Handwritten Answersheet Evaluation Application

[1]ANANDHAN R, [2]GEORGE FERNANDEZ I , [3]SRIMATHI B, [4]SUBAKESHINI R

[1,3,4]A Student, Information Technology, Jerusalem College of Engineering, Chennai.
[2]Assistant Professor, Information Technology, Jerusalem College of Engineering, Chennai. Email:
[1]raja.anand2k@gmail.com, [2]george.contact@gmail.com

Abstract: Automatic evaluation of handwriting answers has been a difficult problem for education system for many years. To speeding up the evaluation remains as the major problem for enhancing the throughput of instructors. This paper shows an easy method for automatically evaluating the handwritten answers from the images. Our main goal is to build a application to evaluate a student's handwritten answer by assigning an evaluation score that is comparable to the human giving scores. Although many essay evaluation systems are available, subjective answer grading is still a tough problem. In the proposed system, we design a application interface using tkinter in order to get handwritten images and answer key input from user, Optical Character Recognition tools are used to extract the keyword printed texts in keyword answer image and Google Vision API tools are used to extract the handwritten texts in student handwritten answer images. In the proposed model evaluates scores based on cosine similarity function. The developed application can be used to evaluate the marks of the unscored handwritten answers. Our System is divided into three modules. The first and second one is extracting the data from the scanned printed text and handwritten answer images and the main module is applying NLTK and cosine similarity function to both extracted text and giving marks in screen.

Keywords: Automated Evaluation, tkinter interface, pytesseract OCR, Vision API, NLTK, Cosine Similarity.

1. INTRODUCTION

The analysis of text answers may be a difficult method that needs nice effort from the evaluators, particularly once the amount of answers to judge is high. Developing Associate in Nursing automatic answer script analysis system is required as a result of human analysis wants concentration and may be biased, whereas Associate in Nursing automatic answer analysis system are effective while not these limitations. relating to text answer analysis, sentence similarity measures are wide accustomed compare student written answers with reference texts. during this paper, we tend to propose an automatic answer analysis system that uses our planned cosine-based sentence similarity measures to judge the answers. circular function measures have verified to be effective in comparison between free text student answers and reference texts.

Handwritten Text Recognition and Evaluating the solution sheet of short answers and comparison with Key word to seek out the result with share of Matches may be a planned technology that's a lot of required during this world as of these days. Before correct implementation of this technology, we've relied on writing texts with our own hands and evaluating manually which may end in errors. It's troublesome for employees to correct an equivalent set of answer sheets manually. manual labour is needed so as to take care of correct organization of the information. Modern-day technology is property individuals store the information over machines, wherever the storage, organization and accessing of information is comparatively easier. Adopting the utilization of written Text Recognition package, it's easier to judge the hand written answer sheets. what is more, it provides additional security to the information. The aim of our project is to acknowledge the handwriting and value the short answers and supply the result.

2. LITERATURE REVIEW

A.R. Lahitani, A. E. Permanasari and N. A. Setiawan[1], Automated Essay Scoring (AES) is one of the development systems for determining a score automatically from text document source to facilitate the correction and scoring by utilizing

applications that run on the computer. In this research, they implemented the weighting of Term Frequency - Inverse Document Frequency (TF-IDF) method and Cosine Similarity with the measuring degree concept of similarity terms in a document. This process results is in a ranking of the document weight that have closeness match level with expert's document.

C. Kaundilya, D. Chawla and Y. Chopra[2], Text extraction can be achieved by applying text detection that identifies image parts containing text, text localization finds the exact position of the text, text segmentation separates the text from its background and binarization process converts the colored images into binary. On this binary image, character recognition is applied to convert it into ASCII text. Text extraction is used in creating e-books from scanned books, image searching from a collection of visual data

D. Vaithiyanathan and M. Muniraj [3], The captured image is analyzed using Google Cloud Vision API Optical Character recognition (OCR). In order to extract text, we use image pre-processing methods to remove any noise or blur in the captured image so that the accuracy can be increased. Further, we include software-based text to speech to convert the text to speech as voice output. In this project we use Raspberry pi interfaced with Night vision camera and speaker to produce the converted text as voice. By applying Google Cloud vision OCR with pre-processing results in high accuracy level of the extracted text through image enhancement methods.

