International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 05, May-2022

# An Innovative Method of Automated Content Evaluation using Soft Computing

Arul Shalom A
Department of Computer Engineering
Thakur College of Engineering and Technology
Kandivali, Mumbai, India

Shiwani Gupta
Department of Computer Engineering
Thakur College of Engineering and Technology
Kandivali, Mumbai, India

***Abstract*— A number of efforts have been taken in the past two decades to develop an automatic evaluation system that is able to evaluate a students' knowledge in the higher levels of the Blooms taxonomy. The evaluation of the student is a subset to the Learning Management System (LMS). The student evaluation is done based on the answers that are given to a particular question by the student. The answer can be of various types and sizes, thus, giving rise to the need for a system that can evaluate any and all answer types and answers of all sizes. Though there are various advances in the answer grading systems, the experts are skeptical to use it due to the risk of accuracy and reliability of the system. This gives rise to the need for a system that is evaluate the students' answer more accurately and is more reliable. In this paper we present a brief study of a few existing systems and propose a system model that is suitable to grade any answer for a particular question.**

***Keywords—Atomatic content evaluation; Automatic short answer grading system; Natural language processing***

## I. INTRODUCTION (*HEADING 1*)

The growth and the influence of the internet over the years have led to the development of various Learning Management Systems (LMS). The evaluation of the students' answer is considered to be a subset task of the LMS. There are many proposed models for the automatic evaluation of the students' answer. However, due to the accuracy and the reliability of the systems make the experts more skeptical on using these systems. The assessment of a students' answer can be done in various ways. This depends on the type of answer and also the size of the answer. Many systems for the automatic grading of the students' have been taken. The initial attempt was made by IBM where they developed a system that could evaluate objective type answers. The evaluation of the objective type answers does not assess the students' knowledge in the higher levels of the Bloom's taxonomy. Therefore, giving rise to a need for short answer evaluation system.

A short answer is considered to be focused on the question and is considered to have not more than one paragraph. However, the answer to every question differs from person to person and thus making the evaluation of short answer a more difficult and time consuming task. During the past decade a lot of research has been carried out in building an automated system that can evaluate short answers. There are many proposed system models but the experts are very skeptic of using the system since they are not reliable and also accurate. With the advancement in the field of artificial intelligence and the field of natural language processing, there rises a need for a highly reliable and accurate system that is able to process the short answers and is able to grade them.

In this paper, we discuss some of the proposed models and also propose a system model that will be able to process and grade the short answers. The proposed system shall be highly reliable. The proposed system gives the subject expert the freedom to give weightage to the key words and phrases in the answer. The answer of the student is compared with the key given by the subject and is graded, based on not only rule based methods but also uses soft computing techniques that allows the system to grade the answers more accurately, thus, making the proposed system more reliable and accurate. The obtained results of the proposed system will be compared with the results obtained from manual evaluation of the subject expert and also results obtained from other implemented models.

## II. HISTORY OF EXISTING SYSTEMS

The literature available on automatic content assessment is very large and over the past two decades there has been a lot of publications in this area of research. Most of the research done in the early part of this decade was mostly based on knowledge engineering Information Extraction based systems and corpus-based systems. The research over the years moved towards machine learning and evaluation based systems. The general form of an automatic content evaluation and grading as shown in Figure.1, is best described by the image given by Steven Burrows, Iryana Gurevych and Benno Stein, in their

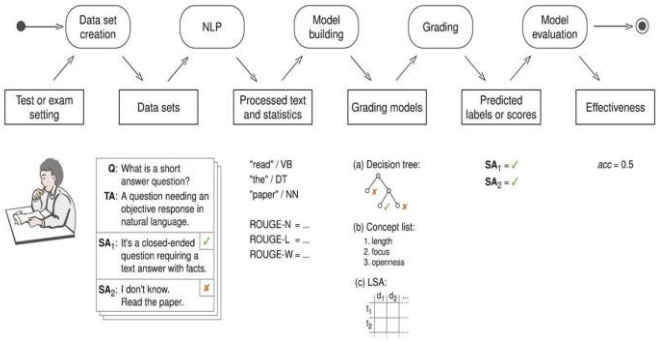


Fig. 1: Structure Pipeline of an Automatic Short Answer Scoring System [5].

study of various existing systems for automatic short answer grading.

The automatic content evaluation system generally has eleven components which consist of 6 artifacts and 5 processes. First and foremost in an automatic content evaluation system, we need to appropriate materials to set the test or exam. Once the test or exam is set, a set of questions

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 05, May-2022

and answers both form the subject expert and also from the students is needed. This is considered to be the data set. Once the data set is ready, the data is processed and processing can be done using various techniques. After the processing of the data sets is done a model of the processed data is built in order to make the evaluation process easier. The models are built are built for the answers' given by the student and the subject expert separately. When the models are ready, the evaluation process begins, where the students' answer is compared along with the model of the answer given by the subject expert. Using various techniques, such as rule based technology is used to compare both the answers and give grades to the students' answer accordingly.

