International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-2, July 2020

One-Word Answer Correction using Deep Learning Models and OCR

K. P. K Devan, Sruthi Prabakaran P, Tamizhazhagan S, Vaishnavi S

Abstract: Examinations/Assessments are a way to assess the understanding of a student on a particular subject. Even today many educational organizations prefer to conduct exams by offline mode (pen and paper). And evaluating them is a time-consuming process. There is no effectual model to evaluate Offline descriptive answers automatically. The traditional method involves staff assessing the content manually. In place of this process, a new approach using image captioning by using deep learning algorithms can be implemented. Handwritten Text Recognition (HTR) can be used to evaluate descriptive answers. One-word Answers captured as images are pre-processed to extract the text features using deep learning models and pytesseract. This paper presents a comparison between the CNN-RNN hybrid model and Optical Character Recognition (OCR) to predict a score for one-word answers.

Keywords: Convolutional Neural Network (CNN), Handwritten Text Recognition (HTR), Optical Character Recognition (OCR), Recurrent Neural Network (RNN).

I. INTRODUCTION

Throughout ages, examinations were conducted in order to know whether a student has grasped the concepts clearly, though it is believed that marks do not define a person's skill and capability. And, almost every school or university prefers the offline mode of examination. Though it has its own advantages, the Offline mode of examinations also has its own disadvantages. Firstly, the papers need to be well preserved and secondly, it requires manual correction. With improvements in science and technology, many scholars have come up with the notion of automatic evaluation. But it is currently limited to the online mode of examination and online correction. Both Handwritten Text Recognition and Optical Character Recognition (OCR) can be used to implement automatic evaluation of offline examinations. Handwritten text recognition (HTR) aims to recognize handwritten characters from papers. HTR can be either online or offline. In Online HTR, a stylus and a writing pad are used to record the contents, and then it is digitalized.

Manuscript received on May 25, 2020.
Revised Manuscript received on June 29, 2020.
Manuscript published on July 30, 2020.
* Correspondence Author
K. P. K. Devan, Associate Professor, CSE Department, Easwari Engineering College, Chennai, Tamil Nadu.
Sruthi Prabakaran P, CSE Department, Easwari Engineering College, Chennai, Tamil Nadu.
Tamizhazhagan S, CSE Department, Easwari Engineering College, Chennai, Tamil Nadu.
Vaishnavi S, CSE Department, Easwari Engineering College, Chennai, Tamil Nadu.

In offline mode, the contents written in a paper are read and converted into digital characters. HTR uses deep learning models to decode the characters present in documents and papers. Optical Character Recognition (OCR) is software that can convert texts and images into the digitalized format. This, in turn, can be manipulated by the machine.

OCR was previously used to read characters from images (printed characters). Slowly, people started to understand the need for an application to digitalize all the physical documents as it cost heavily to preserve them for quite long periods. A lot of resources are required as the fonts used in documents are very different from one another and from fonts that machines can interpret. This paper is divided into sections: sections 2 and 3 provides the motivation for this work and the literature survey, section 4 describes the working principle of CNN-RNN modules and the tesseract module, section 5 describes the output of the two methods and section 6 concludes the paper.

II. MOTIVATION

Even in this digitalization era, we see documents and data being handwritten. Documents are a way to acquire the required information from people for references. But not all data can be maintained in physical documents and often need to be digitalized for preserving. These documents are hard to collect, transport, and preserve. So, many organizations employ people to store data in systems manually, which is both a time and resource-consuming process. Hence a requirement for automatic Handwritten Text Recognition system arises in order to recognize texts in scanned images. So the process of extraction of text from documents(images) is made easier. HTR finds applications in many sectors like banks, postal offices, railways, IT companies, and so on. With an enormous amount of data being generated every day, it would be impossible to store data manually.

