Extraction Method of Handwritten Digit Recognition Tested on the MNIST Database

B. El Kessab¹, C. Daoui¹, B. Bouikhalene², M. Fakir² and K. Moro²

¹Laboratory of modeling and calculation - Faculty of Science and Technology
PB 523, Beni Mellal, Morocco

²Team Information Processing, Faculty of Science and Technology
PB 523, Beni Mellal, Morocco

bade10@hotmail.fr, daouic@yahoo.com, bbouikhalene@yahoo.fr, fakfad@yahoo.fr,
kamalmoro@hotmail.com

Abstract

This paper deals with an optical character recognition (OCR) system of handwritten digit, with the use of neural networks (MLP multilayer perceptron). And a method of extraction of characteristics based on the digit form, this method is tested on the MNIST handwritten isolated digit database (60000 images in learning and 10000 images in test). This work has achieved approximately 80% of success rate for MNIST database identification.

Keywords: Recognition of isolated digit, MNIST digit, neural networks, Extraction of the characteristics, characteristic zones

1. Introduction

The neural networks are widely used for the recognition of characters [1, 2, 3, 4, 5, 6, 7, and 8]. In this work we train and test a neural network classifier using MNIST database, the important step in the recognition is learning, we used descent of the gradient algorithm. In the training process the synaptic weight of the connections between the neurons is modified. The first layers contain attached binary neurons without editable connections. We have proposed the MLP as a classifier used for the recognition of the binary images (black and white pixels). The MLP contains three layers of neurons, the first layer corresponds to the retina. In technical terms it matches the input image. The second layer (hidden layer) corresponds to the extraction of characteristics subsystems. The third layer corresponds to the output system. Each neuron in this layer corresponds to one of the output classes. In the recognition task of handwritten digits, this layer contains 10 neurons corresponding to the digits 0 ... 9 (Figure 1). The original weights of the network at random connections. The weights are changed during the formation of the perceptron. The rule of change of weight corresponds to the learning algorithm.
2. Database

The MNIST database [9] of handwritten digits contains 70000 digits (60000 in learning and 10000 in test) ranging from 0 to 9. The digits have been size-normalized and centered in a fixed-size image with size 28x28. It is free and available on the Web [LeCun (1998)]. Example of the MNIST digit is shown in Figure 2.

3. Preprocessing

The pre-processing phase is an important process for recognition digit. In this work, the images of the digit are extracted from the standard MNIST database, and then noise is reduced using a median filter after thresholding process, finally we normalized the images extracted with size 28 x 28.

4. Extraction

The method used for extraction is based on the form of each digit. Dividing the image on five characteristic zones: West zone, East zone, North zone, South zone and Central zone. These characteristic areas are detected by the dilatation of the image processed in four directions.

4.1 Dilatation of Image

The dilatation is a transformation based on the form of the image. Each digit is dilated in the four directions.
4.1.1 **First Method of Dilatation:** Based on the change of pixels values (black or white) of the processed image. Algorithm is shown in Figure 3.

```
1. Begin
2. For each line
3. For each column
4. {
5. If Image (row, column) = 1
6. Image (row, column + 1) = 1
7. Increment
8. }
9. Find the Dilatation image to The East.
10. End
```

*Figure 3. Dilatation Algorithm to the East*

4.1.2 **Second Method of Dilatation:** Based on the intersection between the object of the image A (white pixels) with a structuring element B. It is defined by the following formula.

\[
\text{Dilatation} (A, B) = \{ x \in \text{Image} / B_x \cap A \neq \emptyset \}
\]

Where A is the object of the image (the white pixels), B the structuring element which is a particular set of Center x, known size and geometry (in this work is a right half). Example of the dilatation of the digit five to the East

*Figure 4. Digit Five and its Dilatation to the East*

And it is the same thing for the other directions West, North and South.

*Figure 5. Dilations of Digit Five to the West North and South*

With these two methods we found the same results for the dilatation of the image in four directions East, West, North and South. After the Dilatation of the image, we used specific intersections of dilated image in four directions for the detection of characteristic areas of each image.
4.2 Detection of the Characteristic Zones of the Image

The characteristic zones can be detected by the intersections of dilations found to the East, West, North and South. We define for each image five types of characteristic zones: East, West, North, South, and Central zone.

