Designing an Automated Intelligent e-Learning System to Enhance the Knowledge using Machine Learning Techniques

G. Deena\(^1\)
Research Scholar, Computer Science & Engineering
Sathyabama Institute of Science & Technology
Chennai, Tamilnadu
India

Dr. K. Raja\(^2\)
Department of Computer Science Engineering
Dhaanish Ahmed College of Engineering
Chennai, Tamilnadu
India

Abstract—The modern digital world requires its users to learn continuously in order to enhance their knowledge in the working environment and the academic sector. This kind of learning is significantly facilitated by the E-Learning platform, which is better than the traditional methods. As E-Learning offers benefits like time and space independence, many learners have made it their choice. However, since an abundant of E-Learning courses are available on websites, learners are confused as to which is the right one to choose. This paper proposes an Automated Intelligent Learning (AIL) methodology which covers the entire Teaching-Learning Process (TLP) to overcome this issue. It enables the selection of suitable topics and framing an appropriate course syllabus and assessment questions for the users. In it, the learner satisfies topic selection based on Bloom’s taxonomy. This enables high-quality knowledge outcomes in the learner. The subject curriculum is framed by using Hierarchical clustering techniques. This helps the user to fix suitable topics and conveniently generate questions using machine learning techniques. The proposed methodology was evaluated by carrying out post and pre-assessment tests on undergraduate students from computer science courses. The performance analysis of the proposed methodology was compared with that of the existing methodology. It was observed that the proposed methodology is effective in applying the topic selection hierarchical method to make a perfect syllabus for the course, and assessment questions. Besides, it was found to enable the learner to learn without any confusion or distraction.

Keywords—e-Learning; teaching learning process; pre-assessment and post-assessment; blooms taxonomy; machine learning

I. INTRODUCTION

In the traditional educational system, the teaching-learning process takes place under the teacher-centric approach, where the teacher delivers as per their preferences. The impact of this method leads to frustration and a lack of interest in the student. Therefore, many students lose or discontinue the course. Later, the online system was developed to solve this issue. Learning has taken a gigantic transformation from classroom teaching to the online learning system. Information Technology has given a drastic change in the educational system by employing Information and communication technology (ICT). For a decade, research has been directed towards enhancing the teaching-learning process with the support of ICT. The modern teaching-learning process is commonly known as Intelligent Tutoring Systems, M-Learning, Pedagogical Agents, U-Learning, and Trainer Tutors [1].

In both the classical educational system and the E-Learning system, a human teacher understands the learning skill of the learners, and based on this, the skilled teacher trains them. However, the teacher has to spend a lot of time to understand the learners in the classical educational system [2]. The Intelligent Tutoring System (ITS) is designed in such a way that it helps to determine the cognitive level of a learner [3][4]. Thus, it teaches the learner with the highest preferences set by the learner.

Now- a- days, all the learning are being moved towards the electronic mode, where the learners feel highly comfortable to learn. The users are provided with the option of making additional choices of courses with the specific and appropriate syllabus. In the traditional system, users took up the classroom teaching i.e., the teacher-centric approach. This consumed a lot of time for enhancing knowledge due to the lack of resources and technology. Further, not all the users can avail of this kind of teaching system to enhance their knowledge, owing to the cost and time factors. Thus, there is a gap between the learner and the learning system in the traditional system. Today, the transition from the traditional to the E-Learning system has made it comfortable for the users to enrich their knowledge [5].

However, even after moving to E-Learning methodology some limitations are faced. If the users are learning the course through the web, they are exposed to unfavorable situations like the inability of course selection. This is due to confusion related to whether or not they are in the correct path of the course syllabus. To make E-Learners more comfortable, bloom-taxonony has been introduced to check whether the syllabus of the particular subject is rightly framed from the basic level to the application level. It ensures that the syllabus is suitable for the learners to continue the course without any confusion.

E-Learning is a universally popular form of online education system widely accepted owing to it’s anytime anywhere learning benefits. Besides, it is also profitable to the
management and the students. The learner can learn the content either individually or collaboratively as a team member. The collaborative method is the usual method selected by the learner, where they select the course as per their choice [6] to enhance their skill. Here come the domains like artificial intelligence, web ontology, and machine learning, which are emerging not only in the learning system but also to all sectors [7] and [8]. E-Learning is now moving towards the automated and integrated intelligence fields to make the user more comfortable in learning the course.

