IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

Typed Character Recognition using ANN

Arindam Saha[1], Nitupon Talukdar[2]

Asst. Prof, Department of Computer Science and Engineering, GIMT, Guwahati, India[1]

M.Tech Scholar, Department of Computer Science and Engineering, GIMT, Guwahati, India[2]

Abstract: In this paper, a Character Recognition system is designed using Multilayer Feed forward Artificial Neural Networks. For the training of neural network Backpropagation Learning algorithm is preferred. Input is typed English alphabets with a space between the individual characters, which are input to the segmentation algorithm and then fed as input to the neural network. A neural network is trained onto the dataset containing 6 different samples for each 26 alphabets for recognition. We set the size of our training images as 40x40 gray scale image. From such training images we create image profile. Each element of the profile vector is connected with one input to the neural network. An implicit approach for character recognition is implemented in this paper for its recognition, which improves the accuracy significantly.

Keywords: Character recognition, ANN, Segmentation, Recognition.

I. INTRODUCTION

Character recognition has been an important and an active area of research in the field of image processing. In this proposed approach, the alphabets are typed manually and saved as word document. The fonts and sizes of the typed alphabets or characters can vary. The typed alphabets of different fonts are compared to different handwriting styles of different people, this is the analogy used here.

II. WORKING PRINCIPLE

We split the problem of recognizing characters into two sub-problems. They are segmentation and recognition.

A. Segmentation

First, breaking an image containing many characters into a sequence of separate images, each containing a single character. For example, breaking the image JUDAS into images, JUDAS

Once the image has been segmented, the program then needs to classify each individual character. So, for instance, we'd like our program to recognize that the first character above, i.e., J is a J.We assume that the letters are typed with a space between each letter and given as an input image to our segmentation algorithm. Then we apply a connected component based technique to find the connected component in the input image which represents a single character. After extraction of single characters, we generate the profile vector for each character. For training data we need to label the instances. Also we used 26 output neurons. Thus we use 26 bits to represent each letter. i.e.

A=00000000000000000000000001
B=00000000000000000000000010
Z=10000000000000000000000000 and so on.

Thus to create data we use the following steps:

1. Generate profile vector from the character image.

2. Append proper label vector to the profile vector. Once we have a labeled training data set, we feed them as input to the neural network. The NN is trained for a certain number of iteration.

B. Recognition

To recognize individual character a three-layer neural network is used. The input layer of the network contains a vector generated from the training images. The size of our training images is set as 40x40 gray scale images. From such training images we create image profile. Each element of the profile vector is connected with one input to the Neural Network. The image is divided into a fixed size grid. For each cell in the grid the average value of the pixels is collected.

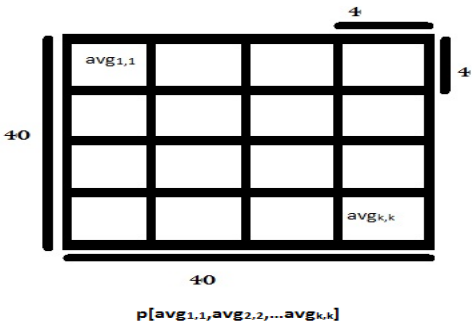

Fig 1 Grid view of the image

The second layer of the network is a hidden layer. We denote the number of neurons in this hidden layer by n. The output layer of the network contains 26 neurons. If the first neuron fires, then that will indicate that the network thinks the character is an A. If the second neuron fires then that will indicate that the network thinks the character is a B. And so on.

What we'd like is an algorithm which lets us find weights and biases so that the output from the network approximates y(x) for all training inputs x. To quantify

DOI 10.17148/IJARCCE.2016.54139

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

how well we're achieving this goal we define a cost function:

$$C(w,b) \equiv \frac{1}{2n} \sum_x \|y(x) - a\|^2.$$

Here, w denotes the collection of all weights in the network, b all the biases, n is the total number of training inputs, a is the vector of outputs from the network when x is the input, and the sum is over all training inputs, x. We'll do that using an algorithm known as gradient descent. Suppose in particular that C is a function of m variables, (v1,…,vm).

Then the change ΔC in C produced by a small change $\Delta v=(\Delta v1,\ldots,\Delta vm)^T$ is $\Delta C \approx \nabla C \cdot \Delta v$ where the gradient ∇C is the vector $\nabla C \equiv (\partial C/\partial v1,\ldots,\partial C/\partial vm)^T$.

Just as for the two variable cases, we can choose $\Delta v=-\eta\nabla C$ and we're guaranteed that ΔC will be negative. This gives us a way of following the gradient to a minimum, even when C is a function of many variables, by repeatedly applying the update rule $v \rightarrow v'=v-\eta\nabla C$. It gives us a way of repeatedly changing the position v in order to find a minimum of the function C.

