Subjective Answer Evaluation System

Kavita Shirsat[1], Akash Hamare[2], Sourabh Pardeshi[3], Vaishnavi Patil[4]

[1]*Assistant professor, Dept. of Computer Engineering, Vidyalankar Institute of Technology, Maharashtra, India*
[2,3,4]*UG student, Dept. of Computer Engineering, Vidyalankar Institute of Technology, Maharashtra, India*

---***---

Abstract - *Exams and assignments play a crucial role in determining the overall academic performance of the students and foster their cognitive learning. However, evaluation of these papers is quite a tedious job for the evaluators as they come in huge numbers. Most of the competitive exams typical comprise of objective and multiple-choice questions. In this ever-increasing modern age, where the world moves towards automation, there is a need for automation in the answer evaluation system. However, there hasn't been developed any system which could assign grades to the descriptive questions. The current system takes extra manpower and the process is laborious. Hence, there is a high need of developing an auto evaluation system which could perform the task of analyzing and assigning precise marks to the given subjective answer. This automation of descriptive answer evaluation process would be helpful for various universities and academic institution to efficiently handle the assessment of exam answer sheets of students. Our objective is to design a Subjective Answer Evaluation Model for the automatic evaluation of the multiple sentence subjective answer. This paper provides an outlook to test the degree of student's learning, by evaluating their answers. Our system uses concepts of natural language processing and Machine learning to achieve the goal. The proposed system uses the techniques of natural language processing for preprocessing the text and then using machine learning algorithms for evaluation of the text and assigning the accurate grades. This system will be useful for educational institutions as the process of evaluation of descriptive answers will be automated to fully examine student's exam answer sheets.*

***Key Words*: descriptive answers, sentence embedding, semantic similarity, NLP, Machine Learning.**

1. INTRODUCTION

The educational system has been continuously transforming from Gurukul to our current day technological learning. The learning process always involved an exchange of information between the guide and the learner. As the technological advancements boomed, the process of e-learning was introduced which reduced the mechanical learning process. Almost, all types of educational, non-educational institutes conduct examinations and the kinds of questions are mostly objective and subjective. Most of the ambitious, competition driven exams are objective based and hence, the process evaluating is effortless because of the machine computation. However, such an approach cannot be implemented in the university examinations as the answer format is subjective. Hence, there is a rising necessity of a system which will automatically grade the descriptive solutions.

Moreover, the standard of Indian education system is gravely hampered because of the weak infrastructure and the ever-rising population. Therefore, the amount of stress encountered by the teachers is unimaginable because of the exorbitant number of answer sheets to evaluate. Manual answer correction is a tiresome and consumes huge amounts of time. Motivation behind automation of descriptive answer evaluation includes fast processing, less manpower, independent of change in psychology of human evaluator, ease in record keeping and extraction. It also ensures uniform evaluation irrespective of any mood swings or change in perspective of human assessor. Subjective answer assessment is considered as one of the excellent ways of evaluation of student's subject understanding and knowledge. Answer assessment is a checkpoint to trace the goal of the learning actions and enhances the execution of the learning process. Developing systems that automate the scoring have eased human interventions to a greater extent. Therefore, an auto-evaluation system embedded in a web portal could ease the process from both student and professors' point of view. The proposed system aims of providing a student- teacher interface for submitting and evaluating the assignment answers respectively.

2. LITERATURE SURVEY

Many features and designs have been proposed for the evaluation of subjective answers. The approaches are mainly based on keyword match, sequence match, semantic and quantitative-analysis. Neethu George, Sijimol PJ, Surekha Mariam Varghese [1] proposed a model for an auto evaluation system which uses techniques of Natural Language Processing for the preprocessing of the answer text. This processed answer text is converted into glove vector representation using embedding layers. The vector form of answer text is then passed to a LSTM-RNN network which is previously trained on 50 similar model answers for that same question. Once the answer's vector is passed through the LSTM-RNN, it is then passed through a dense layer using softmax activation function and hence is classified into one of the grade classes. Toshavi Patil [2] proposed a system in which the preprocessing of data is done in the beginning along with data normalization, stop word removal, stemming and lemmatization. Once the answer is preprocessed, some keywords are extracted from the answer text and then matched with keywords from the model answers by calculating the distance between keywords and number of keywords matched. Finally, the answer is passed through ANN to classify it to one grade class.

