International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Performance Analysis of KNN and SVM Classifiers Using Handwritten Kannada Vowels Recognition

Asha B. R[1], Veena Kumari H.M[2]

[1]M.Tech scholar, Department of ECE, CIT, Gubbi, Tumkur-572216, India

[2]Assistant Professor, Department of ECE, CIT, Gubbi, Tumkur-572216, India

Abstract: ***This work emphasises on the development of Kannada vowels character recognition system using KNN and SVM and performs a recognition performance analysis for both models.The main goal of this project is mainly to compare the performance of two classifiers i.e. KNN (k- nearest neighbor) and SVM (Support Vector Machine) and to obtain their performance plot. A GUI which is integrated with the binaries of KNN/LIBSVM and language rules (stores the set of valid strokes which makes a character) are used, testing is done. The classifiers performance is measured as classification accuracy like correct rate and error rate. Initially the classifiers are being trained with the training samples obtained from various users then the classifiers are tested on a test samples obtained from the users and the performance is being noted and plotted, by observing this plot can tell which classifier performance is more and better suited for the recognition application and the documented text will be converted into machine editable format. Here KNN outperforms well than the SVM. In this method the GUI is developed to show the overall recognition rates and plots. KNN gives 100% accuracy where SVM gives only 92.56% accuracy.***

Keywords: KNN, SVM, Handwritten Kannada vowels Recognition, GUI, Correct rate, Error rate, Performance plot.

1. Introduction

The field concerned with performance study mainly relays on a mix of different research methods. The term 'performance' is actually very broad and includes various performances like sporting events, social, concerts, theatrical events, political and religious events and certain kinds of language use. The study of performance is a challenging and emerging area of various research sectors. The various handwriting recognition systems available, there are two basic handwriting recognition domains distinguished primarily by nature of the input signal-online and offline. In offline system the digitised information is in the static form whereas in the online system, information is acquired during production of the handwriting using equipments such as Tablet PC which captures the trajectory of the writing tool. The information captured undergoes some filtration, pre-processing and normalisation process after which the handwriting is segmented into basic units which are usually a character or part of a character. Finally each segment is classified and labelled. In this system, examine the effectiveness of using K nearest neighbour (KNN) and Support Vector Machines (SVM) for modelling the classifier. KNN has been used for Bangla , Telugu , Tamil, Malayalam and in previous works for Assamese. Support vector machines (SVMs) have also been used for Telugu and Devnagiri scripts while compares the performance between systems developed using KNN and SVM for Telugu script. The classifiers are built individually using KNN and SVM and the recognition accuracies of both the systems are analysed for comparison. In this work, the two coordinate trace between one pen down to Pen up is taken as the basic unit and is termed as stroke. A large number of text/character classification tasks and their comparative experiments are carried out between the KNN and SVM classifiers and the performance accuracies like correct rate and error rate for individual character are tabulated and the performance curve plot is drawn for both the values of KNN and SVM classifiers. In this comparative study the algorithms KNN and SVM are compared with features obtained from the free man chain coding and number of varying training documents on a large number of text classification problems.

2. Proposed System

The block diagram of the proposed system is shown in the Figure 1. The task of recognizing and classifying the characters from an image file or test file, goes through a few processes as illustrated by the figure. The KNN AND SVM will be processing the chain-codes and the output produced would be the identified characters and their associated log-likelihoods. The task is to identify the characters of Kannada vowels and the word processor provides an interface of viewing and editing documents in Kannada. In this work, the sequence of operations carried out is as follows. A page of Kannada text is scanned. The image format used is the bmp format. The input to the system is a scanned image file in BMP format of pure Kannada document. The input image file is preprocessed using minimum preprocessing functions like binarization and thinning. The thinned image will then undergo a feature extraction process of chain-coding. The chain-coded image kept in a file, is then passed through a classifier. The document is then segmented into lines and each line into individual characters. The documented is scanned and a line in the image file is extracted. The extracted line is given as input to the Character Segmentation. Within each line the characters are segmented one by one. The extracted character that is still to be recognized is given as input to the Character Recognizing Module.

