Outcome Based Predictive Analysis of Automatic Question Paper using Data Mining

Simranjeet Kour Bindra
Research Scholar
Department of CSE
GNDEC, Ludhiana
simranjeet.kour74@yahoo.com

Akshay Girdhar
Professor
Department of IT
GNDEC, Ludhiana
akshay1975@gmail.com

Inderjeet Singh Bamrah
Assistant Professor
Department of IT
GNDEC, Ludhiana
inderjeetsinghit@gmail.com

- *Abstract - In recent years, data mining had become very popular to provide a facility in various fields. Large amount of data elements are stored in the database. Various data mining algorithms are used to classify the data elements and perform some operations to find the best solution on the basis of some relational parameters. One of the automatic architecture designs in this research is to find the solution for examination in terms of level wise question paper. The entire examination process is a vital component for direct assessment of an individual learning. So, preparing a complete test paper and the setup is relatively necessary. Currently, the customary technique of making question paper has been handbook. In current scenario the question paper generation is a manual approach leading to unproductive at times owing to bias, repetition and security concerns. The current paper presents an automatic procedure of question paper group which can be modified, streamlined, synchronized and secured. Each task done by this system is automatic, such that storing space, bias and security is not an apprehension any longer. Earlier, the question paper was generated by concerned subject teacher manually and was very time consuming, man power was required and sometimes the question paper lacked accuracy. Outcome Based Education (OBE) designates what students will know and be intelligent to do, as they advance in a program. The information related to student learning was collected for question paper generation and assessment using OBE. This information can be used to predict students ability, advancements in education system, betterment in teaching method, future interest of student etc. Various data mining classifiers like oneR, ZeroR, J48, Naive Bayes, IBk* **are used** *for prediction of the course outcome. The comparative study contains some parameters like time, detection accuracy, classification error etc. to find the performance of the proposed architecture as compared to other techniques to verify the enhancements.*

Keywords – Course Outcome (CO), Program Outcome (PO), Classifiers, Blooms Taxonomy, Outcome Based Education (OBE).

I. INTRODUCTION

Online system turns into a fast growing education method because of its speed and accuracy. Many companies/institutions started conducting their objective test online. This requires less man power and saves time. Through online system, one can evaluate students' performance. Automatic question paper generator helps various school, colleges and universities to automatically generate question papers [1]. This study proposes automatic question paper generation system in which each question paper will contain different set of questions from the database. It helps to generate question according to different levels of Revised Blooms taxonomy. The software performs the task of paper setting according to predefined instruction and generates a word document containing questions which can be downloaded.

Education model is used as a proposal design, i.e. to integrate short term course studies, to implement the instructional terms and to authorize teacher in the classroom. There are generally four models[2] of teaching which are as described below:

- *Information processing model:* Information model deals with the capability of the learner to observe, examine data, to figure out information, design new concepts and solve problems.
- *Personal development model:* This model focuses on the emotional life of an individual and helps the individual to be confident, skilled personality.
- *Social model:* This model mainly pays attention on the relationship of a person with other persons and society.
- *Behavioral model:* This model deals with the behavior of the person.

Question paper is an important tool for direct assessment of a student with the advent of revised

Blooms taxonomy [3]. It becomes easy to perform the assessment based on cognitive. The research introduces a novel approach for automatic question paper generation which further performs outcome based predictive analysis using data mining.

II. OUTCOME BASED EDUCATION

Outcome Based Education (OBE) is a process that focuses on what is to be learned as a result. OBE is not an execution, a set, a method, a craze, a quick fix, a cure, a wonder or an action but it is a transformational way of doing business in education. The different views of OBE are shifting educational program from teaching to learning; services to sensible; satisfied with the process; and instructor training for scholar complaint [4]. Through OBE graduates will be more focused to achieve outlined outcomes. Organization forms the assessment process to obtain the attainment of course outcomes contributive to program outcomes [5].

Fig. 1. Explains the various steps involved in Outcome Based Education.
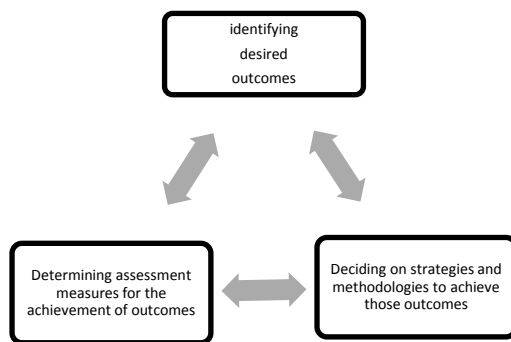

Fig. 1. Outcome Base Education as an Approach[6]

A) Principle of Outcome Based Education

The various principles of Outcome Based Education are described below[7]:

- *Clarity of focus:* As in this principal the teachers need to be focused on their working and train their student more accurately. This belongs to skill development of student and other factors which a teacher can identify and improve them.
- *Designing down:* The design of the curriculum should be clear and concise so that students are able to attain the defined outcomes at the end of a program.

