International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 06 | June 2019 www.irjet.net p-ISSN: 2395-0072

An Automated Approach to Conduct Pune University's In-sem Examination

Sushila Choudhari[1], Shivam Moray[2], Nikita Salunke[3] , Anand Dhawale[4]

[1]*Computer Department, M.E.S.COE*
[2]*Computer Department, M.E.S.COE*
[3]*Computer Department, M.E.S.COE*
[4]*Anand Dhawale, Dept. of Computer Engineering, M.E.S.COE, Maharashtra, India*

Abstract - *Many tests and examinations are carried out to test a candidate's English writing skills. However, reaching certainty about the candidate's knowledge, is a challenge these days. In this paper, a broad overview on the design and implementation of an automated evaluation system for short technical answers by using Machine Learning, along with a comparative study of the various approaches that have been used till current date, is provided. The prevailing system has its own pause in terms of volume, staffing, variation, correctness within the ways of assessing. The proposed system is an evaluation system which can identify the matching keywords in the textual answers and evaluating marks for the same based on some previous knowledge acquired by the machine. A syntactical relation-based feature extraction technique is proposed for automatic evaluation of descriptive type answers. Students are also benefited with a feedback system that can help them improve their score. It also provides a platform to academic institutes to enhance their system which can provide better results in assignment evaluation.*

Key Words: **Machine Learning, Natural Language Processing, Automatic assessment, Short Answer Grading, Similarity measures**

1. INTRODUCTION

The ability to communicate in natural language has long been considered a defining characteristic of Artificial Intelligence. In several cases, the answer rating task prices large human resources, however with less potency and therefore the score given by human authority is generally determined by his data, feeling and energy. There is also an enormous deviation between the scores evaluated by totally different rates. Thus, the correctness of grader rating system can't be guaranteed. In earlier times, the work of assessment in academic terms was heavily reduced by conducting a Multiple-Choice Questions (MCQs) test, whose answers only comprised of a single word or a short text. The modern era demands for the same efficiency in case of short technical answers.

The automatic answer evaluation system has been dealt in numerous forms, like question responsive system, essay grading system over time. All the previous works related to this project throw light upon the key concept of evaluating the natural or linguistic language answers and providing a grade accordingly.

2. RELATED WORK

The basic concept of using computers to increase our understanding of textual features has long been considered as an additional benefit in comprehending written text.

Marti. A. Hearst [1] in her research developed a project called the Essay Grader, where she applied multiple linear regression to determine optimal combination of weighted features. This system was vulnerable and faced many challenges. She then shifted her focus on developing another approach to assess more direct measures of writing quality called Latent Semantic Analysis (LSA).

Huang Houkuan [2] introduced a general process text machine learning at first, and thereby suggested a method for transforming text categorization problem to a series of binary classification problems using Support Vector Machine (SVM). For feature extraction the concept of stemming was used. However, this approach wasn't suitable for larger dataset as training time with SVMs can be high.

P. Selvi [3] presented a system based on combination of novel approach and latent semantic analysis. The novel approach separates composite and primitive features whereas the LSA module determines the number of words after applying stemming. The combination of the above 2 stated approaches shows greater efficiency and proves that combining various algorithms is a possible strategy asses' student's answer.

The TOEFL exam proved to be one of the best examples of Automated Essay Grading (AEG). Siddhartha Ghosh [4] proposed an AEG system which brought significant changes and gave a new shape to Indian Text Categorization and Machine Learning research work. Independent Bayesian Classifiers allow assigning probabilities to documents estimating the likelihood inferring they belong to specific classes. The diagnostic feedback was based on a suite of programs that identify the essay's disclosure structure, recognize undesirable stylistic features, and evaluate and provide feedback on errors in grammar, usage and

mechanics. The proposed framework intended to capture the mental status of the student (Psychometric Analysis).

Automated Marking System for Short Answer Examination (AMS-SAE) [5] is a system that has been developed to grade student's answers based on given marking scheme. Fatimah Dato' Ahmad developed an Automated Text Marker (ATM). ATM is a marking system that uses the language structure to compare the sentences.