III. LITERATURE SURVEY

A number of models/algorithms are available to evaluate online examinations automatically. Recently, Alla Defallah Alrehily [8] used Spearman's correlation to set questions automatically and find the similarity between user answer and the instructor's answer. An accuracy of 89% was obtained compared to 96% obtained in the manual correction. Various correlation algorithms have been widely implemented over the years while experimenting with the automatic evaluation system. Maram F. Al-Jouiea [3] has used Pearson correlation to evaluate essays of school children automatically using Latent semantic Analysis (LSA) and Rhetorical Structure Theory (RST). Hybrid LSA-RST along with Hyperspace Analog to Language (HAL) and Self-Organising Maps (SOM) models were used by many researchers since it gave good accuracy, combined with Pre-processing methods.

Meena. K and Lawrance. R [10] experimented with hybrid LSA-SOM and HAL-SOM models and obtained accuracy >90%. All these models were successful for Online Evaluation. In order to implement Automatic evaluation of offline answers, we need to extract the features before it can be further processed. Deep Learning methods like CNN and RNN have been used to extract the features as it gives much higher accuracy compared to other Machine-Learning methods. Training of models is done prevalently on datasets like IAM, MNIST, NIST, EMNIST, etc. For instance, Rohan Vaidya [11] implemented handwritten character recognition using CNN on the NIST dataset. An accuracy of 94% was obtained in the training phase. A combination of Image segmentation and Deep learning model was used to extract the features conveniently. But the accuracy of the result varies depending upon the datasets used to train the model. Raymond Ptucha [5] presented a comparative model on three different datasets and the results were all different. The result was analyzed based on Word Error Rate (WER) and Character Error Rate (CER). Talking about the usage of OCR, Noman Islam [7] presented a survey on the various systems available in OCR. The sub-processes involved in creating an OCR system is briefed clearly. Similarly, Pratik Madhukar Manwatkar [12] elaborated on the steps to implement OCR and applications of OCR in everyday life.

IV. WORKING

The implementation has been done by training the CNN-RNN model using the IAM dataset. The prediction has been done by collecting handwritten texts of three answers (random) from three students.

The IAM dataset was divided into 500 batches. 95% of the data is used for training and the remaining 5% is used for validation

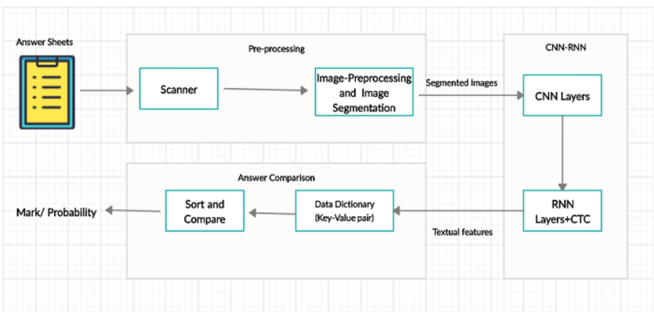


Fig. 1 Functional Architecture

Functional Software Architecture (FSA) in Fig. 1 provides a working view of the software product with no physical or structural features. It is derived from the operational or business model from which the software requirements were specified. The answer sheets are firstly scanned using a scanner and then stored in filed according to the student registration number followed by image pre-processing and image segmentation. Secondly, CNN-RNN and pytesseract modules are used in the process of feature extraction. Lastly, the original answers are compared with the answer decoded from images.

A. CNN-RNN Hybrid Model

(i). Input:

One-word answers written in sheets are scanned using a professional scanner or mobile camera and stored in folders, numbers according to the register/roll numbers. Each of these image files is given as input to the model.

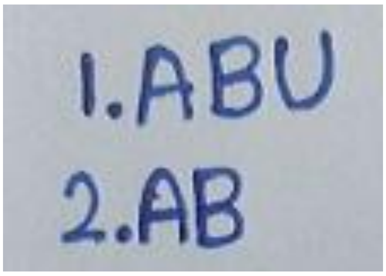


Fig. 2 Input Image

(ii). Pre-processing:

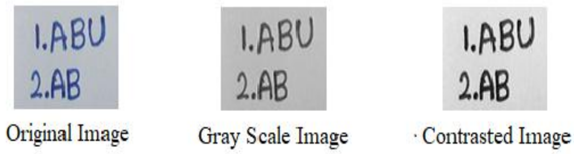


Fig. 3 Image Pre-processing

Before the actual processing, the images are pre-processed. Pre-processing is done so that any unwanted noise in the image (white spaces) is removed, which may reduce the accuracy of prediction. Images are resized to the desired height and are converted to grayscale. The contrast of the image is increased as the next step to improve readability and avoid missing relevant information while processing.