As mentioned earlier, many number of systems have been proposed fpr the evaluation and grading of the content in natural language. There is a large amount of literature available in this field of research. A detailed study on the existing and proposed systems is done by Steven Burrows, Iryana Gurevych, and Benno Stein in the paper titled, "The Eras and Trends of Automatic Short Answer Grading." In this paper, we discuss a few systems and then propose a model that will be much more interactive, realiable and accurate and also serve as a benchmark for other automatic content evaluation system.

### *A. Automarking Web Service:*

The Automarking Web Service system (Quan Meng, Laurie Cutrone) is a computerized marking mechanism that uses the recent natural language processing techniques. It is an open source system that allows the staff to add questions and get the grades for the answers to the question. The system gets the answer from the staff or subject expert and extracts the semantic and syntactic structure of the answer and compares it with the structure that is extracted from the students answer and the grade is given based on the similarity of the structures that has been extracted. The system works on a GNU server and is implemented in Java. The system is developed with very strict constraints. It is capable of processing answers that contains a single sentence. The system also considers the answer to be free of grammatical and spelling mistakes.

### *B. AutoMark:*

AutoMark (Mitchell et al. 2002), this system performs pattern matching as a form of information extraction on the parse tree representations of teacher and student answers that is obtained by the processing of the answers and is used for grading. Two approaches namely the "blind" and "moderated" are described in which the system operates. In the blind approach the system is fully automated and this approach represents the best definition of ASAG. In contrast, the moderated approach includes a step that is human-driven which allows the model to be revised after grading has been performed.

### *C. WebLAS (Web-based Laguage Assessment System):*

This system is a language assessment system delivered entirely over the web. It was developed with keeping in focus the needs of the language assessors that were developed before it. It is written mostly in PERL and is well suited for NLP tasks. The system identifies the important segments in the answer through parsed representation and allows the teacher or subject expert to confirm each segment and apply weights to the segments. The matching between the students and the teachers' answers is done by comparing and matching the regular expression. And the scoring is done by comparing the presence and absence of the segments.

### *D. Auto-Assessor:*

Auto-Assessor (Cutrone et al., 2011), this system focuses on grading canonicalized single-sentence student answers. It is based on bag-of-words coordinate matching and synonyms with WordNet (Pedersen et al., 2004). Coordinate matching in ASAG refers to the matching of every individual term between teacher and student answers. In Auto-Assessor, each word matched exactly is given one point, related words that are obtained from WordNet are given partial points, and the rest are given no points.

### *E. Free Text Authour*

In this system, the teacher answer is used as the model and the system provides the user with an interface for both the input of the teachers answer and the grading of the students answer. The answers given by the teachers are composed into templates that are automatically generated by the system and hence there is no need for the expertise of the user in the natural language processing. The interface allows the teacher to specify the mandatory keywords and also to select the synonyms from the available thesaurus. The score to the students answer is given by comparing the templates and the answer also is determined either as accepted or not accepted.

### *F. ATM – Automated Text Marker*

ATM (Automatic Text Marker) (Callear et al. 2001), this system breaks down the answers by both the subject expert and student into lists of minimal concepts comprising no more than a few words each. Once the answer is broken down the system counts the number of concepts that are common in order to score the answer. Each concept that the answer is broken into is essentially the smallest possible unit in an answer. These concepts are assigned a weight for the purposes of grading. The weights obtained by the students answer are summed up to give the overall score for the students answer.

### *G. C-Rater*

The Concept Rater (c-rater) (Leacock and Chodorow 2003), this system aims at matching as many sentence-level concepts as possible between teacher and student answers. The matching of the concepts is based on a set of rules and a canonical representation of the texts using various techniques such as, syntactic variation, anaphora, morphological variation, synonyms, and spelling correction. In particular, the answers given by the teacher are entered sentence-wise for each and every concept. This simplifies the assessment of the students' answer, since only one concept is considered at a time when grading. This technique avoids the need for an indirect solution, such as dividing the question into multiple parts (Jordan 2009b) and it is argued that this can lead to higher accuracy (Sukkarieh and Blackmore 2009). Furthermore, the natural language input format is

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 05, May-2022

advantageous compared with other systems that require expertise and use of a markup language (Sukkarieh and Stoyanchev 2009).

## III. PROBLEM STATEMENT

The assessment of learning outcomes can be obtained by assessing the results obtained in tests and examinations. These tests and exams can be facilitated by many types of questions based on the size of the answer and also the type of answer. The scoring of the content can be done either by grading the answers manually or automatically by using computational methods. Some questions are more difficult to grade manually than others. The grading of natural language responses to questions that require answers that consists of at least a few sentences can be considered much more difficult, as an understanding of the natural language is required. We believe that the assessment of the students' knowledge on a particular subject cannot be assessed by evaluating the answers given to a particular type of question. Therefore we aim at developing a system that can evaluate an answer book that has various types to responses to different questions. By studying the above systems we find that the scoring is done just by comparing the model that is generated by the system. However, the grades given on comparison may not be the accurate score that the students' answer deserve.