Table 1 Comparative analysis of previous papers

PAPER	DESCRIPTION
1. Handwritten Text Recognition System Based on Neural Network	• A novel approach has been proposed for handwriting recognition system involving segmentation for pre-processing steps and using diagonal based feature extraction technique with neural network for character recognition. • Further, a diagonal based feature extraction technique is used for extracting the features of handwritten alphabets. Here, a character is segmented into parts dynamically for character recognition from the text, which improves the accuracy significantly. • A feed forward artificial neural network is being use d for character classification, which also helps in deciding the threshold value for the character separation from the running text word.
2. Diagonal based feature extraction for handwritten alphabets recognition system using neural network.	1. An off-line handwritten alphabetical character recognition system using multilayer feed forward neural network is described in the paper. A new method, called diagonal based feature extraction is introduced for extracting the features of the handwritten alphabets.
3. Introduction to multi-layer feed-forward neural networks.	2. This paper introduced the basic concept for wavelet Transform, and some applications of Wavelet Transform used in HCR . It also discussed difference between conventional Fourier and modern time-frequency analysis.
4. Document image skew detection: Survey and annotated bibliography	• Algorithms that estimate the angle at which a document image is rotated are surveyed. Four broad classes of technique are identified. • These include methods that calculate skew from a horizontal projection profile, a distribution of feature locations, a Hough transform, or the distribution of responses from local, directionally sensitive masks.
5. Character Recognition Using Neural Network.	3. In the present paper, neural network has been used to recognize the character. In this paper an off-line strategies has been developed for the isolated handwritten English character (A to Z).
6. Scale space Technique for word segmentation in Handwritten Document.	4. In this paper it has developed a novel methodology for segmenting handwritten document images by analysing the extent of “blobs” in a scale space representation of the image.
7. A Simple and Efficient Skew Detection Algorithm via Text Row Algorithm.	5. This paper presents a new, accurate and robust skew detection algorithm based on a method for finding rows of text in page images.

DOI 10.17148/IJARCCE.2016.54139

III.CONCLUSION

Artificial Neural Networks provide many advantages in character recognition because neural networks can work with any data irrespective of the noise and defects present. We consider artificial neural networks as a black box which means that anything can be trained in it. We can increase the recognition rate by collecting as much data as possible to the neural networks. It is not possible to design fully automated system which can handle all kind of variability. Handwriting recognition has become a burning topic in the area of pattern recognition and a lot of improvements are to be made to the system.

REFERENCES

[1] Juan A. Ramirez-Quintana, Mario I. Chacon-Murguia and Jose F. Chacon Hinojos,"Artificial Neural Image Processing Applications: A Survey", Advance online publication: 27 February 2012.

[2] Neeta Nain Subhash Panwar, Malaviya National Institute of Technology Jaipur, "Handwritten Text Recognition System Based on Neural Network", AcademyPublich.org Journal of Computer and Information Technology Vol.2, No.2

[3] J.Pradeep, E.Srinivasan and S.Himavathi, "Diagonal based feature extraction for Handwritten alphabets recognition System using neural network", International Journal of Computer Science & Information Technology (IJCSIT), Vol 3,No 1, Feb 2011.

[4] ChhayaSambhajiGochade, "Handwriting recognition using Neural networking", International Journal of Computer, Information Technology & Bioinformatics (IJCITB) ISSN: 2278-7593, Volume-I, Issue-4 IEEE & IEEE Computational Intelligence Society Hyderabad Section Sponsored!st International Conference on Robust Technology for Human Identification.

[5] Daniel Svozil, Vladimir KvasniEka, JiEPospichal, "Introduction to multi-layer feed-forward neural networks", Chemometrics and Intelligent Laboratory Systems 39 (1997) 43-62 revised 25 February 1997; accepted 6 June 1997

[6] Jonathan j. Hull Ricoh, "Document image skew detection: Survey and annotated bibliography", California Research Centre, 2882 Sand Hill Road, Suite 115, Menlo Park, CA94025

[7] Ankit Sharma, Dipti R Chaudhary, "Character Recognition Using Neural Network", International Journal of Engineering Trends and Technology (IJETT) – Volume4Issue4- April 2013

[8] R.Manmatha and Nitin Srimal," Scale space Technique for Word Segmentation in Handwritten Document", CSE Dept., University of Massachusetts, USA.

[9] Ray Smith, "A Simple and Efficient Skew Detection Algorithm via Text Row Algorithm", Personal Systems Laboratory, HP Laboratories Bristol, HPL-94-113 December, 1994.

[10] Houcine BOUBAKER, MonjiKHERALLAH, Adel M. ALIMI, "New Algorithm of Straight or Curved Baseline Detection for Handwritten writing," 2009 10th International Conference on Document Analysis and Recognition.

BIOGRAPHIES

Mr Arindam Saha is an Assistant Professor of department of Computer Science& Engineering, GIMT Guwahati. His research interest is on Image processing, neural networks and software testing.

Mr Nitupon Talukdar has done his BE in Computer Science & Engineering from Gauhati University and pursuing M.Tech from GIMT Guwahati under ASTU. His research interest is on Image processing and neural networks.