

An Multilingual Numerals Classification Using Soft Computing

Anshu Kumar¹, Akhilesh Pandey²

¹M.Tech scholars Department of Computer Science and Engineering, Suresh Gyan Vihar University, Jaipur

²Asst. Professor Department of Computer Science and Engineering, Suresh Gyan Vihar University, Jaipur

Abstract: This paper presents multi script numerals for off-line hand written Devnagri numbers and English Numbers classification. The main purpose of this research is to find out best recognition result using multiple scripts. This proposed technique uses simple profile technique for finding feature extraction and LDA classifier scheme new neural network for classification. The performance of this technique has been tested with 36000 handwritten numerals randomly selected from CPAR datasets out of which 22000 datasets has been used for training sets and 14000 datasets has been used for test sets and MNIST database where 60000 samples for training and 10000 samples for testing. we found the different result by LDA classifier.

Keywords: CPAR, MNIST, LDA, Neural Network

I. INTRODUCTION

In the today's world one of the critical problems is the system that not recognizes the English and Devanagari handwritten numbers. Numbers in Devnagri is not truthfully and efficiently recognized by the electronic system. For recognizing the numbers various algorithm and research work have been proposed. Numbers of process have been performed for number recognition, it's not like only single algorithm or single technique can do the recognition of the both number to give accurate result. In India, Hindi is considered as our mother tongue so the recognition of the number in Denari is very much important. Therefore to solve complete problem the following system help to solve it in more easy and precise way. But some parts of India People are very frequent in the English. Recognition of the numbers which is handwritten is very much problem in the area of research because also it is office automation and is very important requirement. It provides recognition of the numbers in a very effective and practical way. The writing of the person depends upon its writing style or on their mood. In the process of the recognition all the structure of the numbers as well as topological and statistical information is being observed. Limited variation on the shape and the size is considered in the Hindi and English numbers which is hand printed, the main focus is on the process recognition.

Researchers had been working on the recognition of handwritten numbers from last thirty five years. Public is now looking towards the technology of the recognition of the handwritten numbers or script. Now a days 100% rating for the systems which can recognize handwritten numbers or script is still not created and achieved as humans are not able

to recognize every writer's handwriting without confusion, even they cannot recognize or read their own hand writing in effective manner. Therefore it is writers' responsibility to write any text which should be in readable format.

1.1 Devanagari Script

After English and Chinese language, Hindi language is most common in all over the world and is approximately 500,000 people can read, write and speak this language. Script of Devanagari is the basic script for the various languages present in India. For example, we can say Hindi and Sanskrit. In various other languages close variation of the Devanagari script is used. In the ancient times, Sanskrit was the common language however there is written on material still available in Arabic, French, portugese, Turkish, dravidian as well English. Sanskrit is expressive language and some word of English, portugese, Turkish etc. came out of it. Devanagari script attracted lots of people attention. Pure Indian nature was the Brahmin person who was the devnagri script holder. In the 4th century, script of brahmi gupta was formed. From the script of gupta subsequent kettle come out of it in the 8th century. Later modern script of nagri came to be known as devnagriacharya. Vinoba basu says that the script came out to be known as loknagri because of the common language known by the people in the nation. it is a dialect of Hindi said by few people but it seems not to be fair since other language like Hindi language also written in following script. Devanagari is considered as script whereas Hindi is considered as language.



Figure 1: Sample of Devnagari

Numbers 1.2 English Script

English is a West Germanic language that was first spoken in early medieval in England and is now a global lingua franca. It is spoken as a first language by the majority population of several sovereign state, including the united states, the united kingdom, Canada, Australia, Ireland, New Zealand and a number of Caribbean nation; moreover, it is an official language of almost 60 sovereign state. English sometimes described as the first global lingua franca is the dominant language or in some instance even the required international language of communication, science, information technology, business, aviation, entertainment, radio, and diplomacy.