Considering all the above systems and keeping in mind all the variations that we can have in an answer to the same question. We propose a system that will use soft computing along with the natural language processing tools that will enable us to score the answers given by the students in a more efficient manner. The proposed system is a model which considers any and every answer given by the student is considered to be valid and the grades are given based on the closeness of the answer to the answer given by the subject expert. The proposed system overcomes the incorrect grading of the answers. The proposed system also considers the weightage given to a phrase or concept by the subject expert. Thus, the proposed model will be much more accurate and also reliable and also can serve as a benchmark to the other automatic content evaluation systems.

## IV. PROPOSED SYSTEM

In the proposed work, we consider the questions that are designed satisfy the following criteria's; first, the question must require a response that makes the student recall external knowledge instead of recognizing the answer from within the question. Second, the question must be framed such that it requires a response that is given in natural language. Third, the question framed must be such that they require answers of length roughly between a few phrases to a few paragraphs, which makes the evaluation of the answers considerably easy. Fourth, the assessment of the responses should focus on the content of the response instead of the writing style. With all the above criteria's satisfied the proposed system also works on the grading methodology. The grading of the answer is done by using soft computing techniques that allow us to consider any and every answer for grading. The proposed system also includes a module in which the subject expert is allowed to set the weightage for particular concepts or phrases. These weightage given by the subject expert is added to the score of the students answer by comparing both the answers semantically. If the meaning of the phrase exists then the weightage given by the subject expert is assigned to the students answer. This makes the system more reliable and also improves the accuracy of the system.

## V. PROPOSED METHODOLOGY

The implementation of the proposed work is proposed to be done as a web application. The processing of the natural language answer is proposed to be done using the Stanford parser and other technologies. The entire research work is proposed to be carried out in five phases, as follows;

a) Selection of the dataset

b) Selection of the NLP technique to process the answers given and to build a model that is easy for evaluation.

c) Building models of the processed content to make it easy for evaluation

d) Building a grading model including the soft computing technique along with the natural processed text.

e) Evaluation of the model

## VI. PROPOSED METHODOLOGY

The research work focuses mainly on the evaluation of the short answers. The proposed model deals maximum with textual data, it ignores graphs images, etc.… which however can be incorporated to the proposed system.

a) The proposed system takes in the answer key and the dataset and is able to process the data using the NLP techniques.

b) The processed text is built into models that will help in grading of the dataset.

c) The models built are visible to the subject expert who can assign grades for the keywords or phrases that are present in the dataset.

d) The grading of the dataset is done using rule based methodologies and also soft computing techniques that allows the system to prioritize the schematics of the answer dataset.

The performance and the accuracy of the proposed system will be evaluated by comparing the scores given by the proposed system with the scores that is given to the same data set by the other existing systems and also the scores given by the subject expert manually.

## REFERENCES

[1] Cutrone, L. and Chang, M. (2010). AUTOMARKING: AUTOMATIC ASSESSMENT OF OPEN QUESTIONS. In M. Jemni, D. Sampson, Kinshuk, and J. M. Spector, editors, Proceedings of the Tenth IEEE International Conference on Advanced Learning Technologies, pages 143–147, Sousse, Tunisia. IEEE.

[2] Cutrone, L., Chang, M., and Kinshuk (2011). AUTO-ASSESSOR: COMPUTERIZED ASSESSMENT SYSTEM FOR MARKING STUDENT'S SHORT-ANSWERS AUTOMATICALLY. In N. S. Narayanaswamy, M. S. Krishnan, Kinshuk, and R. Srinivasan, editors, Proceedings of the Third IEEE International Conference on Technology for Education, pages 81–88, Chennai, India. IEEE.

[3] Leacock, C. and Chodorow, M. (2003). C-RATER: AUTOMATED SCORING OF SHORT-ANSWER QUESTIONS. Computers and the Humanities, 37(4), 389–405.

[4] Siddiqi, R. and Harrison, C. J. (2008a). A SYSTEMATIC APPROACH TO THE AUTOMATED MARKING OF SHORT-ANSWER QUESTIONS. In M. K. Anis, M. K. Khan, and S. J. H. Zaidi, editors, Proceedings of the Twelfth International Multitopic Conference, pages 329–332, Karachi, Pakistan. IEEE.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 05, May-2022

[5] Steven Burrows, Iryana Gurevych, and Benno Stein "THE ERAS AND TRENDS OF AUTOMATIC SHORT ANSWER GRADING" International Journal of Artificial Intelligence in Education vol. 25(2015) 60 - 117

[6] Sukkarieh, J. Z. and Stoyanchev, S. (2009). AUTOMATING MODEL BUILDING IN C-RATER. In C. Callison-Burch and F. M. Zanzotto, editors, Proceedings of the First ACL/IJCNLP Workshop on Applied Textual Inference, TextInfer '09, pages 61–69, Suntec, Singapore. Association for Computational Linguistics.

[7] Wang, H.-C., Chang, C.-Y., and Li, T.-Y. (2008). ASSESSING CREATIVE PROBLEM-SOLVING WITH AUTOMATED TEXT GRADING. Computers & Education, 51(4), 1450–1466.