(iii). Image Segmentation:

The pre-processed image is further segmented to get each answer as separate images. Firstly, the image is converted to the required height. It is then processed by applying contour. Contouring is done in order to identify regions that contain useful data. For this process, a filter kernel of size 25 is used. Regions containing features less than the specified area are ignored and only the regions of higher dimensions are processed.

(a) (b) (c)

Fig. 4 Image Segmentation Process

(iv). Feature Extraction:

The segmented images of each answer sheet are stored in separate folders based on the roll numbers. The folders are read one by one and the images are decoded to get the textual features. A hybrid CNN-RNN model is used to extract the features. The CNN model is used for the purpose of image pre-processing. The RNN module along with CTC is used to decode the characters.

CNN Parameters:

- Input image: 101 x 32 pixels
- The CNN module has 5 layers: 5x5 filter kernel in the first three layers and a 3x3 filter kernel in the next two layers.
- Pooling: elu
- Learning rate: 0.02 and adjusts itself based on the number of batches trained
- Epochs: The model trains until an early stopping of 20 epochs. (100 epochs)

RNN Parameters:

- The RNN layer has 2 layers with popular Long Short Term Memory (LSTM)
- The output sequence is mapped to a matrix of size 32x80.

CTC Layer:

- The CTC is using to compute loss value and decode the text.

(v) Answer Comparison:

Each answer sheet (after decoding) is stored in a text file. The correct answers are stored in another text file. The answers are stored in a dictionary with the answer number being the key and value as the answer itself. Each answer is compared with the original answer and the mark is calculated based on the correctness. For experimentation, three handwritten answers for three random questions were collected.

B. Optical Character Recognition (OCR)

i. Input:

The input to OCR is the same as that of the CNN-RNN model. Images scanned using a scanner or captured using a mobile are stored in the folders.

ii. Model:

The following libraries have been used for the purpose of implementation: pytesseract, cv2, and Image from PIL. The images are read one by one from the folder and processed using the 'image_to_string' function. The purpose of this function is to extract the textual features from the image. Text extracted from the images is stored as text in the corresponding answer variables and stored in a text file. This file is compared with the text file that contains the correct answers. The percentage of each student is returned as the output.

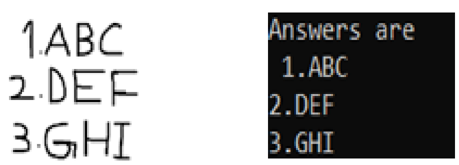


Figure 5. Image to Text

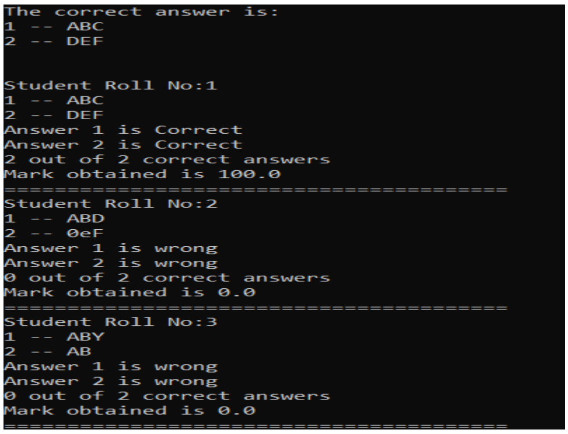
V. OUTPUT

The output for the Deep learning model and pytesseract model is as listed below.

A. CNN-RNN

The Model was trained in CPU (intel core i5- 8th gen). It took two days for training and a character error rate of 10.08% and the word accuracy of 74.81%.

