Handwritten Malayalam Compound Character Recognition System Using Neural Networks

Sherin Mathew G.[1] **and Dr. R. Vijayakumar**[2]

Research Scholar, School of Computer Sciences, Mahatma Gandhi University, Kottayam,1
Professor, School of Computer Sciences, Mahatma Gandhi University, Kottayam,2
E mail: sherinmathewg@yahoo.com1, vijayakumar@mgu.ac.in2

Article History

Received: 1 April 2018
Received in revised form: 30 April 2018
Accepted: 10 May 2018

KEY WORDS:
Handwritten, Compound, Character, Malayalam Recognition, Neural Networks

ABSTRACT

In this paper an effort is made to recognize handwritten Malayalam compound characters using Feed Forward neural networks. A handwritten Malayalam character is resized into 27x27 pixel. The resized character is used for training the neural network. Once the training process is completed the same character is given as input to the neural network with different set of neurons in hidden layer and their recognition accuracy rate for different Malayalam characters has been calculated and compared. The results show that the proposed system yields good.

INTRODUCTION:

The main objective of the proposed system is to find a new solution for Malayalam handwritten compound characters recognition of different styles by improving the design structure of the traditional Artificial Neural Network (ANN). ANNs are used in various areas like association and classification, pattern recognition, forecasting and control applications. Off-line handwritten document recognition is a tedious task, because different persons have different handwriting styles. ANN was a reliable and convenient method for the researchers. So far it proved to be a good recognizer of handwritten digits and printed characters. The major steps involved in recognition of characters include, preprocessing, segmentation, feature extraction and classification. The removal of noise and deviation in handwritten scripts is the mission of preprocessing.

1 Research in Malayalam Handwritten Character Recognition

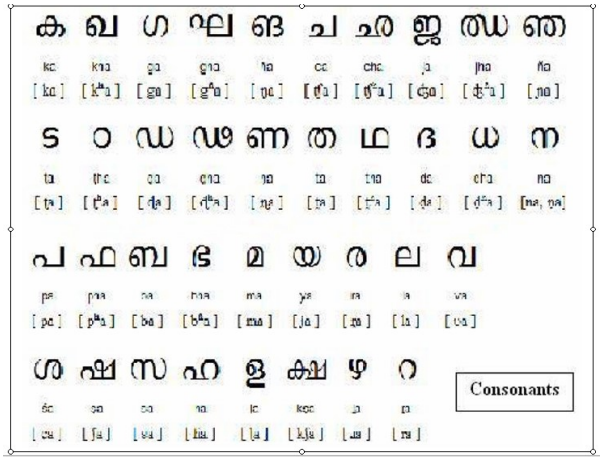
Malayalam is one of the four major Dravidian languages. The first work in offline Malayalam handwritten character recognition using class modular neural network in which fuzzy zoned normalized vector distance features are used (VL, 2007). The obtained accuracy was 78.87%. Other work reported in Malayalam handwritten character recognition are (VL,2007), State space Point Distribution parameters derived from gray scale based SSM of handwritten character samples are utilized to obtain an accuracy of 73.03%. In another work Daubechie wavelet coefficients are used in handwritten character recognition (G, 2006). The accuracy attained for the work using MLP network as a classifier was 76.8% (G, 2008). By using feed forward back propagation network 90% accuracy was obtained (G, 2007). Another work reported using 1D wavelet transform of projection profiles as features

and a multi-level perceptron network is used for classification (R, 2007). Recognition using HLH intensity pattern of characters was proposed by Rahiman (MA, 2010). In (P, 2009), recognition of handwritten Malayalam characters using discrete features, which are extracted from skeletonizsed images and the accuracy was 90.18 %t for 33 classes. By using Canny edge detector and Multilayer perceptron (MLP) the characters are recognized with an accuracy of 95.16% (P, 2010). The handwritten Malayalam character is a promising and offers an excess of chances for research. In another work different features like cross feature, fuzzy depth, distance and Zernike moment feature are extracted for each character glyph. An accuracy of 87.81% was attained for 142 Malayalam characters (V, 2013).