P. A. A. Dumal proposed a system [6] to overcome the issues in existing similar systems. Answer extraction and comparison of similarity were the two methodologies which comprise of the core part of the project. Answer extraction of both student and tutor was done in three steps which were splitting of sentence, tokenizing and Part-of-Speech (POS) tagging. In this project solution the tutors can add questions to the question bank, making assignments and obtaining the final scores automatically. The learners are facilitated too in providing the answers for assessment and getting the feedback immediately.

The automated short answer evaluation system [7] proposed by Sijimol P. J., can identify the text in answer papers and evaluates marks for each short answer based on previous knowledge acquired by the model. In the proposed system Optical Character Recognition tools are used to extract the hand-written texts. The system works based on Machine Learning. It trains a model from the scored short answer paper dataset and a high weightage given key.

Standard key libraries are used to compare the learners' answer in the system [8]. The learning activity gets enhanced since it is a Web-based examination. The main challenge lies in maintaining the security of the system.

Senthil Kumar et al. [13] in their framework obtain the student answer and model answer as input extracting the resource description framework (RDF) triples for each sentence. Ontology is constructed for the RDF sentences obtained, and mapping has been performed to obtain the resemblance. The weights associated with the answers at each vertex determine the closeness of answer. The total score of all vertices justifies the similarity of the student answers with the model answers. According to Resnik theory, the semantic measure of the words depends on the hierarchical level in the WordNet hierarchy.

Michael Mohler [9] devised an evaluation system and a similarity model that tries to improve the grading of candidate answers. The main objective of the system is to construct a dependency graph that automatically assigns a score for the student answer from the connected nodes. A combination of lexical, syntactic and semantic features is used to compute individual weights of the examiner and candidate answer.

The integrated coordinated structure planned by the author [14] portrays the reviewing of the digitized answers written in a customary technique with pen and paper. The question and the appropriate responses are changed over into m*n lattice. The learners obtain their credits completely if addressed effectively, else penalty is obtained. Each question answer is published with its related mistakes pointed, penalties, bonus recommended if any. Students can compare their marks with the fellow mates. The striking masterpiece of the framework is that the student script scores can be viewed online from the learner's end.

3. PROPOSED WORK

In the proposed system an automatic grading system for descriptive answers is presented. The system takes in student's answers as input and asses them based upon the predefined answers. The answer is scored depending upon the context of the answers and the structure of the answer.

The proposed methodology presents an automatic evaluation system for short technical answers. The proposed framework assesses the student's responses with the model dataset's responses. The final score is computed based on several factors like the exact keyword match, the context of the phrases, the structure of the answer and the cosine similarity factor.

The architecture of the proposed system below depicts the various stages in the classification as well as evaluation of answers.

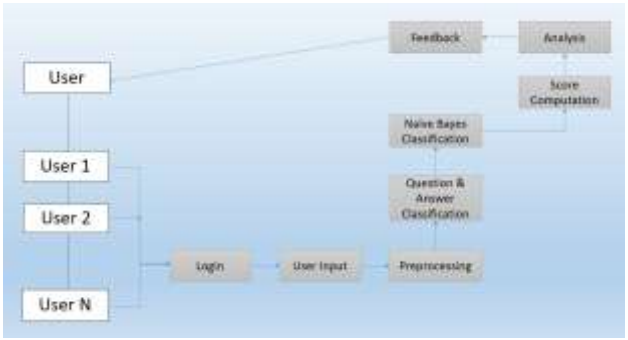

Fig. 1. Architecture of the Proposed System.

The essential side within the assessment course of action is that the critic of the training method that estimates the learners answer region. To assess the performance of the candidate, a completely unique answer assessment system is projected that employs text patterns extracted from the answers to be classified underneath answer classes. This classification methodology focuses on the answers given by the candidates written as per their natural method of representing the answers. It's a difficult task for the machine to place forth the precise inference of the data sent that varies with the kind of queries and therefore the answers. Questions, its corresponding or model answers and student

responses are given as input to the method. The system is organized into training and testing sections.

