Rosenblatt Perceptrons for Handwritten Digit Recognition

Kussul Ernst, Baidyk Tatyana
UNAM, Centro de Instrumentos,
Cd.Universitaria A.P.70-186 C.P.04510
Mexico, D.F.
ekussul@servidor.unam.mx

Kasatkina Lora, Lukovich Vladimir
Cybernetic Center of NAS of Ukraine,
Department of Neural Network Technologies
for Information Processing,
Pr. Acad. Glushkova, 40, Kiev-187, 03680
GSP, Ukraine
kasatkin@boss.dlab.kiev.ua

Abstract

The Rosenblatt perceptron was used for handwritten digit recognition. For testing its performance MNIST database was used. 60,000 samples of handwritten digits were used for perceptron training, and 10,000 samples – for testing. The recognition rate of 99.2% was obtained. The critical parameter of Rosenblatt perceptrons is the number of neurons N in the associative neuron layer. In this work we changed the parameter N from 1,000 to 512,000. We investigated the influence of this parameter on the performance of Rosenblatt perceptron. Increasing of N from 1,000 to 512,000 involves decreasing of test errors from 5 to 8 times. It was shown that large scale Rosenblatt perceptron is comparable with the best classifiers checked on MNIST database (98.9% – 99.3%).

Index Terms — *perceptron, neural network, character recognition, handwritten digit recognition, image classifier, MNIST database, training time, recognition rate.*

Nomenclature

A – the associative layer of perceptron;
S – the sensor layer of perceptron;
R – the reaction layer of perceptron;
T – the threshold of R-layer neurons;
$W_{ij}(t)$ – the connection weight between neuron i of A-layer and neuron j of R-layer in a moment t (before reinforcement);
$W_{ij}(t+1)$ – the connection weight between neuron i of A-layer and neuron j of R-layer in a moment $(t+1)$ (after reinforcement);
E_W – the excitation of the neuron-winner;
E_C – the excitation of the nearest neuron-competitor;
W_S, H_S – the width and height of S-layer;
h, w – the width and height of the rectangle which is located in S-layer;
dx from the range $[0, W_S - w]$; dy from the range $[0, H_S - h]$ – random numbers;
$p = w/W_S = h/H_S$ – the parameters of classifier;
T_E – the superfluous excitation of neuron-winner;
MNIST database – modified US National Institute of Standards and Technology training and test sets;
RSC – Random Subspace Classifier;
LIRA – LImited Receptive Area classifier.

1 Introduction

The perceptrons of F. Rosenblatt [1] were very popular in the latest 50-s and 60-s. Simple device structure and fast training convergence made Rosenblatt perceptrons attractive for researchers. F. Rosenblatt stressed that perceptron wasn't developed to solve any practical task of pattern recognition or artificial intelligence. It was rather the model of human brain, than the applied technical device. However, it was clear that perceptron could be used in practical applications too.

Often Rosenblatt perceptron is considered as one layer perceptron [2, 3]. Three-layered Rosenblatt perceptron usually is mentioned in historical context [4]. But Frank Rosenblatt investigated mainly three-layered perceptrons. It is interesting to build new classifiers on the base of the three-layered Rosenblatt perceptron and examine if they could compete with the modern neural classifiers.

Analyzing the principal deficiencies of perceptrons F. Rosenblatt mentioned the following [1]:

1. An excessively large system may be required.
2. The learning time may be excessive.
3. The system may be excessively dependent on external evaluation during learning.
4. The generalization ability is insufficient.
5. Ability to separate essential parts in a complex sensory field (analytic ability) is insufficient.

These points at present should be revised in context of modern computer capabilities. Up to now computer cannot implement the neural network comparable with human brain, which contains many billions of neurons, but it is possible to simulate the neuron structures containing up to million neurons and even larger ones. In this case it is interesting to know how the number of associative neurons influences on Rosenblatt perceptron performance.

0-7803-7044-9/01/$10.00 ©2001 IEEE

In this study, we consider and describe several modifications of Rosenblatt perceptrons (Sections 2, 3) and experiments with them (Sections 4, 5). These experiments show that it is possible to overcome the abovementioned problems using modern hardware. In the experiments the number of associative neurons was changed from 1,000 to 512,000. The proposed perceptrons were tested on benchmark MNIST data set for handwritten digits recognition [5, 6]. The performance of the modified Rosenblatt perceptron, having 512,000 neurons, on this database is 99.2%. As computer technology improves, larger capacity recognizers become feasible and higher recognition rate becomes possible.

