Handwritten Nastaleeq Script Recognition with BLSTM-CTC and ANFIS method

Rinku Patel[#1], Mitesh Thakkar[*2]

[#]Department of Computer Engineering, Gujarat Technological University
Gujarat, India
*Department of Information Technology, Gujarat Technological University
Gujarat, India

Abstract:- ***A recurrent neural network (RNN) has been successfully applied for recognition of cursive handwritten documents, both in English and Arabic scripts. Ability of RNNs to model context in sequence data like speech and text makes them a suitable candidate to develop OCR systems for printed Nastaleeq scripts (including Nastaleeq for which no OCR system is available to date). In this work, we have presented the results of applying RNN to printed Urdu text in Nastaleeq script. Bidirectional Long Short Term Memory (BLSTM) architecture with Connectionist Temporal Classification (CTC) output layer was employed to recognize printed Urdu text. The propose method use multidimensional BLSTM and ANFIS Method for OCR recognition. The ANFIS approach learns the rules and membership functions from data. ANFIS is an adaptive network. An adaptive network is network of nodes and directional links. These networks are learning a relationship between inputs and outputs. The Recognition error rate is 5.4 %. These results were obtained on synthetically generated UPTI dataset containing artificially degraded images to reflect some real-world scanning artifacts along with clean images. Comparison with shape-matching based method is also presented.***

KeyWords:- URDU character,RNN,BLSTM,ANFIS,CTC

I. INTRODUCTION

Recurrent neural network (RNN) are good at context aware processing and recognizing patterns occurring in time series [1] . The main drawbacks of traditional RNNs are the requirement of pre-segmented input and that the input on the hidden layer either decays or blows-up exponentially [1]. The hidden layer of an LSTM network consists of recurrently connected blocks that in turn contains internal units whose activation is controlled by input, forget and the output gates. The recurrent connections of cells are controlled by the forget gate. So, the network can hold the information as long as the forget gate is switched on. Graves [2] introduced Bi-directional LSTM (BLSTM) architectures for accessing context in both forward and backward directions. BLSTM is a combination of bi-directional neural network (BRNN) and LSTM architectures and it uses two hidden layers, one for forward pass (from left to right) and the other for backward pass (from right to left). Both layers are then connected to a single output layer. Multidimensional LSTM (MDLSTM) for offline Arabic handwriting recognition. They first divided the input image into 3 × 4 sub-images and then scanned them by four MDLSTM layers. They scanned the image in all four directions (right-to-left, left-to-right, top -to- bottom and bottom-to-top) to capture the context [1].

The ANFIS approach learns the rules and membership functions from data .ANFIS are an *adaptive network*. An adaptive network is network of nodes and directional links. Associated with the network is a learning rule - for example back propagation. It’s called adaptive because some, or all, of the nodes have parameters which affect the output of the node. These networks are learning a relationship between inputs and outputs. A class of adaptive networks that is functionally equivalent to fuzzy inference systems.

Urdu is the national language and lingua franca of Pakistan and is considered as one of the important languages of the Indian subcontinent. It belongs to the family of Nabataean scripts and shares many common properties of other family members like Arabic and Persian. Some of its salient features are writing from right to left, presence of huge number of ligatures *(connected set of components with associated dots and diacritics)*, variations in the character's shape depending on its location in a ligature (context), kerning, etc [1].

One of the most important issues in Urdu language is change in shape of a character depending upon its position in a word. Context gives an important role in determining the particular shape of a character at a particular position. The shape of a character is located at initial, middle or final position in a word may differ significantly. Dots and diacritics give meaning and identity to all character in a ligature. Reorganization of dots and diacritics to their base character is also a challenging task in Urdu because of their relatively smaller sizes. There are two prominent writing styles in Urdu: *Naskh*3 and *Nastaleeq*4. Figure 1 shows these two styles. The standard Urdu language written in magazines, newspaper, and books are in Nastaleeq script, while most of Urdu online material is available in Naskh. An important distinction between the two styles is that the Naskh's flow is horizontal from right to left, while the Nastaleeq's flow is diagonal from right top to left bottom. This makes Nastaleeq to occupy less space for a ligature than the Naskh font. The scope of the current work is confined only to Nastaleeq script.

