ISSN No:-2456-2165

Substantiating Precise Analysis of Data to Evaluate Students Answer Scripts

Anshika Singh[1]
Subharti Institute of Technology & Engineering,
Swami Vivekanand Subharti University,
Meerut, India

Dr. Sharvan Kumar Garg[2]
Subharti Institute of Technology & Engineering,
Swami Vivekanand Subharti University,
Meerut, India

Abstract:- Handwriting recognition refers to interpreting and analyzing handwritten text. In rece-nt years, there have- been notable advance-ments in this field, espe-cially in the context of computerize-d assessments. As online e-xams and digital education platforms continue to gain popularity, handwriting recognition plays a crucial role- in evaluating students' written answers. Our proposed system automatically recognizes and scores handwritten responses on answer sheets by comparing them to the correct answers provided by a moderator. To achieve this, the system utilizes Optical Character Recognition (OCR) to convert the handwritten text images into computer-readable text. Additionally, BERT is employed to convert the text into embeddings, and cosine similarity is utilized to take these embeddings as input and provide a final matching confidence score.

Keywords:- *OCR, Google Vision OCR, BERT, Cosine similarity.*

I. INTRODUCTION

In Universities Colleges as well as schools common way of assessment lies in manual evaluation of written examination attempted by the students. The student's response is evaluated based on their understanding of language, concepts, and other relevant aspects. Professors encounter numerous challenges when manually grading handwritten answer booklets. Answer scripts can be in the form of OMRs in case of multichoice question paper which is easier to evaluate using a computer but when we need to evaluate descriptive answer scripts it is trickier due to the nature of handwriting, way of expressing ideas or keywords etc. Hence, the evaluation task demands a significant amount of time and labor.

Ganga Sanuvala et al. presents a model for assessing descriptive responses in tests through a three-module evaluation system: 1) Scanning; wherein an Optical Character Recognition (OCR) technology is employed to scan the page and retrieve student responses and kept in a dataset that is in the form of a text file. 2) Preprocessing; where NLP is utilized to extract a collection of distinct words that correspond to each sentence in the response by conducting a grammatical check, tokenizing the text, removing stop words, checking for synonyms and antonyms, and performing stemming. 3)Learning; comprising both training and testing. During the training phase, a model is constructed by acquiring knowledge from the scored responses dataset and the answer key. The model is employed to assess the ungraded responses during the testing stage. The ungraded responses are transformed into TF-IDF vectors, then cosine similarity matching is executed using the trained model to award scores[1].

Aqil M. Azm et al. developed a system that applies LSA(Latent Semantic Analysis), RST(Rhetorical Structure Theory), which involves two stages: training and testing. In the training phase, pre-scored essays are used to train the Latent Semantic Analysis model. The training set includes essays that were scored by human instructors. During testing stage, a new essay is processed through several steps including pre-processing, checking cohesion, counting spelling mistakes, comparing essay length, and applying RST. The essay's overall score is calculated based on the weighted sum of the scores assigned to semantic or conceptual analysis, spelling mistakes and writing style[2].

Muhammad Farrukh Bashir et al. presented a method that employs machine learning (ML) and natural language processing (NLP) techniques like WMD(word mover's distance), Cosine similarity, MNB (Multinomial naive Bayes) to evaluate subjective answer responses. Assessment of responses involves employing solution statements and relevant keywords, while a ML model is constructed to forecast the grades of the replies. The comparison score is determined by assessing the solution sentence against each answer sentence using keyword weighting and similarity distance calculations. The process of keyword-weight computation involves identifying keywords in both the solution sentence and the matching answer sentence. The keyword-weight number, which falls within the range of 0 to 1, is derived by computing the proportion of keywords in the answer sentence relative to the solution sentence and dividing it by 100. The computation of similarity distance is performed using either the Word Movers Distance (WMD) method or the Cosine Similarity (CSim) approach. The present comparison score is computed by amalgamating the similarity weight and keyword-weight when the similarity distance descends below a certain threshold of 30% keywords present. The keyword-weight is considered only when the similarity distance is over the threshold, but only if it is below 30%. The overall score is calculated by taking the average of the current comparison scores for all solution sentences[3].