E-Learning is a learner-centric approach and a highly enabled computer-based system for enabling learning and knowledge enhancement. E-Learning has transformed the educational paradigm through its recent technologies. This learning system may be in any of the specified forms such as Virtual Education (VE), Web-Based Learning (WBL) or Computer Based Learning (CBL). The learning content is supplied via the tape-recorded format which may be in audio, video, text or image set-up, and can be supplied either in the CD form via the internet.

In this current era, the data storage in internet technology increase day by day. It is not easy for the learner to retrieve their specific data and information. In [9], Zhang proposed a personalized E-Learning system based on Google Web Toolkit to avoid confusion during the retrieval of information from the internet.

In the Automated Intelligent E-Learning system proposed by this paper, the preferences in the learner model are given to the learner’s choice, the learning material is held by the domain model and the assessment of the model generates the assessment questions.

The rest of the paper is organized in the following manner: Section II discusses the related works, Section III presents the system architecture, Section IV presents the performance analysis, and Section V deals with the conclusion and future enhancement.

II. RELATED WORK

In research filed, many researchers have been currently focusing on the personalized adaptability learning system, which was initially given by the author Brusilovsky & Peylo [10]. In this system, these authors suggested the determination of the cognitive factors and learner’s learning style by a psychological test [11].

As per the adaptive educational system, the personalization of the subject for each student is provided as per the individual’s knowledge level, skill, needs, and background. The adaptive system supports and guides the learner to enhance knowledge. Currently, many systems are designed using intelligent techniques and methods [12].

Artificial intelligence has several concepts related to Naive Bayes rule, Fuzzy Logic rule, etc. These can also be integrated with the frequent words search with relevance to form an appropriate course syllabus [13] and [14].

An instructional design process was used to assess the cognitive and learning style of a learner for web-based learning [15]. Personalization was enabled in an online educational system by designing the system to adapt the learner characteristics through ontology for the development of the content. This system collected the personal data of the learner and gave the recommendation to the teacher [16].

Apriori Algorithm was used in recommender E-Learning System to find the association between the features of the searching key element of the individual learning model. [17][18][19][20][21].

In [22], the author used an agent called mod-knowledge to completely track the learner’s knowledge, and the internal and external structure of the learner using the machine learning algorithms. Authors in [23], called the cognitive skill as thinking skills. Higher the domain cognitive skill, the easier the problem-solving ability is. The sequence or the learning path would vary on the basis of the learner’s performance.

The learning styles of an individual may also vary due to factors like the learner’s mindset, the learning time, and the complexity of the course content. Learning Styles deliver a more prominent effect on the learning exhibitions, particularly in E-Learning situations.

In [23], authors Tee, T. K. et al., utilized the Learning style index designed by Felder and Silverman to examine the learning styles of students in a Malaysian vocational college’s Business Management and Hospitality programs.

In [24], L.M. Al-Saud et al. proposed a methodology to decide the favored learning style of dental students at Saudi Arabia’s King Saud University located at Riyadh. The experimental outcomes stated that there were multimodal learning inclinations in more than half of the students.

The domain model was constructed using fuzzy logic in [25]; the personalization model was adapted to fetch the learning content as per the learner’s preference using the similarity index between the prerequisite model and the domain model. The subjective style was characterized as a person's predictable way of dealing with sorting out and handling data amidst considerations [26] [27].

An author in [28], applied the data mining techniques of the Bayesian network to discover the knowledge of the domain model. The computer system learner interacted with the learner to understand the knowledge level of the learner. The Bayesian network was utilized to design the student model and the domain model.

An existing methodology has shown the easy way to help continuous learning in an employee of an organization, but still, there is a lack of understanding of the concept [29]. The present and future situation of the organization seeks to improve their business further by making the learners continue their learning without any distractions [30].

Most of the existing e-Learning courses are available with multiple course contents, but the focus has not been directed on how effectively these help the user in enhancing knowledge.

In the proposed system, the complete teaching-learning process has been designed to determine the student’s
knowledge level to train them effectively using Bloom's taxonomy followed in the domain model. Assessments were carried out before and after the course to determine their level of understanding of the concept.

III. SYSTEM ARCHITECTURE

To make the learner a more efficiently knowledgeable person, we have developed an Automated Intelligent E-Learning a system, which covers the complete set of the teaching-learning process. It identifies the cognitive knowledge level, frames the syllabus, and conducts the assessment to the user. The learner model accesses the students for their flexibility of accessing the learning content. Learning contents are developed by experts as per the hierarchical clustering techniques. Each file is represented as the learning element. The target of the assessment model is to generate the procedural and declarative question using Natural Language Processing techniques. The proposed system's architecture is given in Fig. 1.

