AUTOMATIC ANSWER EVALUATION USING MACHINE LEARNING

S.NAHIDA Associate Professor, Department of CSE, NRI INSTITUTE OF TECHNOLOGY, Vijayawada, A.P, India.

MANDADA SAMEMI[2], TIRUMALA SAI HAREESHA[3], GUDLURU VENKATA SIVA SAI PAVAN KUMAR[4], NALLURI PRAMOD[5] Student, Department of CSE, NRI INSTITUTE OF TECHNOLOGY, Vijayawada, A.P, India.

ABSTRACT

Nowadays, as we are moving towards automation there is a need for an automatic descriptive answer evaluation system. Manual evaluation is a time and energy-consuming task. Presently, we have automated systems for objective type, single sentence answers and descriptive answers with less accuracy level.Automaticevaluation of answer scripts has been found very useful from our experiments, and often the assigned marks is the same as manually scored marks. In this paper, our goal is to develop an automated answer evaluation system using machine learning. The system will evaluate the answer based upon the number of words and number of letters from extracted text from the pre-processed data. After that, we have to implement the Natural Language Processing(NLP) for cleaning the extracted text. The text summarization is a process of creating a short, accurate summary of the longer text. Then, we have to implement a deep learning algorithm such as Artificial Neural Network (ANN). The experimental results will be shown like accuracy and marks based on a number of words and number of letters from 0 to 10.

INTRODUCTION

Optical character Recognition, conversion of images of text into a character. We can segment the each characters in one word. Using classification algorithm, to recognize

the text .The process of Recognition a text using classification algorithm. Result is compared with the text that should appear in that specific part of the image.Optical character recognition or optical character reader (OCR) is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo) or from subtitle text superimposed on an image (for example: from a television broadcast). Widely used as a form of data entry from printed paper data records whether passport documents, invoices, bank Statements, computerized receipts, business cards, mail, printouts of static data, computerized receipts, business cards, mail, printouts of static-data, or any suitable documentation- it is a common method of digitizing printed texts so that they can be electronically edited, searched, stored more compactly, displayed on-line, and used in machine processes such as cognitive computing, machine translation, (extracted)text-to-speech

key data and text mining. OCR is a field of research in pattern recognition, artificial intelligence and vision. Early versions needed to be trained with images of each character, and worked on one font at a time. Advanced systems capable of producing a high degree of recognition accuracy for most fonts are now common, and with support for a variety of digital image file format inputs.Some systems are capable of reproducing formatted output that closely approximates the original page including images, columns, and other non-textual components.Optical character recognition (OCR) technology is a business solution for automating data extraction from printed or written text from a scanned document or image file and then converting the text into a machine-readable form to be used for data processing like editing or searching. Optical character Recognition (OCR) is a conversion of scanned or printed text images,

handwritten text into editable text for further processing. This technology allows machine to recognize the text automatically. It is like combination of eye and mind of human body. An eye can view the text from the images but actually the brain processes as well as interprets that extracted text read by eye. In development of computerized OCR system, few problems can occur. First: there is very little visible difference between some letters and digits for computers to understand. For example it might be difficult for the computer to differentiate between digit "0" and letter "o". Second: It might be very difficult to extract text, which is embedded in very dark background or printed on other words or graphics.

EXISTING SYSTEM:

In existing system, nowadays, as we are moving towards automation there is a need for an automatic descriptive answer evaluation system. Manual evaluation is a time and energy-consuming task. Presently, we have automated systems for objective type, single sentence answers and descriptive answers with less accuracy level. In this paper, our goal is to develop an automated answer evaluation system using machine learning. The system will evaluate the answer based upon the matched keywords and the minimum length of the answer provided by the moderator. The scanned handwritten answer sheet will be given as input to the proposed system. The system will use an artificial neural network with back-propagation algorithm.

PROPOSED SYSTEM:

In proposed system, the imagedataset was taken as input. Then, we have to implement the pre-processingstep. In this step, we have to resize the original image and convert the image into gray scale.After that, we have to extract the text from the pre-processed image by using mean standard deviation. After that, we have to extract the text from input image by using the pytesseract and the extracted text will be stored in text format. Then, we have to implement the Natural Language Processing for cleaning the extracted text. Then, we can calculate the number

of words and number of letters from extracted text. After that, we have to implement the deep learning algorithm such as Artificial Neural Network (ANN). The experimental results shows that some performance metrics such as accuracy andevaluate marks based on number of words and number of letters.

IMPLEMENTATION

MODULES :

INPUT:

The dataset contains the images in the form of '.jpg' or '.png'

In this step, we have to read or load the input image by using the imread () function.

The input image is used to detect or recognize the text.

In our process, we are used the tkinter file dialogue box for selecting the input image.

PREPROCESSING:

In our process, we have to resize the image and convert the image into gray scale.

To resize an image, you call the resize () method on it, passing in a two-integer tuple argument representing the width and height of the resized image.

The function doesn't modify the used image; it instead returns another Image with the new dimensions.

Convert an Image to Grayscale in Python Using the Conversion Formula and the matplotlib Library.

