Offline handwriting recognition using neural networks

Nitin Sharma	*Pranshu Agarwal*	*Dr. Upasana Pandey*
nitin741996@gmail.com	*pranshuagarwal9997@gmail.com*	*coe.upasana@gmail.com*
IMS Engineering College, Ghaziabad, Uttar Pradesh	*IMS Engineering College, Ghaziabad, Uttar Pradesh*	*IMS Engineering College, Ghaziabad, Uttar Pradesh*

ABSTRACT

Handwriting recognition system can be defined as the ability of a computer to intelligently recognize the handwritten text by the user. The handwriting recognition is considered to be an open research topic due to the fact that every human being have a unique handwriting as well as there are numerous languages which are being used worldwide in written form. A neural network can be designed for recognizing handwritten texts in any language required.

As different people have different handwriting styles,the task of recognizing the handwriting is a difficult task. As of today, the OCR (Optical Character Recognition) Engines are primarily focused on machine printed texts or handwritten texts which are written in block letters. By recognizing the texts in the handwritten text images by user into letter codes is becomes usable digitally in computers and text processing applications.

Keywords*: Offline Handwriting Recognition (HWR), Neural Networks, Machine Learning, Optical Character Recognition (OCR), Pattern Recognition.*

1. INTRODUCTION

Handwriting Recognition systems can be defined as the ability of the computer to efficiently recognize the handwritten text by the user from the sources like touch screens, papers, photographs, etc. The main component of a handwriting recognition systems is optical character recognition (OCR). However, a complete handwriting recognition system also handles scanning, formatting and segmentation of the text into characters. Handwriting recognition systems is of two types: Offline Handwriting Recognition Systems and Online Handwriting Recognition Systems.

In online handwriting recognition systems the text is recognized simultaneously as the user writes it using the hardware pen connected with the system. There are three elements to an online handwriting recognition system:

- A pen or stylus to write with.
- A touch screen surface, which is connected with the computer.
- An application to interpret handwritten inputs.

In case of an Offline Handwriting Recognition System the text is not simultaneously recognized as the user writes it but the text which is already written of a paper, photograph by the user is recognized. The text document is first scanned and provided to the computer system for its interpretation and recognition of the text written.

Recognition of handwritten text offline is considered to be harder than online handwriting recognition systems due to the reason that in online recognition the features of the handwritten texts can be extracted from both the trajectory of the pen/stylus as well as the resulting image, while in case of offline recognition of text only image of the text is available.

Nonetheless, the process of recognition of handwritten text is almost similar in both offline and online recognition: extracting the features from the images, then the features are matched with the sequence of labels which the neural network learns during its training process.

Coming to the recognition of texts, we can understand this concept by taking an example of our own brain only. Consider the following image of handwritten text:

We can effortlessly recognize that "HELLO HOW ARE YOU ALL ?" is written. This is possible because we carry a supercomputer in our head as our brain consists of millions of neurons which are trained to recognize the handwritten texts.We will be using multilevel neural networks approach to make our computer understand handwritten notes. The idea is to take a large data set of handwritten English characters containing both lower case as well as uppercase letters as well as the special symbols and the numerical values. That data set will be used to train our neural network. After training is done the system will be able to recognize the handwritten notes efficiently. Furthermore, we can increase the accuracy of the system by training it on a efficient and larger database.

2. WORKING TECHNOLOGY

A. OCR- Optical Character Recognition

OCR is the mechanism of translation of images of handwritten, typed or printed texts by the means of scanner into a machine editable text. It is generally used for converting the paper books and documents into electronic files. When we scan any paper, the scanner only produces the image file or a photo of the page. This image is not readable for texts by computer that is the computer cannot understand the letters written or typed on the page. So, we cannot edit the texts of that image as we can do in any word processor. For that specific function we use OCR software to convert it into a word processor file or text to enable us to do editing on those printed or written texts. For our Offline Handwriting Recognition system we propose the use of Neural Network. We intended to use the nearest neighbour OCR algorithm
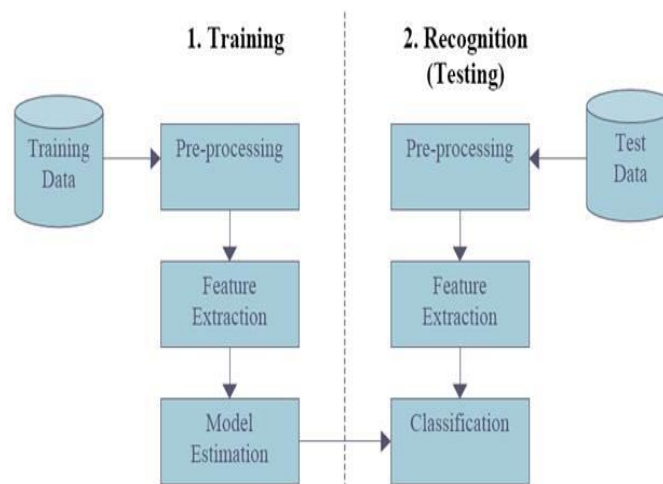
Error!