J. O. Contreras, S. Hilles and Z. B. Abubakar [4], The study aims to apply the preliminary approach forautomatically generating the domain concept ontology in essays using OntoGen and applied natural language processing algorithms using NLTK (NaturalLanguage Tool Kit) that enhance the teachers essay grading.

3. PROBLEM STATEMENT

Reading through all this information is a very tedious task. Thus, there can be chances when teachers miss to read important point in the large paragraphs of explanation. However, simple evaluations so many cases takes a lot of time and does not guarantee a dependable outcome. In the same way its difficult for a common man to read the whole answer sheet and crosscheck the answers in the answer key and textbooks.

The current way of checking subjective paper is adverse. Evaluating the Subjective Answers is a critical task to perform. When human being evaluates anything, the quality of evaluation may vary along with the emotions of Person. In Machine Learning, all result is only based on the input data provided by the user.

4. PROPOSED SYSTEM

Answer key image Text Recognition System using Teserract OCR. Handwritten answer text extract from image using Google Vision API in Google Cloud Platform. In evaluation process changing the both answer key text and handwritten text into vector format and apply cosine similarity to measure the score of the answer. To improve the evaluation process using NLTK(Natural language Toolkit) used for removing stopwords in the answer text.

4.1 Tesseract OCR

The dual encryption is the process of pushing the original information to modify its structure for twice time using private and public keys. Dual decryption is the reverse of dual encryption where we can attach projective hash function to regulate the original data from the encrypted value. The regulation includes hiding, deleting and making system reserved data in allotted memory.

4.2 Google Cloud Vison API

The Google Cloud Vision API allows developers to easily integrate vision detection features within applications, including image labeling, face and landmark detection, optical character recognition (OCR), and tagging of explicit content using the Vision API with Python.

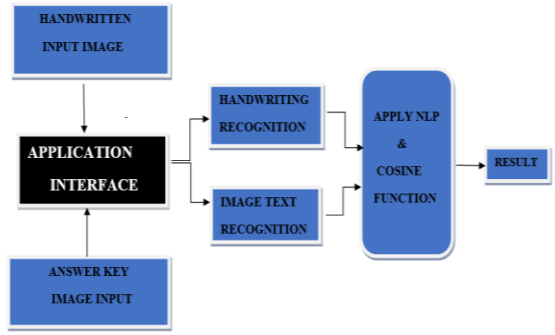
4.3 Natural Language Processing

Natural Language Processing, or NLP for short, is broadly defined as the automatic manipulation of natural language, like speech and text, by software. In their text on applied natural language processing, the authors and contributors to the popular NLTK Python library for NLP describe the field broadly as using computers to work with natural language data.

4.4 Cosine Similarity

Cosine similarity is a metric used to measure how similar the documents are irrespective of their size. Mathematically, it measures the cosine of the angle between two vectors projected in a multi-dimensional space. ... The smaller the angle, higher the cosine similarity

5. ARCHITECTURE



The above architecture clearly explains our proposed system we build a application interface in order to get handwritten answer image and answer key image input further it process and extract the text and apply nltk and cosing function we get a score marks obtained by students.

6. METHODOLOGY

In the proposed system, we tried to bring a automated evaluation system using the graphical user interface and the modules are following below.

6.1 User Interface

In Information Technology, the User Interface is everything designed into an information with which human being can interact to the computing systems. The proposed project implements the high end user interface with which user can reliably interact with systems without any uncomforted. The design of the project is developed using tkineter user interface which is lighter in loading dynamic application pages.

6.2 Answer key extraction

Pytesseract or Python-tesseract is an Optical Character Recognition (OCR) tool for python. It will read and recognize the text in images, license plates, etc. Here, we will use the tesseract package to read the text from the given image.

