Automated Assessment of Students Responses to the Questions using Various Similarity Techniques

Mr. Varun Agarwal
School of Computer Engineering and Technology
MIT Academy of Engineering
Pune, India

Ms. Rutuja Sutar
School of Computer Engineering and Technology
MIT Academy of Engineering
Pune, India

Shweta Tiwari
School of Computer Engineering and Technology
MIT Academy of Engineering
Pune, India

Pankaj Choudhary
School of Computer Engineering and Technology
MIT Academy of Engineering
Pune, India

Abstract— **In current education, distance education or online degree program is the current trend through which students get their degree online, so assessment will be also online. The student knowledge about concepts can be evaluated only through assessments and student assessment is essential to measure the performance of individual students. There are many assessment procedures that are carried out. The method of computer-assisted objective testing is not sufficient because they do not produce any qualitative data and for quantitative assessment, we require a lot of manpower, so to minimize this automation can be done specifying similarity score judged differently based on the question types. For instance, semantics is not a key factor for short type answers. This paper describes the governance of automated student assessment systems on the basis of various similarity algorithms by taking some reference answer and applying preprocessing and then using cosine similarity, fuzzyWuzzy logic to check the similarity with the student response and calculate the similarity between them. Responses were collected from students from our college and some reference answers were picked by the faculty member, to assess them. This all helps in reducing manpower and will lead one more step towards automation.**

Keywords- Computer-assisted testing, semantics, similarity algorithms.

I. INTRODUCTION

In this era, the whole world is moving towards automation [3, 8] so there is a need to change the traditional approach of an answer evaluation system. It is a hectic task for a teacher to evaluate individual student’s answers and grade them accurately and also sometimes it’s not fair enough hence the evaluation of theory and allotting the marks requires new computer-assisted techniques. The objective questions can be graded using computer-assisted techniques but it is not sufficient to assess the overall performance of students so to overcome this problem various similarity algorithms can be used to evaluate subjective answers of students.

II. TYPES OF ASSESSMENT

A. Objective question assessment:

It is the formative assessment that has a single correct answer. One of the common forms of computer-aided assessment is online quizzes or exams and this includes all objective questions. It cannot check in-depth knowledge of a student. The tests do not evaluate the candidate's languages or writing skills. Computer-aided assessment is more feasible in this assessment.

B. Subjective question Assessment:

It is the formative assessment that includes a short-answer essay, extended-response essay, problem-solving and performance test items. It evaluates students' understanding of subjects and concepts.

III. USE OF COMPUTER-ASSISTED SYSTEMS

It is envisaged that computer-aided assessment will play an increasingly important role in learning. [3] The most common kind of questions used in the computer-assisted system is objective test questions where answers selected are compared with predefined sets of answers. Computer-assisted evaluation of essays is continuing to research topic as we know that computer-assisted evaluation is not prone to human error but it includes many limitations.

A. Problem formulation and evaluation

Computerized assessment can make assessment more interesting immersive and interactive as it provides quick feedback. The assessment of objective questions is an easy task and it is used on a large scale but the assessment of subjective questions still remains a challenge. As we know the majority of online exam questions are objective and many systems assess them and give quick feedback but many of the systems which include subjective questions cannot assess the solution for it automatically. In this proposed paper we have discussed various methods and algorithms [5] that analyze and represent the associative patterns among them.

B. *Feasibility*

The various dimensions of human communities, tools, and methods of teaching –assessment have changed due to widespread information technology and its influence. As we know that assessment is an important and critical activity in the education system. For in-depth assessment, we have to use subjective questions. In this paper, we have proposed and implemented various techniques to assess the subjective answers automatically. This includes similarity assessing concepts such as cosine similarity [16], fuzzy logic [15], Jaccard similarity etc. We have checked the feasibility of these concepts to assess all types of subjective questions. Our approach is based on the similarity between a student's answers and reference answers.