English originated in the dialects of north sea Germanic that were carried of Britain by Germanic settlers from various part of what are now the Netherlands northwest Germany and Denmark. The old English was latter transforms by two waves of invasion. The first was speakers of the North Germanic language and the second was by speakers of romance language old Norman in the 11th century.



Figure 2: Sample of English Number

Miniature form of optical number recognition is the optical character recognition. Therefore the method used here help machine to indicate number mechanically as optical mechanism. Due to this, Humans are able to identify objects in many numbers. Optical machine is the eyes whereas the activities of input are look up by the brains. The advance technology of the system OCR faced by the technologist of this system had made easier to recognize the variable. Suppose a person is allowed to read an unfamiliar language of the page, he/she may not be able to recognize the words however if the

numerical value is present in those statement, it can be easily recognized and explained by the person because the number the number present are used by the people all over the world. Therefore this concludes that Is only used to acknowledge the numbers. Secondly, the size of the alphabetical and numerical signs is almost same. It is hard to study the word which is printed which appears pitch dark, backdrops either it is mentioned above the graphics or words. The paper document i.e. document present in paper form can be easily read by human being but it creates problem for the computer to read which is a machine to learn straight report. Therefore OCR system changes the paper document into the process able shape of computer.

Machine searched pictures are reshaped and to sign in process able form of computer like ASCII of the work which is hand written and alphabet numbers are the following process of the system.

Pattern recognition is one of the areas of the OCR and character which is handwritten is processed due to the motivation of having improvement in the machine and man communication. For character recognition few commercial products are currently available. Though many research had been done but the product which can perform recognition of the handwritten data are still not available. However to solve the problem, neural networks which is artificial in nature is used due to the high interest level which is recently observed. During past generation mostly the neural network development for the feature extraction make use of either approaches of statistical or pattern matching. In the field of machine learning or artificial intelligence one of its basic aims was to enable the computer to accomplish the task in such a way that which seems very natural to the people.

1.3 Application of number recognition

There are various numbers of recognition of number applications. They are as follows:

Work-specific readers: there is number of petition on the recognition of the number for the application of high volume where input of high level is needed is praised by many of the readers. High price input is needed managing in one desired field consumes less time. Though document of same kind held size of same kind as well as layout. It is easier for the scanner of picture to concentrate on the whole information.

Address readers: Address reader's means delivering mail related to the system. It's target on the size of the mail as well as ZIP code and sorts the postal mail.

Forum reader: 2 graphics has been divided for the forum reading method. First one is the instruction which is printed and second one is field data. Only that part of the forum is mentioned in the system whose data is printed.

Check reader: here is the following mode of reading the picture of check is taken and goodness of the price is identified as well as information of the accounts present on the check and hence this data is used to cross check the outcome.

Bill processing system: the purpose of bill processing system generally is used to study the inventory documents, payments slips and utility bills. On the document a certain region is focused by the system where the information is located which is inspected. For example the value of the payment and number of account.

Passport readers: with the help of custom inspection the returning American passengers speed up is done by the automated readers of the passport. The date of birth, number of passport present of the traveler is read by the reader and cross checked with the records of the database which contains information of the smugglers and felons.

General purpose page readers: the page readers have two categories that are low end page readers and high end page readers. Compared to low end page readers more advance is high end page readers. The low end page is compatible with scanners of the flat beds which generally don't come with scanner in it. The low end page readers are generally used in environment of the office with work station desktop, which in the throughput of the system is less demanding. A sacrifice in the accuracy of the recognition is made in order to handle a document of broader rang. Some software of the OCR for the improving accuracy of the recognition allows the user to adopt engine for the recognition of the customer data.