Fig: An OCR Process

B. Neural Networks

A neural network is a powerful data modeling tool that is able to capture and represent complex input/output relationships. The motivation for the development of neural network technology stemmed from the desire to develop an artificial system that could perform "intelligent" tasks similar to those performed by the human brain. Neural networks resemble the human brain in the following two ways: they acquire knowledge through learning, and the knowledge is stored within inter-neuron connection strengths known as synaptic weights. A neural network is a powerful data modeling tool that is able to capture and represent complex input/output relationships. The motivation for the development of neural network technology stemmed from the desire to develop an artificial system that could perform "intelligent" tasks similar to those performed by the human brain. Neural networks resemble the human brain in the following two ways:

- A neural network acquires knowledge through learning.
- A neural network's knowledge is stored within inter-neuron connection strengths known as synaptic weight.
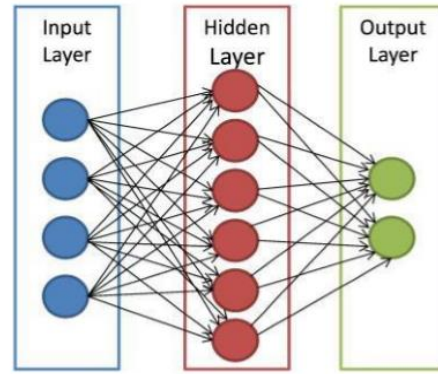
Fig: Neural network

3. BACKGROUND AND LITERATURE SURVEY

We have to survey from the other author research paper and they have to measure some related work of our research project. They are:

A) Rahul KALA, et.al: proposed work on Offline Handwriting Recognition method with used Genetic Algorithm. In this research describe to a piece of paper and then convert it into text. They are used genetic algorithm to implement Offline Handwriting Recognition.[1]

B) Brandon Maharrey COMP 6600 Artificial Intelligence Spring 2009 they survey about A Neural Network Implementation of Optical Character Recognition that measure that neural network is also use in OCR for the handwritten notes or words.[2]

C) Sang Sung Park, Won Gyo Jung, Young et.al: teams they are implemented Optical Character Recognition System Using BP Algorithm they told her They use OCR (OCR : Optical Character Recognition) technique which is that saving relevant documents to DB after extracting text   by using OCR. That is, text should be entered to DB after classifying segments one by one in realized whole document after doing character recognition through OCR. In this paper, in order to solve this problem, we constructed OCR system that saves abstracted characters to DB automatically after extracting only equivalent and necessary characters from a large amount of documents by using BP algorithm.[3]

D) Made Edwin Wira Putra, Iping Supriana Suwardi both has implement the Structural off-line handwriting character recognition using the purpose of those model is to give the ability in improving recognition accuracy without relying in normalization technique. They are use in graph technique. The graph consists of several edges that indicate the connected vertices. The vertices are joining and to form a curve that make the character. The curve is extracted by analyzing the character's chain code, and its string feature is created using some principle.

E) Krupa dholakia has to define about the handwritten character recognition technique are divided into some subparts such as preprocessing, segmentation, feature extraction, classification and post processing.[4]

4. PROPOSED METHODOLOGY

Data Collection

The data that we are using to train our program is a collection of English Alphabets (Both lower case and upper case), special symbols and numbers. Each image used in the data set are in 20x15 pixels.
The first step is training the system on the data set. Character recognition is done using OCR.
The highest accuracy in an OCR is achieved using SVM algorithm. So we will be training the algorithm using SVM.
SVM gives 97.2% accuracy when used for Optical Character Recognition when we provide a proper training set to the system.

Before going to the recognition part, we must understand that the handwriting can be classified into 5 forms:

In boxed discrete handwriting form the characters are enclosed in a box and each box contains a single character only. So, it us very easy to extract each character in the document.