The training section includes:

I. Question Classification

II. Answer Classification

The prediction or testing section includes:

III. Answer evaluation and Score computation

I. Question Classification

In this section the model classifies the questions prior to classifying the answers. The aim is to extract the headword of each question so that this headword can be compared with the corresponding student's response, and the answers meaningfulness can be understood.

There are various steps to be carried out in the processing of the question module:

The primary stage is to find out the type of question asked, so that it can provide assistance in predicting the nature of the answer. Key feature 1 corresponds to the question-type analysis. The question type here refers to categorizing the question into 4 types: Factual, Inductive, Analytical and Application based types. The feature extraction is carried out in the following manner:

The factual type of questions can be identified by the keywords who, where, what or which. These keywords can be isolated from the questions by entity recognition or using stemming process.

Factual questions

Example:

What is a Socket?

What do you mean by SCTP?

The tag 'what' to the question in the above questions imply that the answer expected is the explanation of the entity mentioned in the question.

Inductive questions

Example:

Why is IP Protocol considered as unreliable?

How does TCP perform error control?

From these question tags such as the 'why' or 'how', it is evident that here the questions require the student to present the detail facts as well as other entities too.

Analytical questions:

Example:

Distinguish between…

Compare…

These questions not just ask the students to present facts but also want to them to find out the relation between the entities mentioned in the question. It can test the analytical part of the student's answers.

The figure given below shows how the questions get classified into their corresponding question-type. It represents the Key Feature 1. It results with an accuracy of 92.8% in correctly classifying the question-type.

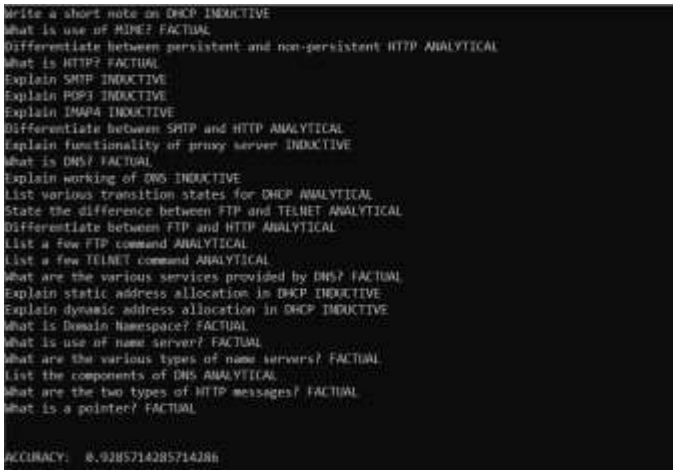

Fig. 2. Key Feature 1: Question Type

The next feature, i.e., Key Feature 2 corresponds to the identification of the headword from the question. Headword is a word that points out to the explanation or the main keyword in the question asked. Parsing of the sentence can be used to find out the headword in the question.

The purpose of extracting the headword is to test for the linguistics relations between the phrases and also the question tag extracted from the question. The figure below displays the headwords extracted from some queries. These words justify clearly what the candidate is asked for to put in writing in their answers.

Key feature 3 depicts the variety of answer the student is predicted to jot down. These options can facilitate to supply meaningful relationship between the question and answers.

The planned work will reason exploitation using the Naïve Bayes classification. It uses a supervised approach and probability to decide on the simplest likelihood of the answers. Mean or Variance of the values will be accustomed calculate associated values.

Pseudo code for headword extraction:

```
headword_extraction(question_list)
while question_list NOT empty
        sent = preprocess(question_list)
        if sent[i] = where || why then
                feature2 = sent[i+2]
        else if sent[i] = define || how then
                feature2 = sent[i+1]
        else if sent[i] = between then
                feature2 = sent[i+1]
        else if sent[i] = what then
                if sent[i+1] = (is || are) && (a || an || the) then
                        feature2 = sent[i+3]
                else if sent[i+1] = is || are then
                        feature2 = sent[i+2]
                else if ( sent[i+1] = is || are ) && ( sent[i+3] = used ) && ( sent[i+4] = for ) then
                        feature2 = sent[i+2]
                else if ( sent[i+1] = do || does ) && ( sent[i+3] = mean ) then
                        feature2 = sent[i+2]
                else if ( sent[i+1] = cause || causes ) then
                        feature2 = sent[i+2]
                else if ( sent[i+1] = does || are ) && ( sent[i+3] = stands ) && ( sent[i+4] = for ) then
                        feature2 = sent[i+2]
                end if
        else if sent[i] = who then
                feature2 = sent[i+2]
        end if
end while
preprocess(sent)
        tokenize the question
        return sent
```