There are data about different classifiers performance on this database. The best classifier on this database shows 99.3% [6].

2 Rosenblatt perceptrons

3-layer Rosenblatt perceptron contains sensor layer S, associative layer A and reaction layer R (Figure 1).

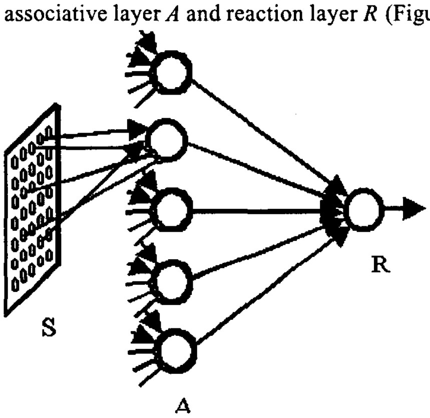

Figure 1: 3-layer Rosenblatt perceptron

Many investigations were dedicated to perceptrons with one neuron in layer R (R-layer) [1]. Such perceptron can recognize only two classes. If output of R neuron is higher than predetermined threshold T the input image belongs to class 1. If it is lower than T the input image belongs to class 2. The sensor layer S (S-layer) contains two-state $\{-1, 1\}$ elements. The element is set to 1 if it belongs to object image and set to -1, if it belongs to background.

Associative layer A (A-layer) contains neurons with 2-state $\{0, 1\}$ outputs. Inputs of these neurons are connected with outputs of S-layer neurons with no modifiable connections. Each connection marked by arrow (Figure 1) has the weight 1 (positive connection); connection marked by circle has the weight -1 (negative connection). Let the threshold of such neuron equals to number of its input connections. This neuron is active only in the case if all positive connections correspond to the object and negative connections correspond to background.

The neuron R is connected with all neurons of A-layer. The weights of these connections are changed during the perceptron training. The most popular training rule is increasing the weights between active neurons of A-layer and neuron R if the object belongs to class 1. If the object belongs to the class 2 corresponding weights are decreasing. It is known that such perceptron has fast convergence and can form nonlinear discriminating surfaces. The complexity of discriminating surface depends on the number of A-layer neurons.

3 Descriptions of the Rosenblatt perceptron modifications

We investigated several changes in perceptron structure to create the neural classifiers for handwritten digit recognition. For this purpose we used MNIST database [7]. This database contains 60,000 handwritten digits in the training set and 10,000 handwritten digits in the test set. Each black and white digit image is presented by 20*20 pixel box. The image contains gray level and is centered in 28*28 image by computing the center of mass of the pixels, and translating the image so as to position this point at the center of the 28*28 field. For our experiments we transformed gray level image to the binary image using the threshold operation.

For the first modification of simple Rosenblatt perceptron ten neurons were included into R-layer. The neuron from R-layer having the highest excitation determines the class under recognition. This rule is used always in the process of recognition. In the training process this rule should be changed. Let the neuron-winner has excitation E_w, its nearest competitor has excitation E_c. If

$$(E_w - E_c) / E_w < T_E \qquad (1)$$

where T_E is the superfluous excitation of the neuron-winner, the competitor is considered as a winner. It is the second modification of Rosenblatt perceptron.

The connections between A-layer and R-layer of Rosenblatt perceptron could be negative and positive. We used only positive connections. In this case training rule is following.

1. During recognition process we obtain excitations of R-layer neurons. The excitation of neuron R_j corresponding to correct class is decreased by the factor $(1 - T_E)$. After this the neuron having maximum excitation R_k is selected as winner.
2. If $j = k$, nothing to be done.
3. If j does not equal k,

$$W_{ij}(t+1) = W_{ij}(t) + a_i,$$

where $W_{ij}(t)$ is the weight of connection between i-neuron of A-layer and j-neuron of R-layer before reinforcement, $W_{ij}(t+1)$ is the weight after reinforcement, a_i is the output signal (0 or 1) of i-neuron of A-layer.

$$W_{ik}(t+1) = W_{ik}(t) - a_i, \quad \text{if } (W_{ik}(t) > 0),$$
$$W_{ik}(t+1) = W_{ik}(t), \quad \text{if } (W_{ik}(t) = 0),$$

where $W_{ik}(t)$ is the weight of connection between i-neuron of A-layer and k-neuron of R-layer before reinforcement, $W_{ik}(t+1)$ is the weight after reinforcement.

