International Journal of Scientific Research in Engineering and Management (IJSREM
Volume: 03 Issue: 06 | June -2019 ISSN: 2590-1892

AUTOMATIC EVALUATION OF ANSWERSHEET AND MARKING SYSTEM

Shubhanshu Pandey[1], Shivendra Singh[2], Tanveer Chougule, Shrinivas Chavan[4]
Prof. Chaitanya Bhosale[5]

Department of Computer Engineering, Dr.D.Y. Patil School of Engineering, India

ABSTRACT:

Only objective type questions are assessed and according to that marks are given to the student, in traditional Online Examination System. These technique lacks the capability of evaluating descriptive answers. To avoid such enormous amount of man hours for evaluation of examinations we are implementing this system. To develop off-line short answer automated assessment system which is capable of evaluating the descriptive answers using OCR, data mining and machine learning

Keywords: Text Mining, Automatic assessment, Machine learning, Examination System, Pattern Matching Teacher, Staff, Students, OCR.

1. INTRODUCTION:

Evaluations work is not in proper manner and time consuming. To avoid this problem automated evolution environment is developed. The main aim is to describe an automated answer marking system that can be utilized and improve quality of teaching and learning of the subjects Question paper is prepared by the teacher staff in the desired format and submitted to the appeared Candidates. Candidates give the exam and submit the hand-written answer-sheet to the teacher staff. The teacher staff collects the answer-sheet from student and submitted to the system with the help of scanning. In case of a system failure, the work is saved backup storage. The problem like Noise removing, Binarization, Rescaling, Edge Detection is solved by using OCR.

. With the help of OCR, character get extracted which defines each character by the presence or absence of key features,. Pattern recognition plays important role in handwritten character recognition. With the help of pattern recognition evaluation of marks is done. This information is saved in the hardware like computer or laptop because of large storage device.

Problem definition:

To develop offline short answer automated assistant system which capable of evaluating the descriptive answers using OCR, data mining and machine learning

Project scope: The view of project is to improve the existing O_-line Examination Evaluation by developing automated system that is capable of evaluating the descriptive answer Hardware cost includes the cost of system which is to provide to the examiner. System includes OCR, data mining and machine learning.

2. General Recognition scheme:

This work was partially inspired by previous work on recognition without explicit segmentation into characters [1, 2, 3]. However, it aims at developing a general sentence recognition system, possibly not confined to hand printed writing. The sentence recognition scheme we propose is based on a bottom-up strategy:

Information is first extracted at low (image) level, in the form of character Hypotheses, then lexical knowledge is applied to obtain candidate word sequences, to which grammatical knowledge can be finally applied. Character hypotheses are generated by a process which avoids traditional image segmentation, by scanning the word image with a sliding window where feature extraction and neural-based classification of the computed feature vector are performed at each location. We can call this approach character spotting. Characters are simultaneously located and recognized within the word. This approach was successfully applied to Automatic Speech Recognition [4], and can be adapted to cursive handwriting as well, provided a robust set of primitives (graphemes, sub-graphemes and grapheme groups) is defined. Character recognition is of course error prone, due to the uncertain location of units; lexical information intervenes to overcome this problem. Instead of character recognition, central to the system is word recognition, to which both character classification and lexical knowledge concur.

3. The Word Recognition method

3.1 Word Level Pre-Processing

In traditional recognition approaches, the basic unit to be detected and classified is the character. The written lines are segmented into characters by several different methods. Character-level preprocessing consists of deslanting and size normalization. In our approach, where there is no explicit segmentation into characters, word is the typical basic unit, and preprocessing comprises noise filtering and baseline straightening.

Noise filtering can consist of the following: point noise filtering, removal of fixed structures, contour smoothing, filling or reconstruction of small areas impaired by the image acquisition process.

Baseline straightening is performed by estimating the inferior line or word guideline (baseline); this can be modeled by one or more contiguous line segments.

The baseline is straightened by progressive vertical shifts of pixel columns.

The straightened word is then segmented into regions (“pseudo-characters”) either completely connected or composed of connected parts with overlapped projections. Each “pseudo-character” is then separately normalized, either to a standard size, or to a standard height while maintaining the original aspect ratio; the normalization type is chosen according to the latter. Finally the word image is rebuilt by joining the normalized “pseudo-characters”.

