HANDWRITTEN NUMERIC RECOGNITION USING SUPPORT VECTOR MACHINE TECHNIQUE IN MACHINE LEARNING

[1]K.V.K.SASIKANTH,[2]Y.VINEETHA,[3]P.SUPRAJA,[4]K.V.MOHAN KRISHNA,[5]SAGAR MANDAL
[1]Assistant Professor, [2,3,4,5]UG Student, Department of Computer Science and Engineering,
Godavari Institute of Engineering & Technology, Rajahmundry, AP

Abstract: Handwritten Numeral recognition plays a vital role in postal automation services. This is an important but very hard practical problem. Digit recognition is used in post offices, in banks for reading cheques, for license plate recognition, for street number recognition. The digit recognition can be divided into two groups, printed digit recognition and handwritten digit recognition. Recognition of printed digits is easier compared to the handwritten digit recognition. On the other hand, there are numerous handwriting styles for the same digit; hence more effort is required to find the accurate handwritten digit. In this project, we propose using SVM for recognition of handwritten digit. SVM is Machine Learning Technique. Support Vector Machine (SVM) is one of the most successful classifiers. Many applications use SVM for solving the classification problem, especially those for handwritten digit recognition. The SVM is used to improve classification accuracy. Our proposed algorithm will be tested on standard MNIST dataset for handwritten digit recognition. Total dataset size is of 70,000 datapoints. Among 70,000 datapoints, 60,000 datapoints are for train dataset and 10,000 datapoints are for test dataset. Each image size in 28*28 pixels**.**

Keywords: svm, mnist dataset, digit recognition, etc.

1. Introduction

This paper presents a technique developed for recognition images of handwritten digits. This is a task of paramount importance for businesses and enterprises of most diverse areas. A practical example would be automatic bank check processing. This type of business requires a solution that provides high successful rates in handwritten digit recognition in order to correctly process the values of the checks. The need for a method with low error rates is clear, since an error in interpreting the value of a check can cost to the client or to the bank high amount of money. Handwritten digit recognition is a task which presents an elevated level of difficulty due to the high variety of different writings. The constant quest for classifiers with smaller error rates is
justified by the high costs associated with misclassifying a character.

Support Vector Machine (SVM) is a supervised machine learning algorithm. SVM is one of the powerful techniques for Classification, Regression & Outlier detection with an intuitive model. Support Vector Machine was originally invented by Vapnik & Chervonenkis. At that time, the algorithm was in early stages. Drawing hyperplanes only for linear classifier was possible. Many applications use SVM for solving the classification problem, especially these for handwritten digit recognition. Later in 1992 Vapnik, Boser & Guyon suggested a way for building a non-linear classifier. SVM is widely used for classification objectives. The objective of the support vector machine algorithm is to find the best hyperplane in an N-dimensional space that classifies the data points distinctly.

2. Related Works

K. Gaurav, Bhatia P. K. [1] Et al, this paper deals with the various pre-processing techniques involved in the character recognition with different kind of images ranges from a simple handwritten form-based documents and documents containing colored and complex background and varied intensities. In this, different preprocessing techniques like skew detection and correction, image enhancement techniques of contrast stretching, binarization, noise removal techniques, normalization and segmentation, morphological processing techniques are discussed. It was concluded that using a single technique for preprocessing, we can't completely process the image. However, even after applying all the said techniques might not possible to achieve the full accuracy in a preprocessing system.

In [2], Renata F. P. Neves have proposed SVM based offline handwritten digit recognition. Authors claim that SVM outperforms the Multilayer perceptron classifier. Experiment is carried out on NIST SD19 standard dataset. Advantage of MLP is that it is able to segment non-linearly separable classes. However, MLP can easily fall into a region of local minimum, where the training will stop assuming it has achieved an optimal point in the error surface. Another hindrance is defining the best network architecture to solve the problem, considering the number of layers and the number of perceptron in each hidden layer. Because of these disadvantages, a digit recognizer using the MLP structure may not produce the desired low error rate.