ISSN No:-2456-2165

Sijimol P J et al. proposed a system named HSAES(Handwritten Short Answer Evaluation System) which is a system specifically created to automatically detect and assess short responses written on answer papers. The system employs Optical Character Recognition (OCR) technologies to retrieve handwritten texts and utilizes NLP to obtain keywords from human-evaluated sample datasets of answer papers(handwritten) and answer keys. The model assesses scores by utilizing sentence cosine similarity metrics, where every sentence in an examined answer paper is assigned a corresponding score or marks. The established approach can be utilized to assess the grades of ungraded brief responses. The study emphasizes the necessity of employing automated essay scoring approaches to enhance the caliber of students' writing and diminish the laborious and expensive process of essay evaluation. The model is trained by selecting text files at random and utilizing a heavily weighted response key text file. The computation of semantic similarity between sentences can be achieved by a method that relies on cosine similarity. Mobile Agents are a highly efficient paradigm for distributed applications. The objective of the suggested system is to automatically evaluate descriptive text responses through the use of three modules: scanning, training, and testing. The scanning phase processes answer paper PDF files and extracts pertinent characteristics from the answers. It employs a model to assess the marks of descriptive answers. Meanwhile, the training phase identifies crucial sections from the answer sheets and maps each processed answer and key to vector spaces using TF-IDF score and cosine similarity score. The learning phase entails generating a trained model by acquiring knowledge from the scored responses dataset and the answer key. The testing phase evaluates unanswered questions by utilizing the acquired knowledge in the training model. It transforms these questions into TF-IDF vectors and then compares their similarity using cosine-based similarity. The grades are determined based on the scores of the sentences that are most comparable to the unscored replies[4].

To address the issue of manual assessment of answer booklets, our paper proposes a system based on BERT and cosine similarity which will result in reducing the time and effort of an examiner.

II.METHODOLOGY

Our proposed system (Figure 1) employs sophisticated artificial intelligence techniques for various objectives- Optical Character Recognition (OCR) is employed to convert handwritten answers from images into textual data; Text Embedding is utilized to convert regular text into a numerical representation by utilizing a deep learning model, which aids in the system's comprehension of the student's intended message while placing emphasis on significant terms; The concept of vector similarity is employed in the system to evaluate the correctness of answers. This is done by assessing the numerical representations of the student's response in relation to the correct answer. There are numerous methods for assessment of handwritten answer scripts majorly relying on matching keywords, comparing sequence of words and performing quantitative analysis. Evaluating subjective or descriptive answers is still a challenging task which remains unresolved and even trickier in case of handwritten.

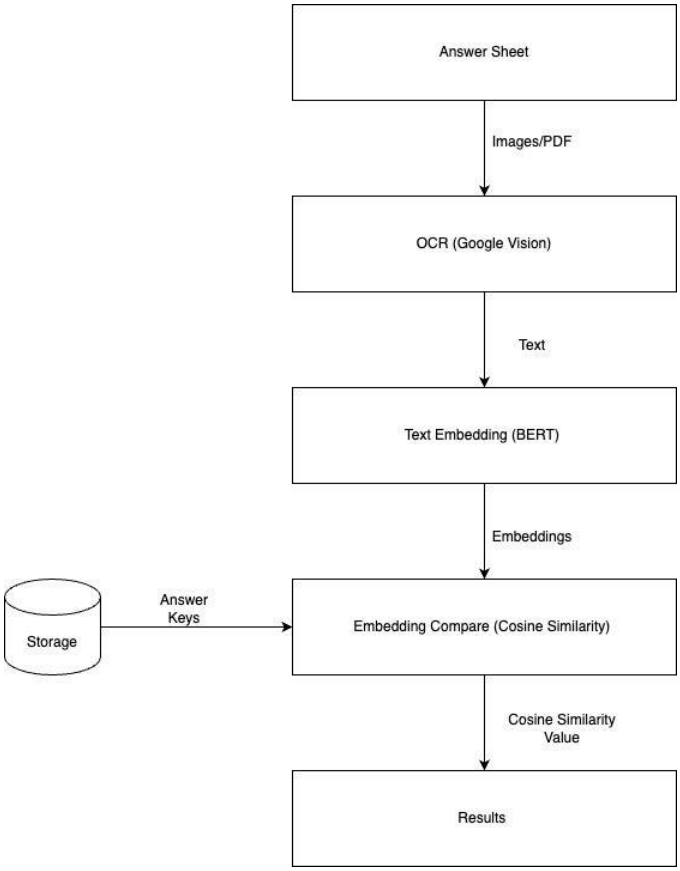

Fig1: Proposed system

Initially, We will collect scanned answer scripts submitted by the students which are then fed to Optical Character Recognition. OCR is an automated technique used to identify and interpret the characters included in digital images by performing feature extraction process and classification stages. It is a highly reliable technology in the field of pattern recognition and AI(Artificial Intelligence) which is employed to convert handwritten text including words, letters, or characters into a digitized format that can be easily edited, searched, and stored more efficiently[5,9].