IV. RELATED WORK

There are many pieces of research going into the proposed area. We review the literature regarding the topic to explore our current knowledge about the area in which we are studying the approaches that have been proposed to solve this problem. Nabin Maharjan Et al. [1] proposed an approach to assessing subjective answers and takes context into account. They have developed the probabilistic Gaussian mixture model using the DT-Grade corpus with four different levels of answer corrections. Their best performing model achieved a significant improvement of 9% in terms of accuracy. In 2010, Xinming Hu Et al. [3] explored an approach to automated assessment for subjective assessment based on the latent semantic indexing (LSI). The use of LSI reduced the influence of synonymy [2] and polysemy [2] and the reference unit vector unit is introduced to alleviate the problem of trickiness. Even though the results graded by this system are not equivalent at all. The system proposed by [4] Navjeet Kaur Et al. accesses a text by computing a percentage based on keywords matching between the students' answers and actual answer and irrelevant words are removed. It is used for summative assessment of short responses.

Recently the enhancement in the knowledge of natural language processing and machine learning encouraged several researchers to use these techniques in the assessment of short and long essay type answers. In 2018 Prince Sinha Thakur, Et al. [5] created an application system which provides an automatic evaluation of answer based on keyword provided to the system as input after scanning the students' answer and will evaluate the algorithm on the basis of number keyword matched and length of the answer. V Senthil Kumaran Et al. [6] accentuated that ontology mapping tries to find semantic correspondences between similar elements of different ontologies. The performance of the pre-processing part [8] can be increased, using a dynamic structure [8]. The information is always accessible and avoids a constant reading from disk due to the information concerning an exam, such as questions/answers teachers and student's answers are carried to a dynamic structure. Which will help to increase the accuracy of the system. To surpass the limitation Fátima Rodrigues Et al. [8] created paraphrases of reference answers it will provide different correct variations, for the same question, with a vocabulary more wide-ranging and less deterring that will consent a more accurate assessment. Miguel Santamaría Lancho Et al. [7] performed an experiment and after summarizing all the results and come up to inference that all from tools the G-rubric.[7] They have proposed the utility and satisfaction graph of G-rubric tool which is able to give accurate and formative feedback for short and open-ended questions. In 2016 Automated Essay Grading System (AEGS) [10] that provides automated grading and evaluation of the student, essays proposed which rely on natural language processing and neural network grading engine. The similarity measures such as WordNet [13] String match and spreading process to calculate similarity [2], [5] are applied to the graphical form of students' subjective answers. WordNet [13] [14] is applied to the initial input to overcome the problem of lexicon ontology. The outputs of LSA [14] are mapped using the Soft computing technique and fuzzy logic. Checking of grammar and also antonyms checking is followed to preprocess the answers [15]. In the experiment carried out by Stig Johan Berggren et al. [11], sci-kit learn library is used to minimize multinomial loss and 'lgbfs' solver linear regression model[10,11] is also used.

V. APPROACH

A. *Data Set:*

The data that we have used for our research is basically the data extracted from 100 Information technology students through college application and are scored manually by faculty. This data was available to us in .txt format which we converted in .csv format because of faster processing time in .csv. Question asked to them was "What is API?" and we had taken some reference answers to evaluate the students.

B. *Preprocessing module:*

In this preprocessing method we are comparing student's answers with some reference answers and matching of keywords from this finds the similarity [13] between student's answers and reference answers. This kind of system may suffer from some problems of word checking to solve this we performed some preprocessing on the data. Following are some preprocessing tasks:

Removal of stop words: stop words are the most common words which do not affect the semantic meaning of the sentence [3]. These words are filtered out before processing as they are not meaningful in the quotation of an answer.

· Lemmatizing: Lemmatize means sorting the word so as to group together modulated forms of the same word. The goal of performing lemmatizing is to generate root from the inflected word.

ISSN: 2278-0181

- Stemming: It is the process of producing root word the same as lemmatizing but in this process, stem might not be an actual word.

- Removal of punctuation: This task removes all the punctuation such as full stop, comma from the answer.

- Tokenizing: In this process, we have split the string and sentences into a list of tokens.