A. Brakensiek, J. Rottland, A. Kosmala, J. Rigoll [3] et al, in this paper a system for off-line cursive handwriting recognition is described which is based on Hidden Markov Models (HMM) using discrete and hybrid modelling techniques. Handwriting recognition experiments using a discrete and two different hybrid approaches, which consist of a discrete and semi-continuous structure, are compared. A segmentation free approach is considered to develop the system. It is found that the recognition rate performance can be improved of a hybrid modelling technique for HMMs, which depends on a neural vector quantizer (hybrid MMI), compared to discrete and hybrid HMMs, based on tired mixture structure (hybrid - TP), which may be caused by a relatively small data set.

Salvador España-Boquera et al [4], in this paper hybrid Hidden Markov Model (HMM) model is proposed for recognizing unconstrained offline handwritten texts. In this, the structural part of the optical model has been modelled with Markov chains, and a Multilayer Perceptron is used to estimate the emission probabilities. In this paper, different techniques are applied to remove slope and slant from handwritten text and to normalize the size of text images with supervised learning methods. The key features of this recognition system were to develop a system having high accuracy in preprocessing and recognition, which are both based on ANNs.

In [5], a modified quadratic classifier-based scheme to recognize the offline handwritten numerals of six popular Indian scripts is proposed.

3. Handwritten Digit Recognition Using Support Vector Machine

Digit recognition means the recognition of digits that are either written or printed. Digit recognition is used in post offices for sorting the mail, in banks for reading checks, for license plate recognition, street number recognition, etc. The digit recognition can be categorized into two groups, printed digit recognition and handwritten digit recognition. Recognition of printed digits is easier compared to the recognition of handwritten digits because printed digits have regular shape and the difference between images of the same digit are just in the angle of view, size, color, etc. Handwritten digit recognition is difficult as there are numerous handwriting styles which mean that same digit can be written in different styles which takes more effort to find similarities between instances of same digit.

Classification is the most important part of digit recognition. Classification in the field of computer science represents prediction of class or label for an object based on its similarity with previous objects. To perform classification, we have many classifiers. One of the most used and successful classifiers is Support Vector Machine (SVM). Support Vector Machine classifies the data points and forms different classifiers of similar data points.

Let's imagine we have two tags: red and blue, and our data has two features: x and y. We want a classifier that, given a pair of (x, y) coordinates, outputs if it's either red or blue. We plot our already labeled training data on a plane as shown in Fig1.

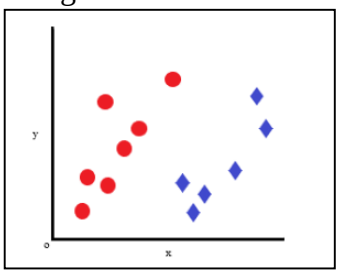


Figure1. Data (Two Classes)

A support vector machine takes these data points and outputs the hyperplane (which in two dimensions it's simply a line) that best separates the tags as shown in Fig2. This line is the decision boundary. Anything that falls to one side of it we will classify as blue, and anything that falls to the other as red.

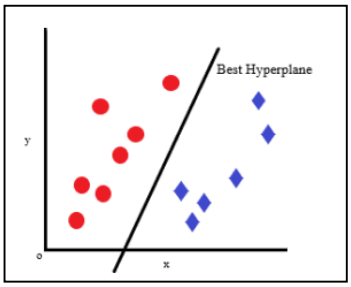


Figure2. In 2D, the Best Hyperplane is a simple line

As shown in the below Fig3, many hyperplanes are possible. But what exactly is the best hyperplane? For SVM, it's the one that maximizes the margins from both tags. In other words: the hyperplane (remember it's a line in this case) whose distance to the nearest element of each tag is the largest.

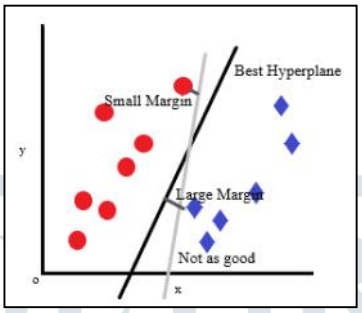


Figure3. Different hyperplanes possible

4. Working Principle of Handwritten Numeric Recognition

The handwritten recognition is divided into six phases which are image acquisition, pre-processing, segmentation, feature extraction, classification and post processing. The block diagram of the basic character recognition is shown in Fig4.