4.2.1. Extraction of East Characteristic Zone: A point of the image (Figure 6) belongs to the East characteristic zone (Figure 7) if and only if:

- This point does not belong to the object (the white pixels in image).
- From this point, moving in a straight line to the East, we do not cross the object.
- From this point, moving in a straight line to the south, north and west one crosses the object. The result of the extraction is illustrated in (Figure 7).

![Figure 6. Five Digit Image](image)

![Figure 7. East Characteristic Zone (EZ)](image)

4.2.2. Extraction of West Characteristic Zone: A point of the image (Figure 8) belongs to the West characteristic zone (Figure 9) if and only if:

- This point does not belong to the object (the white pixels in image).
- From this point, moving in a straight line to the West, we do not cross the object.
- From this point, moving in a straight line to the south, north and East one crosses the object. The result of the extraction is illustrated in (Figure 9).

![Figure 8. Five Digit Image](image)

![Figure 9. West Characteristic Zone (WZ)](image)

4.2.3. Extraction of South Characteristic Zone: A point of the image (Figure 10) belongs to the South characteristic zone (Figure 11) if and only if:

- This point does not belong to the object (the white pixels in image).
- From this point, moving in a straight line to the South, we do not cross the object.
- From this point, moving in a straight line to the North, East and West one crosses the object. The result of the extraction is illustrated in (Figure 11).
4.2.4. Extraction of North Characteristic Zone: A point of the image (Figure 12) belongs to the North characteristic zone (Figure 13) if and only if:

- This point does not belong to the object (the white pixels in image).
- From this point, moving in a straight line to the North, we do not cross the object.
- From this point, moving in a straight line to the East, West and South one crosses the object. The result of the extraction is illustrated in (Figure 13).

4.2.5 Extraction of the Central Characteristic Zone: A point of the image (Figure 14) belongs to the Central characteristic zone if and only if:

- This point does not belong to the limit of the object.
- From this point, moving in a straight line to the South, North, East and West we cross the object. The result of the extraction is illustrated in (Figure 15).

Each digit is characterized by these five characteristic areas. But the problem is it is that the digit two and five for example they also have same characteristic zones West and East (the number of white pixels), even if they are different, the West (East) zone of the digit five is located at the bottom (top) and West (East) zone of the digit two is located at the top (bottom).
To remedy this problem. We surround the object of digit and we divided the two zones West and East in two parts the part that is at the top and bottom part. Then we calculate the distance between the first line and the west characteristic zone, and between the last line and the west characteristic zone. If the West zone is near to the first line so this zone belongs to the part who is at the top, if this zone is near to the last line so this zone belongs to the part which is at the bottom, otherwise If the two distances are about equal we divided the zone by the central axis of the lines.

We used this method only for the two zones West and East because all digits they have these two zones, then the other digits do not have the characteristic zones north and south except the digit four containing a North characteristic zone. Instead of working with five components we find seven components.

NWZ₁, NWZ₂, NEZ₁, NEZ₂, NNZ, NSZ, NCZ

With,

NWZ₁: Number of pixels of value 1 in the West characteristic zone which is at the top
NWZ₂: Number of pixels of value 1 in the West characteristic zone which is at the bottom
NEZ₁: Number of pixels of value 1 in the East characteristic zone which is at the top
NEZ₂: Number of pixels of value 1 in the East characteristic zone which is at the bottom
NNZ: Number of pixels of value 1 in the North characteristic zone
NSZ: Number of pixels of value 1 in the South characteristic zone
NCZ: Number of pixels of value 1 in the Central characteristic zone.

Therefore the extraction vector will be defined as follows:

\[
\text{Vext} = [WZ₁, WZ₂, EZ₁, EZ₂, NZ, SZ, CZ]
\]
With, **Npixels**: Number of pixels in the image from the MNIST database size 28 x 28.

\[
\begin{align*}
WZ_1 &= NWZ_1 / (N\text{pixels}). \\
WZ_2 &= NWZ_2 / (N\text{pixels}). \\
EZ_1 &= NEZ_1 / (N\text{pixels}). \\
EZ_2 &= NEZ_2 / (N\text{pixels}). \\
NZ &= NNZ / (N\text{pixels}). \\
SZ &= NSZ / (N\text{pixels}). \\
CZ &= NCZ / (N\text{pixels}).
\end{align*}
\]

5. **Neural networks**

The neural networks [10, 11, 12, 13, 14] based on properties of the brain to build systems of calculation best able to resolve the type of problems as human beings live know resolve. They have several models one of these models is the perceptron.