Out of 6 decoded answers, four were decoded correctly. Figure 6 displays the final output of the model. Answers 1 and 2 of student roll number 2 and 3 were decoded incorrectly. Firstly, the correct answer is displayed followed by actual decoding. The Answer of students in the order of their roll numbers is decoded and whether or not the answer is correct is displayed following which their score and mark are displayed.



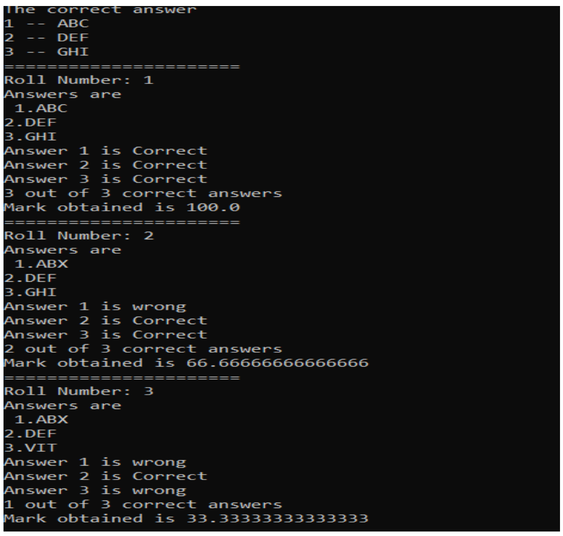
```
The correct answer is:
1 -- ABC
2 -- DEF

Student Roll No:1
1 -- ABC
2 -- DEF
Answer 1 is Correct
Answer 2 is Correct
2 out of 2 correct answers
Mark obtained is 100.0
========================================
Student Roll No:2
1 -- ABD
2 -- 0eF
Answer 1 is wrong
Answer 2 is wrong
0 out of 2 correct answers
Mark obtained is 0.0
========================================
Student Roll No:3
1 -- ABY
2 -- AB
Answer 1 is wrong
Answer 2 is wrong
0 out of 2 correct answers
Mark obtained is 0.0
========================================
```

Figure 6. CNN-RNN Output.

B. Pytesseract

The model is able to decode accurately the answers and successfully display the results. But, this model works efficiently only when the answers are written in uppercase and in a clear white sheet.



```
The correct answer
1 -- ABC
2 -- DEF
3 -- GHI
====================
Roll Number: 1
Answers are
 1.ABC
2.DEF
3.GHI
Answer 1 is Correct
Answer 2 is Correct
Answer 3 is Correct
3 out of 3 correct answers
Mark obtained is 100.0
====================
Roll Number: 2
Answers are
 1.ABX
2.DEF
3.GHI
Answer 1 is wrong
Answer 2 is Correct
Answer 3 is Correct
2 out of 3 correct answers
Mark obtained is 66.66666666666666
====================
Roll Number: 3
Answers are
 1.ABX
2.DEF
3.VIT
Answer 1 is wrong
Answer 2 is Correct
Answer 3 is wrong
1 out of 3 correct answers
Mark obtained is 33.33333333333333
```

Figure 7. Pytesseract output

VI. CONCLUSION

Almost all available models are used to evaluate Online Answers. Our proposed method combines Offline HTR and evaluation methods to evaluate handwritten one-word answers. In this paper, a comparative study is done using Deep learning models and OCR to implement the first part of feature extraction. The CNN-RNN hybrid model trained for accuracy of 74.81% with the fixed parameters and a character error rate of 10.08%. Two out of the six answers evaluated were decoded incorrectly. The pytesseract decoded the answers correctly without any error. But, the deep learning methods obtained good results also for cursive handwriting besides uppercase words. Whereas, the OCR method obtained only satisfactory results for cursive handwriting. Overall, the model can be improved by training the deep learning model with a different dataset other than IAM, adding more hidden layers, and adjusting the learning rate (alpha) of the model. In the future, the method can be extended to evaluate descriptive answers by combining methods like SOM and HAL. A separate application can be put together to implement the model with improved UI that is more user friendly. Also, a separate answer sheet similar to the OMR sheet can be used to have QR codes to replace the difficulties in storing answer sheets based on the student register number

