PERSONALIZED VIRTUAL SCHOOL ENVIRONMENT USING CLASSIFIER ALGORITHM AND SEMANTIC ADVISOR-ASSISTING FRAMEWORK

JAIMIN MANIYAR, JENILA LIVINGSTON LM*

School of Computer Science and Engineering, VIT University, Chennai, Tamil Nadu, India. Email: jenila.lm@vit.ac.in

ABSTRACT

As we knew that every individual is different than each other and their brain levels are also different, some students are very bright and some are not, so there is a need to develop such system which will teach them as per their thought process and learning habits. This research aims to develop a personalized virtual school environment for students which help them to learn the things same as they learn in school physically, but the thing here is each student will be treated differently as per his/her ability. Hence, this system would be a virtual school for any student, which resembles the CBSE school system in India. In this virtual school, all things are present such as teachers, homework, games, and exams which present in actual school except one which is the physical classroom because in virtual environment the course and things are personalized for each student as per his/her thoughts and brain level, so here the student can seat in his/her home and can learn at any time with help of computer. The system is a web-based environment and machine learning would be used for doing personalization on family and knowledge context of student. This web system contains four components: (1) Classification based on family context, (2) classification based on knowledge context, (3) learning material selection algorithm, (4) web-based learning system on top of above three, are discussed in this paper.

Keywords: Personalized e-learning, Virtual school, Machine learning, Intelligent tutoring system.

INTRODUCTION

From ancient to modern era education describes the key role in bringing up the society to next level. In ancient time, Indian student went to any sages' ashram for studying purpose. As time changes, the ashrams converted to school now student goes to school for learning all the things. But nowadays the technological advancements open whole new world called e-learning. In e-learning, student can learn anything from anywhere through web. The system is made in such a way that it should be useful for all students based on their interest. However, this is missing in both e-learning and school, i.e., we will not able to do the personalization when we are trying to generalize some students by considering their interest and profile. Hence, this research work provides a remedy for the above said problem.

Basically the ways of learning are varied for each individual because each one is unique. The Indian students learning style are affected by family background, school in which they are studying, where they live and so on. Hence, here we can design an intelligent tutoring system which will tutor based on individuals interest. Different pace of every student can be monitored and handled by web system without interrupting other students learning process [1].

For doing personalization in proposed system classification is done on students based on various criteria. In criteria, we basically set competencies [2] that can be reached by every learner. These competency criteria then drive to expect learning outcome. Here, the goal for learning system would be after taking tutoring from system the student group belongs to particular criteria will get almost same result in evaluation. To develop such a personalized learning system, we need to design an architecture which runs as per student level. The proposed architecture and system is expected to provide a way in enabling learner to reach expected outcome which compiles their characteristic.

BACKGROUND

This section describes various background study required for the personalization, machine learning, material selection approach, and education.

Learning material selection

For selecting any type of content, majority of currently available algorithm either based on information retrieval or any machine learning technique. Information retrieval approach finds the similarities between the groups of document or retrieves the information from the query itself to search the material. Machine learning algorithm was done some analytical and combinational evaluation on the database of document to find the appropriate material.

Standard-based education

Standard-based education [1] describes the competency of the knowledge and skill which is required by student at particular stage in learning. It also provides the structure of the learning curriculum. In other words, we can say it determines what type of material and learning style should be adopted. Now competency also acts as guiding mechanism during assessment and grading of student. We can say that assessment should be done with taking standard and competency in mind. We know that assessment tells us the level achieved by the learners with different competencies.

Personalization and techniques in personalized e-learning

To enhance the effectiveness of e-learning personalization would be very much useful. Personalization means for everyone who use that it should work according to person's interest that uses that environment. Personalized e-learning can be achieved by doing various type of personalization based on students' interest, profile and current schooling system. With this personalization can be done for each learner based on their learning activity, prior knowledge, etc. are different. Mudra and Jenila [3] developed a digital virtual learning environment for learning a Java course based on students interest using multiple intelligence and provided content navigation algorithm to provide content personalization.

Developing learner model

The biggest challenge in e-learning system is to make learner attentive for the learning content because most of the learner are not familiar to read from the computer screen. So to make learner more active during

the teaching time, a school teacher will ask something or give some live demo to student during the lecture. This can be done here also the system should detect that learner is active or not by applying some programming rules. So as per this learner model should be consists of such a bunch of information by which every learner should look at it with more concentration. Learner model can be made up of assumption about the learner, past experience of the learner with system and most important individual preferences of learner.

Personalization with learning activity

Learning Activity would be different for each student. It is because each student has different learning style, different context, etc. To achieve the personalization based on activity, we can record the each activity of student and then that activity can be used for predicting the student's activity. Moreover, we can make such type of learning activity which used various contexts at each activity step to predict next level of learning activity.