ApTeSa is a proposed tool in the paper by Dharma Reddy [3] which was used by MLR institute of Technology. This system makes use of keyword matching and phrases matching. On the basis of number of matches for keywords and phrases, marks are assigned for that answer. In the method proposed by Dr. Sheetal Rathi[4], the professor gives the keywords, maximum marks and minimum length of the answer. Then, keywords are extracted from the student's answer and are compared by the keywords stored in the dataset. Depending on the percentage matching of keywords and phrases, the marks are allocated to the answer. In the paper written by Wallace Dalmet [5], they proposed a different way of grading students answers using Deep Learning. According to this article, use of Siamese deep neural networks could be useful in evaluating the similarity in the answers. Siamese Neural Network here consists of two Manhattan LSTM networks which take vectors as input and provide a similarity score between the two outputs of the two Manhattan LSTM. The input sentence is converted into a vector via word embeddings. Thus, the two sentences are converted into vectors and then passed to the Siamese network to get a similarity score. Similarly, this can be used to grade student answers by comparing it to the Model answer and getting the similarity score.

3. LIMITATIONS OF THE EXISTING SYSTEM

The major drawback of some above mentioned solutions is the sole dependency on the keywords matching. Furthermore, the absence of semantic analysis of the answer with respect to the model answer could yield inefficient results as semantically similar answers may not have the same keywords. Another drawback in one of the systems mentioned above[1] is that this system asks for approximately 50 model answers to train the neural network and then evaluate the student's answers. In such a case, the professor needs to submit many answers for one question which is a time-consuming and tedious task. Moreover, this process of providing sample answers and training the model would be repeated for every question, which is not feasible.

The System using the Siamese network for evaluating the answer[5] also comes with one major drawback. Although there are some datasets which can be used to train the Siamese neural network like SNLI dataset, Quora Question Pairs dataset, these datasets contain short sentences which are very generic in nature. On the other hand, for evaluating academic answers the neural network needs to be trained on more domain specific sentence pairs and this need to be done for multiple domains. The biggest hurdle in achieving this is the building of the huge domain specific dataset which would efficiently evaluate answers from multiple domains. Thus, even though the approach is intelligent, it may not efficiently grade the answers.

4. PROPOSED SYSTEM

In the proposed System we make use of Natural Language Processing Techniques to extract the semantics of the answer text and then use machine learning algorithms to predict the score of the submitted answer. This system uses a centralized database which will store the uploaded answers from both students and faculties. The system extracts answer texts from the uploaded files and passes them to the Evaluation Model. The Evaluation Model takes the Model answer and Student's answer as input and outputs the grades ranging from 1-9. Our model makes use of 3 factors for evaluating the answer, namely Similarity index, Grammar and Question Specific Parameters(QSP). These factors are explained in detail below:

4.1 Similarity index:

This factor is to give the similarity between the two provided answers in a range of 1-6 (1 being highest). The two answer texts cannot be directly compared in raw form, hence they are converted to a vector form which is a logical representation of the text. This conversion should take into account the semantic and contextual meaning of the sentences. Hence, we use Universal Sentence Encoder (USE) which is a sentence embedding tool which converts sentences into a 512 dimensional embedding vector with taking care of semantic and contextual meaning. This tool specifically targets transfer learning to other NLP tasks. The USE gives similar embedding vectors for similar sentences.

Once the answer is converted into a 512 dimensional vector, we make use of cosine similarity to get the similarity between the answer texts. The cosine similarity gives output between 0-1 which is scaled to the range 1-6 and is labelled as Similarity Index.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

4.2 Grammar

The System will check for grammatical errors in the answer texts provided by the student. The score for grammar is allocated on the basis of the number of grammatical errors. This score is binary , that is , 1 for good grammar and 0 otherwise.

4.3 Question Specific Parameters (QSP)

There are some key sentences which are very specific to the particular question and need to be in the answer, like definition, facts and figures. Since these things are very important and should be included in the student's answer, so we take such text as an input from the faculty and then check for these texts in the students answer. This is done with the help of algorithms like Longest Common Subsequence which looks for a particular text subsequence(QSP) in large paragraphs(answer texts). Depending on the QSPs included in the student's answer, we assign a value ranging from 1-6 (1 being highest).