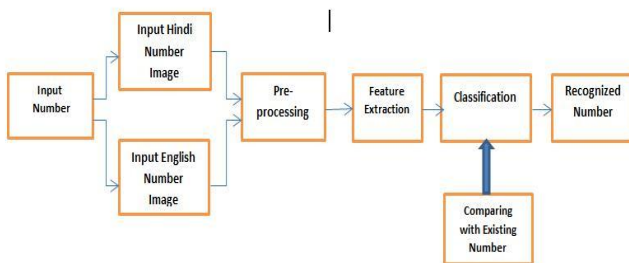


Figure 3: Block Diagram of proposed algorithm

II. PRE STUDY

In today's era the challenge in language is the neatly handwritten characters having 100% accuracy. No other optical character like roman Arabic can match this type of

achievement. In this thesis, we have proposed a system for offline recognition of handwritten Devanagari numbers. In year 1974 the first research was done based on Devanagari characters and numbers. After that again research was done on printed Devanagari character recognition using different types of neural networks.

One of the examples for this is offline numeral cognizance which is having lots of application. The most important application for this is swatting of postal zip codes in address written. This application is having widely used in post offices all around the world. This type of systems has a capability of automatic sorting billions of mails that are posted everyday making easy efforts for human kind. However the research which has been done in this area doesn't meet the requirements. So, a reliable numeral cognizance system is needed that will gain more accuracy in this area.

In recent, the different Indian scripts are much admired by existence of character modifiers. Therefore algorithm designed for them are not associated with Indian scripts. Different types of Indian scripts for many OCRs have been supposed. However none of these has a capability to store handwritten Devanagari text for composite characters and mantras.

Printed Devanagari character recognition was implemented using Kohonen neural network that is KNN. However other types of neural network were also used. These results are also implemented for bangla language. There are also some features which are implemented like concavities and intersections. Similar type of research was done on Gujarati which gave less success and for Gurumukhi script reasonable result were reported. Sinha et al. [7] reported different aspects of Devanagari script recognition whereas chatterjee and sethi has reported Devanagari numeral recognition with the help of structural approach. The various primitives which were used are left and right slants, vertical and horizontal segments. A decision tree based analysis is performed on the presence and absence of these primitives and their interconnections.

In reference [4] a similar type of strategy is implemented to inhibit hand printed Devanagari characters. A neural network based approach for different isolated characters was also implemented [8]. An implementation for recognizing handwritten Devanagari numbers based on structural descriptors has also been done [7]. Moment features, density and segment were also used as classifiers with feed-forward super structure of kohonenmodules [11].

The handwritten characters are represented as sequence of strokes in which their features is extracted and classified. For construction of stroke recognition engine support, vector machines are used. These systems have been presented after testing the system on Telgu and Devanagari scripts. In this process system for offline recognition of handwritten Devanagari numbers has been implemented. There are several attempts that has been done for OCR of Indian printed characters for very few of them are for handwritten numbers.

However, Hanmandlu and Murthy have proposed a fuzzy model based recognition of handwritten Devanagari numbers with 92.67% accuracy. Bajaj et al [7] used three different kinds of features namely moment features, Density features and Descriptive components for grouping of Devanagari numbers. They implemented advanced multi classifier connection architecture for development of recognition accuracy with 89.60% accuracy. Bhattacharya et al [7] advance layer perception; a neural network based Devanagari handwritten numbers with accuracy 91.28% recognition was implemented.

A different devnagri number recognition analysis with 900 different writers with their writing of 4 samples of each number in unconstrained way has been implemented. Then Database centre of pattern analysis and Recognition was created and that database is used for classifying the Devnagri numbers. An accuracy of 92.73% has been achieved through the composite of multiple classifiers that focuses on either local offline properties or global offline properties. Further advancements had been performed by used of other classifiers. We have also analysed the use of writer immature model to boost up accuracy of recognition. This research has been reviewed in terms of number classification, pattern classification as well as learning. In this thesis, we have compared the important results and discussed the possible directions for future research in this area.

III. IMPLEMENTATION

The major problem is to recognize the Devanagri and English numbers and it depends on the ability of the human being to identify the handwritten numbers with the little attempt. The Devanagri Script and English numbers can be recognized by using the following objective to produce the result in the main system-

- Every input belongs to a particular class for recantation for each and every pattern.
- To design the database algorithm to recognize the handwritten numbers in Devanagari and English.
- To utilize the classifier propagation and the cascaded feed forward to obtain the best.