We can also convert an image to grayscale using the standard RGB to grayscale conversion formula that is imgGray = 0.2989 * R + 0.5870 * G + 0.1140 * B.

FEATURE EXTRACTION:

Variance and Standard Deviation are essentially a measure of the spread of the data in the data set. Variance is the average of the squared differences from the mean.

mean () function can be used to calculate mean/average of a given list of numbers. It returns mean of the data set passed as parameters. Arithmetic mean is

the sum of data divided by the number of data-points.

TEXT RECOGNITION:

In our process, we have to recognize the text by using the pytesseract.

Python-tesseract is an optical character recognition (OCR) tool for python. That is, it will recognize and "read" the text embedded in images.

Python-tesseract is a wrapper for Google's Tesseract-OCR Engine.

NLP TECHNIQUES:

NLP is a field in machine learning with the ability of a computer to understand, analyse, manipulate, and potentially generate human language.

Cleaning (or pre-processing) the data typically consists of a number of steps:

Removepunctuation:Punctuation can provide grammatical context to a sentence which supports our understanding.

Tokenization: Tokenizing separates text into units such as sentences or words. It gives structure to previously unstructured text. eg: Plata o Plomo-> 'Plata','o','Plomo'.

Stemming: Stemming helps reduce a word to its stem form**.**

*Padding***:** In any raw text data, naturally there will be sentences of different lengths. However, all neural networks require to have inputs with the same size. For this purpose, padding is done.

DATA SPLITTING:

During the machine learning process, data are needed so that learning can take place.

In addition to the data required for training, test data are needed to evaluate the performance of the algorithm in order to see how well it works.

In our process, we considered 70% of the input dataset to be the

training data and the remaining 30% to be the testing data.

Data splitting is the act of partitioning available data into two portions, usually for cross-validation purposes.

One Portion of the data is used to develop a predictive model and the other to evaluate the model's performance.

Separating data into training and testing sets is an important part of evaluating data mining models.

Typically, when you separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing.

CLASSIFICATION:

In our process, we have to implement the deep learning algorithmsuch as Artificial Neural Network (ANN).

Artificial Neural Networks (ANN) arepart of supervised machine learning where we will be having input as well as corresponding output present in our dataset.

Our whole aim is to figure out a way of mapping this input to the respective output.

ANN can be used for solving both regression and classification problems.

EVALUATE ANSWER:

In this step, we can evaluate the answers from the extracted text based on length of answers and length of words.

We can put the marks from o to 10 based on length of answers and length of words.

PERFORMANCE METRICS:

The Final Result will get generated based on the overall classification and prediction. The performance of this proposed approach is evaluated using some measures like,

Accuracy

Accuracy of classifier refers to the ability of classifier. It predicts the class label correctly and the accuracy of the

predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

AC= (TP+TN)/ (TP+TN+FP+FN)

CONCLUSION

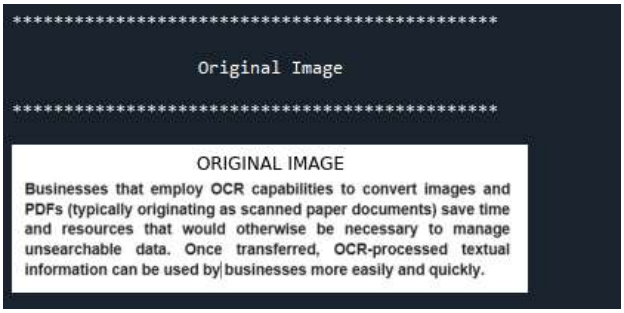
We conclude that, the image dataset was taken from dataset repository. We are implemented the NLP techniques and classification algorithms (i.e.) deep learning algorithm. Then,deep learning algorithms such as artificial neural network (ANN). Finally, the result shows that the accuracy for above mentioned algorithm and evaluate the answer and put marks from 0 to 10 based on length of letters and length of words.

FUTURE ENHANCEMENT:

In the future, we should like to hybrid the two different machine learning. In future, it is possible to provide extensions or modifications to the proposed classification algorithms to achieve further increased performance. Apart from the experimented combination of data mining techniques machine algorithms can be used to improve the detection accuracy. Finally, the sentiment analysis detection system can be extended as a prevention system to enhance the performance of the system.