4.4 Model

We built a small dataset with combinations of numeric values of these three parameters and corresponding marks to be given for such combinations. This dataset is then used to train a model using Naive Bayes machine learning algorithm. The trained model will then take these three parameters as input and give a predicted output between 1-9 which would be considered as a grade evaluated by the model. Hence, after the model receives an answer text, it extracts the numeric values of above mentioned three parameters and then predicts the marks/grade that should be given to that answer.

5.RESULTS

5.1 USER INTERFACE

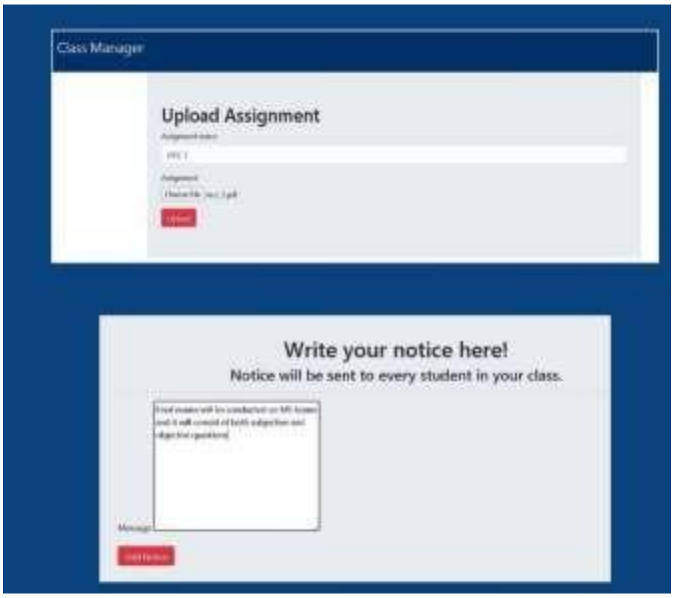

Fig 1. Assignment and notice uploading interface for the teacher

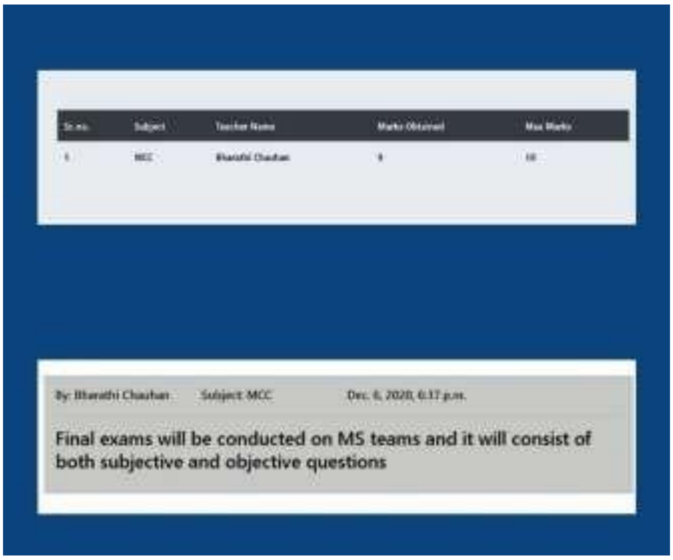

Fig 2. Viewing marks allocated and checking notice

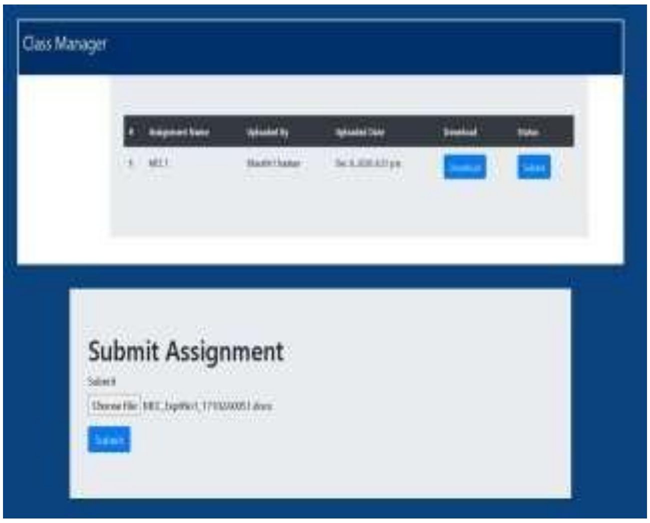

Fig 3. Assignment viewing and submission interface for the student

5. 2 MODEL OUTPUT

MODEL ANSWER	STUDENT SUBMITTED ANSWER	SIMILARITY INDEX	GRAMMAR	QST	FINAL GRADE ASSIGNED BY PROF	FINAL GRADE ASSIGNED BY AUTO EVALUATION MODEL
Learning is a process that helps in nurturing and developing the minds of young people	Learning helps in the process of developing and nurturing the growth of young people	1	1	1	9	9
Thermal energy is a process of converting heat energy into electrical energy. It is cost-friendly and does not cause pollution	Thermal energy is the conversion of heat energy into electricity	1	1	3	7	8
Astronomy is the study of the sun, moon, stars, planets, comets, gas, galaxies, gas, dust and other non-Earthly bodies and phenomena	Astronomy is the overal study of the sola system and other non-Earthly bodies	3	1	1	6	5
Mental health and physical health go hand in hand. Both of them should be given equal importance	Mental health is very important than physical health. In every field such as sports,etc., mental strength has an edge over other.	2	1	3	7	7

6.CONCLUSION

In this review paper, we discuss the implementation of a new technique which will automatically perform the grading of subjective answers. Universities always judge the performance of the students based on their scores in the assignments and the final exams. Now, while a majority of the exam types is online multiple-choice questions, online testing machines are available to grade them. On the contrary, the end term exams are always descriptive and hence, there is a high need of a system which can automatically grade these brief answers without taking much time. Thus, our proposed system provides an interactive platform for all the educational institutes and assigns accurate grades to the subjective answers submitted by the student. It attempts to grade the student's answers based on the following three parameters: similarity index, grammar and Question Specific Parameters(QSP). The model solution provided by the faculty will be compared with the solution submitted by the student and based on the above three parameters, appropriate grades will be assigned. Such systems can be helpful in many online evaluation platforms and college portals as it saves time and the trouble of checking bundles of answer sheet.