C. Evaluation module

After preprocessing both reference answers and responses, our final task to score the students based on the similarity with the reference answers and for that we used some similarity techniques:

Fuzzywuzzy:
It is a ratio function that computes the standard Levenshtein distance similarity ratio between two sentences or sequences. There is fuzz. Token function in python which is having an important advantage over ratio and partial ratio. They tokenize the string and preprocess them by turning them to lower case and gets rid of the punctuation, but in the case of fuzz.token_sort_ratio(), the string tokens get sorted alphabetically and they joined together then fuzz ratio is applied to get the similarity percentage.
Jaccard Similarity:
The measurement is referred to as the number of common words. More common words mean both objects should be a similarity.

$$jaccard\ similarity = \frac{(A \cap B)}{(A \cup B)}$$

The value ranges in this between 0 to 1. The value 1 represents that both the sentences are identical and 0 represents that there is no common similarity between them.
The limitation of this method is that it does not handle the synonym scenario.
Cosine Similarity:
Cosine similarity between two sentences can be found as the dot product of their vector representation.

$$Cosine\ Similarity = \frac{\sum_{i=0}^{n} \mathrm{Ai\ Bi}}{\sqrt{\sum_{i=0}^{n} \mathrm{Bi}^2}\ \sqrt{\sum_{i=0}^{n} \mathrm{Ai}^2}}$$

$$similarity\ = \frac{\mathrm{A.B}}{||\mathrm{A}||.||\mathrm{B}||}$$

For cosine it is noted that it is very much related to the common words, that is a greater number of common words similarity increases. More of the part was done through stemming and lemmatizing but the other part was a question. So, one thing that we noted which resulted in the increase of similarity is number of reference answer you provide.
Figure 1: x-axis=Question number of students
y-axis=Similarity score

As we can see that when similarity score is calculated with one reference answer then there is a large deviation of score from normal assessment.

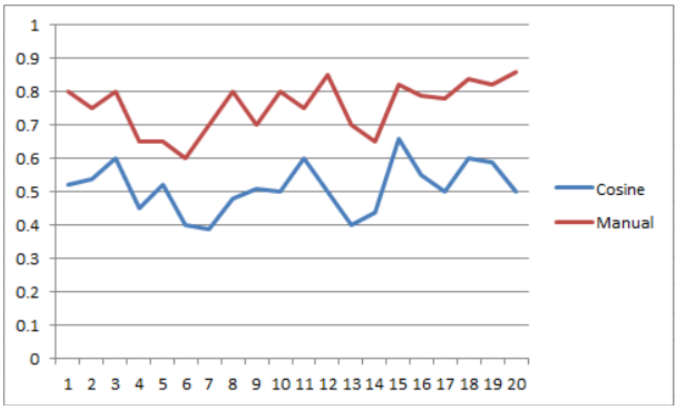

Figure 1

Figure 2: x-axis=Question number of students
y-axis=Similarity score
For this we can that as 2 reference answer are taken there is a change in the deviation.

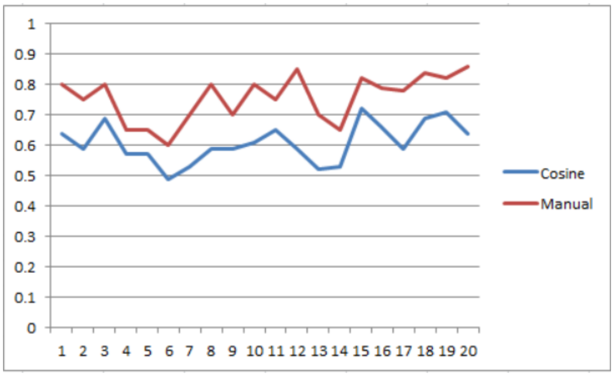

Figure 2

Figure 3: x-axis=Question number of students
y-axis=Similarity score
This shows we have reached up to a certain level where similarity score is coming nearer to manual score resulting in having less error.

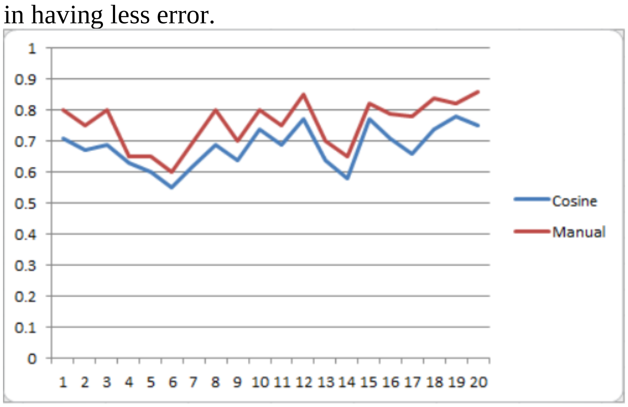

Figure 3

D. Conclusion:

TABLE I.