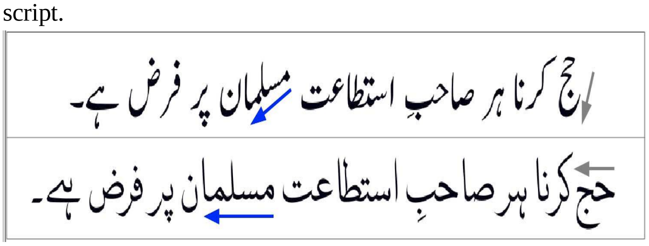


Fig. 1. Two commonly used styles for Urdu scripts. Nastaleeq script (above) is used for Urdu publications, while Naskh(below) is used for Web-viewing. Arrows show the direction of reading flow.

Urdu script consists of 45 basic characters. Five (05) characters can only occur in isolation, 10 can occur in first position or at last position, 2 characters can occur only at the end of a ligatures, and only 1 character can occupy position in middle; it can't be located in any other position. Remaining 27 characters may occur in isolation, at the beginning, at the end or in the middle of a ligature. Moreover, there are 26 punctuation marks, 8 honorific marks, and 20 digits. Some common punctuations (like %,<,>, parentheses, etc.) and English numerals are also used in Urdu publications frequently; so they are also included in the list of possible characters/class-labels (in terms of machine learning terminology). Characters belonging to above-mentioned eight categories are shown in Figure 2. So, in total there are 99 individual labels. Moreover, if we take the shapes of various characters as a separate label, then there are 191 labels. The last column in Figure 2 details the number of classes in each category as per their number of shapes depending on their position in a ligature.

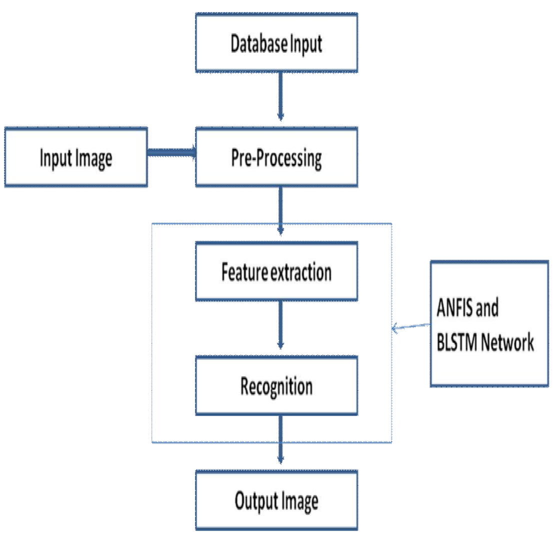
Category	Urdu Characters	Classes
Characters that can occur only in isolation. (CAT-I)	آ، ژ، ئے، ؤ، ء	5
Characters that can not occur in the beginning or in middle of a ligatures. (CAT-II)	ا، د، ڈ، ذ، ر، ڑ، ز، ل، و، ے	20
Characters that may occur in isolation, at the beginning, at the end or in the middle of a ligature. (CAT-III)	ب، پ، ت، ٹ، ث، ج، چ، ح، خ، س، ش، ص، ض، ط، ظ، ع، غ، ف، ق، ک، گ، ل، م، ن، ہ، ھ، ی	108
Characters that can occur only in isolation or in the middle of a ligature. (CAT-IV)	ئ	2
Characters that can occur only at the end of a ligature. (CAT-V)	ۃ، ۂ	2
Honorific marks (CAT-VI)	ﷲ، ﷺ، ﷽، ؑ، ؓ، ؒ، ؔ، جل جلالہ	8
Punctuation marks (CAT-VII)	‹‹، (،)، [،]، {، }، /، >، <، !، %، ؛، :، '، "، ؟، ۔، ٖ، ٗ، ٘، ٍ، ً، ٌ	26
Numbers (CAT-VIII)	0123456789, ۰۱۲۳۴۵۶۷۸۹	20

Fig. 2. Character categorization.

The recognition of cursive characters is an active research field. Work is evaluated by [8] for cursive character recognition using Support Vector Machine (SVM) which is based on segmentation. They performed experiments on isolated characters and computed local and global features of it. Another work in relation to cursive script is proposed by Nagata [3]. They presented an OCR approach for cursive characters of a language which has a large character set (like Chinese, Japanese etc). They used approximate character shape similarity and a word segmentation algorithm with support of language model. Graves [4] evaluated multidimensional LSTM (MDLSTM) for offline Arabic handwriting recognition. They first divided the input image into 3 × 4 sub-images and then scanned them by four MDLSTM layers. They scanned the image in all four directions (right-to-left, left-to-right, top-to-bottom and bottom-to-top) to capture the context. Sankaran and Jawahar [6] applied BLSTM networks for Devanagari script OCR problem. Frinken et al [7] applied BLSTM networks to word spotting problem by modifying the CTC token-passing algorithm.