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our Head of Department of Computer Engineering Dr.Sachin Bojewar for his support and guidance and giving us this opportunity to explore more on this domain. Our sincere gratitude towards our mentor and project guide, Mrs. Kavita Shirsat who has been constantly supporting us and her continuous engagement in throughout the course of this work. It would have been difficult without her guidance and valuable time. Last but not the least, we also thank each and every individual who had supported us in providing appropriate insights to this domain.

REFERENCES

[1] George, Ng'Ochoi & Sijimol, P. & mariam varghese, Surekha. (2019). Grading descriptive answer scripts using deep learning. International Journal of Innovative Technology and Exploring Engineering. 8. 991-996.

[2] Toshavi Patil, " Automatic Assessment of Descriptive Answers for Online Examination using Semantic Analysis", IJITEE

[3] Tetali, D.R. & Kumar, G. & Ramana, L.. (2017). A python tool for evaluation of subjective answers (aptesa). International Journal of Mechanical Engineering and Technology. 8. 247-255.

[4] Bharadia, Sharad & Sinha, Prince & Kaul, Ayush. (2018). Answer Evaluation Using Machine Learning.

[5] Dalmet, Wallace & Das, Abhishek & Dhuri, Vivek & Khaja, Moinuddin & Karamchandani, Sunil. (2020). Siamese Manhattan LSTM Implementation for Predicting Text Similarity and Grading of Student Test Papers. 10.1007/978-981-15-1002-1_60.

[6] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, Ray Kurzweil, "Universal Sentence Encoder".

[7] Baoxun Wang, Bingquan Liu, Xiaolong Wang, Chengjie Sun, and Deyuan Zhang. 2011. Deep Learning Approaches to Semantic Relevance Modeling for Chinese Question-Answer Pairs. <i>ACM Transactions on Asian Language Information Processing</i> 10, 4, Article 21 (December 2011), 16 pages. DOI:https://doi.org/10.1145/2025384.2025389

[8] Dimitrios Alikaniotis, Helen Yannakoudakis, Marek Rei, "Automatic Text Scoring using neural network".

[9] Shweta Deotare, Rubeena A. Khan, "Automatic Online Subjective Text Evaluation using Text Mining".

[10] Nandini, V., Uma Maheswari, P. Automatic assessment of descriptive answers in online examination system using semantic relational features. *J Supercomput* 76, 4430–4448 (2020).

[11] Sadiq, Ahmed & Kadim, Khudhair. (2015). Text Similarity Based on Modified LSA Technique. 5. 64-70.