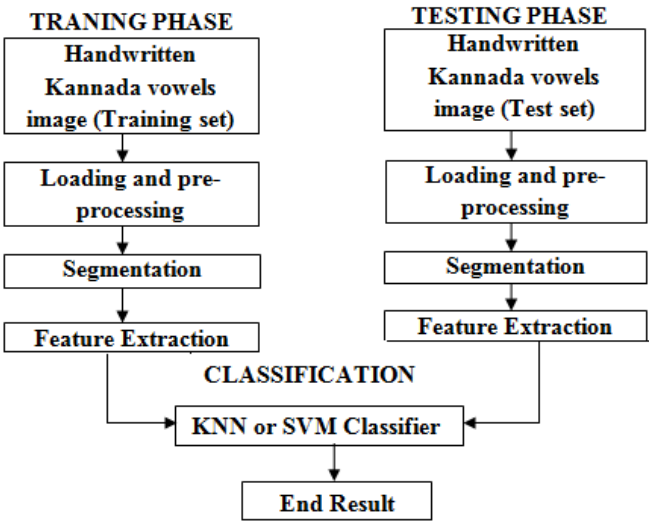

Figure 1: Block diagram of proposed system

The correct rate and error rate of correctly classified test samples are also obtained. In this project KNN will have good correct rate and all the samples are correctly classified but SVM will have some misclassification of characters so correct rate is not so good there is some amount of error rate will occur. The KNN will have good recognition rate than the SVM in this project.

Basically, the steps are divided into four major blocks they are pre-processing, segmentation, feature extraction and classification.

2.1 Pre-processing

The raw data of handwritten characters, no matter how it is acquired, will be subjected to a number of pre-processing steps to make it useable. The pre-processing phase aims to extract the relevant textual parts and prepares them for segmentation and recognition. The main objectives of pre-processing are i) noise reduction, (ii) normalization of data and (iii) compression in the amount of information to be retained. In noise reduction alone there are hundreds of available techniques which can be categorized into three major groups of filtering, morphological operations and noise modelling. Filters can be designed for smoothing, sharpening, thresholding, removing slightly textured background and contrast adjustment processes. Various morphological operations can be designed to connect broken strokes, decompose the connected strokes, smooth the contours, prune the wild points, thin the characters, and extract boundaries.

The basic steps of pre-processing of image or any document are

- Removal of noise
- Binarization
- Thinning
- Morphological operations

2.2 Segmentation

The very important part of handwritten character recognition system is the segmentation part. By performing the good segmentation process the performance of recognition can be widely increased. In an image depending on the application the important parts are being segmented and extracted for further processing steps. This phase will divide the given input image in to its constituent parts as regions or characters or as objects. The error will occur in this segmentation stage due to the touching characters which will affects the classifiers functioning. Due to inappropriate scanners the scanning resolution will be altered for the good input documents. Segmentation technique is divided into the following modules:

1. Line Segmentation
2. Letter Segmentation
3. Boundary segmentation

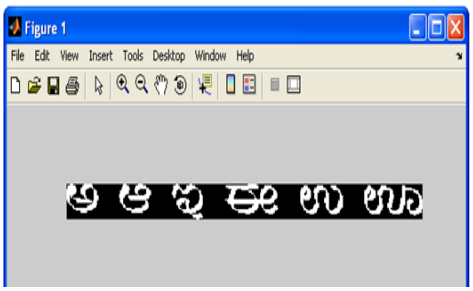

Figure 2: Line Segmentation

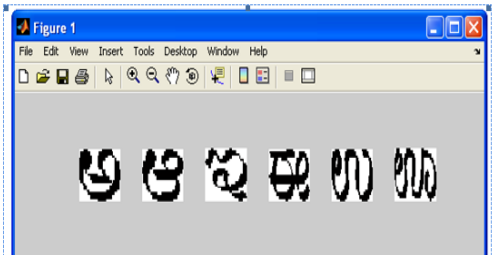

Figure 3: Character Segmentation

2.3 Feature Extraction

The feature extraction phase in a handwriting recognition system is agreed by many to hold a very important role. Feature extraction can be defined as the process of extracting distinctive information from the matrices of digitized characters. In CR applications it is important to extract those features that will enable the system to discriminate between all the character classes that exist. Coding is a category where the strokes of the character are mapped into chain-codes. One of the most popular coding schemes is Freeman's chain code, even though there are many versions of chain coding.

Chain Code Generation

The main chain coding algorithm used for the recognition system is shown in Figure 4. Here the 4- neighborhood or 8-neighborhood methods are basically used to compute the features. The one-dimensional model of the image is obtained by tracing the contour edges of the character image and representing the path by Freeman chain codes. The objective of the edge tracing would be to get chain codes that would traverse an image of a handwritten character as naturally as it would as it was written. The challenge of the chain-coding process lies very much on the way the image would be traversed and the starting point of the traversing method. A same image will produce a different chain-code if it starts

from a different point or traverses in a different direction. Consistency is required in order to minimize variations in chain-codes of the same character. The image is traversed using the connected component analysis algorithm. It then performs a traversal of the skeleton, segmenting it into strokes separated by points that have one or more than two neighbors (since these points are either endpoints or junctions where different strokes meet)obtain the chain code features.

Algorithm to obtain chain coding with 8-neighborhood is as follows:

Step 1: Firstly find out the non zero values in the starting point of the character and store that.

Step 2: secondly assign 0-7 total of eight directions to character.

Step 3: Then travel along all the 8-neighbors.

Step 4: Add this in to list of the chain code.

Step 5: Then Move on to next position one by one.

Step 6: Check whether we have reached first point or not if not then repeat the 3rd step.

A simple example of chain code extraction for one character is being shown in the Figure 4

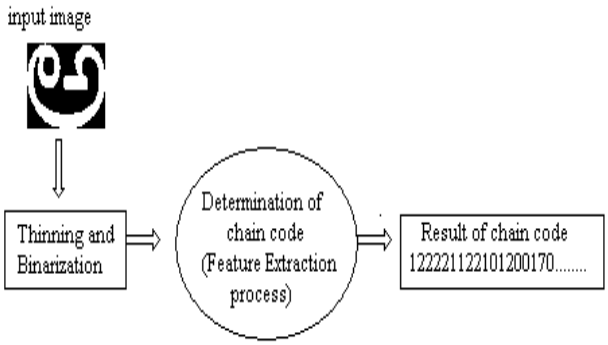

Figure 4: Chain code extraction method

2.4 Classification

A brief of the classification phase would be required to give a more meaningful account of the work mentioned in this paper. The classification phase of the system is based on the concept of classification rules developed using Mat lab (Version 7.4.0). The feature values are passed to KNN classifier once and then to SVM classifier next which generates the different set of groups which will get classified into different classes with the accurate correct rae and error rate is being displayed. The different groups provide the classified class. The classified class provides the detected character. The detected single characters can be again combined in the KNN classifier into continuous form as initial test sample but in the SVM classifier it is not possible to combine more than two characters because it is a binary classifier. In SVM there is a chance of miss classification of the characters with error rate but in KNN the error rate is absolutely zero and no miss classification of characters will occur. Finally the plot of correct rate and error rate for KNN and SVM are plotted. By looking this plot one can say which classifier works well for this application. Correct rate is 100% for KNN that is all the characters are correctly classified without misclassification.