The aim of current work is to further extend the research towards reliable OCR for Nastaleeq script. The next section describes preprocessing and feature extraction step. Configuration and training procedure for our MD-BLSTM network and ANFIS is outlined in Section III. Section IV presents the experimental evaluation of MD-LSTM network for Nastaleeq script and the results are discussed in Section V.

Proposed Character recognition System



II . PREPROCESSING AND FEATURE EXTRACTION

Baseline information of a text line is an important feature for common distinguishing characters. So it is necessary to normalize the input images to a specific height. Currently, there are no Nastaleeq-specific normalization methods reported. In the current work, each text-line image was rescaled to a fixed height. Raw pixel values are used as features and no other sophisticated features were extracted. A 30 × 1 window is traversed over the text-line image and the resulting MD sequence is fed to ANFIS network for training.

III. ANFIS CONFIGURATION WITH MD - BLSTM NETWORK

As mentioned earlier, ANFIS with MD-BLSTM architecture with CTC output layer was employed to evaluate RNN for Urdu script. A publicly available RNN library [9] was used for evaluation .Implementation of multidimensional BLSTM networks is provided in this library along with CTC output layer. For the training of the network, ANFIS is used forward pass and a backward pass algorithms. We now look at each layer in turn for the forward pass. The forward pass propagates the input vector through the network layer by layer. In the backward pass, the error is sent back through the network in a similar manner to back propagation. Size of hidden-layer, learning rate and momentum are other tunable parameters.

For training purpose, the normalized gray-scale input text line image was scanned from left to right and top-to-bottom to extract the features. The corresponding transcriptions were reversed to make it consistent with the input image (Urdu is read from right to left). Normalized text-line images along with their transcriptions were fed to MD-BLSTM network, which performed the forward propagation step first. Alignment of output with associated transcriptions is done in the next step and then finally backward propagation step was performed. After each epoch, training and validation error were computed and the best results were saved. When there was no change in training and validation errors for a pre-set number of epochs, the training stopped. Training and validation errors were recorded and the network was evaluated on test set. Hear we used four parameters, namely input-image size, hidden-layer size, learning rate and the momentum. The input image height was set to 40 and was not altered. Momentum value was also kept fixed at 0.5. Best parameters for hidden-layer size and learning rate were 200 and 0.01 respectively. For this network with best parameters, training and validation errors as a function of number of epochs are shown in Figure 3. This network took 22 epochs to converge. Hear use line dataset for present work. This line is divided into three sub categories training (70%), validation (35%), testing (25%).However, it can be seen that the validation error is minimum after 16 epochs (marked as dotted-line in Figure 3). This network is returned as the best network and Recognition Error is decreases with increasing the number of unit in hiddenlayer.

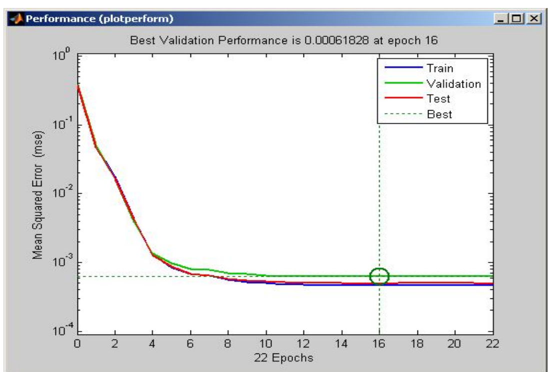


Fig 3 Training vs. validation error

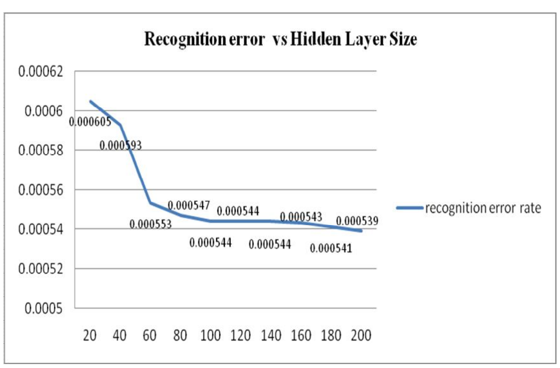


Fig 4 Recognition Error Rate vs. Hidden Layer

Here, train MD-BLSTM network with different hidden layer sizes 20,40,60,80,100,120,140,160,180,200 and then find training time with different hidden layer size.

