A New Data Base Scheme Arabic Handwriting Recognition by Hopfield Neural Networks Algorithm

Hasan M. Al Shalabi, Mowafak F. Hasan and Abbas M. Ali
Faculty of Computer Engineering and Information Technology
Al Hussein Bin Talal University, Ma`an, Jordan

Abstract: A new scheme for Arabic handwriting recognition by means of mat lab software is proposed. It can be seen as construction of neural networks that associate the hand writing Arabic text with the stored Arabic text in a specific data base are nearly equivalent. This result indicates that the new scheme have the potential to be robust in the presence of noisy data input.

Key words: Hopfield, Neural Network, P.E., Frames, Preprocessing, Database

INTRODUCTION

There are many languages using Arabic characters such as Fairs, Curds and Urdu. These characters have specific properties in their shapes as well as flexibility in the handwriting. For example the character these and many others have the same meaning in all these language so it is important to get away that will share all these shapes in nearly one shape or font stored in the data base. Work on recognizing Arabic letters is still limited until recently. Nevertheless, Somaya AL-Na'adeed *et al.* [1] deal with the processing hand writing text recognition research. Tim Klassen demonstrates a framework for giving good recognition accuracy to Arabic letter [2]. Almuallim and Yamaguchi proposed a structured recognition technique for Arabic hand write work [3]. Al-Sadown and others proposed a technique for recognition of hand print Arabic letters using Neural Network [4]. Much advantages of the Arabic hand writing can be seen in the bank cheque [5] and postal addresses [6] as well as signature [7].

In this research, we used neural network to share all the fonts in only one font and an explanation of neural network property for pattern association, this will address the problem of different fonts. Hopfield neural network is typically used since it has its efficiency property.

This study deals with the design of Artificial Neural Network (ANN) to store and retrieve as well as the processing of handwritten Arabic characters. Because the Arabic letter recognition involves many steps that can not be separated from each other, so we need to separate the text into many tokens and therefore the data base organized according to this idea. The token will be introduced to the (NN) and this will be recognized a according to this procedure. It is so important that the Arabic handwriting space considered in it such as if two token or words will be separated by spaces or blanks, this will allow the preprocessor programs that recognize the words or tokens to introduce it to the (NN).

The present study will be discuss how the forms are designed. There are different approaches to form dropout. Some approaches use separate cleaning steps, while others uses combined cleaning method for both foreground and background [8], as before these steps, will consider morphological algorithm in this research.

ANN and Association: Artificial Neural Network is an information processing system that has certain performance characteristics in common with biological neural network [2].

Information processing occurs at many simple elements called neurons, Fig. 1. Where signals are passed between neurons over connecting links, each connecting link has an associated weight which in a typical neural net multiplies the signal transmitted, each neuron applies an activation function (usually non linear) to its net input (sum of weighted input signals) to determine its output signal [3].
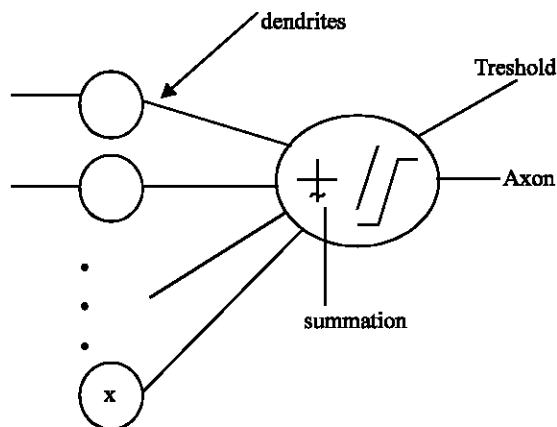

Fig. 1: Processing Element How Processing the Signals

One of the important characteristics that most NN impose is that adaptive learning where the ability to learn to do tasks based on the data given for training [8].

Hopfield Neural Network: We are all familiar with the idea of association an English word may have an associated Arabic word, our ability to recognize a form is an example of the capability of the capability of human memory to respond to new situations.
An iterative net, the net is fully interconnected neural net, the sense that each unit is connected to every other unit. The net has symmetric weights with no self connections.

Wij=Wji where i,j=1,2,3, … ,n and Wii=0.

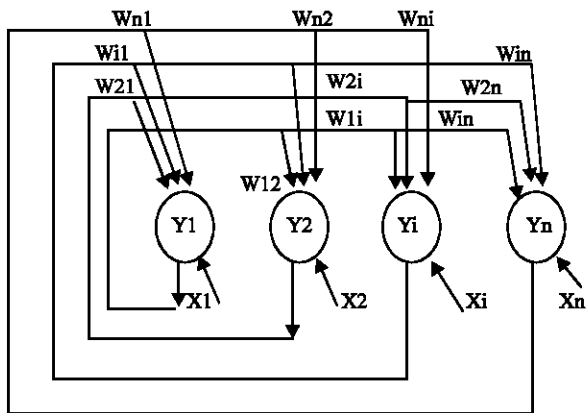The architecture of the neural net.