In case of discrete handwriting form each character is having some space from other character. So, they can be extracted using the gap between them.

In run-on handwriting type even though the pen is lifted after each character, but there is overlapping of characters over each other. Which makes it hard to extract the characters individually.

Same is in the case of cursive where all characters are connected to each other and in unconstrained which is the combination of all the other handwriting types.

So, to extract the different characters in the document what we can do is we can divide the document into small blocks of 20x15 pixels, so that on an average each block contains a character which will be recognized using the OCR of our program.

Now the handwriting recognition is done in 3 steps:

- Pre-processing
- Feature Extraction
- Recognition of Characters.

Pre-processing

In the pre-processing part we will be providing the hand written documents as scanned images to the computer which will be processed to recognize the characters.
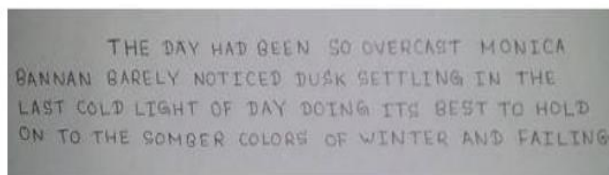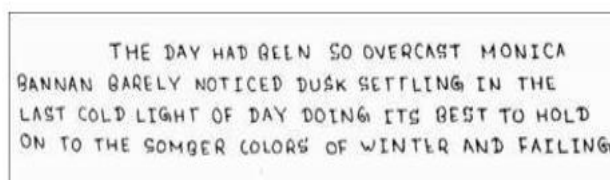

Fig: the image to be scanned.

Fig: the filtered scanned copy

The scanning of images can be done using scanners or using phone cameras and then it will be given as a input to the system.

Feature Extraction

The recognition rate mainly depends upon the features of the character. The scanned image of the input is divided into segments. Each segment contains a single character to be recognized. Then the features of each character in each segment are extracted with the help of the data stored in the program during the training period of the program.

Fig: segmented image.

Recognizer

After feature extraction the recognition system is used to recognize the handwritten characters provided as input to the system. The accuracy of the result depends upon the amount of training provided as well as the technique used to train our system on training data.

5. CHALENGES AND LIMITATIONS

- Existing systems does not include scanning of handwritten documents. Systems provide questionnaire and user have to choose options according to their own understanding.
- The result may be incorrect or not 100% correct.
- To get the correct results proper training should be provided to the program otherwise the results will be incorrect.

6. CONCLUSION

In this paper we have introduced the offline handwriting recognition, to provide in improving the ability of recognition accuracy. We will be making a machine learning program using python programming language and will provide training to it using neural networks. This will provide an ability to the computer to understand the handwritten inputs provided to it. User can use this project to understand the handwriting which is not easily understandable.

7. REFERANCES

[1] Rahul KALA, Harsh VAZIRANI, Anupam SHUKLA,and Ritu TIWARI "Offline Handwriting Recognition using Genetic Algorithm" International Journal of Computer Science Issues, Vol. 7, Issue 2, No 1, March 2010 ISSN (Online): 1694-0784 , ISSN (Print): 1694-0814
[2] Brandon Maharrey "A Neural Network Implementation of Optical Character Recognition" Technical Report Number CSSE10-05 COMP 6600 – Artificial Intelligence Spring 2009
[3] Sang Sung Park, Won Gyo Jung, Young Geun Shin, Dong-Sik Jang "Optical Character Recognition System Using BP Algorithm" IJCSNS International Journal of Computer Sci 118 ence and Network Security, VOL.8 No.12, December 2008
[4] Krupa Dholakia "A Survey on Handwritten Character Recognition Techniques for Various Indian Language" International Journal Of Computer Application (0975-8887) Volume 115No. 1, April 2015
[5] https://en.wikipedia.org/wiki/Handwriting_recognition
[6] https://en.wikipedia.org/wiki/Optical_character_recognition
[7] Abbas Mustafaoglu - "Offline handwriting recognition using neural networks"School of computer science, University of Manchester.
[8] Pooja Yadav, Nidhika Yadav - "Handwriting recognition system - A review",International Journal of Computer applications (0975-8887) volume 114, no. 19 march 2015
[9] Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks by Alex Graves and Jurgen Schmidhuber http://papers.nips.cc/paper/3449-offline-handwriting-recognition-with-multidimensional-recurrent-neural-networks.pdf