3. Experimental Results

In this paper initially the training Kannada vowels are obtained from different peoples because different people will write the Kannada characters in different font style. Some of the writings are easily understood by everyone and some writings are very difficult to understand hence data is collected from various people. They are written in paint tool and saved as bitmap images. On these bitmap images pre-processing, segmentation and free man chain coding is computed to extract the chain code features. These chain code features are implemented in the MATLAB environment. The chain code features are then normalized and used to train the KNN classifier and SVM classifier separately. The accuracy obtained for KNN is 100% with error rate of 0% and all the characters are correctly classified and corresponding vowels are displayed. But in SVM over all accuracy is 92.56% correct rate and error rate is 7.44% is obtained due to this error rate misclassification of characters will occur is as shown in figures.

Figure 5: GUI developed for comparison of KNN and SVM for 5 Kannada vowels

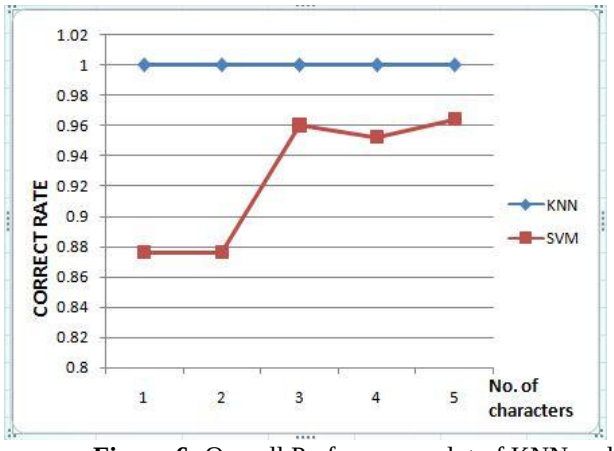

Figure 6: Overall Performance plot of KNN and SVM for correct rate

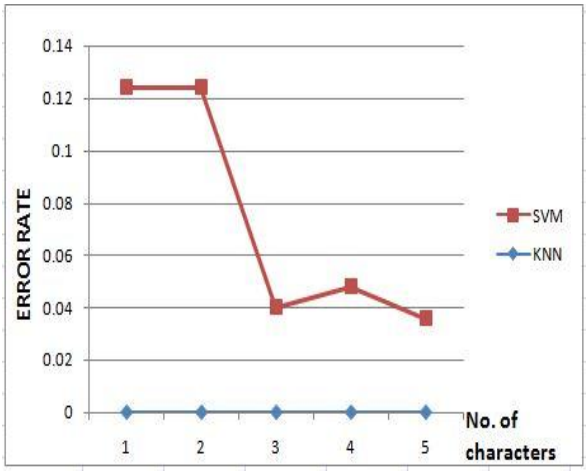

Figure 7: Overall Performance plot of KNN and SVM for error rate

Table 1: Shows the correct rate and error rates of KNN and SVM for the corresponding Kannada vowels

5 CHARACTERS	KNN		SVM	
	CORRECT RATE	ERROR RATE	CORRECT RATE	ERROR RATE
1)	100%	0%	87.6%	12.4%
2)	100%	0%	87.6%	12.4%
3)	100%	0%	96%	4%
4)	100%	0%	95.2%	4.8%
5)	100%	0%	96.4%	3.6%

4. Conclusion

An efficient and simple Handwritten Kannada vowels recognition system for performance comparison of KNN and SVM classifiers are investigated in this project. Here a new method of feature extraction algorithm is used i.e. chain coding technique. The most important part is the selection of feature extraction method which will increases the recognition accuracies like correct rate and error rate. With the use of this method the KNN and SVM are trained separately in the training phase and in the test phase the test characters are tested one by one in KNN and SVM and the corresponding correct rates and error rates are obtained. Over all the KNN classifies well than the SVM and can achieve 100% correct rate and 0% error rate for all the characters tested in KNN classifier but Characters tested with SVM classifiers will give over all correct rate of 92.56% and 7.44% error rate. The performance plot for correct and error rates of KNN and SVM are plotted as correct rate versus no of characters and error rate versus no of characters. This method will find application in postal address recognition, mail sorting, banking, historical documents conversion to machine editable format, medical etc.