IV. EXPERIMENTAL EVALUATION

This section discusses the results of evaluating BLSTM architecture on printed Urdu script.

A. *Database:* - Sabbour [10] is used synthetic database' called UPTI *(Urdu Printed Text Images)*-dataset, was used for evaluation. This Urdu dataset consists of 10, 063 synthetically generated text lines and more individual character and different shapes.14 sets were generated by varying three parameters, namely, *jitter, sensitivity* and *threshold.* This dataset contains both ligatures and lines versions; hear lines dataset was used for the present work. These lines were divided into three sub-categories, training (70%), validation (35%) and testing (25%). The ground-truth of these text-line images was also available.

B. Parameter Selection: - In the present work, two parameters namely *learning rate, momentum* and *number of hidden-layers* were evaluated for their respective effect on the recognition accuracies. First, the most appropriate number of hidden-layers was determined keeping learning rate constant at 0.01. We trained MD-BLSTM networks with hidden-layer of sizes 20, 40, 60, 80, 100, 120, 140, 160,180 and 200. The comparison of respective recognition-errors on test set is shown in Figure 4. The training time as a function of hidden-layer sizes is shown in Figure 5. From Figure 4 and Figure 5, we can deduce two points; first that increasing the number of hidden-layer sizes decreases the recognition error but at the same time, training for network with large number of hidden-layers requires more time. Moreover, it is also noted that increase in training-time is almost linear, while increase in hidden-layer sizes does not increase accuracy more than 5% when the hidden-layer size is from 120 to 200. So, it was decided to select 120 as the optimal hidden-layer size for the present work.

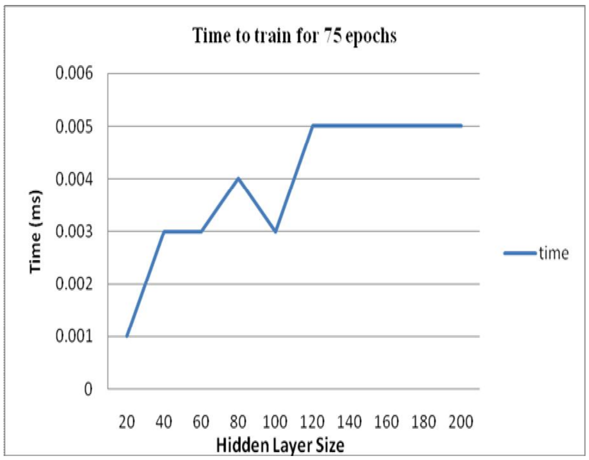


Figure 5 Training time as a function of hidden layer size

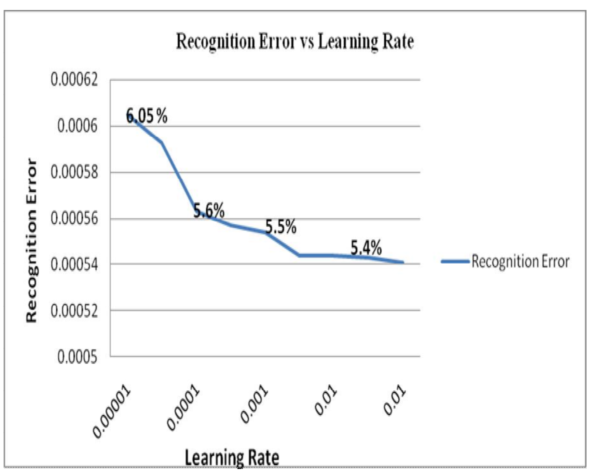


Figure 6 learning rate 0.01 is given the lowest recognition error

Hear, keeping the 120 hidden layer size and learning rate varied between 0.01,0.001, 0.0001,0.00001. The comparison of respective recognition error on test set is shown this graph. Learning Rate 0.01 is more suitable for best recognition error rate we can find 5.4% recognition error rate.

C. Results: - As mentioned in Section IV, ANFIS with MD-BLSTM networks have been evaluated for Hand written Nastaleeq Script Recognition with BLSTM-CTC and ANFIS method and find recognition error rate 5.4%.As mentioned in Section I that there have not been many OCR systems available for Hand written Urdu Nastaleeq script. Only shape matching based OCR system proposed by Sabbour et al [10] is reported in recent times. They evaluated their system on clean printed text as well on some of the artificially degraded versions of the clean dataset. They achieved 11.2% letter error rate on clean images. And offline printed Urdu Nastaleeq script recognition proposed by Adnan UI-Husan[1] and they achieved 13.6% recognition error rate. They also reported error rates for various degradation effects on individual basis. There is no error rate reported for mixed dataset that we used in our evaluations. Moreover, they did not consider the case where ligature shape variations are not considered (where we achieved 5% error rate).