The Optical character recognition system is created by utilizing the different Indian languages to achieve the required aim and also to make the complete study for the endless procedure. To produce a number recognition system by finding a neural network concept to obtain the 100% of result.

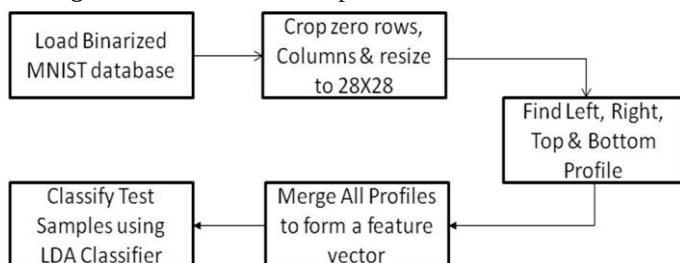


Figure 4: Proposed algorithm for MNIST database using LDA classifier

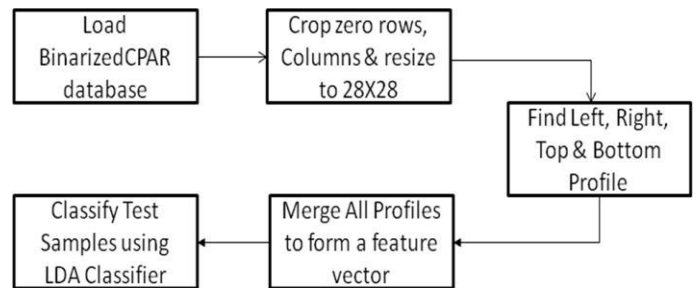


Figure 5: Proposed algorithm for CPAR database using LDA classifier

IV. RESULT

For Recognition of the English and Devenagari Image Sample we use the two databases:

1. MNIST database

This database is Contain the total 70000 handwritten sample from the different writer. In this database 60000 samples for the training purpose and 10000 samples for the testing.

2.CPAR database

- Currently have more than 80000 Numerals and 1.25 lakh Devnagari character.
- More than 5000 Hindi pangram for document recognition.
- Numerals from CPAR database are used in small scale for this project.
 - More than 1300 Numerals
 - About 120 variations of each numeral
- From the CPAR Database we use the 40000 sample for recognition of numbers.
- We use 3000 dataset for the training and 2030 dataset for the testing.

	0	1	2	3	4	5	6	7	8	9	10		%
0	1087	6	1	0	0	0	2	17	4	4	1	1122	96.88
1	1	1024	2	4	1	4	0	14	0	0	20	1070	95.7
2	2	6	1026	55	2	9	7	1	11	18	4	1141	89.92
3	2	1	63	1014	6	15	6	7	3	6	3	1126	90.05
4	0	4	4	11	1042	27	10	4	4	11	9	1126	92.54
5	0	5	10	24	50	1038	7	4	0	6	2	1146	90.58
6	0	2	10	5	11	8	1005	20	13	30	20	1124	89.41
7	10	4	3	2	4	2	7	1104	2	3	0	1141	96.76
8	3	3	8	5	1	3	7	0	1095	9	4	1138	96.22
9	0	0	14	7	5	3	20	1	2	1043	3	1098	94.99
10	3	34	4	3	3	1	13	2	6	11	688	768	89.58
									Total Samples			12000	92.97

Figure 6: Confusion matrix of CPAR database

	0	1	2	3	4	5	6	7	8	9	%acc	
0	914	0	1	0	14	3	11	0	33	4	960	93.26531
1	0	1092	3	4	14	0	5	0	17	0	1125	96.21145
2	3	0	879	38	4	14	24	3	61	1	1032	85.17442
3	11	2	48	858	2	1	3	24	38	23	1010	84.5505
4	17	5	1	0	837	0	7	0	20	95	862	83.23422
5	2	1	19	33	6	767	12	4	40	8	892	85.56655
6	17	5	1	1	14	44	957	0	0	0	958	90.50104
7	2	9	7	27	10	2	0	730	10	181	1028	75.87549
8	57	4	8	3	8	18	3	3	837	28	974	85.91429
9	16	6	0	10	15	3	0	40	32	887	1009	87.50882
	Total Samples											10000
	Average Percentage Accuracy											87.16421