The perceptron with these changes we call Random Subspace Classifier (RSC) [8 – 10]. Each A-layer neuron of RSC has random connections with S-layer. To install these connections it is necessary to enumerate all elements of S-layer. Let the number of these elements equals to N_S. To determine the connection of the A-layer neuron we select the random number from the range [1, N_S]. This number determines S-layer neuron, which will be connected with the mentioned A-layer neuron. The same rule is used to determine all connections between A-layer neurons and S-layer neurons. Frank Rosenblatt proposed this rule [1]. Our experience shows that it is possible to improve the perceptron performance by modification of this rule.

We connect A-layer neuron with S-layer neurons randomly selected not from the whole S-layer but from the rectangle (h * w), which is located in S-layer (Figure 2).

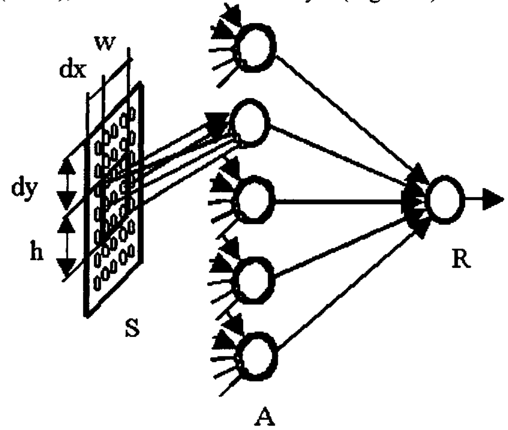

Figure 2: Random Subspace Classifier

The distances dx and dy are random numbers selected from the ranges: dx from [0, $W_S - w$] and dy from [0, $H_S - h$], where W_S, H_S stand for width and height of S-layer. We call such perceptron LImited Receptive Area (LIRA) classifier [11]. Very important features of this classifier are ratios w/W_S and h/H_S. The analysis of LIRA parameters is made in this paper.

We used also another version of Rosenblatt perceptron. In this version S-layer is composed from initial image and images of special features. Each feature image represents all short lines extracted from initial image with determined inclinations (Figure 3). We call this classifier FEATURES [11].

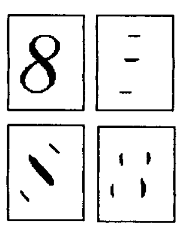

Figure 3: Example of S-layer of classifier FEATURES

The FEATURE classifier is very similar to LIRA classifier, except for the details of S-layer structure.

To set connections between A-layer neuron with S-layer neurons it is necessary to enumerate the pixels of all S-layer images and select random pixel for each new connection.

4 Handwritten digit recognition results

We carried out preliminary experiments to estimate the performance of Rosenblatt perceptrons. On the base of preliminary experiments we selected one perceptron and carried out final experiments to obtain maximal recognition rate. In preliminary experiments we changed the A-layer neuron number form 1,000 to 128,000. For perceptron LIRA we changed also the ratio $p = w/W_S = h/H_S$ from 0.2 to 0.8. We selected the parameter $T_E = 0.1$.

For each set of parameters we made 10 training cycles on MNIST training set. After that we estimated the recognition rate on MNIST test set. The recognition rates obtained in the preliminary experiments are presented in Table 1.

Table 1: The recognition rates
of different classifiers (RSC, LIRA, FEATURES)

A-layer neuron number	Error number					
	RSC	LIRA				FEATURES
		p = 0.2	p=0.4	p=0.6	p=0.8	
1000	1864	3461	1333	1297	1355	862
2000	1027	1705	772	772	827	503
4000	622	828	452	491	532	352
8000	451	482	338	335	388	262
16000	337	330	249	247	288	223
32000	270	245	205	207	246	189
64000	217	218	186	171	190	164
128000	195	207	170	168	190	146

The connections of A-layer with S-layer are formed randomly. To estimate the effect of this randomization we made three versions of each classifier, wh ich differ only by random connections between A-layer and S-layer. The recognition rates for each classifier (Q_1, Q_2, Q_3, – error numbers of classifiers with different random connections) and mean recognition (Q) rates are presented in Tables 2 – 7.