Data Collection

Here we collect data, for this we are using MNIST dataset. This dataset has 60000 training images and 10000 testing images. This data contains the digits written in different styles.

Normalization

Here we perform normalization on the data since each image have different value. For example, gray scale images have the value ranging from 0 to 288. Using normalization, we convert the gray scale image to binary image with each image having the value ranging between 0.0 and 1.0.

Feature Extraction

The main aim of feature extraction phase is to extract that pattern which is most pertinent for classification. Feature Extraction is done by Support Vector Machine (SVM). These features are used to train the system.

Classification

When input image is presented to the system, its features are extracted and given as an input to the trained classifier like support vector machine or artificial neural network. Classifiers compare the input feature with stored pattern and find out the best matching class for input.

Recognition

It refers to the processing of the output from shape recognition. Language information can increase the accuracy obtained by pure shape recognition. For handwriting input, some shape recognizers yield a single string of characters, while others yield several alternatives for each character, often with a measure of confidence for each alternative.

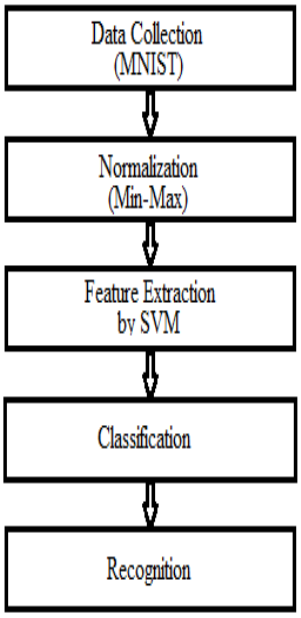
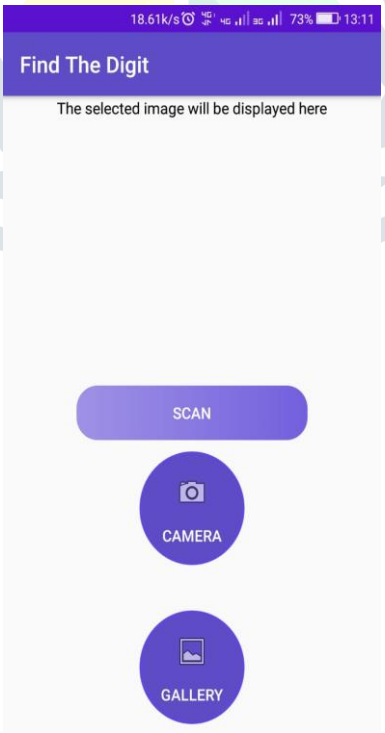


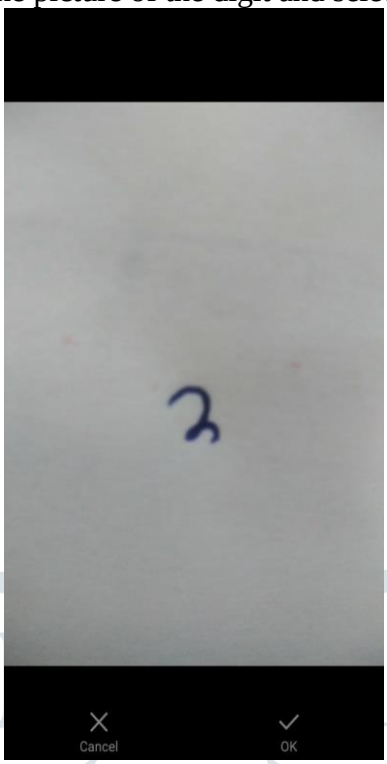
Figure4. Block Diagram of Handwritten Recognition

5. Results

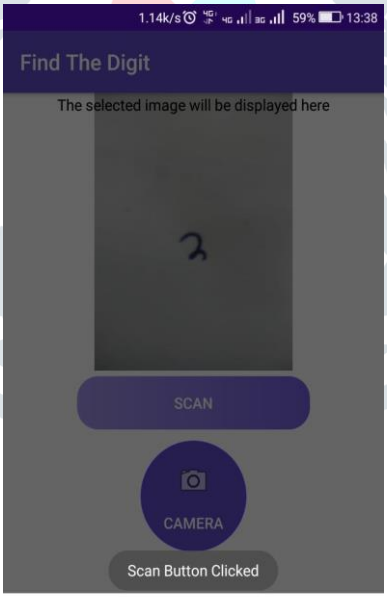
Step1: Open Mobile App



Step 2: Select Camera Option. Take the picture of the digit and select ok.