Analysis of student record to identify pattern in performance

Cross-industry standard process for data mining (CRISP-DM) model [4] can be used for analysis of pattern in student's performance. For analyzing student records for performance, the first step need is to understand the business. It includes the retrieval of business objective, assessing the situation, finding of DM goals, and producing project plan. The model building phase includes understanding of data for CRISP-DM and then preprocessing of data. And also in their approach, they had used PASW 13.0 modeling tool from SPSS to find the patterns in performance.

Random forest, C4.5 decision tree and XGboost tree

The above title describes the three methods which are used in proposed system. In general, the forest means collection of trees and wild animals. In terms of machine learning algorithm, the definition remains the same random forest is a bunch of decision trees and each decision tree has some feature for making decision [5] C4.5 decision tree [6] is an advanced version of ID3 tree developed for reducing of error pruning and it also improves the computational efficiency. XGboost is gradient boosting method used for classification and regression.

RELATED WORK

There are many e-learning websites and massive open online course systems are available presently on internet, e.g., Udemy, edX, and NPTEL. In current web learning scenario two different types of system are available: (1) Adaptive web system and (2) intelligent tutoring system. In adaptive system, the materials are given to student based on various context of student such as past studies and demographic data but not on feedback given by student on particular topic. In opposite to this intelligent tutoring system adaption is based on the response of student to particular material, i.e., both systems are individually trying to find some good content to teach but both have some problem. Hence, the new system can be proposed with an architecture which combines the part of both things student context as well as response of student.

Based on above discussion, there are two different system exists K12 education [7] and e-tutor [8]. E-tutor is a personalized e-learning system for college students and K12 education. The proposed system provides virtual school environment for students of kindergarten (KG) to 12th.

E-tutor

E-tutor is an intelligent tutoring system which learns the optimal sequence of teaching material to be shown based on the learners context and feedback. System uses regression to find the best material from the given material and past experience of student with that course. By using these two contexts, e-tutor algorithm search the appropriate material and student can learn from it. After learning again result and feedback is taken from individual student which is used for finding next material. In this system, the decision taken for material selection is based on the

part teach to the student and it also depend on the previous knowledge of student on same topic which he/she is learning right now. This type of approach fails if we talk about the school student who does not know anything about the current topic. For this kind of problem, we should think about a solution which will not relies on students' knowledge on current topic instead it should be more user friendly, so they can learn easily.

Working of e-tutor

E-tutors goal is to find the optimal sequence of teaching material for student based on student context related to topic and feedback to past sequence of material. To achieve this thing regression is used on the sequence of teaching material and in e-tutor the exploration and exploitation made balanced to find material via regression. E-tutor uses the sample mean reward estimate for current student with the help of final exam score of student with context X on the same sequence of material. Now how exploration and exploitation keep balanced in system? For that e-tutor asks the question and receives the feedback. After that, some parameters of e-tutor are taken if it holds for the current student than by using that e-tutor will give final exam and final obtained score to the regret equation made for e-tutor.

Result achieved by e-tutor

To check pilot performance, e-tutor was deployed in one of the university environment to teach the digital signal processing. This pilot study was made on students who already learned the course some years ago. The students were refreshing the concept of discrete Fourier transform by the system. With three type of different material in sequence, the topic was taught to batch of 500 students in online system. With help of e-tutor algorithm, they achieved 15% more average than random rule method or fixed rule method to select the material.

K12 education

K12 education is a web-based virtual school system based in US. This system brings schools to the student's bedroom. K12 education is providing virtual school for KG to standard 12th as per US schooling system. There are so many schools which provide free as well as paid virtual schooling in system. They design a school working in such a way that it will fit to students habit and convenience. Same as physical school all things are divided in to lectures and student can learn on their own pace.

History

K12 was found in 1999 to answer a call to growing need of US parents about their children teaching. Parents' problem at that time was the traditional model is not suitable for their children. Hence, K12 is started to provide tailored course and one-to-one online learning to students across the country. Their major mission was to transform the style of learning by which each student learn something.

How K12 provide online learning?

K12 offers best in class curriculum that enables mastery of core concept related to any topic or skill. They have wide range of courses for different type of students such as virtual academies, insight schools, and blended school and they have so many supplement courses also. Now to learn all these courses you need just a computer and internet connection. Now online lessons and interactive activities are given to students by OLS in K12. At last student is assessed based on standardize test.

LEARNEASE

LearnEASE is a proposed web-based personalized virtual school which would be a combination of above two systems e-tutor and K12 education. LearnEASE is an online virtual school like K12 education based on Indian CBSE schooling system which uses personalization in backend. For doing personalization in proposed system machine learning is used effectively. Architecture for LearnEASE is mentioned in Fig. 1.