Fig. 2: Hopfield Neural Network P.E. Activation

We state the algorithm to be reapplied for the present work, where there are several versions of the discrete Hopfield net.

A set of binary patterns S(p)= 1,…, p
Where S(P)= (S1(p),….,Si(p),..Sn(p),

Step 0: initialize weights to store patterns (use hebb rule)
    While (activations of the net are not converge) do step 1 - 7
        Step 1: for each input vector X do step 2 - 6
            Step 2: set initial activations of net equal to the external input vector X .
            Ti=Xi , (i= 1, ……,n).
        Step 3: to steps 4-6 for each unit Yi (units should be updated in random order).
        Step 4: compute net input

$$Yin \ = \ Xi \ + \ \sum YiWsi$$

Step 5: determine activations (output signal)

$$Yi = \begin{cases} 1 & if & Yi \geq \theta_i \\ Yi & if & Yi = \theta_i \\ 0 & if & Yi \leq \theta_i \end{cases}$$

Step 6: broadcast the value of Yi to all other
        units (this update the activation vector)
    Step 7: Test for convergence.

The text image will be converted to binary patterns using matlab, which provided the algorithm development and tokenize to a set of token then these tokens will be entered to the above algorithm for comparison with the entire data base.

Arabic Word Storage: As seen in Fig. 3, the DATABASE that stores the Arabic words and meaning in English is contains thousands of words that possible written in the cheque that will be supported by the customer to the bank to proceed the procedure of getting money.

	Arabic word	Meaning in English
1		In
2		AND
3		IS
4		FROM
5		TO
6		THAT
7		THOUSAND

Fig. 3: Shows the Database for the Original Fonts

Form Design and Recognition the Words: The form is already design by the office like (bank) it will be the first three lines that the bank name and the title of the cheque, we focus on the quantity written by the customer [5].
The color of the forms was chosen to be filtered from the cheque and also the line which is exist in the cheque will be subtracted from it, to be prepared for the text translated into typed words, however there are disadvantages of time-consuming and taking up unnecessary storage space, with the morphology system to detect the words (tokens) in the image sentence, the words will be compared with the stored word in the database using Hopfield neural network. It is much better to compare the first letter from the token to speed up the comparison. The Fig. 4 showing the difference handwritten fonts for the word

Fig. 4: Shows the Difference of Word Images

Retrieving Images: Every word image is saved with its name to identify the writer common file types in bitmap format are JPEG, GIF, TIFF and WMF.

To be easy retrieve of handwritten images, the Arabic handwritten data was stored alphabetically for retrieving the images the system will compare the entered image tokenized ("hand written") and the store image according to the Hopfield Neural Network (Fig. 5).

			#	
		#	#	#
				#
#	#	#	#	#

-1	+1	-1	-1	-1	-1	+1	+1	+1	-1

Fig. 5: Shows the Representation of Fonts

$$Y = \begin{bmatrix} \dots & \dots & \dots \end{bmatrix}$$

$$W = \begin{bmatrix} 0 & -1 & & & \\ & 0 & & & \\ & & 0 & & \\ & & & & \\ & & & & 0 \end{bmatrix}$$

Fig. 6: Hopfield Weight Matrix

Fig.7: Processing Phases of the Image

Preprocessing Procedures: It is very useful to do some procedures that to process the handwritten word image, this part contains some processing for the image such that removing some of these variations in the images which do not affect the identity of the word [4] in this system the word image is loaded, then will be segmented into frames (Fig. 7).

CONCLUSION

A new scheme for Arabic handwriting recognition by means of matlab software have been proposed based on well known neural networks but using Hopfield algorithm admittedly, the database presented in this study assumes an artificial neural networks to store the Arabic word only one time, so a free handwriting words in specified area of writer is mimicked. With regard to computational requirements the new database is competitive and in particular, is more efficient.

REFERENCES

1. Somaya Al-Ma`adeed, Dave Elliman and Colin Higgins, 2004. A database Arabic handwritten text recognition research. IAJIT, 1: 1.
2. Tim Kalssen, 2002. Towards neural network recognition of hand written Arabic letters. MSc. Thesis, Dalhousie University.
3. Almuallim, H. and S. Yamagunchi, 1987. A method of recognition of Arabic cursive handwriting. IEEE Tras. Pattern Analysis and Machine Intelligence (PAMI-9), pp: 715-722.
4. Al-Sadown, A. H. and S.S. Fischer, 1996. Hand-printed Character Recognition System Using Artificial Network. Pattern Recognition, 29: 663-675.
5. Howell, D., 2000. Getting to Grips with Graphics File Format. Computer Publishing Issue.
6. Laurence Fauselt, 1994. Fundamentals of Neural Networks. Prentice-Hall.
7. Downton, A. and S. Impedoro, 1997. In Progress in Handwriting Recognition. World Scientific, U.K.
8. Haykin, S., 1994. Neural Networks: A Comprehensive Foundation. Prentice-Hall, New-Jersey, USA.