Neural Network with Ensembles

Ernst Kussul, *Member, IEEE,* Oleksandr Makeyev, *Member, IEEE*, Tatiana Baidyk, and Daniel Calderon Reyes

***Abstract*— Pattern recognition systems usually have a relatively small number of patterns to be recognized. As a rule the number of handwritten symbols, number of phonemes or number of human faces are of the order of some dozens. But sometimes the pattern recognition task demands much more classes. For example, a continuous speech recognition system can be created on the base of syllables; a handwriting recognition system will be more efficient if the recognized units are not different letters, but triplets of letters. In these cases it is necessary to have various thousands of classes. In this paper we will consider the situation of the recognition problem that demands many thousands of classes to be recognized. For such problems we propose the use of the neural networks with ensembles. We give a short description of this type of neural network and calculate its storage capacity.**

I. INTRODUCTION

Pattern recognition systems usually have a relatively small number of patterns to be recognized. For example, the numbers of handwritten characters, phonemes, human faces rarely exceeds one hundred. But sometimes it is necessary to create a recognition system that can handle many thousands of classes to be recognized. Moreover, it is a good idea to preserve the class similarity. If two classes having a large value of similarity measure are confused, the error will be considered as not significant, and if the classes with a large difference are confused, the error is considered significant. The neural network with ensembles can represent many classes of objects. Each class in this network corresponds to an ensemble of neurons. Ensemble of neurons or neural ensemble is the subset of neurons connected with excitation connections that have a high value of synaptic weights. The neural ensemble has some properties that are very important for the creation of approximate reasoning systems:

- the neural network can store a large number of neural ensembles, much bigger than the number of neurons in the network;
- if a significant part of neurons in the neural ensemble is excited, the other neurons of the ensemble will become excited too, so the neural ensemble can be considered as an information unit similar to the concept or pattern;
- the neural ensemble has the internal structure that corresponds to the internal structure of the concept or pattern and in this sense it differs from a concept or pattern identifier that contains no information about the concept's or pattern's internal structure.

In this paper we will briefly describe the neural ensembles and will consider some ideas of using them for situation analysis. The storage capacity of ensemble neural networks will be evaluated.

II. SITUATION DESCRIPTION

Let us consider the situation presented in Fig. 1. A man wants to cross the street. He does not see the car that is moving in the street.

Fig.1. Street situation 1

The situation can be dangerous or not depending on many factors. Some of them are:

- speed of the car,
- acceleration of the car,
- speed of the man,
- distance between the car and the man,

Manuscript received January 30, 2010. This work was supported in part by the Projects PAPIIT IN110510-3, PAPIIT IN119610-3, and ICyTDF 332/2009

E. Kussul is with the Center of Applied Research and Technological Development, UNAM, Mexico city, C.P.04510, D.F., Mexico (corresponding author: phone:52-55-56228602 ext.1204; fax: 52-55-55500654; e-mail: ekussul@servidor.unam.mx).

O.Makeyev is with Department of Electrical and Computer Engineering Clarkson University, Potsdam, NY 13699, USA (makeyevo@clarkson.edu).

T. Baidyk, is with the Center of Applied Research and Technological Development, UNAM, Mexico city, C.P.04510, D.F., Mexico (tbaidyk@servidor.unam.mx).

D. Calderon Reyes is with the Center of Applied Research and Technological Development, UNAM, Mexico city, C.P.04510, D.F., Mexico (cyvirt@hotmail.com).

- direction of the man’s look.

Some minor factors also can affect the situation such as, for example:

- type of the car (compact, SUV, truck),
- man’s age,
- weather conditions.

We assume that all these factors may be recognized by a computer vision system based for example on the LIRA classifier or PCNC [1]. Next we intend to present the examples of approximate reasoning for evaluation of the situation:

Case 1.

“The speed of the car is high”, “the car is not braking”, “the man walks rapidly”, “the distance between the car and the man is small”, and “the man does not see the car”.
The conclusion is: “The situation is very dangerous”.

Case 2.

“The speed of the car is high”, “the car is braking”, “the man walks rapidly”, “the distance between the car and the man is not very small”, and “the man does not see the car”.
The conclusion is “the situation is dangerous but less dangerous than in the case 1”.

Case 3.

The situation of the case 3 is shown in Fig. 2.

Fig.2. Street situation 2

All the factors related to the left car are the same as in the case 1, but in front of the man there is another car that does not let him go rapidly. Moreover, this car attracts the attention of the man to the safety of street crossing, so this situation can be considered as less dangerous.

Each situation can be presented by one neural ensemble. Such ensemble will have complex internal structure. This structure may include not only the situation description but the conclusion, how dangerous is the situation. It is clear that the number of situations can be very large, so the storage capacity must also be very large. When the neural network contains a large number of situations, each new situation can be considered as dangerous as the most similar known situation. The most similar situation will be found not by parametric comparison of the new situation with all the known situations but in the process of associative transition from one ensemble to the most similar ensemble.