Detailed architecture

In LearnEASE architecture, we have three different entities such as student, tutor, and teaching system. Student is a main entity who interacts with the teaching system directly. Teaching system is a web system which teach student with different e-learning methodologies. Tutors are entity who will provide material in repository. Now teaching system give appropriate material from study material repository made by tutors and teaches that to student. Student will learn from the e-content projected and after learning a topic, student is being evaluated. The student score is stored in the database and used for further process.

Background context

Basically background context is repository of student's background. This repository is used as an input for context classifier algorithm. This repository basically contains the demographic data about student such as name and age. Fields for background context is listed in Table 1. Background context would be updated whenever new student enrolled in the system.

Study context

Basically study context is repository of students past study record and marks scored in tests taken by system. This repository is used as an input for context classifier algorithm. This repository basically contains the data about student such as past performance and IQ. Fields for study context is listed in Table 1. In study context, there is one more parameter e-learning ready that parameter is used to check whether student is ready (know) to use the web-based learning system completely or not. This parameter is included because the propose system targeted school students as users.

Context classifier algorithm

Context classifier algorithm is a classification algorithm which takes both background context and study context as an input. Here, context classification outputs a level of student, first level classification from level 1 to 5 as given in Table 2. For context classification proposed system uses random forest classifier based on C4.5 decision tree [9]. Random forest is an ensemble learning method based on decision tree which is used for classification and regression. Here in proposed system C4.5 decision tree is used as base in random forest. C4.5 is an extended version of ID3 algorithm and also referred as statistical classifier. Here, all context data (demographic data) are there which is use as an input so C4.5 would be best for taking best feasible decision. Random forest can gives best classification when we have more number of categorical variables in data so combination of both is used in context classification to achieve best results.

Final classifier algorithm

Final classifier algorithm is a classification algorithm which takes output of context classifier and students past knowledge data as an input. Here, final classification outputs the final level (second level classification) of student from level 1 to 5. For final classification proposed system will use XGboost classifier [10]. This is required because demographic data (context data) affect the student's performance and produce the actual student level as an output. Actually this classification is required because as per the student overall knowledge and some other parameter, we can check how student is good at thinking related to study. XGboost algorithm [9] is a gradient boosting algorithm and gives best classification when we have more number of numerical variables in data.

Learning material finding algorithm

Now after completing both classifications, we know the level of student ranging from level 1 to 5. The system has the repository of study material ranging from level 1 to 5. The job of learning material finding algorithm is to find the best material from the repository as per the students level and displayed to the student. For designing this algorithm, the proposed system uses semantic advisor-assisting framework to select learning materials [11]. Hence, if there exists more than one material

Table 1: Fields for background and study context

Background context	Study context
Age	E-learning ready
Gender	Pretest score
School	IQ
City	Chapter 1-20 score
Social status	Total
Mother qualification	
Father qualification	
Family income	
Siblings	

Table 2: Levels description

Level	Description
Level 1	Expert
Level 2	Very good
Level 3	Good
Level 4	Poor
Level 5	Very poor

(text, audio-visual, slideshow) for same student level then score will be given by the framework to material and the framework plays the key role in assigning the material. After assigning material, it goes to the teaching system.

Semantic advisor-assisting framework

This framework [11] does the semantic and syntactic analysis of summary and title of the material which is uploaded by the advisor or tutor of the course. Hence, basically this framework does the semantic analysis to find the best material. This framework first processes the title and finds the keywords from that and at the same time it also performs syntactic and orthographic processing for summary. Then, the framework uses BPMRC method to categorize the keywords and do the scoring. Based on the score any material is chosen for teaching a student.

Teaching system

Teaching system is a web-based system which used for teaching a course to a student. Teaching system works differently for all individuals. Teaching system works based on the material assigned by the previous step. After teaching phase, the small evaluation is done to check the knowledge of the student.

HOW PROPOSED SYSTEM WILL WORK?

The proposed system works actually as a web-based system. The LearnEASE system works as web-based module to run properly and effectively all the parts such as classification, datasets, and learning material finding algorithm. The basic running methodology is depicted in Fig. 2.

Initial setup - find student context

As part of the initial setup for running the system and and to train the model for machine learning algorithm (classification) the system needs a well-defined dataset. Hence, data collected from the local schools which follow the CBSE school system in India, and then data are prepared as per fields of the background and study context. The data are preprocessed with the help of standard data preprocessing algorithm.

Apply classification - find brain level

After data are preprocessed now the part comes to train the model for classification algorithm. So based on data the processed model is trained. Initially, the web system finds the students level when any new student enrolled to the system. Here, model is also retrained time to time. When initial student improved her/his ability, the model is
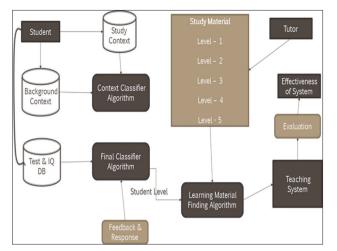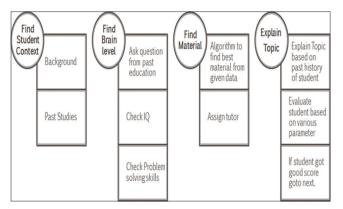
Fig. 1: Architecture of LearnEASE

Fig. 2: Working methodology of LearnEASE

retrained based on the new data. Hence, the system will become better and better as student improves his/her ability.