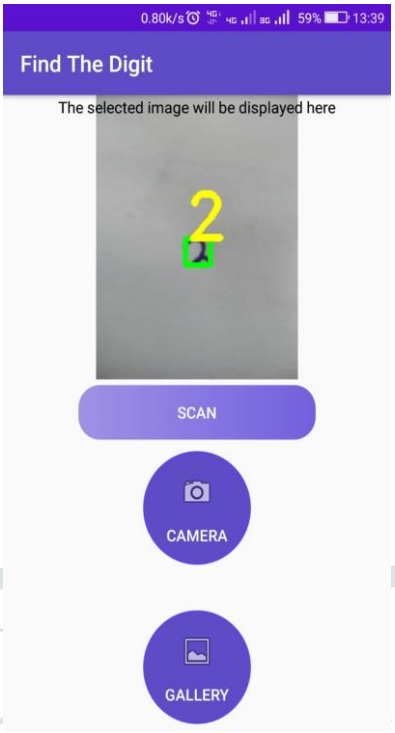


Step 3: Select Scan Button



Uploading, please wait...

Step 4: Output



Internal conversion of the input digit:

Step1:

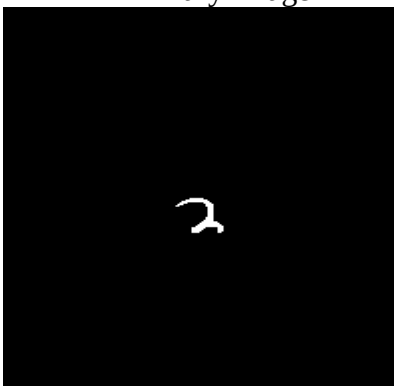
Gray Scale Image

Step2:

Image Smoothening

Step3:

Binary Image



6. Conclusion

This paper proposed hand written digit recognition using support vector machine. In the existing system, spatial information is lost which is important for classification. Using, Support Vector Machine, we overcome this problem which provides best classification. So, with Support Vector Machine Technique we can recognize the digit more accurately.

7. Future Scope

The proposed recognition system is implemented on handwritten digits taken from MNIST database. Handwritten digit recognition system can be extended to a recognition system that can also able to recognize handwritten character and handwritten symbols. The proposed work can be extended to work on broken digits. The proposed work can be further extended to improve the digit recognition accuracy.

8. References

[1] K. Gaurav and Bhatia P. K., "Analytical Review of Preprocessing Techniques for Offline Handwritten Character Recognition", 2nd International Conference on Emerging Trends in Engineering & Management, ICETEM, 2014.

[2] Renata F. P. Neves, Alberto N. G. Lopes Filho, Carlos A.B. Mello, CleberZanchettin, "A SVM Based Off-Line Handwritten Digit Recognizer", International conference on Systems, Man and Cybernetics, IEEE Xplore, pp. 510-515, 9-12 Oct 2011, Brazil.

[3] A. Brakensiek, J. Rottland, A. Kosmala and J. Rigoll, "Offline Handwriting Recognition using various Hybrid Modeling Techniques & Character N-Grams", Available at http://irs.ub.rug.nl/dbi/4357a84695495

[4] Salvador España-Boquera, Maria J. C. B., Jorge G. M. and Francisco Z.M., "Improving Offline Handwritten Text Recognition with HybridHMM/ANN Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 4, April 2011.

[5] U. Pal, T. Wakabayashi and F. Kimura, "Handwritten numeral recognition of six popular scripts," Ninth International conference on Document Analysis and Recognition ICDAR 07, Vol.2, pp.749-753, 2007.

ACKNOWLEDGEMENT

We have great pleasure in expressing our gratitude to Sri K.V.V.Satyanarayana Raju, Founder & Chairman, Chaitanya Group of Institutions, Sri K. Sasi Kiran Varma, Vice Chairman, GIET Group of Institutions, Smt. Lakshmi Raju Executive Director, GIET, for their kind support in providing us an opportunity to do research in this college.