A. Learner Model

Wolf has identified the Student model as a significant model in Intelligent Tutoring System (ITS). The learner model is an important part of adaptive E-Learning systems. This is commonly referred to as a user or student model. Generally, the learner has to spend too much time in identifying the appropriate learning content. In general, two types of information are collected from each learner, namely the domain-independent and domain-dependent information. Domain-specific the information pertains to the learner's subject knowledge. The pre-requisite knowledge of the subjects has to be identified and stored for further process. The domain-independent knowledge implies their cognitive behavior, their style of learning, their current understanding capability level, and their preferences of learning. In [31], author Paramythis & Loidl-Reisinger specified that this model sustains the learner's account. In the online educational system, the learning style improves the learner model.

1) Cognitive style: Cognitive learning styles are data handling tendencies of a person. Individuals are different in their capacities. Cognizance depicts a man's normal capacity of seeing, reasoning, the critical thinking and recalling. The subjective style is usually depicted as an identity measurement that affects the qualities, social collaboration and states of mind, and can characterize as a person's predictable way to deal with data [32].

The cognitive skill is an essential part of the process of learning information. The author in [33], states that cognitive style implies the mechanism of processing the received information into knowledge and that it is totally different for each individual. When the cognitive level is high, the learner concentrates more on learning the subject at high speed. Otherwise, there are chances of quitting the course.

In the proposed system, the cognitive skill of the learner has to be identified and based on that skill; the content will be supplied for further training or coaching of the material. Before taking up the course, learner pre-requisite subject knowledge has to be identified by conducting a simple assessment. The assessment questions are generated automatically as per Bloom's order. Bloom's taxonomy is the best classifier that classifies information into two viz., the lower level and the higher level. Lower level covers factors including Knowledge, Understanding, and Application; and higher-level covers Analysis, Synthesis and Evaluation of the concept. The Learner model is given in Fig. 2.

![Intelligent E-Learning System](image)
In [34], the author used the demographic information of the learner, specifically the name, age, experience, and qualification. However, the proposed system, the learner characteristics like learning style and subject knowledge are considered. A pre-assessment was conducted using a familiar diagnostic method like multiple-choice questions to check the knowledge level, but not to classify learners. When a learner crosses the threshold level of the mark, they will proceed to the next level of the material. If not, the current topic will be continued to refresh information and give the learner more clarity of the content.

Each learner has an individual learning style that has the most important role in the learning process to help learn the material more effectively [35]. Felder-Silverman [36] is the most appropriate method used in the E-Learning system to determine the learning style. This method classifies learners into any one of the styles including the active, the sensing, the visual, the verbal and the sequential type learner.

2) Felder-Silverman learning style: Styles are the psychological facet of the learner which enables learning. This is important in the online educational system for enhancing the technology supporting system. The learners grasp and process the information in different ways. Felder-Silverman model uses any of the techniques to motivate learning, to generate solid information, explicit illustration, provide balanced material for practical problem solving, computer-assisted instruction and scientific method [37]. Once the learning style has been determined, suitable content can be mapped to the learner. Felder suggested four types of learning styles, of which the proposed system includes two viz., the visual and the verbal for our system.

- Active Learner: Learner recognizes the information only by doing the concept learned from the learning material. This type of learner joins a group to work in it as a team member.
- Sensing Learner: Learner prefers to consider the solid learning material and facts. They like to solve the problem in fixed methods, are patient with available details, and are fine at learning by performing hands-on work and rote facts.
- Visual Learner: Learners learn the best by way of what they have seen. They focus more on study materials like demonstrations, pictures, flow charts and visual aids.
- Verbal Learners: Learners learn best from vocal and written explanations delivered by experts. They like to follow storytelling techniques during learning.
- Sequential Learners: Learners learn in a sequential ladder and progress in a linear fashion. Even at times when they have not completely comprehended the concept, they still are capable of doing something with the concepts learned.

Our Proposed System has been developed for the visual and verbal type of learners, who are higher in number in the online educational system. The readily available printed materials are given as the source. The Learner characteristics are identified from the cognitive skill identification and Learning style using the pre-assessment. Once the learner characteristics are identified, the learner becomes ready to learn the suitable content without any disappointment and boredom. The learner’s knowledge of the subject will be thus enhanced, and that can be compared with the existing methodologies.