- *High expectations:* The teacher needs to design the teaching standard with some interesting and challenging aspects. Teachers should also encourage the students in their qualities and way of learning something from the outside world.
- *Expanded opportunities:* As learning level of all students is not same, some of the student learn quickly and achieve high standards while some of them need time to achieve that. Teacher should try hard to enhance the skills of their student.

B) Components of Outcome Based Education

The various components of Outcome Based Education are described below:

- *Course Outcome:* Course outcome (CO) describes student ability to perform after course completion. Course outcomes are defined by the course instructor [7] .
- *Program Outcome*: Program Outcome (PO) are the defines the level of knowledge and intelligence of a student after graduation from a program [8].
- *Program Educations Objectives:* It is broad declaration describing levels of progresses which are achievable within few years of completion of degree.

C) Blooms Taxonomy

Dr. Benjamin Bloom discussed the endorsement of advanced procedures for rationalizing the educational area, which incorporates studying and assessing ideas, procedures, measures, and values, rather than impartial memorizing truths (repetition knowledge). It is most repeatedly used when scheming instructive, exercise, and learning procedures [9]. Three domain of knowledge are:

- *Cognitive*: psychological skills (awareness)
- *Affective*: growing in spirits or sensitive zones
- *Psycho-motor*: physical or corporeal skills.

Revised Blooms taxonomy includes six levels [3]*:*

- *Remembering*: It is the basic level. It recalls the data.
- *Understanding*: This level helps to understand the information.
- *Applying:* In this information is used to solve the problem.

- *Analyze:* This level is used to draw conclusions.
- *Evaluate:* This level is used to form an opinion.
- *Create:* It is the highest level. In this something new is designed.

III. RELATED WORK

Traditional system suffered from a number of drawbacks like low level of accuracy, redundancy and inability to measure the cognitive levels of a student learning with efficacy. To resolve this problem researches have been made in the direction of automatic question paper generation involving the assessment based on cognitive level.

Cena et al. [10] presented an automatic paper generation system with B/S standard in J2EE. It works with various management modules which help to manage user data and other services of the system. classify data, question data storage and management user data storage, paper and question selection etc are the parts of some of the module. The proposed mechanism helps in designing and managing different situations with inputs required to generate question paper. It also manages an automatically generated paper for the better efficiency of the system. The user of this system can enter some terms like difficulty status, subject etc. It collects the outside input and compare with its' knowledge base with some rules to form the output as question paper. **Omar et al.** [11] proposed system working on some predefined rules for the generation of final output. The whole process is managed by the natural language processing (NLP) with the formation of verbs and some very important keywords from the knowledge base. The keywords and other parameters of the output paper depend upon the paper type in this system. It helps system to find the particular subject and the type of output required by the user. **Pandey et al.** [12] designed a very intelligent system which worked on the basis of text file input to form question paper with the help of Bloom's taxonomy. The process of fetching input from the outside world is used to design the paper for user on the basis of their requirements. The author proposes an approach which enhances the efficacy in terms of judgment for system learning based on cognitive level. **Naik et al.** [13] presented a software system which helps to find the solution on the basis of various input rules. Some of the major rules like difficulty level of the output, type of selection form the stored data, and type of cluster can help to form a better quality output for the users. System process the input rules from the user and compare with the stored data set to form a question paper in the form of document file. The

proposed system is easy to understand due to the interface parameter selection and more reliable as compared to currently used/ developed systems. **Fei et al.** [14] provided artificial intelligent system which can help to provide required question paper from their knowledge base. The whole process is trained on the basis of text classification and provides results with back propagation processing of neural network to form a better learning system. Scaled conjugate gradient algorithm is used to create a knowledge base of the system and train neural with that. Upon user input for text requirement, trained data is used and system extracts their features for the user query and provides results which are based on extraction of the stored features. The research proposes a high accuracy based question paper generation scheme. **Panicker et al.** [15] showed that question can be categorized by using machine learning approaches such as Support Vector Machine (SVM) and Naive Bayes (NB) with good results if only a great amount of data is involved. In this case, the training set consists of 1500 questions for each 20 categories of the newsgroup. SVM got precision 0.95 for both category 1 and 2 while NB obtained precision 1 for category 1 and 0.58 for category 2. **Bednarik and Kovacs** [16] combined several distinct tools for developing a prototype to generate questions automatically. Proposed approach selects some keywords from the input data and processes on the other content on the basis of clustering approach. All the output structure is grammatically processed by the system to form an automated question paper. **Nalawade and Ramesh** [17] proposed a flexible system offering a semantically tagged question repository for automatic question paper generation in the normal course teachers are responsible for conducting an assessment of a students learning outcome. The onus of question paper generation is on teachers which make the question paper generation task very challenging and biased. Authors proposed different sets of attributes like cognitive level, complexity level, and question type for generating questions paper automatically.