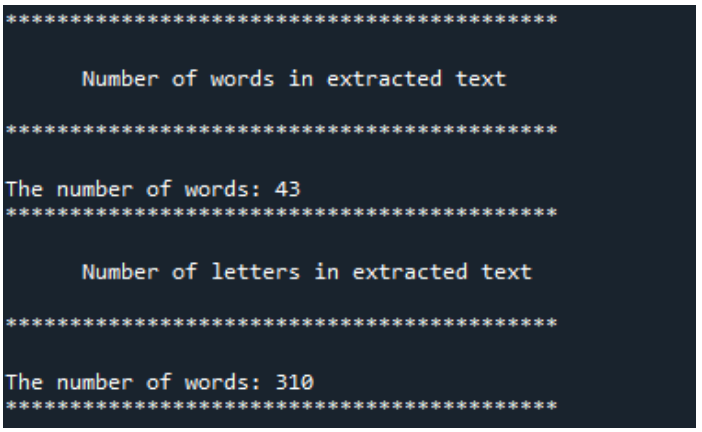
SAMPLE SCREENSHOTS

1. Data selection:

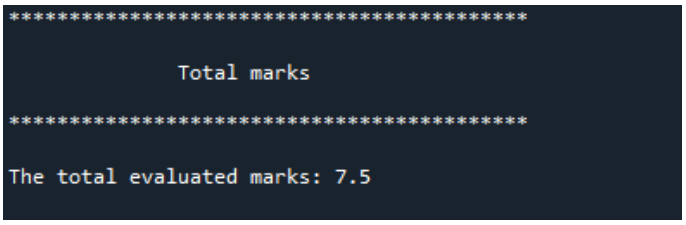


2. Preprocessing:

5. Calculate length:



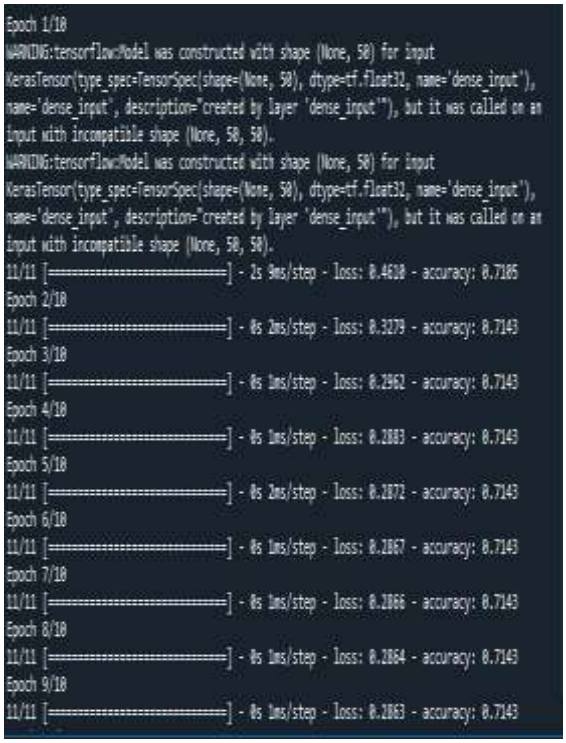
6. Evaluate Marks:



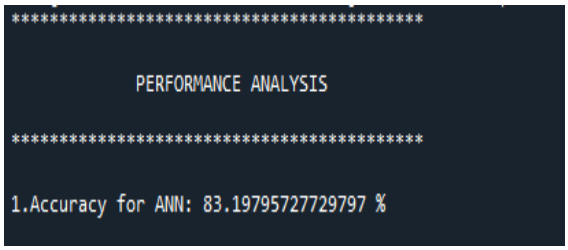
7. ANN:

3. Feature extraction:

4. Text extraction:



8. Performance metrics:



REFERENCES

[1] Prince Sinha, SharadBharadia and AyushKaul, “Answer Evaluation Using Machine Learning”, March 2018. Available: https://www.researchgate.net/ publication/333856264_Answer_Evaluation_Using_ Machine_Learning [Accessed: September 19, 2020]

[2] Honey Mehta, Sanjay Singla and AartiMahajan, "Optical Character Recognition (OCR) System for Roman Script & English Language using Artificial Neural Network (ANN) Classifier", in International Conference on Research Advances in Integrated Navigation System (RAINS - 2016), April 06 - 07, 2016.

[3]B.Vani and M. ShyniBeaulah, "High accuracy Optical Character Recognition algorithms using learning array of ANN", in International Conference on Circuit, Power and Computing Technologies [ICCPCT] 2014, pp. 1474 - 1479.

[4] SameekshaBarve, "Optical Character Recognition Using Artificial Neural Network", in International Journal of Advanced Research in Computer Engineering \& Technology, Volume 1, Issue 4, June 2012, pp. 131-133.

[5] J. Pradeep, E. Srinivasan and S. Himavathi, "Neural Network-based Handwritten Character Recognition system without feature extraction", in International Conference on Computer,

Communication and Electrical Technology (ICCCET) 2011, 18th \& 19th March, 2011, pp. 40-44.

[6] SkAsifAkram, MousumiSaha and TamasreeBiswas, "Evaluation of descriptive answer sheet using machine learning", in International Journal of Engineering Sciences \& Research Technology (IJESRT), April 2019, pp. 184-186.

[7] V. Lakshmi and Dr V. Ramesh, "Evaluating students descriptive answer using natural language processing and artificial neural networks", in International Journal of Creative Research Thoughts (IJCRT), Volume 5, Issue 4, December 2017, pp. 3168-3173.

[8] S. Impedovo, L. Ottaviano and S.Occhinegro, “Optical character recognition”, International Journal Pattern Recognition and Artificial Intelligence, Vol. 5(1-2), pp. 1- 24, 1991