	Question	Headword
2	What is a socket?	socket
3	What are the 5 basic service primitives of Transport layer protocol?	primitives
4	Explain RTP	RTP
5	What is QoS?	QoS
6	What is the significance of SYN flag in the TCP header?	significance
7	Write a short note on DHCP	DHCP
8	What is use of MIME?	MIME
9	Differentiate between persistent and non-persistent HTTP	persistent
10	What is HTTP?	HTTP
11	Explain SMTP	SMTP
12	Explain POP3	POP3
13	List various transition states for DHCP	transition
14	Explain functionality of proxy server	proxy
15	What is DNS?	DNS

Fig. 3 Key Feature 2: Headword Extraction

Key feature 3 depicts the variety of answer the student is predicted to jot down. These options can facilitate to supply meaningful relationship between the question and answers.

The planned work will reason exploitation using the Naïve Bayes classification. It uses a supervised approach and probability to decide on the simplest likelihood of the answers. Mean or Variance of the values will be accustomed calculate associated values.

II. Answer Classification

Students will write the answers through the GUI provided and it will be stored in the system. Each answer will have to go through a grammar check. Here spelling and grammar mistakes present in each answer will be checked. Following the general rule, we expect that the length of answer i.e. the sentences in the answer will be double the marks assigned to the question. For example, if the question is of three marks then the student will have to write an answer that has at least six sentences.

In this part of classification, we will check if the skill of student in writing skill and presenting of facts through his answers. There are 2 kinds of cognitive-based non-fiction matter pattern for textual structures specifically, deep structure and surface structure. Here we are only going to focus on the surface structure. Surface structure identifies pattern, structures and phrases from the answers. It refers to the method during which student organizes his/her answer. The system notes the structural pattern and hunts for signal words to judge the content utilized by the learner. For this analysis the classification is focused on five categories:

1. Chronological ordering

It shows ordering of the texts from start till the end.

Example: Not long after, Next, Before, First, etc.

2. Descriptive

The descriptive details asked in the question are identified.

Example: For instance, Such as, In addition, Furthermore, etc.

3. Comparative

The ideas are described to reader through the answers.

Example: Different from, As opposed to, etc.

4. Problem/Solution

The question asked wants the students to address a problem. Words describing the solution are to be found in the answer.

Example: Problem, The solution is, One answer is, A solution to the problem is, etc.

5. Cause and effect

The question demands the student to write the answers which show the occurrence of an event and the causes or effects of it are asked.

Example: Accordingly, Because, To, Since, For this reason, Also, Not only, etc.

For simplification purpose, the 5 above stated classes are collectively categorized into 3 broad answer types. They are:

1. Factual type (Chronological ordering and Descriptive)
2. Analytical type (Comparative and Problem/Solution)
3. Inductive type (Cause and effect)

The process of feature extraction is applied on the answers also. The features under comparison are the non-textual features that could be recognized with the help of signal words in the answer input by the student.

The key features to be extracted are as follows:

1. To find the answer type
2. To enclose the bigram count in the answers
3. To check for the grammatical and syntactical errors

III. Answer Evaluation and Feedback

The evaluation of the answers is performed by connecting the model answer with the student answer. The comparison can extract all the required features from the student answers. The solution is then additionally classified into the classes of description, cause and result, etc. The solution written by the student is compared with the model answer given. The factors for the analysis of the solution here is, firstly the answer type and also the question type ought to match. Secondly, the keywords or facts that are asked within the question must be addressed. Signal words further get checked to predict the category of the answer.

Cosine similarity and Jacquard similarity are utilized to ascertain the comparison and relatedness between the sentences. The final score is given to the student along with the feedback based on the analysis of the answers.