Setup web system

Now after data preprocessing and initial classification model, we can setup a web system. In setting up a web system, we need following things. (1) A best user interface (UI) with good user experience (UX), (2) learning material finding system, (3) teaching system to teach different type of study material, (4) an evaluation system specifically designed for LearnEase architecture, and (5) A response and feedback system.

UI and UX

The most important part of any web system is UI and UX so for the LearnEASE the UI and is designed in such that it is comfortable to student as well as student gets attract to learn from the system.

Learning material searching

A web-based algorithm needs to be designed for the semantic advisor framework which stores and retrieves the data from web system database server. To convert the framework algorithm into web-based algorithm, we need a web service to help the LearnEASE system to choose the right material. Hence, when we request to the web service with the student level and other parameters it produces the appropriate output which helps the next step to run or proceed.

Teaching system

Teaching system gets input from previous step. Now task of this module is to run the material, i.e., teach that material through system. However, learning material can be of different types such as portable document format, video, audio, and slide show, etc., so system takes care of the type and start the teaching.

Evaluation system

After the teaching process is completed by the teaching system student will be redirected to this system. Task of this system is to evaluate the student to check whether student get the idea what is been taught. The evaluation is done based on various parameters such as (1) knowledge checking, (2) time for learning, and (3) ability of the student to relate the content with real world or student can implement or explain same things in simple word.

Response and feedback system

Response is based on various activity of the student that automatically stored in the database and then that data are used for making the changes or in existing system methodology. Feedback is taken time to time revision from the parents of students also.

Assessment system

Now at end we need a good assessment system which assesses the student based on various parameters. From this, we can achieve the effectiveness of LearnEASE system. For doing assessment LearnEASE system focuses on the practical approach of assessment after teaching a learning material. The assessment module is designed in such a way that it will relate to the real word problem and based on that test. Now score achievement in this assessment plays important role for next material to be shown to the student. The ultimate goal of the system is to know the achievement mastery of the student on a particular topic.

CONCLUSION

In this paper, a state of the art online education system is proposed which teaches student based on their brain level, past knowledge, and background details. Above said parameters help the proposed system to find best suitable materials for student. After finding the material online virtual teaching system teaches the student as per the material. Proposed system also able to adapt the new things based on feedback and responses of the student.

REFERENCES

1. Laksitowening KA, Hasibuan ZA. Personalized e-learning architecture in standard-based education. International Conference on Science in Information Technology (ICSITech); 2015. p. 110-14.
2. Nitchot A, Gilbert L, Wills GB. Towards a competence based system for recommending study materials (CBSR). 10th IEEE International Conference on Advanced Learning Technologies; 2010. p. 629-31.
3. Mudra S, Jenila LL. Automatize content personalization by agents in digital learning of java course. Int J Appl Eng Res 2015;10(9):23063-71.
4. Hoe AC, Ahmad MS, Hooi TC, Shanmugam M, Gunasekaran SS, Cob ZC, *et al*. Analyzing students records to identify patterns of students' performance. International Conference on Research and Innovation in Information Systems (ICRIIS); 2013. p. 544-47.
5. Breiman L, Cutler A. Description of Random Forest. Berkely University; 2016. Available from: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm.
6. Hamilton H. Description of C4.5 Decision Tree - Tutorial. University of Regina; 2012. Available from: http://www.2.cs.uregina.ca/~dbd/cs831/notes/ml/dtrees/c4.5/tutorial.html.
7. K12 Inc. K12 - Home Page. Information for K12 project; 2016. Available from: http://www.k12.com.
8. Tekin C, Braun J, Schaar VM. eTutor: Online learning for personalized education. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2015. p. 5545-9.
9. Moitra S. Random Forest Classifier based on C45 Decision Tree. Project Work - University Institute of Technology, Burdwan. Available from: https://www.github.com/xw19/RandomForestClassifier.
10. Atsushi K. Xgboost Predictor Java. XGBoost predictor for online prediction tasks, Project work – Smart News Inc., Tokyo. Available from: https://www.github.com/komiya-atsushi/xgboost-predictor-java.
11. Mahmoudi MT, Taghiyareh F, Rajavi K, Shokri F, Khamnian L. Semantic advisor-assisting framework to select learning materials. 6th National and 3rd International Conference of E-learning and E-teaching; 2012. p. 59-64.