References

[1] Cheng-Lin Liu, Hiroshi Sako, Hiromichi Fujisawa, " Performance evaluation of pattern classifiers for handwritten character recognition", International journal on Document Analysis and Recognition, September 2001.
[2] B.V.Dhandra, Gururaj Mukarambi, Mallikarjun Hangarge, "Handwritten Kannada Vowels And English Character Recognition System" International journal of image processing and vision science, Volume 1, Issue 1,2012.
[3] Shashikala Parameshwarppa, B.V.Dhandra, " Handwritten Kannada Characters Recognition Using Curvelet Transform", International Journal Of Computer Applications, National Conference On Digital Image And Signal Processing, DISP 2015.
[4] Sandhya Arora, Mita Nasipuri, L.Malik, D.K.Basu,"Performance Comparision Of SVM And ANN For Handwritten Devanagari Character Recognition", International Journal Of Computer Science Issues, Volume 7, Issues 3, May 2010.
[5] Mahesh Jangid Kartar Singh, Renu Dhir, Rajneesh Rani,"Performance Comparision Of Devanagari Handwritten Numerals Recognition", International Journal Of Computer Applications, Volume 22-No-1, May 2011.
[6] Baheti M.J, Kale K.V,"Gujarati Numeral Recognition: Affine Invariant Moments Approach", International Conference On Recent Trends In Engineering & Technology, March 2012.
[7] Sanghamitra Mohanty, Himadri Nandini Das Bebartta,"Performance Comparision Of SVM And KNN For Oriya Character Recognition", International Journal Of Advanced Computer Science And Applications.
[8] Deepjoy Das, Rituparna Devi, SRM Prasanna, Subhankar Ghosh, Krishna Naik,"Performance Comparision Of Online Handwriting Recognition System For Assamese Language Based On HMM And SVM Modeling.
[9] R Sanjeev Kunte and R D Sudhaker Samuel,"A Simple And Efficient Optical Character Recognition System For Basic Symbols In Printed Kannada Text", Volume 32, October 2007.
[10] Sandeep Saha, Nabarag Paul, Sayam Kumar Das, Sandip Kundu,"Optical Character Recognition Using 40 Point Feature Extraction And Artificial Neural Network", International Journal Of Advanced Research In Computer Science And Software Engineering, Volume 3, Issue 4, April 2013.
[11] Mamatha H.R, Karthik S, Srikanta Murthy K,"Classifier Fusion Method To Recognize Handwritten Kannada Numerals", 2010.
[12] Azizah Suliman, Mohd. Nasir Sulaiman, Mohamed Othman," Chain Coding And Pre Processing Stages Of Handwritten Character Image File", Electronic Journal Of Computer Science And Information Technology, Volume 2, Issue 1,2010.
[13] B.M.Sagar, Dr.Shobha G,Dr.Ramakanth Kumar P,"converting printed kannada test image file to machine

editable formate using database approach", international journal of computers, Volume 2, Issue 2,2008.
[14] Amrita Hirwani, Sandeep Gonnade,"Handwritten Character Recognition System Using Neural Network", International Journal Of Advance Research In Computer Science And Management Studies, Volume 2, Issue 2, Feb 2014.
[15] Rajbala Tokas, Aruna Bhadu," A Comparative Analysis Of Feature Extraction Techniques For Handwritten Character Recognition", International Journal Of Advanced Technology & Engineering Research.
[16] Suruchi G.Dedgaonkar, Anjali A.Chandavale, Ashok M.Sapkal,"Survey Of Methods For Character Recognition", International Journal Of Engineering And Innovative Technology, Volume 1, Issue 5, May 2012.

Author Profile

Asha B.R, M. Tech final year student of Department of Electronics. She is pursuing his Master's Degree in Electronics from VTU. He has completed his B.E from VTU.

Mrs. Veena Kumari H.M, Assistant Professor of Department of Electronics and Communication Engineering, CIT, Gubbi, Tumkur. She has completed her B.E and M.Tech from VTU.