Table 2: Preliminary experiments
with classifier FEATURES

A-layer neuron number	Error number			
	Q_1	Q_2	Q_3	Q
1000	872	873	841	862
2000	466	494	548	503
4000	357	338	360	352
8000	264	271	252	262
16000	241	198	229	223
32000	168	192	206	189
64000	161	174	157	164
128000	156	147	135	146

Table 3: Preliminary experiments with classifier RSC

A-layer neuron number	Error number			
	Q_1	Q_2	Q_3	Q
1000	1663	1951	1977	1864
2000	977	1020	1085	1027
4000	626	639	602	622
8000	445	431	474	451
16000	339	313	358	337
32000	265	284	262	270
64000	212	219	221	217
128000	200	190	196	195

Table 4: Preliminary experiments with classifier LIRA ($p = 0.2$)

A-layer neuron number	Error number			
	Q_1	Q_2	Q_3	Q
1000	3381	3398	3603	3461
2000	1863	1548	1703	1705
4000	776	838	871	828
8000	450	465	530	482
16000	324	349	318	330
32000	251	240	245	245
64000	204	224	225	218
128000	204	201	215	207

Table 5: Preliminary experiments with classifier LIRA ($p = 0.4$)

A-layer neuron number	Error number			
	Q_1	Q_2	Q_3	Q
1000	1244	1357	1393	1333
2000	726	762	828	772
4000	476	438	441	452
8000	341	329	344	338
16000	247	252	248	249
32000	207	198	210	205
64000	181	192	186	186
128000	169	177	164	170

Table 6: Preliminary experiments with classifier LIRA ($p = 0.6$)

A-layer neuron number	Error number			
	Q_1	Q_2	Q_3	Q
1000	1297	1268	1328	1297
2000	785	773	757	772
4000	501	463	510	491
8000	326	331	347	335
16000	256	250	235	247
32000	213	195	212	207
64000	169	169	176	171
128000	167	169	167	168

Table 7: Preliminary experiments with classifier LIRA ($p = 0.8$)

A-layer neuron number	Error number			
	Q_1	Q_2	Q_3	Q
1000	1377	1306	1382	1355
2000	832	801	847	827
4000	544	519	533	532
8000	370	395	400	388
16000	266	295	304	288
32000	230	262	245	246
64000	187	198	184	190
128000	185	196	189	190

To achieve the best recognition rate we have selected the classifier LIRA with the following parameters. $T_E = 0.15$; $p = 0.5$. In the preliminary experiments we created 3 positive and 3 negative connections for each A-layer neuron. In the final experiments we created 3 positive and more than 3 negative connections (Table 9).

To increase the training set the distortions of input images are recommended [5, 6]. Distortion models can be used to increase the effective size of a data set without actually requiring to collect more data. We used 12 distortion variants: 8 shifts (Table 8) and 4 skewing.

Table 8: Input image distortions (shifts)

X	–2	0	2	0	–1	–1	1	1
Y	0	–2	0	2	–1	1	–1	1

The skewing angles were selected –26°, –13°, 13° and 26°. The results of the final experiments are presented in the Table 9.

Table 9: Final experiment result

Number of A-layer neurons	Number of negative connections	Error number			
		Q_1	Q_2	Q_3	Q
256000	5	79	84	82	82
512000	7	74	79	84	79

The error number 79 corresponds to 99.21% of recognition rate. Further improvements are to be expected with increasing of A-layer neuron number.

5 Discussions

When Frank Rosenblatt perceptrons are discussed they keep in the mind as a rule one-layered perceptron, which has many drawbacks. But Frank Rosenblatt paid attention 3-layered perceptrons too. He considered his perceptron as a model of a human perception. As concerned with the pattern recognition application of the perceptron Frank Rosenblatt mentioned the following deficiencies:

1. An excessively large system may be required.

2. The learning time may be excessive.
3. The system may be excessively dependent on external evaluation during learning.
4. The generalization ability is insufficient.
5. Ability to separate essential parts in a complex sensory field (analytic ability) is insufficient.

Let us analyzed these deficiencies.

1. The first one is connected with hardware possibilities. It is clear that Frank Rosenblatt and his colleagues didn't have sufficiently powerful computers to model large neural networks comparable with human brain. Modern computers also are not comparable with human brain (as concerned with number of neurons in the model). But they permit to model perceptrons having hundreds of thousands associative neurons. We investigated influence of the neuron number on the perceptron performance in the task of handwritten digit recognition. Our experiments show that it is possible to decrease the number of errors almost ten times by increasing the neuron number from 1,000 to 128,000. The number of neurons in human brain is much larger. Therefore it is possible to suppose that excellent capabilities of human brain partially depend on the huge number of its neurons. The proliferation of powerful, inexpensive computers guarantees rapid progress in this field.