A brief analysis of the answers can give a detailed view of the student's strong or weak areas in the test. This will create a good learning environment for the students. The feedback to the student can be provided in the form of graphs, charts or pie diagrams.

In the figure shown below, a model answer is compared against the student's response and the Cosine similarity and Jacquard similarity are computed.

4. EXPERIMENTAL SETUP

The evaluation of the answers is performed by connecting the model answer with the student answer. The comparison can extract all the required features from the student answers. The solution is then additionally classified into the classes of description, cause and result, etc. The solution written by the student is compared with the model answer given. The factors for the analysis of the solution here is, firstly the answer type and also the question type ought to match. Secondly, the keywords or facts that are asked within the question must be addressed. Signal words further get checked to predict the category of the answer.

The proposed system was tested on a dataset of more than 200 questions and 20 students are expected to attempt the test in one session. This test is based on an engineering subject of stream Computer Science. The measure of correctness and accuracy of a student's response is calculated on the basis of the given categories:

1. Student answer's resemblance with the model answer
2. Answers containing only keywords
3. Answers whose word ordering is changed
4. Vague and contradictory answers

The similarity score is computed by evaluating and comparing the model against the answers fed in the system, namely model answers. The grammatical syntax, synonyms, and the context of the answers are also taken into consideration. Similarity measure is computed using Cosine similarity along with Jacquard similarity. The final score reviewed would comprise of all the above stated metrics.

5. RESULTS

The final result's computation is an additive measure of keyword match, similarity score and context of the answer. The classification modules of the system show 88% and 93% accuracy in classifying the question type and the answer type respectively.

The graph below shows a comparative analysis of a student's response versus the model answers fed into the system. Each question is evaluated against the marks assigned to it. Each question has different marks associated with it. Questions 1 to 4 are of 2 marks each, questions 5 to 7 are of 3 marks each, questions 8 and 9 are of 4 marks each and question 10 is of 5 marks.

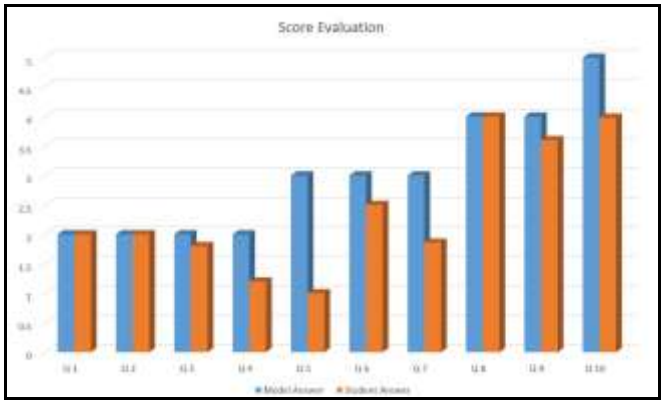

Fig. 4 A student's comparative Score Analysis

6. CONCLUSION

The present-day evaluation system confronts more challenges in grading the answers written by the students. There are a lot of issues associated with the scheme of the manual evaluation. Substantial resources are required as well as it a very tedious and time-consuming task. To purge these difficulties out of the current evaluation system, an automated evaluation system has been proposed for evaluating descriptive type of short answers. If the answer written by the student falls into the category of the question which is been asked, then it reflects that the student has certain knowledge in expressing about the topic. Further research can be made to improve the grammar for a language which includes the different ways in which students write.

The algorithm proposed has a large execution time which needs to be reduced considerably for efficient execution. The answers are currently stored in text files. This can be replaced with something that takes less space. The logic of this algorithm is very naïve. It

simplifies very complex tasks and does it in a very simple way. As a result, it takes too long to execute. Also, it cannot handle very complex sentence structures. English is a very confusing language to be analyzed completely. Same sentence can be constructed in a number of ways. So the method proposed for contextual matching, especially considering the antonyms, will not work for some complex sentence structures.