2. Approaches of Handwritten Character Recognition

Malayalam language is derived from the Grantha script, which is the descendant of Ancient Brahmi. The character set consists of 52 letters which includes 15 vowels and 37 consonants. The complete character set of Malayalam is depicted in Figure 1. The set also consists of 12 vowel signs. These vowels are called as dependent vowels as they are validated unless present in some combination with a consonant or a conjunct. The Malayalam script exhibits no inherent symmetry and thus making the recognition task very tedious (M, 2011).

Compound characters are special type of characters formed as combination of two or more (pure) consonants as shown in Figure 2. The shapes of the compound characters are different from the shapes of the constituent characters. These shapes are normally has complex orthographic structure and some of them are difficult to recognize when isolated from the context. Two types of compound characters occur in Malayalam: vertically compound (ൟ, ൣ, ൢ) and horizontally compound (ഞ്ഞ, ത്ത, ന്ന). There exist many similar characters with little difference in their shape.



ISSN-2455-0736 (Print)
ISSN-2456-4052 (Online)

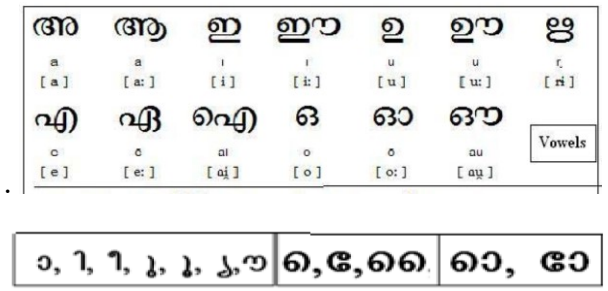


Figure1. Consonants ,Vowels & Vowel Signs

ക	kka	പ്ല	pla	ഹ	hma
ക്ല	kla	ബ്ബ	bba	ഹ	hna
ക്ഷ	kSa	ബ്ല	bla	ന്ധ	ndha
ഗ്ഗ	gga	മ്പ	npa	ത്സ	tsa
ഗ്ല	gla	മ്മ	mma	ജ്ജ	jja
ങ	n"ka	മ്ല	mla		
ങ്ങ	n"n"a	യ്യ	yya	സ്ഥ	stha
ച്ച	cca	റ്റ	RRa	ന്ത	ntha
ഞ്ച	n~ca	ല്ല	lla	ഞ്ഞ	jn~a
ഞ്ഞ	n~n~a	ള്ള	LLa	ത്ഭ	tbha
ട്ട	TTa	വ്വ	vva	ഗ്മ	gma
ണ്ട	NTa	ശ്ല	shla	ങ്ക	gda
ണ്ണ	NNa	ഷ്ണ	shsha		
ത്ത	tta	സ്ല	sla	ശ്ച	shca
ത്ഥ	ttha	സ്സ	ssa	ണ്ഡ	NDa
ദ്ദ	dda	ഹ്ല	hla	ണ്ഢ	NDha
ദ്ധ	ddha	സ്സ	sRRa	ന്മ	tma
ന്ത	nta	ഡ്ഡ	DDa	ന്ന	tna
ന്ദ	nda			ക്ത	kta
ന്ന	nna			ഗ്ന	gna
ന്മ	nma			ന്റ	nRa
പ്പ	ppa	ച്ച	ccha	ഷ്ട	STa

Figure2. Sample Compound Characters

3. Character Matrix

A character matrix is an array of black and white pixels; the vector of 0 is represented by black, and 1by white. If the size of the matrix is very big, there may be few problems like training of neural network may take days, and results may take hours. In addition, the computer's memory may not be able to handle enough neurons in the hidden layer needed to efficiently and accurately process the information. However, the number of neurons may just simply be reduced, but this in turn may greatly increase the chance for error. A character matrix size of 27X27 was created and it is shown in Figure 3.