REFERENCES

1. Piyush Patil, Sachin Patil, Vaibhav Miniyar, Amol Bandal, "Subjective Answer Evaluation Using Machine Learning", International Journal of Pure and Applied Mathematics, Volume 118 No. 24, 2018.
2. V. Nandini, P. Uma Maheswari, "Automatic assessment of descriptive answers in the online examination system using semantic relational features", The Journal of Supercomputing, Springer, 2018.
3. Maram F. Al-Jouiea, Aqil M. Azmia, "Automated Evaluation of School Children Essays in Arabic, 3rd International Conference on Arabic Computational Linguistics", ACLing, Elsevier, 2017, pp:19-22.
4. Haiqing Ren, Weiqiang Wang, Chenglin Liu, "Recognizing online handwritten Chinese characters using RNNs with new computing architectures", Pattern Recognition, Elsevier, 2019, pp:179-192.
5. Raymond Ptucha, Felipe Petroski Such, Suhas Pillai, Frank Brockler, Vatsala Singh, Paul Hutkowski, "Intelligent character recognition using fully convolutional neural networks", Pattern Recognition, Elsevier, 2019, pp:604-613.
6. J.Ignacio Toledoa, Manuel Carbonell, Alicia Fornés, Josep Lladós, "Information extraction from historical handwritten document images with a context-aware neural model", Pattern Recognition, Elsevier, 2019, pp:27-36.
7. Noman Islam, Zeeshan Islam, Nazia Noor, "A Survey on Optical Character Recognition System", Journal of Information & Communication Technology-JICT, Vol. 10 Issue. 2, 2016.
8. Alla Defallah Alrehily, Muazzam Ahmed Siddiqui, Seyed M Buhari, "Intelligent Electronic Assessment for Subjective Exams", ACSIT, ICITE, SIPM, 2018, pp: 47-63.
9. Meena.K, Lawrance Raj," Evaluation of the Descriptive type answers using Hyperspace Analog to Language and Self-organizing Map", IEEE International Conference on Computational Intelligence and Computing Research, IEEE, 2014.
10. Meena.K, Lawrance.R, "Semantic Similarity Based Assessment of Descriptive Type Answers", IEEE, 2016.
11. Rohan Vaidya, Darshan Trivedi, Sagar Satra, Prof. Mrunalini Pimpale, "Handwritten Character Recognition Using Deep-Learning", Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies, IEEE, 2018
12. Pratik Madhukar Manwatkar, Dr. Kavita R. Singh, "A Technical Review on Text Recognition from Images", IEEE 9th International Conference on Intelligent Systems and Control (ISCO), IEEE, 2015.

AUTHORS PROFILE

K. P. K. Devan, Associate professor, Computer science and Engineering Department, Easwari Engineering College, Chennai, Tamil Nadu. He completed his UG in BE CSE from Madras University and PG in M.Tech. from SASTRA University. He has published around 20 papers in National and International journals/conferences. He is an active member of ISTE and IETE. He has a teaching experience of 20 years and his area of interest is Natural Language Processing and Big Data Analytics.

Sruthi Prabhakaran P, final year student, computer science and engineering department, Easwari Engineering College, Chennai, Tamil Nadu. Currently pursuing UG in B.E CSE. She has completed a course on machine learning with real-time projects and fundamentals of Python Programming. She has attended various workshops and In-plant training from reputed organizations like BSNL and AICTE SmartBridge.

Tamizhazhagan S, final year student of Computer Science and Engineering Department, Easwari Engineering College, Chennai, Tamil Nadu. Currently pursuing UG in B.E CSE. He has undergone an internship on DOTNET. Attended workshop on Android application development and World's biggest International Hands-On Gaming and security workshop. He has completed a course on Python and Machine Learning.

Vaishnavi S, final year student, Computer Science, and Engineering Department, Easwari Engineering College, Chennai, Tamil Nadu. Currently pursuing under graduation in, B.E CSE. She has attended several workshops on android development, machine learning, and IoT. She has completed a course on Android application development, cloud computing, data analysis, and Machine Learning. She has learned to use Tableau and R programming for data analysis. area of expertise is Python, Java and C program.