IV.PROPOSED MODEL

The work flow of proposed framework is shown in the Fig 2 and is implemented in the following steps:
- Review Bloom's Taxonomy based developed techniques in automatic systems.
- Study Revised Bloom's Taxonomy for its integration into cognitive level based assessment.
- Study the procedure of attainment calculation in Outcome Based Education (OBE).
- Design the table structure to store question required for assessment along with the course

outcome (CO) - program outcome (PO) attainment data.

- Design and construct a framework to enter the question according to the predefined format.
- Develop a procedure to extract keywords, levels from the entered question which lead to the extraction of course outcome (CO) and program outcome (PO).
- Set up a repository of CO-PO attainment data.
- Perform course outcome based on predefined matrices like processing time, correctly classified instances, incorrectly classified instances, Kappa statistics, mean absolute error, root mean square error using classification algorithms like oneR, ZeroR, J48, Naive Bayes, IBk.

The initial step is to form a repository of data along with structures of the table. In the repository of the data, the keywords of the Revised Blooms taxonomy must be mapped with the levels of Blooms taxonomy. Once the database is designed and the interface for entering the questions is developed.The interface also incorporates the ability to extracts keywords,levels from the question which further extract course outcome and program outcome for mapping purpose.
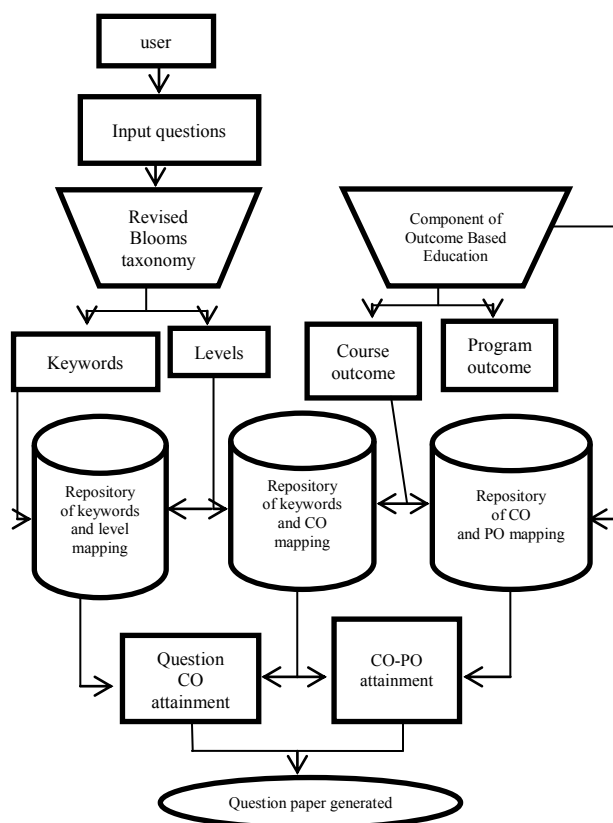

Fig. 2. Workflow Diagram for automated question paper generation with CO and PO attainment

The process for classification of course outcome attained data and predictive model evaluation is illustrated in Fig.3.
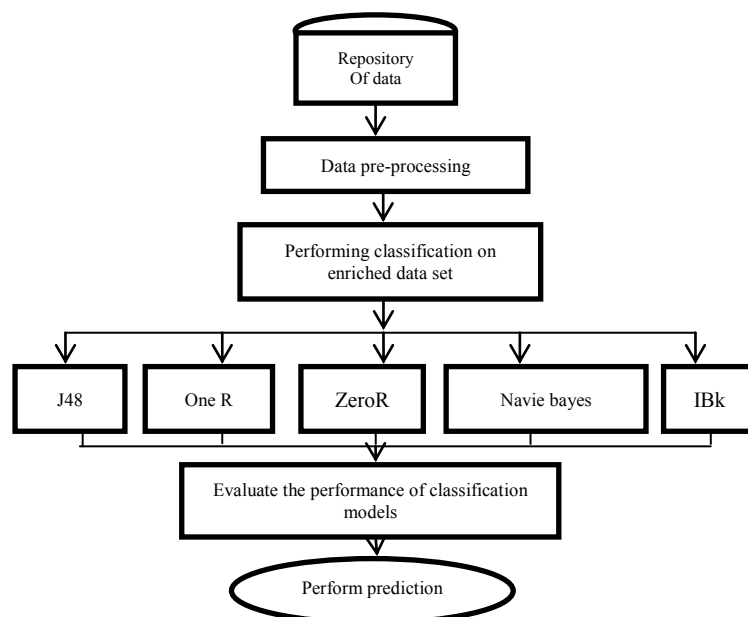

Fig. 3. Workflow diagram for classification of CO attained data and predictive model evaluation.