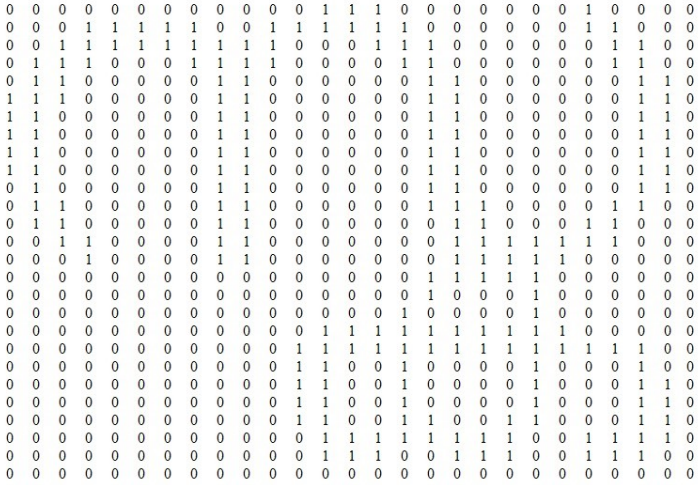


Figure3. Character Matrix of 'ska'(ಸ)

4. Artificial Neural Networks

The functions of a human brain can be simulating to that of a computer using Neural Networks. As an initial step recognition of a character by a human brain has to be identified. By practice a human brain can understand the characters in any form whether overlapped script, broken character, touched character, loosely configured character, and mixed script. Humans are able to recognize a wide collection of handwritten and typed characters through the continuous learning. These human learning activities are simulated by the system models like Artificial Neural Networks (ANN). The network generates outputs at once at the output layer, and the running time is independent of the number of outputs. The neural network works as a multi-layer decision function, and the extracted features are stored in the hidden layers arbitrarily as shown in Figure 4. Because of neural networks implement feature extraction implicitly, no need to indicate or be conscious of the relationships between neurons and features by the user. ANN is a computational model with an interconnected group of neurons, being inspired by the structure of biological neural networks. A neural network is an adaptive system that modifies its weights and biases based on an external training set and a real output that are predefined during the learning phase in pattern recognition. Present neural networks are nonlinear statistical data modeling tools used to symbolize complicated relationships between inputs and outputs (F, 1997), (PD, 1997), (H, 1996) and (TV, 2008).

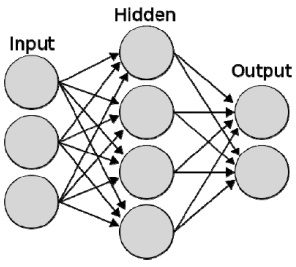


Figure 4. Neural Network Example

In the above diagram, at each layer, we have

$$Y_j = f(y_j) = f(\sum_{i=0}^{n-1} w_{i,j} X_i + b_j) \qquad (1)$$

where *f(.)* is the activation function and *y*j is the weighted summation. The output of each layer depends on the input of the previous layer and the weights between both layers. The formula may be different due to different actual connections. Normally, a sigmoid or hyper tangent function is applied after each layer to make a binary decision. The functions that establish nonlinearity to the network are called activation functions. So there is no need to add higher order items to the weighted sum. That is the reason to use neural networks in this proposed system to recognize handwritten characters.

The original binary, 128 * 128, images are resized to a smaller size using an anti-aliasing technique. The gray levels of the resulting images come from the relative portion taken by black and white pixels in a particular area. Once the neural network has been trained successfully it is then required to identify the same corresponding character of 27x27 pixel. Also the network should also be able to handle noise

TABLE I
Four Neural Network Systems used for Character Recognition

Networks	1	2	3	4
Number of Layers	2	2	2	2
Number of neurons in hidden layer	10	25	50	100
Number of Neurons in output layer	18	18	18	18
Learning rate	0.1	0.1	0.1	0.1

TABLE II
Performance comparison of four neural based character recognition systems

Handwritten Compound Characters	Percentage of Accuracy Rate			
	Network 1	Network 2	Network 3	Network 4
	82.34	83.30	84.10	83.90
	81.40	82.90	82.90	82.90
	88.00	89.90	90.23	90.10
	87.90	88.90	90.20	90.20
	83.30	88.10	89.20	89.12
	82.30	84.10	85.20	84.98
	81.34	83.90	84.20	84.10
	79.60	81.20	82.30	82.10

