Annals of DAAAM for 2011 & Proceedings of the 22nd International DAAAM Symposium, Volume 22, No. 1, ISSN 1726-9679
ISBN 978-3-901509-83-4, Editor B. Katalinic, Published by DAAAM International, Vienna, Austria, EU, 2011
Make Harmony between Technology and Nature, and Your Mind will Fly Free as a Bird
Annals & Proceedings of DAAAM International 2011

OCR SYSTEMS BASED ON NEURAL NETWORK

PALKA, J[an] & PALKA, J[iri]

Abstract: *This paper deals with the recognition of handwritten text. It is mainly discussing improving nowadays OCR systems. In detail is this article focused on the possibilities of implementing the neocognitron network in this improvement. Next part deals with the problems of document processing, recognition of individual characters and subsequent search for whole words against the dictionary. Main goal of this work is to invent new principles in the field of processing hand written text especially focused on text with language specifics like diacritics*

Key words: *OCR, MNIST, recognition, hand-written text, neural network*

1. INTRODUCTION

The core of the most OCR systems are, algorithms based on neural networks, hidden Markov models, minimum distance classifiers or support vector. These methods usually have a decent success rate in recognition of machine-written data. The handwritten data recognition is more complicated due to the variability of individual manuscripts. The best method for recognizing handwritten characters seems the neural network. There are many different structured neural networks, so I've chosen for further testing the multilayer neural network with error back spread. As the training set suitable for testing and verification of network capacity, for recognizing and classify a pattern was chosen database MNIST.

2. MNIST DATABASE

This is the database of handwritten digits, containing 60,000 patterns for training and 10 000 patterns for testing purposes. These patterns were obtained from approximately 250 different authors. The database is suitable for testing learning algorithms for pattern recognition and classification (handwritten digits). Numbers are normalized in size and they are centered. The advantage in this database is because the patterns in the database do not require pre-processing or other treatment. Individual patterns in the database are stored as grayscale image, which is then normalized to 20x20 pixel size while maintaining aspect ratio.

3. MULTI-LAYER NEURAL NETWORK (PERCEPTRON)

Multi-layer neural network is composed of at least three layers of neurons: input, output and at least one inner layer. Neurons in the output and the inner layer have a defined threshold, which corresponds to the weight value assigned by the connection between the neuron and a fictitious neuron, whose activation is always 1 between two adjacent layers of neurons occurs of full interconnection of neurons, each neuron in lower layer is connected to all neurons higher layers.

For learning neural network is necessary to have a training set containing elements describing the problem being solved and a method that can fix these patterns in a neural network by the values of synaptic weights. Each training set pattern describes how neurons are excited in the input and output layers.

The most common adaptation algorithm of multilayer neural networks is the backpropagation method, which allows adaptation over the neural network training set. The actual algorithm consists of three stages: a forward spread of the input signal, reversing the spread of abnormality and updating the weight values on connections. During the forward spread, the signal is received by each neuron in input layer, and mediates its transfer to all neurons in the inner layer. Each neuron in the inner layer calculates its activation, and sends a signal to all neurons in output layer. Each neuron in the output layer calculates its activation, which corresponds to its actual output n-th neuron after the presentation of input pattern. (Haykin, 2009).

Experimentally it was found that it is not necessary to increase the number of hidden layers to values greater than 3. Specifying the number of neurons in the hidden layers is a classic problem. The idea that a larger number of neurons provide a greater percentage is misleading. With the increasing number of neurons in the hidden layer (or as the number of hidden layers) increases the nonlinear behavior of the network, but also a growing demand for learning (number of patterns, learning time). Too large network tends to be overfitting, etc. Too much focus on unimportant details occurring in the training set can happen, but these details may not are relevant for the resolution of problems. There is no general recommendation on how to choose the number of neurons. One of the heuristics used says that the number of neurons in the hidden layer should be twice the number of neurons in the input layer.

4. TESTING PERCEPTRON NETWORK

The Matlab algorithm was developed that performs multi-layer neural network. It is possible to choose the number of hidden layers, number of neurons in the hidden layers, types of transfer functions, adjust the methods of learning and many other parameters.

There were created two different types of neural networks. The first included a single network with number of output neurons corresponding to the number of classification groups. The second set featured a mesh of simple networks, which have only one output. The number of combinations needed was given by $\frac{n!}{2(n-2)!}$, where n is the number of characters. The proposed model was compared across a set of subnets. This method was much more time consuming to learn, especially when large numbers of characters.

A key attribute of the character recognition using neural networks is actually present on the network input (feature extraction)(Gonzales; Woods, 2002). The main task of the extraction is to obtain a set of characteristics that maximize the success of classification and which will have the minimum possible number of elements. During the solution to this problem has been progressively tested various options, covering both statistical (zoning, design, sections, zero-crossing) and structural (histograms, Fourier transform) characteristics were tested and different torque characteristics. I started from the simplest, it means using conventional 1-bit image representation, the case of representing the feature vector by 0 or 1 according to the model and then continued with gradual adaptation of the model. It has taken a number of test measurements in order to find a combination that would be most appropriate in terms of correct recognition. An important feature of the proposed network was the ability to learn and correctly classify unknown patterns. In this case, the total network learned fairly, in presenting models from training sets were successfully classified over 90%, but after the presentation of unfamiliar patterns (the test set), this percentage dropped below 50%, which was totally unsatisfactory. Further testing found that the network is incorrectly classified as small as a result of input pattern such as displacement, angle, scale, or different rotation. To eliminate such commonly occurring phenomena has been created an algorithm that removed the gradient and normalized in some way presented patterns (Cappelini, 1997).