V. CONCLUSION

In this paper Describe hybrid of MD-BLSTM – CTC and ANFIS method for Handwritten Nastaleeq Script recognition. The context-capturing property of RNN makes it a better candidate for Nastaleeq scripts like Arabic, Urdu, Persian, etc, than other neural networks based methods. Use multidimensional BLSTM networks would localize the position of dots and diacritics better, thereby further lowering the error rates. And find the recognition error rate 5.4%. also Not used any language model because this method used token passing algorithms (CTC).ANFIS is an adaptive neurofuzzy network which allows the usage of neural network topology along with fuzzy logic. It not only includes the characteristics of both methods, but also avoids disadvantages of both fuzzy logic and artificial neural network. ANFIS combines both neural network and fuzzy logic; it is capable of handling complex problems.

REFERENCE

[1] Adnan Ul-Hasan, Saad Bin Ahmed, Sheikh Faisal Rashid, Faisal Shafait and Thomas M. Breue , "Offline Printed Urdu Nastaleeq Script Recognition with Bidirectional LSTM Networks" in 12th International Conference on Document Analysis and Recognition, 1520-5363/13 $26.00 © 2013 IEEE DOI 10.1109/ICDAR.2013.212.

[2] A. Graves, "Supervised sequence labelling with recurrent neural network."Ph.D.Dissertation, Technical University Munich, 2008.

[3] M. Nagata, "Japanese OCR Error Correction using Character Shape Similarity and Statistical Language Model." in *Int. Conf. on Computational Linguistics*, 1998, pp. 922–928.

[4] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*, ser. Studies in Computational Intelligence. Springer, 2012, vol.385.

[5] ——, "ICDAR 2009 Arabic Handwriting Recognition Competition." In *ICDAR*. IEEE Computer Society, 2009, pp. 1383–1387.

[6] N. Sankaran and C. V. Jawahar, "Recognition of printed Devanagari text using BLSTM Neural Network." in *ICPR*. IEEE, 2012, pp. 322–325.

[7] V. Frinken, A. Fischer, R. Manmatha, and H. Bunke, "A Novel Word Spotting Method Based on Recurrent Neural Networks." *IEEE Trans.Pattern Anal. Mach. Intell.*, vol. 34, no. 2, pp. 211–224, 2012.

[8] F. Camastra, "A SVM-based cursive character recognizer." *Pattern Recognition*, vol. 40, no. 12, pp. 3721–3727, 2007.

[9] A. Graves, "RNNLIB: A recurrent neural network library for sequence learning problems." [Online]. Available: http://sourceforge.net/projects/ rnnl.

[10] N. Sabbour and F. Shafait, "A Segmentation Free Approach to Arabic and Urdu OCR," in *DRR XX (Part of the IS&T/SPIE 25th Annual Symposium on Electronic Imaging)* , Feb. 2013.

[11] H. S. Baird, "Document Image Defect Models ," in *Structured Document Image Analysis*, H. S. Baird, H. Bunke, and K. Yamamoto, Eds.New York: Springer-Verlag, 1992.

[12] Emanuel Inderm¨uhle, Volkmar Frinkeny and Horst Bunke" Mode Detection in Online Handwritten Documents Using BLSTM Neural Networks"

[13] Raman Jain , Volkmar Frinken , C.V. Jawahar ,and R. Manmatha," BLSTMNeural Network basedWord Retrieval for Hindi Documents"

[14] Sorousha Moayer, Parisa A. Bahri" Hybrid intelligent scenario generator for business strategic planning by using ANFIS". www.elsevier.com/locate/eswa, Expert Systems with Applications 36 (2009) 7729–773

[15] Prof. Sheetal A. Nirve ,Dr. G. S. Sable "Optical character recognition for printed text in Devanagari using ANFIS" International Journal of Scientific & Engineering Research, Volume 4, Issue 10, October-2013 236 ISSN 2229-5518

[16] Sheikh Faisal Rashid, Marc-Peter Schambach, Jörg Rottland, Stephan von der "Low Resolution Arabic Recognition with Multidimensional Recurrent Neural Networks".