5.1 **The Perceptron (Rosenblatt, 1958-1962)** [15, 16] (Figure 20) is the simplest module in neural networks. It is composed of two layers of entry and exit, entry called the retina layer and the layer of the output is called the response. These two layers are connected between them by the coefficients of weights \(W_{ij}\).

![Figure 20. The Perceptron](image)

x: Input layer.
y: Output layer
I: The number of neurons in the input layer x.
J: The number of neurons in the output layer y.
\(W_{ij}\): Weight or synaptic coefficients between the ith entry cell and the jth output cell

5.2 **The multi-layer Perceptron (MLP)** [17, 18, 19] (Figure 21) is a set of interconnected neurons which has the role of shape recognition, decision, or memory problems. It is also used in signal processing, the treatment of vision, speech, and robotics.
The MLP is composed:

- With a layer of input x: the input neurons contains I neurons.
- With a layer of output y: output neuron contains J neurons.
- Hidden layers (intermediate): located between the input layer and the layer of output, which having no contact with the outside world. Each hidden layer is composed of a number of neurons. Connections are all oriented the input layer to output layer, i.e: each neuron is connected to all the following layer neurons. The number of neurons in the network is:
  - Seven neurons in the input layer (the number seven corresponds to the values found in the vector of extraction).
  - Ten neurons in the output layer (the number ten corresponds to the numbers of digit using in recognition).

The number of neurons in the hidden layer is chosen after its three conditions:

- Equal the number of neurons in the input layer.
- Equal 75% of number of neurons in the input layer.
- Equal the square root of the product of two layers of exit and entry.

We followed these three conditions; we varied the number of neurons in layer hidden between four and nine neurons. The method used for learning is gradient back-propagation algorithm [20, 21, 22, 23, and 24].
Table 1. Details of the Neural Networks

<table>
<thead>
<tr>
<th>Layers</th>
<th>Neurons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>7</td>
</tr>
<tr>
<td>Hidden</td>
<td>7</td>
</tr>
<tr>
<td>Output</td>
<td>10</td>
</tr>
<tr>
<td>Constant of learning</td>
<td>α = 0.9</td>
</tr>
<tr>
<td>Squared error</td>
<td>$E = \frac{1}{2} \times (t - o)^2$</td>
</tr>
<tr>
<td>Activation function</td>
<td>$F(x) = \frac{1}{1+e^{-x}}$</td>
</tr>
</tbody>
</table>

5.3 Learning Algorithm
1. Initialization of the network settings.
2. Calculation of the State of the network
3. Calculation of error = fct (output - output desired)
4. Calculation of gradient back-propagation algorithm
5. Modification of synaptic weights
6. Stopping criterion, on the error, and the recognition rate of validation database (we stop the algorithm if the rate of recognition of validation it does not increase). After obtaining the optimal parameters of the network, we used a database validation consists 25% of learning entry database initially to validate the results.

5.4 Methodological Issues
- Choice of learning function?
- When stop learning?
- Validation of the results?
- Choice of structure: number of layers, cells, numbers of neurons?
- Quality of the base of learning?

6. Classification
After extracting the characteristic data of the input image, we will classify its data with neural networks (the multilayer perceptron). We must compute the coefficients of weight and desired outputs.

6.1 Experimental Results
For the classification by neural networks, we put values of characteristic vectors in the neurons of input layer, and we known the desired output, we forced network to
converge towards a final specific State (supervised learning). Each digit is characterized by a vector of seven components extraction. For the training of the network (multi-layer perceptron MLP), 1 set of 10 images, and 100 sets of 1000 images, We started by a set of ten images, to find the best parameters that maximizes network (Figure 21).