5.CONVOLUTIONAL NEURAL NETWORK NEOCOGNITRON

It is a multi-layered hierarchical neural network. Its advantage is the ability to correctly identify not only learned images but also images that arise in partly moving, rotating or other deformations. This hierarchy is based on that the network will detect lower levels of simple signs. With each level these symptoms are more and more complex. Each level always contains three layers: the S-layer, the C-layer V-layer. The only exception is the zero level, containing only the input layer, which stores the input information. The individual layers consist of a number of areas which are under the layers of either S, C or V. The area consists of two-dimensional array of cells. Neocognitron network consists of four basic types of cells: S-cells, C-cells, V-cells and receptor cells. The basic elements, individual cells, have been working with real non-negative values. Neocognitron networks are characterized by a high density of connections between individual cells. Each cell is linked to a group of cells by attachment area to the previous layer. The attachment area is mostly sized on 3x3 or 5x5 cells. Cells input layer or previous layers of C-levels are linked to cells in the S-layer and V-layer. In addition, the each cell is linked to the S-cell layer. Finally, the S-cells are associated with C-cells each network link has a certain weight (Jain, 1988).

6. TESTING NEOCOGNITRON NETWORK

The Matlab algorithm was developed that perform convolutional network. Again, it was possible to choose the number of hidden layers, the number of cells in different layers, types of transfer functions and many other parameters. In this case of testing the MNIST database and own created database was used. Example learning curve when using a database MNIST: RMSE parameter represents mean square error, parameter CR success of process classification, which in the course of learning is changing (increasing). It can therefore be seen if the network during learning is improving or are no longer learning. For time reasons, the test sample was relatively small, about 500 patterns. Because of this the resulting success of the learned network classification can vary, so after each learning of the network the testing were made. Result of this is before mentioned percentage.

	Success rate at the end of learning	Success rate in the final test	Learning duration
500 samples	86 %	85,6 %	0,4 hour
5000 samples	91 %	91,9 %	2 hours
50000 samples	97 %	97,4 %	10 hours

Tab. 1.Convolutional network (100 neurons), 500=500 patterns from one classification group

If we assume a logarithmic dependence on number of patterns, so it could be calculated that for 100% success rate would be needed in this case about 19 000 patterns from each class. This number, however, in terms of practical implementation is inappropriate, regardless of the learning curve required to learn the network. This calculation is only indicative because of the small number of measurements. Another factor that has influence on the successful classification is the quality of the training set (Javidi, 2002). If we classify numbers for example, a very common problem is the correct resolution number "7" and "2".

7. CONCLUSION

Tests and their results suggest that an important factor for correct classification using neural networks is not only the parameters of neural networks, but also sufficiently large training set and a chosen approach to the image processing of submitted designs. Best results in terms of successful classification were achieved by using a convolutional neural network, since it is resistant to translation, rotation or scaling the input pattern. Perceptron network was significantly more sensitive to these changes. For increasing the percentage success was given appropriate preprocessing, which, however, significantly increased required time.

Tests were made on two databases (MNIST and custom database of samples, approximately 60 thousand of patterns). The greatest attention was paid to the recognition of handwritten digits, when after appropriate learning, the network can correctly classify over 97% of unknown patterns (database MNIST). The same network configuration was then applied to learn from own database (about 264 patterns), which achieved greatest success rate was 84%. In the next part, the networks were tested on the capitals of the Czech alphabet (including diacritical marks) and in combination with numbers, the results ranged between 70% and 80%. Depending on number of classification groups, the percentage success is different. From research is obvious that acquired principles are not simply usable at different languages with different character set.

8. ACKNOWLEDGEMENTS

This research was supported by the Internal grant agency (IGA/29/FAI/11/D).

9. REFERENCES

Jain, A.K. (1988). *Fundamentals of Digital image processing*, Prentice Hall, 978-0133361650, New Jersey, USA

Cappelini, V. (1997). *Time Varying Image Processing,* Elsevier Science, 978-0444823076, Florence, Italy

Sun, D. (2007). *Computer Vision Technology for Food Quality Evaluation*, Academic Press, 978-0123736420, Lincoln, USA

Gonzales, R.C.; Woods, R.E. (2002). *Digital Image processing,* Prentice Hall, 978-0201180756, Miami, USA

Javidi, B. (2002). *Image Recognition and Classification,* CRC Press, 978-0824707835, Connecticut, USA

Haykin, S. (2009). *Neural Networks and Learning Machines,* Prentice Hall, 978-0131471399, Ontario, Canada