Hence, the logic of contextual matching of this algorithm needs to be enhanced. Since, it was out of scope of the project, we haven't considered the fact that the students may be asked to write programming codes in the answers. To be able to check those, we would need to create a compiler and then compare the output and would require many other complex things. Also, students currently do not have any way, in this project, to draw diagrams. So a graphical user interface will have to be created and image processing will need to be used to check for the correctness of the diagrams drawn. To conclude, this project forms only the basis of a functioning system. People can add onto it to create a fully functional grading system.

REFERENCES

[1] Hearst, Marti A. "The debate on automated essay grading." *IEEE Intelligent Systems and their Applications* 15.5 (2000): 22-37.

[2] Jun, Hu, and Huang Houkuan. "An algorithm for text categorization with SVM." *TENCON'02. Proceedings. 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering*. Vol. 1. IEEE, 2002.

[3] Selvi, P., and N. P. Gopalan. "Automated writing Assessment of Student's Open-ended Answers Using the Combination of Novel Approach and Latent Semantic Analysis." *Advanced Computing and Communications, 2006. ADCOM 2006. International Conference on*. IEEE, 2006.

[4] Ab Aziz, M. J., Dato'Ahmad, F., Ghani, A. A. A., & Mahmod, R. (2009, October). Automated marking system for short answer examination (AMS-SAE). In *Industrial Electronics & Applications, 2009. ISIEA 2009. IEEE Symposium on* (Vol. 1, pp. 47-51). IEEE

[5] Ab Aziz, Mohd Juzaiddin, et al. "Automated marking system for short answer examination (AMS-SAE)." *Industrial Electronics & Applications, 2009. ISIEA 2009. IEEE Symposium on*. Vol. 1. IEEE, 2009.

[6] Dumal, P. A. A., et al. "Adaptive and automated online assessment evaluation system." *Software, Knowledge, Information Management and Applications (SKIMA), 2017 11th International Conference on*. IEEE, 2017.

[7] Sijimol, P. J., and Surekha Mariam Varghese. "Short Answer Scoring System Using Neural Networks." (2018).

[8] Nandini, V., and P. Uma Maheswari. "Automatic assessment of descriptive answers in online examination system using semantic relational features." *The Journal of Supercomputing*(2018): 1-19.

[9] Nehete, Charusheela, et al. "Checkpoint--An Online Descriptive Answers Grading Tool." *International Journal of Advanced Research in Computer Science* 8.3 (2017).

[10] Qureshi, Jameel, and M. Rizwan. "A PROPOSAL OF ELECTRONIC EXAMINATION SYSTEM TO EVALUATE DESCRIPTIVE ANSWERS." *Science International* 27.3 (2015).

[11] Paul, Dimple V., and Jyoti D. Pawar. "Use of syntactic similarity based similarity matrix for evaluating descriptive answer." *Technology for Education (T4E), 2014 IEEE Sixth International Conference on*. IEEE, 2014.

[12] Chakraborty, Udit Kr, Rashmi Gurung, and Samir Roy. "Semantic Similarity Based Approach for Automatic Evaluation of Free Text Answers Using Link Grammar." *Technology for Education (T4E), 2014 IEEE Sixth International Conference on*. IEEE, 2014.

[13] SenthilKumaranV,SankarA(2015)Towardsanautomatedsystemforshort-answerassessmentusing ontology mapping. Int Arab J e-Technol 4(1):17–2

[14] Fernández-Iglesias MJ, González-Tato J, Mikic-Fonte FA (2013) Blended eassessment: migrating classical exams to the digital world. Comput Educ 62:72–87

[15] Papri Chakraborty (2012), "Developing an Intelligent Tutoring System for Assessing Students' Cognition and Evaluating Descriptive Type Answer", IJMER, pp 985-990

[16] Mita K. Dalal, Mukesh A. Zave (2011), "Automatic Text Classification: A Technical Review", International Journal of Computer Applications, pp.37-40

[17] Asmita Dhokrati, Gite H.R.2, Mahender C.N.3 (2012),"Computation Linguistic: Online Subjective Examination Modeling", Advances in Computational Research, pp-31-33.

[18] Wael H. Gomaa,Aly A. Fahmy (2013), "A survey of text similarity aproaches",International Journal of Computer Applications.