Table 2. Experimental Results

<table>
<thead>
<tr>
<th>Handwritten digit sets</th>
<th>Numbers of digits</th>
<th>Validation Database</th>
<th>Test Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Set</td>
<td>10</td>
<td>61.50</td>
<td>40.22</td>
</tr>
<tr>
<td>10 Sets</td>
<td>100</td>
<td>71.90</td>
<td>68.98</td>
</tr>
<tr>
<td>50 Sets</td>
<td>500</td>
<td>82.00</td>
<td>77.43</td>
</tr>
<tr>
<td>100 Sets</td>
<td>1000</td>
<td>82.00</td>
<td>79.28</td>
</tr>
<tr>
<td>Numbers of images for test</td>
<td>1.000</td>
<td>Images</td>
<td>60.000 Images</td>
</tr>
</tbody>
</table>

Table 3. Recognition Rate for each Digit with 100 Sets of Images

<table>
<thead>
<tr>
<th>Types of digits</th>
<th>Recognition rate of test Database</th>
<th>Error rate</th>
<th>Reject rate</th>
<th>Execution Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>86.45</td>
<td>13.55</td>
<td>0.00</td>
<td>91.64</td>
</tr>
<tr>
<td>1</td>
<td>94.39</td>
<td>05.61</td>
<td>0.00</td>
<td>112.58</td>
</tr>
<tr>
<td>2</td>
<td>88.73</td>
<td>11.27</td>
<td>0.00</td>
<td>95.63</td>
</tr>
<tr>
<td>3</td>
<td>77.02</td>
<td>22.98</td>
<td>0.00</td>
<td>76.71</td>
</tr>
<tr>
<td>4</td>
<td>76.12</td>
<td>2388</td>
<td>0.00</td>
<td>84.29</td>
</tr>
<tr>
<td>5</td>
<td>84.10</td>
<td>15.90</td>
<td>0.00</td>
<td>70.77</td>
</tr>
<tr>
<td>6</td>
<td>78.81</td>
<td>21.19</td>
<td>0.00</td>
<td>85.91</td>
</tr>
<tr>
<td>7</td>
<td>77.12</td>
<td>22.88</td>
<td>0.00</td>
<td>90.33</td>
</tr>
<tr>
<td>8</td>
<td>79.03</td>
<td>20.97</td>
<td>0.00</td>
<td>85.37</td>
</tr>
<tr>
<td>9</td>
<td>49.64</td>
<td>50.36</td>
<td>0.00</td>
<td>93.90</td>
</tr>
</tbody>
</table>

Table 4. Recognition Rate for All Digits
7. Conclusion

In this work, the neural network (the multilayer perceptron and back propagation) is proposed for the classification of the standard base MNIST isolated digit. An extraction technique is used in the phase of extraction of characteristics before implementing the classification of the digits. The recognition rate is 80.00% with a Test database containing 60,000. The method of extraction shows enough good results.

Acknowledgements

We are grateful to Yan LeCun for the MNIST database, which is a standard used by anyone in the world, who wants to test its classifier.

References

Authors

B. El Kessab

Received his Master’s degree in 2009 from Faculty of Sciences and Technology University Sultan Moulay Slimane Beni Mellal Morocco, currently working on his Ph.D. in the modeling and calculation Laboratory at Sultan Moulay Slimane University. His current research interests include pattern recognition, image analysis, document processing and automatic processing of natural languages using hidden Markov models and neural networks.

C. Daoui

Received his Ph.D degree on mathematics in 2009 from Mohamed V University Rabat Morocco. Currently is a professor in Faculty of Sciences and Technology, University Sultan Moulay Slimane Beni Mellal Morocco. His research topics are: the mathematics, operational research and pattern recognition.

B. Bouikhalene

Received his Ph.D degree on mathematics in 2001 and Master’s degree on Science of Computer and Telecommunications in 2007 from the University Ibn Tofel Kenitra. Currently is a professor in the Sultan Moulay Slimane University Beni Mellal Morocco. His research topics are: the pattern recognition, artificial intelligence and mathematics and its applications.

M. Fakir

Received his Master’s degree on Electrical engineering from Nagaoaka University of Technology in 1991 and Ph.D. degree on Electrical engineering Cadi Ayyad University Morocco. it is a team Hitachi Ltd., Japan between 1991 and 1994. Currently is a professor Faculty of Sciences and Technology, University Sultan Moulay Slimane Beni Mellal Morocco. His research topics the recognition and artificial intelligence.

K. Moro

Received his Master’s degree in 2009 from Faculty of Sciences and Technology University Sultan Moulay Slimane Beni Mellal Morocco, currently working on his Ph. D in Sultan Moulay Slimane University. His current research interests include pattern recognition, image analysis, document processing and automatic processing of natural languages using hidden Markov models and neural networks.