V. RESULT AND DISCUSSION

Proposed research inculcates the automatic question paper generation based on outcome based education (OBE) and Revised Bloom's taxonomy. Predictive analysis of attainment data for course outcome is performed using various classification algorithms. The overall calculation is based on the keyword mapping, level mapping and course outcome (CO) mapping in this research.

The prediction of the educational model outcome is performed following steps:

- Creating repository of CO attainment data based on CO-PO mapping.
- Calculate parameters like processing time, correctly classified instances, incorrectly classified instances, Kappa statistics, mean absolute error, root mean square error and design a comparison between them on the basis of accurate classified data.

A. Attainment Prediction

The entire data is stored in the system in different repositories. During the automatic question paper generation process the data from these different repositories is fetched and processed resulting in the generation of attainment data. Various classification techniques are performed to form the attainment prediction. Naïve Byes, J48, ZeroR, IBk and oneR are the classifiers used.

Table 1. compares various classification techniques on the basis of attained information. Among all the classifiers J48 provides the best accuracy of 90%.

Table 1. Comparative analysis of classification techniques performed on attainment data

Classifier	Processing Time(sec)	Correctly classified instances (%)	Incorrectly classified instances (%)	Kappa statistic	Mean absolute error	Root mean square error
J48	0.02	90	10	0.7692	0.146	0.3018
Naive Bayes	0.01	85	15	0.625	0.1896	0.3497
OneR	0	70	30	0	0.3	0.5477
ZeroR	0	70	30	0	0.4238	0.4594
IBk	0	68.3333	31.6667	0.2339	0.3312	0.5566

B. Receiver Operating Characteristics (ROC) Analysis

This curve is an outline of the True Positive Rate (TPR) with the False Positive Rate (FPR) for the dissimilar possible cut-points of a diagnostic test. ROC is having features:

- It defines the trade-off between true positive and true negative (any enhance in recall will be accompanied by a decrease in precision).
- The nearest the curve belongs the left-hand border and then the top-border of the curve - space, the more enhance the test.
- The nearest the curve originates in the 45-degree oblique of the curve-space, the less improved the test.
- The field under the curve is a measure of textual accuracy.

ROC curve for two classifier label yes and no is generated. Classification model are considered to be optimal if the area under the curve for ROC is closer to 1. Fig. 4. shows that the area under curve is 0.8677 for label yes and Fig. 5. shows that the area under curve is 0.8677 for label no.
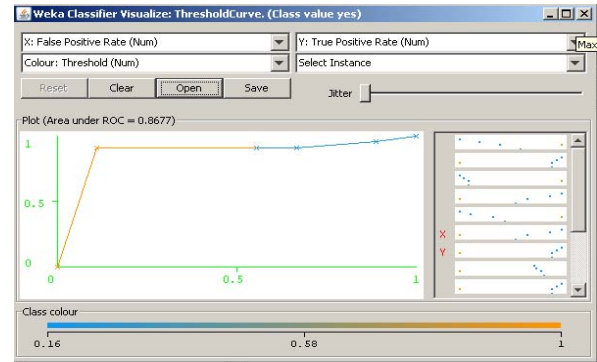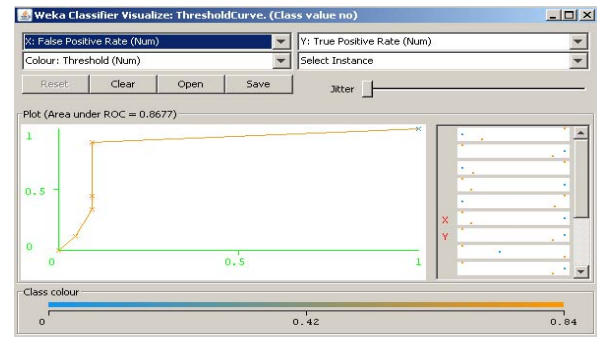