We may utilize a CNN or RNN or any deep learning model to develop a customized Optical Character Recognition (OCR) system. We require the training data and develop our own model which can process each character individually and subsequently combine them into a cohesive word. The accuracy of the model depends on various criteria like diversity of dataset, size of dataset, pre and post processing of data and to create such an accurate model is a challenging task due to lack of availability of resources like cost and time. So, it is advisable to utilize existing OCR models like “tesseract” which is trained on large datasets catering different languages. However, the level of accuracy is still not meeting our minimum standards. Furthermore, a significant amount of pre and post processing is still necessary for the misclassification of some characters. A better option is to use publicly accessible OCR APIs like Google Cloud Vision OCR, Amazon Textract, Microsoft

ISSN No:-2456-2165

Azure Computer Vision API which are cost-effective as compared to any user made OCR model. They undergo training using extensive datasets and are capable of processing numerous languages, including Indian languages. It enables us to perform OCR with exceptional precision and little time consumption. Google Cloud Vision OCR API gives 98% accuracy and support 200+ languages while both Amazon Textract and Microsoft Azure Computer Vision API result in 95% accuracy and support 200 and 120+ languages respectively. Therefore, we choose to utilize the Google Cloud Vision API for OCR.

The Google Cloud Vision API is a machine learning service hosted on the cloud that assists developers in comprehending the information of photos and movies. The system employs a range of machine learning methodologies, such as deep learning, to scrutinize photos and videos and derive valuable insights from them. The Vision API provides an extensive array of functionalities, encompassing: Image classification, Text extraction, Landmark recognition, Facial detection and analysis etc. The Google Cloud Vision API is a dynamic service that undergoes continuous development, with Google frequently introducing novel functionalities and enhancements. It is an invaluable resource for developers seeking to construct intelligent applications capable of comprehending and manipulating photos and videos[6].

Our next step is Text Embedding, which involves transforming ordinary text into a numerical format using Word2Vec, LSA, LDA or more advanced deep learning models like BERT. The main purpose of utilizing text embedding is to encapsulate the fundamental semantic meanings and connections between words or phrases, or texts in a way that is suitable to process and analyze via machine learning algorithms. This helps the system understand the student's intended answer's conceptual meaning while giving importance to important terms. Here, each individual word or token is paired with high-dimensional numerical vectors. Each dimension in this arrangement corresponds to a certain attribute or aspect of the word's meaning.

BERT embeddings are compact vector representations of words and phrases acquired by the BERT model through pre-training. BERT is a language model that has been pre-trained using transformers, which is capable of encoding text bidirectionally. It learns to comprehend the contextual connections among words in a given text. BERT embeddings differ from earlier word embedding approaches in that they are contextual, capturing the meaning of a word by considering its position in a sentence and the words surrounding it[7].

BERT embeddings are generated by training the BERT model on a vast collection of text data, enabling it to learn how to forecast the masked words inside the text and determine if two sentences are consecutive. As the model acquires the ability to execute these tasks, it cultivates a profound comprehension of the connections between words and phrases. The BERT embeddings encapsulate this comprehension by utilizing real-number vectors to symbolize the semantic essence of individual words or phrases.

BERT embeddings are applicable for several NLP applications, including text categorization, question answering, and sentiment analysis. For instance, BERT embeddings can be employed to train a text classifier that can differentiate between spam and ham emails, or to build a question answering model that can respond to queries on a specific passage of text.

After creation of BERT embeddings which transformed text into vectors with size 786. Now, vector similarity is employed in the system to evaluate the correctness of answers. It can be calculated using various similarity measures like Euclidean distance, Cosine similarity etc. A similarity measure employs these embeddings and produces a numerical value that quantifies their similarity. [8]. Cosine similarity is preferred over Euclidean distance and Jaccard similarity for measuring text similarity due to its distinct advantages that are specifically suited to the unique properties of textual data. Firstly, it standardizes vectors, rendering them unaffected by the magnitude of vectors, which is essential in text analysis where document lengths differ significantly. Euclidean distance, on the other hand, can be sensitive to high-dimensionality and often suffers from the "curse of dimensionality," which is typical with text data. Cosine similarity is particularly effective in dealing with high-dimensional and sparse data, which is commonly seen in text documents. It achieves this by emphasizing the angle between vectors, which accurately measures the similarity of texts based on word usage, a more relevant metric than the actual distance between vectors. Moreover, text data is intrinsically sparse, as the majority of documents only include a small portion of the entire lexicon. Cosine similarity effectively addresses sparsity by taking into account vector components that are not zero. Unlike Jaccard similarity, which focuses on set intersections and unions and may not consider word frequencies, cosine similarity also incorporates semantic meaning to some degree, making it a versatile and appropriate option for various natural language processing applications.

