strict. Hence we can understand the student performance should not follow the strict class range as we discussed in the example.

## ACKNOWLEDGMENT

We express our deep felt gratitude to our dissertation Coordinator **Dr. A. S. Ghotkar** for her support. **Mr. Santosh Dwiwedi** for continuous guidance. I also thank to my parents for their belief in me.

## REFERENCES

- [1] Sabine Graf and Kinshuk, "Advanced Adaptivity in Learning Management Systems by Considering Learning Styles" IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technologies Technology-Workshops, 2009
- [2] Zhaoli Wang, Xinhui Tang, Delai Chen, "A Resume Recommendation Model for Online Recruitment 11th International Conference on Semantics, Knowledge and Grids, 2015.
- [3] Mohamed Amine Chatti, Simona Dakova, Hendrik Ths, and Ulrik Schroeder, Tag-Based Collaborative Filtering Recommendation in Personal Learning Environments IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. 6, NO. 4, 2013
- [4] Lakshmi Sreenivasa Reddy, D, M. Rao Batchanaboyina, D.V.V.S. Phanikumar, "Learning Styles Vs Suitable Courses" IEEE International Conference in MOOC, Innovation and Technology in Education (MITE), 2013.
- [5] Vijaya Bhaskar Velpula, S. Sailaja, "Finding Peculiar Students from Student Database using Outlier Analysis: Data Mining Approach", 2014.
- [6] Xin Chen, Mihaela Vorvoreanu, Krishna Madhavan, "Mining Social Media Data for Understanding Students Learning Experiences" IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. 7, NO. 3, 2014.
- [7] Dr. A. Padmapriya, "Prediction of Higher Education Admissibility using Classification Algorithms", 2012.
- [8] Silvia Quarteroni and Suresh Manandhar, "User Modelling for Adaptive Question Answering and Information Retrieval" American Association for Artificial Intelligence", 2006.
- [9] Vaibhav M. Kale, Arvind W. Kiwelekar, "An Algorithm for Question Paper Template Generation in Question Paper Generation System", 2013.
- [10] Julia Enthoven, "An Adaptive System for Standardized Test Preparation", 2015.