B.Sc(Honors) 4th year Project Proposal
Department of Information and Communication Technology
Islamic University, Kushtia-7003

Project Name: Handwritten Digit Recognition with Pattern Transformations and Neural Network Averaging.

Abstract:

Recently there has been a considerable improvement in applications related with isolated handwritten digit and letter recognition supported on the use of deep and convolutional neural networks and other combinations which make use of ensemble averaging. The proposal of the present work is based on a relatively modest sized Neural Network trained with standard Back Propagation and combined with a set of input pattern transformations. Applying ensemble averaging on the trained Neural Networks gives an encouraging error rate of 0.34% measured on the MNIST dataset.

Introduction:

Some recent works on Neural Networks applied to handwritten character recognition show an interesting improvement driven by the use of new neural models. Convolutional Neural Networks and Deep Neural Networks are both rather complex structures, which allow reaching a highly respectable performance measured on the popular MNIST Database [1][2]: 0.4% [3] and 0.35% [4]. Combining these architectures with committees or integrating them with ensemble-like structures allows to further improve down to 0.27% [5] or even 0.23% [6]. Using committees with a traditional MLP displays an error rate of 0.39% [7]. Other interesting works which are based on different approaches reach an error rate of 0.40% [8] and 0.32% [9] respectively. The present work, derived from [10][11], shows a promising approach which advocates for the use of the standard Back Propagation learning algorithm with a relatively modest sized Multilayer Perceptron (MLP). A specific alternative pattern deformation is combined with other usual transformations (displacement and rotation), with an input size reduction and an additive input noise schedule. This helps to avoid local minima and stalling during the learning process.

The outcome is a creditable mean error rate of 0.46% tested on the MNIST Database. Applying ensemble averaging on a collection of trained Neural Networks helps to reduce the final error rate down to 0.34%.

Data Processing:

The MNIST Database contains 70000 digitized handwritten numerals distributed in ten different classes. The whole dataset is divided into 60000 images for training purposes, and the remaining 10000 are reserved for the test set. The graylevel values of each pixel are coded in this work in the [0,1] interval, using a 0 value for white pixels and 1 for black ones. An important point for managing a high performance in the learning process is the construction of a useful training set. The 60000 different patterns contained in the MNIST database can be seen as a rather generous set, but evidence shows that the usual learning algorithms run into serious trouble for about one hundred or more of the test set samples [10]. Therefore, some strategy is needed in order to increase the training set cardinality and variability. Usual actions comprise geometric transformations such as displacements, rotation, scaling and other distortions. Here a specific set of transformations is combined with an image size reduction and an additive input noise schedule. In this work displacements and rotations are combined with an alternative deformation procedure that yields rather good results. A problem with which the Back Propagation algorithm tackles is the relative high input dimensionality for the original 28x28 sized digits. Using downsized images helps to reduce the error rate in a small amount. Therefore, a second version of both the training and test sets are generated where each pattern is downsized through interpolation to 20x20 pixels. Each digit is randomly shifted zero, one or two pixels both in the horizontal and in the vertical axis. The final performance is rather sensible to the probability distribution of the different displacements. Finding an optimal probability distribution is a cumbersome task. An interesting possibility is to design different displacement schemas in order to reduce the error correlation of the trained networks, which in turn can induce an improvement with the ensemble averaging procedure. This is shown further on in the Experimental Results Section. The most important transformation relies on the so called deformation, which involves pulling or pushing each of the four image corner pixels in a random amount along the vertical and horizontal axis. The rest of the pixels are proportionally displaced simulating an elastic behavior. This leads to a combination of partial stretching and/or compression of the image. Fig. 1 illustrates this process. For the full sized images the displacement interval of the corner pixels is [-5, +5] (distance measured as pixels). For the 20x20 sized images, the best results are achieved with displacements in the order of [-4, +4] pixels. In parallel with the deformation, a rotation is applied around the image center selecting a random angle between -0.15 and +0.15 radians. For technical reasons, the deformation and the rotation need to be computed in an inverse way.

Submitted by:

Name: Md Sumon Mia
Roll:1418013
Reg.No:1328
Dept: Information and communication Technology.
Islamic University,Kushtia-7003,Bangladesh.