B. Domain Model

The central part of the learning process is the domain model. The learning materials are organized in a specific manner, where the learner's choice may be the text, image or video format as per the learning style predicted by using LSI. The materials are developed by a domain expert and converted into the digital format. The materials are classified as per bloom's taxonomy from the low-level content to high-level content.

Hierarchical clustering is the unsupervised algorithm it contains a set of a similar item in a group. The syllabus is organized using the hierarchical clustering technique as follows. The similar topics come under the subject, and the sub-topics are related to the main topics, which form a cluster. As per hierarchy, the learning material is divided into units, the units are divided into topics, and the topics are divided into sub-topics, and so on as given in Fig. 3. This type of organized material is stored in the database, where each content is represented as an object. The object is utilized whenever there is a need for the material.

Learning content accessed by the learner is as per the learner characteristics [38] such as the knowledge level, the cognitive style, and the learning style. This way, the basic concept of the subject is well understood by the learner, and the learner keeps continuing to the next level without any frustration owing to learning. Before and during the learning process, a diagnostic assessment will be conducted for which the important sentences have to be identified automatically using the Latent Semantic Analysis (LSA) [39].

In our proposed system, the content of topics is stored in a hierarchical order. This content undergoes some pre-processing to get rid of the stop words and the least occurring words from the input file to generate some meaningful questions. The domain model has to follow the information retrieval techniques such as Sentence Scoring, Input Matrix and Singular Value Decomposition (SVD).
1) **Latent semantic analysis:** Latent Semantic Analysis involves the application of natural language processing and the techniques of information retrieval. This unsupervised algorithm helps in coming up with the most significant sentences from the course material without any human interference.

With the supplied keywords from the phrases, it helps in determining the relevant learning content with a high degree of correlation to the target sentence. Arithmetic processes are used to calculate the similarity of the sentences from the material with the support of Singular Value Decomposition. Analyzing the relationships between a set of words and a set of concepts is done by Latent Semantic Indexing (LSI). It also helps in judging the quality of the content in the sentences.

2) **Singular value decomposition:** Singular value decomposition (SVD) matrix finds the correlation between each word in the sentences by analyzing the phrases and their relation to the learning material. The original matrix is created from the orthogonal factors and large words in the input file by linear combination. The original matrix is decomposed into three matrices:

\[
A = USV^T
\]

A is the m*n Matrix, row and column value
U and V is an orthogonal matrix
S is the diagonal Matrix

3) **Sentence scoring:** Sentence tokenizer has been used to frame individual sentences from the input topic file. Sentence scoring has been applied to evaluate the similarity between the two non-zero vectors of the words by using cosine similarity method. The value of each sentence is allocated using the number of occurrences of the words in the sentences. For every sentence, the vector coordinate is evaluated with the help of the following formula.

\[
\text{Similarity}(q_n, d_n) = q_n \cdot d_n
\]

The similarity of the sentences is supplied to the input matrix to generate the matrix.

4) **Input matrix:** The original input matrix is represented as ‘d’. It is decayed into three individual sub-matrices viz., U, V and Σ. Now, the query matrix is raised with the available terms, the keywords and the noun. The most important word used for summarizing the material is the noun. Query matrix ‘Q’, consists of the significant noun related terms, is given by

\[Q = [n_k1, n_k2... n_kn]\]

Term Frequency (TF) of the word is set in the binary format, where ‘1’ fixes for the word’s presence in the sentence, and ‘0’ for the nonexistence of the word in it. Based on this, the important sentences are identified and supplied to the assessment model.

The diagnostic assessment questions are generated automatically using the machine learning techniques in the form of multiple-choice questions or fill in the blanks and the procedural type. Sentence selection is the critical point in framing the questions.

**C. Assessment Model**

In the educational system, assessment plays the highest role in determining the learning skill level of the learner. Assessment can be of different types, typically the procedural and the declarative. Declarative types of questions are highly used to test the learner’s basic skill level. The learner will answer the question in one word. The questions will have multiple choice answers or fill up the blank space form of questions [40]. This type of question is easily implemented to test the skill level of the learner to check whether he/she belongs to the low level, middle level or high level.

Each time before taking the course, the learners are advised to take up the pre-requisite test to know their basic knowledge in the subject. Using Bloom’s taxonomy, pre-assessment questions are generated from low order to high order thinking [41]. In traditional E-Learning, assessment questions were framed by a teacher who takes much time for it. Since the learner has to wait until the teacher evaluates his knowledge level and creates a question paper, it leads to time complexity.