Fig. 4. ROC curve for class label yes

Fig. 5. ROC curve for class label no

VI.CONCLUSION AND FUTURE SCOPE

Education system is a vital activity for educational foundations to measure student learning outcome. Examination process helps in accessing the course outcome (CO) - program outcome (PO) attainment corresponding to different curricula design and developed by an education system. Formulating the exam queries is time-consuming and the superiority of question paper is despoiled. Outcome based education (OBE) has gathered massive recognition in higher learning. OBE is a procedure that emphases on what are to be learned - the outcome. CO-PO is the vital elements of OBE process. This research proposes an automatic question paper generation tool incorporating different cognitive levels of revised blooms taxonomy. Different cognitive levels are selected based on the keywords which are extracted from the entered questions. The selected levels then form the basis for the selection of CO-PO mapping. Classification algorithms are applied on the CO-PO attainment data to perform outcome based predictive analysis.

The future scope, automatic question paper tool can be further extended for the following cases:
- Multiple choice questions (MCQ) for the quiz can be incorporated.
- Achromous java script and Xml (AJAX) based keywords and levels identification.

REFERENCES

[1] N. Myller, "Automatic Generation of Prediction Questions during Program Visualization," *Electronic Notes in Theoretical Computer Science*, vol. 178, no. 1, pp. 43–49, 2007.

[2] E. D. Leslie Owen Wilson, "The Second Principle." [Online]. Available: http://thesecondprinciple.com/teaching-essentials/models-of-teaching/.

[3] P. Armstrong, "CENTER for TEACHING." [Online]. Available: https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/.

[4] "Aligning and building curriculum." [Online]. Available: http://gototheexchange.ca/index.php/curriculum-at-course-level/course-learning-outcomes.

[5] P. Polimetla, "Assessment of Engineering programs," in *ieee Transaction on Eduaction* pp. 22–26, 2014,.

[6] "authorSTREAM." [Online]. Available: http://www.authorstream.com/Presentation/maimai245431-2626290-outcomes-based-education/.

[7] "Center for Education Innovation." [Online]. Available: http://cei.ust.hk/teaching-resources/outcome-based-education/institutional-resources/obe-principles-and-process.

[8] "ASCD LEARN.TEACH.LEAD." [Online]. Available: http://www.ascd.org/publications/curriculum_handbook/413/chapters/An_Overview_of_Outcome-Based_Education.aspx.

[9] A. Shaikh, A. Handore, and P. P. Kudal, "Quality Question Paper Generator on Role Based Hierarchy model," *International Journal of Research in Advent Technology*, vol. 5, no. 2, pp. 6–9, 2017.

[10] G. Cen, Y. Dong, W. Gao, L. Yu, S. See, Q.Wang, Y. Yang, H. Jiang, , "A implementation of an automatic examination paper generation system," *Mathematical and Computer Modelling*, vol. 51, no. 11–12, pp. 1339–1342, 2010.

[11] N. Omar, S. Haris, R. Hassan, H. Arshad, M. Rahmat, N. Zainal and R. Zulkifli,, "Automated Analysis of Exam Questions According to Bloom's Taxonomy," *Procedia - Social and Behavioral Sciences*, vol. 59, no. 1956, pp. 297–303, 2012.

[12] S. Pandey and K. C. Rajeswari, "Automatic Question Generation Using Software Agents for Technical Institutions," *International Journal of Advanced Computer Research*, vol. 3, no. 13, pp. 307–311, 2013.

[13] K. Naik, S. Sule, S. Jadhav, and S. Pandey, "Automatic Question Paper Generation System using Randomization Algorithm," *International Journal of Engineering and Technical Research (IJETR)*, vol. 2, no. 12, pp. 192–194, 2014.

[14] T. Fei, W. J. Heng, K. C. Toh, and T. Qi, "Question classification for e-learning by artificial neural network," in *ICICS-PCM 2003 - Proceedings of the 2003 Joint Conference of the 4th International Conference on Information, Communications and Signal Processing and 4th Pacific-Rim Conference on Multimedia,* vol. 3, pp. 1–5, 2003.

[15] A. D. Panicker, A. U, and S. Venkitakrishnan, "Question Classification using Machine Learning Approaches," *International Journal of Computer Applications*, vol. 48, no. 13, pp. 1–4, 2012.

[16] L. Bednarik and L. Kovács, "Implementation and assessment of the automatic question generation module," in *3rd IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2012 - Proceedings*, pp. 687–690, 2012.

[17] G. Nalawade and R. Ramesh, "Automatic Generation of Question Paper from User Entered Specifications using a Semantically Tagged Question Repository," in *IEEE 8th International Conference on Technology for Education Automatic*, pp. 148–151, 2016.