Output using Google Vision OCR:

Sample Input Answer:

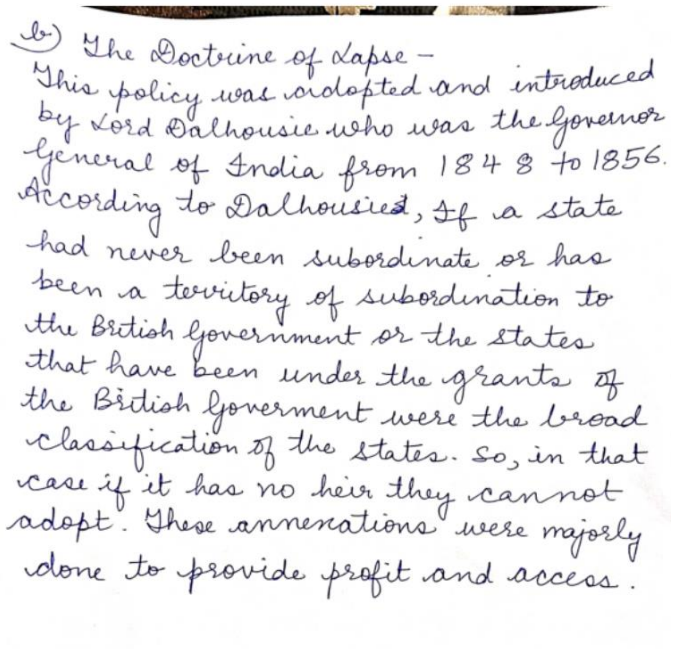
b) The Doctrine of Lapse –
This policy was adopted and introduced by Lord Dalhousie who was the governor general of India from 1848 to 1856. According to Dalhousies, If a state had never been subordinate or has been a territory of subordination to the British government or the states that have been under the grants of the British goverment were the broad classification of the states. So, in that case if it has no heir they cannot adopt. These annerations were majorly done to provide profit and access.

Fig 2. Sample handwritten Answer

Sample Output By Google OCR:

Answer 1-

b) The Doctrine of Lapse -

This policy was ordopted and introduced by Lord Dalhousie who was the Governor General of India from 184 8 to 1856. According to Dalhousies, If a state had never been subordinate or has territory of subordination to the British Government or the states that have been under the grants of the British Goverment , verment were the broad. classification of the states. So, in that been a case if it has no heir they cannot adopt. These annerations were majorly done to provide profit and access.

III. CONCLUSION

We applied our model on more than 100 answers and got OCR accuracy more than 90+% using google vision API. We pass the BERT embedding of the user handwritten answer and the correct answer embeddings. We find the cosine similarity for both of them. As cosine similarity works on vector direction we get average cosine similarity is 0.72 when the answer is not matched. And the cosine similarity is 0.81 when the answer is matched. Results with cosine similarity suggest that there is very little difference between positive and negative matched answers. Also we can't use cosine similarity for the partial scoring for the answer as its not matching percentage. Our work can be further extended to calculate the score for answers in terms of percentage. It also find its applications in grading of answers in languages other than English.

REFERENCES

[1]. G. Sanuvala and S. S. Fatima, "A Study of Automated Evaluation of Student's Examination Paper using Machine Learning Techniques," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, pp. 1049-1054, 2021. doi: 10.1109/ICCCIS51004.2021.9397227.

[2]. Aqil M. Azmi, Maram F. Al-Jouie, Muhammad Hussain, "AAEE – Automated evaluation of students' essays in Arabic language", Information Processing & Management, vol. 56, pp. 1736-1752, 2019. https://doi.org/10.1016/j.ipm.2019.05.008.

[3]. M. F. Bashir, H. Arshad, A. R. Javed, N. Kryvinska and S. S. Band, "Subjective Answers Evaluation Using Machine Learning and Natural Language Processing," in IEEE Access, vol. 9, pp. 158972-158983, 2021, doi: 10.1109/ACCESS.2021.3130902.

[4]. Sijimol P J, Surekha Mariam Varghese, " Handwritten Short Answer Evaluation System (HSAES)", International Journal of Scientific Research in Science and Technology(IJSRST), Print ISSN : 2395-6011, Online ISSN : 2395-602X, vol. 4 , pp.1514-1518, January-February-2018. https://ijsrst.com/IJSRST1841325

[5]. Singh, A., Garg, S.K. (2023). Comparative Study of Optical Character Recognition Using Different Techniques on Scanned Handwritten Images. In: Sharma, D.K., Peng, SL., Sharma, R., Jeon, G. (eds) Micro-Electronics and Telecommunication Engineering . Lecture Notes in Networks and Systems, Springer, Singapore 617. https://doi.org/10.1007/978-981-19-9512-5_38

[6]. Cloud Vision API. https://cloud.google.com/vision

[7]. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. https://arxiv.org/abs/1810.04805

[8]. https://developers.google.com/machine-learning/clustering/similarity/measuring-similarity

[9]. https://aws.amazon.com/what-is/ocr/