3.2 Character Level Measure

Character spotting is performed by feature extraction and MLP-based classification within a fixed size sliding window scanning horizontally the normalized word image. Scanning is carried out in two

alternate modes: a sacchadic mode, where the window "jumps" from pseudo-character to pseudo-character, and an exhaustive mode, where the window shifts continuously (i.e. with regular speed) within the current pseudo-character. At each window position a measure operation is performed, consisting in the following: a vector of features [8] for character representation is generated, and a 26-class MLP classifier [5, 6], trained on uppercase handprint, is fed with the feature vector; the MLP output is symbolic, i.e., when the class with the highest response is scored above a given threshold, the ASCII corresponding to that class is issued; otherwise, a "don't know" symbol ("_") is produced. When the pseudo-character width is identical to the window size, only one measure is taken; the output symbol is replicated a number of times proportional to the window size. The sequence of character hypotheses is then filtered to remove incongruities.

4. Literature survey:

In [1] deals with taking scanned copy of a document as an input and extract texts from the image into a text format using Otsu's algorithm for segmentation and Hough transform method for skew detection. The system was confined to recognize English alphabets (A-Z, a-z) and numerals (0-9). OCR technique has been implemented to recognize characters..

In [2], This paper presents detailed review in the field of OCR. Various techniques are determined that have been proposed to realize the center of character recognition in OCR system. OCR translates images of typewritten or handwritten characters into the electronically editable format and it preserves font properties. Different techniques for pre-processing and segmentation have been researched and discussed in this paper.

In [3], OCR becomes necessary first step for all applications that consider typewritten or handwritten manuscripts as input. We need to train our classifier in case if we are using data mining techniques for such purposes. There are several established generic classification techniques that can be used together with feature extraction mechanisms but it is important to know which of them do better under which circumstances. We evaluate three approaches for OCR from handwritten manuscripts and we also studied their results. We considered a case study where we need to identify cases with probability of dyslexia.

In [4], This paper aims to create an application interface for OCR using artificial neural network as a back end to achieve high accurate rate in recognition.The proposed algorithm using neural network concept provides a high accuracy rate in recognition of characters. The proposed approach is implemented and tested on isolated character database consisting of English characters, digits .

In [5], machine learning approach is used to solve this problem using text mining. Measuring the similarity between, words, sentences, paragraphs and documents is an important component in various tasks such as text summarization, information retrieval, automatic essay scoring, document clustering, and machine translation and word-sense disambiguation. JSON is used for transferring data between web application and server, serving as an alternative to XML in this system.

In [6], this paper represents an algorithm for implementation of OCR to translate images of handwritten or typewritten characters into electronically editable format by preserving font properties. By applying pattern matching algorithm, OCR can do this. The recognized characters are

stored in editable format. Thus OCR make the computer read the printed documents removing noise.

5. SYSTEM REQUIREMENT:

- **Hardware Requirements:**
  - Windows 64/32 bit system
  - Desktop Requirements
  - RAM - 8GB
  - Processor i3/i5
  - Hard Disk – 500GB
- **Software Requirements:**
  1. Operating system: XP/7
  2. Database MySQL/H2
  3. IDE Tool: Eclipse

6. CONCLUSION AND FUTURE SCOPE:

In this paper, we presented different kind of variations in recommended systems proposed in various reference papers we explained the system architecture of various papers. We investigated the different types of techniques which are used automatic evaluation of answer sheet and marking system.

The view of project is to improve the existing Off-line Examination Evaluation by developing automated system that is capable of evaluating the descriptive answers. For future work, we will try to this system can be adopted by all universities as well as level organizations and also overcome the drawbacks of current system.

7. ACKNOWLEDGEMENT:

We wish to express our thanks to all who helped us directly or indirectly in making this paper. We are thankful to our project guide **Prof. Chaitanya Bhosale** for her valuable guidance. We also wish to thank our **HOD. Prof.Soumitra Das** and Director **Dr. S.S. Sonawane** for their kind Support.

REFERENCES:

[1] Neha Agrwal,ArashdeepKaur, An algorithmic approach for text recognition from printed/Typed text images,2018

[2] Hiral Modi,M.C.Parikh
A review on optical character recognition technique,2017

[3] Syed Hasan tanvir, Tamim Ahmed Khan, Abu Bakar Yamin 2016, Evaluation of OCR algorithm and feature extraction technique

[4] Mrs B wani , *Ms. M. Shyni Beaulah, Mrs. R. Deepalakshmi,*2014,High accuracy optical character recognition algorithm using ANN

[5] Pooja kudi,Amitkumar Manekar, Kavita daware,Tejaswini Dhatrak ,2014 ,Online examination using short text matching algorithm

[6] Faisal Mohammad, Jyoti Anarase, Milan Shingote, Pratik Ghanwat,2014,Optical charater recognition using pattern matching

[7] HimSubrataadri Nath Saha, Supratim Auddy 2017, Data extraction from exam answer sheets using OCR with adaptive calibration of environmental Threshold parameters.

****