Similarity Algorithms	Cosine Similarity	FuzzyWuzzy	Jaccard Similarity
Our Method	74.7%	65.3%	57.5%
Manual	83.2%	83.2%	83.2%
Error	8.5%	17.9%	25.7%

From above table we can infer that our method is showing higher accuracy using cosine similarity and lower accuracy is shown by Jaccard similarity. So, cosine similarity can be used for short answer evaluation and a greater number of reference answer should be provided that will help in attaining greater accuracy

REFERENCES

[1] Automated Assessment of Open-Ended Student Answers in Tutorial Dialogues Using Gaussian Mixture Models by Nabin Maharjan, Rajendra Banjade, Vasile Rus, Department of Computer Science/Institute for Intelligent Systems the University of Memphis, Memphis, TN, USA {nmharjan, rbanjade,vrus}@memphis.edu.

[2] Synonymy and polysemy by JiweiCi , University of International Business and Economics, Beijing, People's Republic of China

[3] Automated Assessment System for Subjective Questions Based on LSI byXinming Hu and Huosong Xia College of Economics and Management, Wuhan University of Science and Engineering, Wuhan, Hubei, China huxinming@sina.com.

[4] Automated Assessment of Short One-Line Free-Text Responses In Computer Science Navjeet Kaur & Kiran Jyoti Guru Nanak Dev EngineeringCollege, Ludhiana, Punjab, India. Email:navjeet.kaur23@gmail.com,kiranjyotibains@yahoo.com.

[5] Answer Evaluation Using Machine Learning, Prince Sinha, Ayush Kaul Thakur Sharad Bharadia, Dr. Sheetal Rathi, March 2018.

[6] Towards an automated system for the short-answer assessment using ontology mapping by Senthil Kumaran and A Sankar PSG College of Technology, Coimbatore,India.International Arab Journal of Information Technology · January 2015

[7] Using Semantic Technologies for Formative Assessment and Scoring in Large Courses and MOOCs by Miguel Santamaría Lancho, Mauro Hernández, Ángeles Sánchez-Elvira Paniagua, José María Luzón Encabo and Guillermo de Jorge-Botana, Journal of Interactive Media in Education, 2018(1): 12, pp. 1–10, DOI: https://doi.org/10.5334/jime.468.

[8] AUTOMATIC ASSESSMENT OF SHORT FREE TEXT ANSWERS by Fátima Rodrigues and Lília AraújoGECAD – Knowledge Engineering and Decision-Support Research Center.

[9] An Exploration of Automated Grading of Complex Assignments by ChaseGeigle,ChengXiangZhai,DuncanFerguson,L@S 2016 Automated Assessment

[10] A Hybrid Scheme for Automated Essay Grading Based on LVQ and NLP Techniques by Shehab, Mohamed Elhoseny, and Aboul Ella Hassanien. The Scientific Research Group in Egypt (SRGE).

[11] Regression or classification? Automated Essay Scoring for Norwegian by Stig Johan Berggren ,Taraka Rama,Lilja Øvrelid Building Educational Applications workshop, 2019.

[12] Subjective Answer Evaluation System by Aditi Tulaskar, Aishwarya Thengal, Kamlesh Koyande,International Journal of Engineering Science and computing April,2017

[13] E-Learning: an Approach to Evaluate Subjective Questions For Online Examination System Using Data Similarity: Research Paper by Manisha Malyal, Ms. Sudheshna, Dr. Soniya, in International Journal of Information Technology and Management | IT & Management .

[14] Computerized Evaluation of Subjective Answers using Hybrid Technique. Innovations in Computer Science and Engineering, Edition: 1, Publisher: Springer-Verlag Singapur, Editors: H. S. Saini, Rishi Sayal, Sandeep Singh Rawat.

[15] Handwritten Short Answer Evaluation System (HSAES) by Sijimol P J, Surekha Mariam Varghese,2018 IJSRST | Volume 4 | Issue 2 | Print ISSN: 2395-6011.

[16] INTELLIGENT ELECTRONIC ASSESSMENTFOR SUBJECTIVE EXAMS Alla Defallah Alrehily, Muazzam Ahmed Siddiqui, Seyed M Buhari Faculty of computing and information technology, King Abdulaziz University,Saudi Arabia, Jeddah