6.3 Handwritten Answer extraction

In handwritten answer extraction using traditional OCR does not give the better result. So in our proposed system we use the google cloud Vision API service provided by Google Cloud Platform. Vision API is a large pretrained
Machine learning module in order to get a handwritten text from student answer sheet.

We need a service account and private key as a JSON file in order to access the Google cloud platform.

Provide authentication credentials to your application code by setting environment variable GOOGLE_APPLICATION_CREDENTIAL.
Replace [PATH] with the file path of the JSON file that contains your service account key.
The Vision API can detect and extract text from images:
DOCUMENT_TEXT_DETECTION extracts text from an image (or file); the response is optimized for dense text and documents. The JSON includes page, block, paragraph, word, and break information.

6.4 Evaluation Module

Implementation of this Module using cosine similarity function. It is useful in classifying data on the number of objects that have a certain similarity, as research clustering based on cosine similarity measure.
Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them.
Similarity = (A.B) / (||A||. ||B||) where A and B are vectors.
Cosine similarity and nltk toolkit module are used in this program. To execute this program nltk must be installed in your system. In order to install nltk module follow the steps below –
1. Open terminal.
2. pip install nltk
3. python3
4. import nltk
5. nltk.download('all')
Functions used:
nltk.tokenize: It is used for tokenization. Tokenization is the process by which big quantity of text is divided into smaller parts called tokens. word_tokenize(X) split the given sentence X into words and return list.
nltk.corpus: In this program, it is used to get a list of stopwords. A stop word is a commonly used word (such as "the", "a", "an", "in").
Then, use cosine_similarity() to get the final output. It can take the document term matrix as a pandas data frame as well as a sparse matrix as inputs.

7. CONCLUSION AND FUTURE SCOPE

This paper presents an approach for the design and implementation of automated handwritten answersheet evaluation system. The proposed system works based on TesseractOCR and vision API services and cosine similarity function and NLTK took kit. We get an similar result as compared to human evaluator.

The future scope in this project is that we can improve the evaluation process in semantic analysis using genism library and soft cosine similarity metrics.

ACKNOWLEDGMENTS

The authors would like to thank for the faculty members of Department of Information Technology, Jerusalem College of Engineering. The work is made possible by the Head of Department, Department of Information Technology and Jerusalem College of Engineering.

REFERENCES

[1] A. R.Lahitani, A. E. Permanasari and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," 2016 4th International Conference on Cyber and IT Service Management, Bandung, Indonesia, 2016, pp. 1-6, doi: 10.1109/CITSM.2016.7577578.

[2] C. Kaundilya, D. Chawla and Y. Chopra, "Automated Text Extraction from Images using OCR System," 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2019, pp. 145-150.

[3] D. Mulfari, A. Celesti, M. Fazio, M. Villari and A. Puliafito, "Using Google Cloud Vision in assistive technology scenarios," 2016 IEEE Symposium on Computers and Communication (ISCC), Messina, Italy, 2016, pp. 214-219, doi: 10.1109/ISCC.2016.7543742

[4] D. Vaithiyanathan and M. Muniraj, "Cloud based Text extraction using Google Cloud Vison for Visually Impaired applications," 2019 11th International Conference on Advanced Computing (ICoAC), Chennai, India, 2019, pp. 90-96, doi: 10.1109/ICoAC48765.2019.246822.

[5] J. O. Contreras, S. Hilles and Z. B. Abubakar, "Automated Essay Scoring with Ontology based on Text Mining and NLTK tools," 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), Shah Alam, Malaysia, 2018, pp. 1-6, doi: 10.1109/ICSCEE.2018.8538399.

[6] K. P. N. V Satya and J. V. R. Murthy, "Cluster based cosine similarity measure", no. 3, pp. 508-512, 2012.

[7] S. Srihari, J. Collins, R. Srihari, H. Srinivasan, S. Shetty,and J. Brutt-Griffler, “Automatic scoring of short handwrittenessays in reading comprehension tests,” 2008.

[8] V. Rowtula, S. R. Oota and J. C.V, "Towards Automated Evaluation of Handwritten Assessments," 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia, 2019, pp. 426-433, doi: 10.1109/ICDAR.2019.00075.