	78.98	80.30	81.20	80.90
	82.67	82.90	83.20	82.80
	85.70	86.60	87.70	87.60
	89.90	90.10	90.30	89.90
	88.60	89.20	89.90	89.20
	91.01	91.90	92.00	91.89
	89.90	90.20	91.10	91.02
	87.70	88.30	89.10	88.90
	89.12	90.20	91.20	91.80
	82.40	83.20	84.30	84.10
	82.20	84.10	85.90	85.70
	81.30	84.40	85.20	85.17
	87.90	89.20	89.90	89.80
	84.80	87.90	88.20	88.18
	82.30	84.80	85.20	84.70
	83.20	85.90	88.90	87.90
	86.10	88.80	89.00	88.40
Overall %	84.80	86.41	87.23	87.01

5. Conclusion

A neural network based Malayalam Compound Character recognition system has been introduced in this paper. The pixel values derived from the resized characters using image processing techniques have been directly used for training the neural network. As a result, the proposed system will be less complex compared to other methods of character recognition systems. Of the several neural network architectures used for classifying the Malayalam characters, the one with a hidden layer having 50 neurons has been found to yield the highest recognition accuracy of 87.23%. However, time efficiency and recognition accuracy still require some improvements.

BIBLIOGRAPHY

1. VL, Lajish. “Handwritten character recognition using perpetual fuzzy zoning and class modular neural networks”. Proc. 4th Int. National conf. on Innovations in IT, 2007, 188 – 192.

2. VL, Lajish. “Handwritten character recognition using gray scale based state space parameters and class modular NN”. Proc. 4th Int. National conf. on Innovations in IT, 2007, 374 – 379.

3. G, Raju. “Recognition of unconstrained handwritten Malayalam characters using zero-crossing of wavelet coefficients”. Proc. of 14th International conference on Advanced Computing and Communications, 2006, pp 217 – 221

4. G, Raju. “Wavelet transform and projection profiles in handwritten character recognition- A performance analysis”. Proc. Of 16th International Conference on Advanced Computing and Communications, Chennai 2008, pp 309-314.

5. G, Raju and K, Revathy. ”Wavepackets in the recognition of isolated handwritten characters”. Proceedings of the World Congress on Engineering 2007 Vol IWCE 2007, July 2 - 4, 2007, London, U.K

6. R, John, G, Raju and DS, Guru. “1D Wavelet transform of projection profiles for isolated handwritten character recognition”. Proc. of ICCIMA07, Sivakasi, 2007, 481-485, Dec 13-15.

7. MA, Rahiman. ”Isolated handwritten Malayalam character recognition using HLH intensity patterns”. 2010 Second International Conference on Machine Learning and Computing.

8. P, Binu Chacko and P, Babu Anto. “Discrete Curve Evolution Based Skeleton Pruning for Character Recognition”. Seventh International Conference on Advances in Pattern Recognition, 2009

9. P, Binu Chacko and P, Babu Anto. “Pre and Post Processing Approaches in Edge Detection for Character Recognition”. 12th International Conference on Frontiers in Handwriting Recognition , 2010

10. M, Abdul Rahiman and MS, Rajasree. “An Efficient Character Recognition System for Handwritten Malayalam Characters Based on Intensity Variations”. Proc. Of *International Journal of Computer Theory and Engineering, Vol. 3, No. 3, June 2011*

11. V, Vidya, TR, Indhu , VK, Bhadran and R, Ravindra Kumar. “Malayalam Offline Handwritten Recognition using Probabilistic Simplified Fuzzy ARTMAP”. *Advances in Intelligent Systems and Computing Volume 182, 2013, pp 273-283.*

12. F, Kimura, Kayahara, NY, Miyake and M, Shridhar. “Machine and Human Recognition of Segmented Characters from Handwritten Words”. *4th International Conference on Document Analysis and Recognition*, Ulm, Germany, pp. 866-869, 1997.

13. PD, Gader, M, Mohamed and JH, Chiang. “Handwritten Word Recognition with Character and Inter-Character Neural Networks”. *IEEE Transactions on Systems,Man, and Cybernetics-Part B: Cybernetics*, Vol. 27, pp. 158-164, 1997

14. H,Yamada and Y, Nakano. “Cursive Handwritten Word Recognition Using Multiple Segmentation Determined by Contour Analysis”. *IEICE Transactions on Information and Systems*, Vol. E79-D, pp. 464-470, 1996.

15. TV, Zant, DL, Schomaker and K, Haak. “Handwritten-Word Spotting Using Biologically Inspired Features”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 11, November 2008.