Figure 7: Confusion matrix of MNIST database

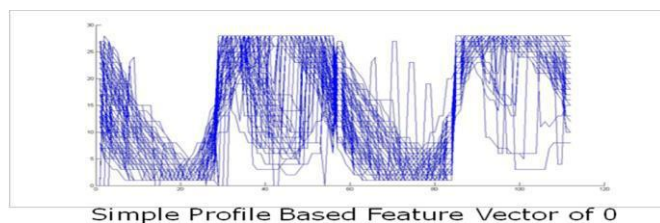


Figure 8: Feature vector of '0'

V. CONCLUSION

Offline handwritten Devnagari number reformation is inspiring task, Since of the huge expanse of these resemblances in hominid being script But also some what's yielded a unified numbers, identifying lines chiefly grounded on the records to be familiar. Mean though handwritten Devnagari statistics could be dissimilar figure and scope and script panache of dissimilar workers so for that gratitude of the Devnagari amount so difficult since instant disparity is also a big set-up in the number.

- Occasionally statistics exist overlay and combined.
- Similar operator usage the dissimilar pen and mark revolution of the news paper by of the line and

pressure equal and disposition of the individual is the symbols panache is otherwise in dissimilar kind.

- The writer can inscribe the amount in dissimilar typeface.

Ended all the details are slow here. A dataset of Devnagari amount consuming back circulation neural system for exercise and problematic. Then the consequence is not up to spot. So additional the certain novel dataset through ancient dataset before over skilled then exam the system formerly .We became the abundant extreme proved product of the previous dataset. This process every day however over and after added completely the dataset and two the upshot is abundant restored on the comprehensive dataset.

Reference:

- U. Pal, B.B. Chaudhri, Indian Script character recognition: a survey Pattern Recognition 37(2004) pp 1887-1899.
- J. J. Hull, S.N. Srihari, E. Cahen, C. L. Kuan, P.Cullen and P. Palumbo, "A black-board approach to handwritten ZIP code recognition" in Proc. United States Postal Service Advance Technology Conf. 1988.
- R. Plamondon and S. N. Srihari, Online and Offline handwritten character recognition: A comprehensive survey, IEEE Trans. On Pattern Analysis and Machine Intelligence Vol. 22 pp 62-84, 2000.
- I. K. Sethi and B. Chatterjee, "Machine Recognition of constrained Hand Printed Devengari", Pattern Recognition Vol 9, pp 69-75, 1977.
- M. Hanmandlu, O. V. Ramana Murthy, Fuzzy model based recognition of handwritten numerals, Pattern Recognition 40(2007) pp 1840-1854.
- Reena Bajaj, LipikaDey, and S. Chaudhury, "Devnagari numerals recognition by combining decision of multiple connectionist classifiers", Sadhana, Col.27, part I pp-59-72, 2002.
- U. Bhattacharya, B. B Chaudhri, R. Ghosh and M. Ghosh, "On Recognition of Handwritten Devengarinumerals", In proc. Of workshop on learning Algorithms for Pattern recognition, Sydney, pp 1-7 2005.
- R.M.K Sinha, H.N. Mahabala, Machine recognition of Devnagari script, IEEE Trans. Syst Man Cybern 9(8) (1979) 435-441.
- I.K.Sethi, B.Chatterjee, Machine recognition of handprinted Devnagari numerals, J Inst, Electron. Telecommun. Eng 22(1976) 532-535.
- V. J. Dongre, V. H. Mankar, A Review of Research on Devengari Character Reconition, Vol 12 No 2 (2010).
- M. Hanmandlu, K.R Murali Mohan, H.Kumar, Neural based handwritten character recognition, in

Proceeding of fifth IEEE Internation Conference [1999] pp
241-244.