To avoid time complexity like that in the existing methodologies, the proposed system generates questions automatically using machine learning techniques without any question bank concept. Each and every time the learner fetches the new set of questions without any human intervention. Since the answers to the questions are also evaluated immediately, immediate identification of subject knowledge is facilitated. When the learner’s score crosses the threshold level, they step into the next topic. Otherwise, the current topics are reinforced. The process continues to the end without making the learner lose interest or get disturbance.
In the proposed system, an input file is fetched from the domain model and sent for the document preprocessing, where the entire file is split into the individual sentences, and the stops words such as ‘is, was, be, of, on, because and so on are removed. With the available list of words, the keywords, which may be the most frequently occurring words in the sentence, are identified. And finally, noun identification, which is the most important part of the preprocessing, is done. This noun is used to frame questions to test the skill level of the learner.

The problem to be solved is known as stem, which is nothing but the assessment question. The machine learning techniques are used to generate the Fill up the blank questions and the multiple choice questions (MCQ) [42]. A new set of questions are generated every time using the randomization algorithm.

1) *Declarative type question*: There are two types of declarative questions in general, viz., Fill up the blanks type and Multiple Choice Question (MCQ) type. In the proposed system, the noun is replaced by the blank space in the fill-up the blank. A sample question is given in Fig. 4.

In Fig. 4, the input sentence is supplied to the question generator, where the preprocessing takes place, and finally, the important noun keyword is replaced by the blank space to generate the assessment. Each time the learner gets a new set of questions to test the skill level.

2) *Multiple Choice Questions (MCQ)*: Another declarative type question is MCQ. Stem is selected by sentence tokenizer and sentence scoring; and the distractors are generated by Named Entity Recognizer (NER), which is a Natural Language Processing (NLP) Technique. NER splits the information into small pieces as an entity name and entity type. `tagged_tree.leaves()` and `tagged_tree.label()` are operated to identify the chunked data. Noun filter is used to extract the noun, and Rapid Automatic keyword extraction (RAKE) is used to extract the keywords for generating the distractor and the correct answer. The keyword or the noun is the correct answer, and the other distractors are selected by entity type and their entity names. The answers are instantly evaluated, and the marks are shown to the learner. The marks are not to classify the learner, but help to them upgrade their knowledge of the learning content. The sample MCQ is given in Fig. 5, which is implemented in python language.

![Fig. 4. Fill up the Blank Question.](image1)

![Fig. 5. MCQ Generator.](image2)

**IV. PERFORMANCE ANALYSIS**

In the proposed Intelligent E-Learning system, the learner’s attributes are identified via the individual’s cognitive skill and learning style in the learner model. The domain model keeps the syllabus in the hierarchical clustering method, which helps the learner to fetch the content easily. The relevant topics are grouped together as subtopics. The material is processed using Latent Semantic Analysis of NLP techniques to identify the important sentences, and to generate the questions. Finally, the assessment is automatically carried out without any support from the teacher or trainer, by employing NLP techniques. As a result, the suitable course content is supplied to the learner as per the bloom's hierarchy. The learner continues the course with full interest, and the subject knowledge in the particular domain is highly increased as compared to the existing traditional E-Learning systems. The satisfaction level is highly increased by the proposed system. The system was tested with the different categories of school students, and the performance is shown in Fig. 6. It shows the accuracy of the proposed system is increased.

![Performance Analysis](image3)

**Fig. 6. Performance Analysis.**
Each learner is assessed with three different learning systems such as the traditional teaching method highly teacher-centric, traditional E-Learning as student-centric approach and finally with proposed system. The different learners are given in the x-axis; their marks are given in y-axis for the grade scale of 5 point. Each learner is trained and assessed with all type of teaching method and their result is given in the performance analysis. Finally, achieved the learners are highly satisfied with the proposed system.

V. CONCLUSION

The proposed AIL methodologies have overcome the problem of learners in selecting the course material by determining the learner’s cognitive skill and learning style using machine learning techniques. The syllabus was framed in hierarchical ordering, and the assessment questions were generated automatically using machine learning techniques. The subject knowledge and satisfaction level of the learners are highly increased through the proposed system. From the analysis, we have concluded that by using the proposed methods, the course outcome can be enhanced in terms of better understanding and suitable topics selection as per the blooms taxonomy. In future, the E-Learning system can be designed using any other advanced method like big data, deep learning automatically, etc.

REFERENCES