2. The training time of modern classifiers for MNIST database can be several weeks [5]. In our experiments with perceptron, which contained 128,000 neurons, training time was approximately 6 hours. In final experiments training time was approximately 40 hours. It is necessary to take into account that distortions increase the training set by the factor 13. So the training time of the Rosenblatt perceptron is not excessively large.

3. At present many databases including MNIST don't contain errors of external evaluations. So the learning process doesn't depend on this factor.

4. The generalization ability of Rosenblatt perceptron is really insufficient. But very often it is possible to correct this drawback increasing data set by distortions of initial images. This technique is widely used for other classifiers too [5 – 7].

5. Ability to separate essential parts in a complex sensory field (analytic ability) is insufficient too. In principle this property could improve the recognition rate. For the simplest pattern recognition problems (as for the handwritten digit recognition) this property is not very important.

Our experiments show that 3-layered Rosenblatt perceptrons could be efficiently used for the real world pattern recognition problems. On the base of 3-layered Rosenblatt perceptron we designed the handwritten digit recognition system. This system is inferior only to Boosted LeNet-4 with distortions in recognition rate and exceeds other systems, which were tested on MNIST database.

6 Conclusions

Frank Rosenblatt proposed his perceptrons almost half century ago. It was very popular to discuss the possibilities of proposed perceptrons. At the first time the insufficient technical level didn't permit to use them in practical application. Marvin Minsky works prevented their widespread in 70-th and earlier 80-th. After that the main attention was paid to one-layer Rosenblatt perceptrons with limited abilities. Frank Rosenblatt developed 3-layered perceptrons too. We have shown that 3-layered Rosenblatt perceptron could be used as the base of modern pattern recognition systems. The performance of handwritten digit recognition system based on these principles is 99.2% and comparable with the best known recognition systems.

7 References

[1] Rosenblatt, F., Principles of Neurodynamics, Spartan books, New York, 1962.

[2] S.N. Diggavi, J.J. Shynk, I. Engel, N.J. Bershad, "On the Convergence Behavior of Rosenblatt's Perceptron Learning Algorithm", *1992 Conference Record of the 26 Asilomar Conference on Signals, Systems and Computers,* Vol. 2, 1992, pp. 852-856.

[3] S.N. Diggavi, J.J. Shynk, N.J. Bershad, "Convergence Models for Rosenblatt's Perceptron Learning Algorithm", *IEEE Transactions on Signal Processing,* Vol. 43, Issue 7, July 1995, pp. 1696-1702.

[4] G. Nagy, "Neural Networks – Then and Now", *IEEE Transactions on Neural Networks,* Vol. 2, Issue 2, March 1991, pp. 316-318.

[5] L.Bottou, C. Cortes, J.S. Denker, H. Drucker, L. Guyon, L.D. Jackel, J. LeCun, U.A. Muller, E. Sackinger, P. Simard, V. Vapnik, "Comparison of Classifier Methods: a Case Study in Handwritten Digit Recognition", *Proceedings of 12th IAPR International Conference on Pattern Recognition,* 1994, V. 2, pp.77-82.

[6] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based Learning Applied to Document Recognition", *Proceedings of the IEEE,* Vol. 86, N 11, November 1998, pp. 2278-2344.

[7] http://www.research.att.com/~yann/ocr/mnist/index.html

[8] E. Kussul, T. Baidyk, V. Lukovitch, D. Rachkovskij, "Adaptive High Performance Classifier Based on Random Threshold Neurons", *Proceedings of 12-th European Meeting on Cybernetics and Systems Research (EMCSR-94), Austria,* April 1994, in R.Trappl (ed.): Cybernetics and Systems'94, World Scientific Publishing Co.Pte.Ltd, Singapore, pp.1687-1695.

[9] E. Kussul, T. Baidyk, "Neural Random Threshold Classifier in OCR Application", *Proceedings of the Second All-Ukrainian International Conference "UkrOBRAZ'94,* Ukraine, December 1994, pp.154-157.

[10] E. Kussul, L. Kasatkina, D. Rachkovskij, D. Wunsch, "Application of Random Threshold Neural Networks for Diagnostics of Micro Machine Tool Condition", *International Joint Conference on Neural Networks,* NJ Alaska, 1998, Vol. 1, pp. 241-244.

[11] E. Kussul, L. Kasatkina, V. Lukovich, " Neural Classifier for Handwritten Symbol Recognition", *Control Systems and Computers,* Vol. 10, N 4, 1999, pp. 77-86 (in Russian).