III. Neural ensembles

The neural network can contain some associative neural fields for complex ensembles formation [1]. To describe the activity of the neural field, we introduce the binary vector $V = \{v_1, \ldots, v_n\}$. If neuron a_i in the field A is active, we set $v_i = 1$, and if a_i is not active, we set $v_i = 0$. We say that vector V is encoding the activity of neural field A.

Many different methods of input information encoding for neural networks have been developed [2], [3], [4], [5]. In this paper, we will consider the so-called stochastic coding [6], [7]. These coding methods relate to distributed coding with a low level of activity (sparse coding) [8], [9], [10]. This means that in the neural network, any encoded object is represented not with one neuron, but with many neurons. The low level of activity means that the number of neurons in the neuron ensemble is much less than the total number of neurons in the neural field.

During training, two active neurons are connected with trainable connection. The weight of the connection increases, if both neurons are exited in the same time interval. If the same vectors are input to the network several times, the weights of connections forming between the active neurons will increase. Thus, in the network the sets of neurons having higher weights of connections than the mean weight of connection in the remaining network are formed. Such sets are named the neural ensembles [11].

In an ensemble, it is possible to distinguish the nucleus and the fringe [12], [13]. Neurons of the ensemble that have weights of connections higher than the mean weight value in the ensemble correspond to the nucleus. The most typical information about the presented object corresponds to the nucleus. The individual properties of the representatives of the object class correspond to the fringe. If we select different quantities of neurons with the highest activity, for example, decreasing the quantity by assigning a higher threshold of neural activity for selection, then we can achieve different levels of concretization in the description of the object. For example, if the nucleus of the formed ensemble is named “dangerous street situation,” the extended description (taking into account the neurons entering the fringe of the ensemble) can contain the information “the speed of the car is high, the car is not braking, the man walks rapidly, the distance between the car and the man is small, the man does not see the car.” Existence of the object description at different levels makes it possible to speak about existence of the neural network hierarchy similar to “class – element of a class”.

The neural ensemble is the basic information element of all hierarchical levels of the neural network. It is formed from the elements of lower hierarchical levels and can correspond to the feature, to the description of an object, to the description of a situation, to the relation between the objects, and so forth. Its internal structure reflects the structure of the corresponding object. The fact that the excitation of the part of the ensemble results in the excitation of the whole ensemble allows us to consider the ensemble as a united and indivisible element in one hierarchical level. However, when it is transferred to the lower hierarchical levels, it is divided to its components.

Assume, for example, that it is necessary to build the description of a dangerous street situation, which includes the car and the man. Each element of the situation has its own description. For example, the car can move. The properties of the movement are speed and acceleration. Let each of the named properties be coded in the form of a subset of neurons in the associative field of the lower level of neural network. Then, the neural ensemble corresponding to the car can be formed at the higher hierarchical level. The neurons that describe car's speed and acceleration at the lower level will enter into the ensemble that corresponds to the car at the higher level. To ensure that the size of the ensemble at the higher level is not too large only part of the neurons from the lower level ensembles falls into the higher level ensemble. For example, during the construction of the ensemble that corresponds to a car, only parts of each of the ensembles describing speed, acceleration, type of the car and so on are included in the ensemble "car". We term the procedure for selecting a part of the neurons to be transferred to the higher level as normalization of the neural ensemble. The ensemble is formed in such a way that, using the neurons that entered into the ensemble at the higher level, it would be possible to restore the ensembles of the lower level due to the associative restoration of the entire ensemble from its part. This type of organization of the associative-projective neural networks makes it possible to form the hierarchy of "part – whole."

The formation of neural ensembles is ensured by a change in the synaptic weights between the neurons of one associative field. The degree of an ensemble formation is characterized by the synaptic weights between the neurons that belong to an ensemble. The better the ensemble is formed, the higher the synaptic weights are between its neurons. Formation of neural ensembles in the associative field is performed during the training process. Different training algorithms can be used. The Hebbian training method works very well. In our case, we use Hebb's modified rule.

In this paper we will give a brief description of an ensemble neuron network for representation of situations. It is very important to develop neural networks that can store a large number of situations in compact form. Each situation will correspond to one neural ensemble. We developed a computer program to estimate the number of neural ensembles that can be stored in the neural network with a determined number of neurons. Experimental results show that the number of ensembles can be much larger than the number of neurons. We propose to use ensemble neural networks to store large numbers of situations and recognize them to increase human safety.

IV. ENSEMBLE NEURAL NETWORKS

Let us consider a neural network that contains N neurons. The activity (excitation) of the neurons is described by the following equations:

$$E_i = \sum_{j=1}^{N} w_{ji} y_j \ , \qquad (1)$$

$$y_i = \begin{cases} 1, & if \quad E_i > \Theta_i \\ 0, & if \quad E_i \le \Theta_i \end{cases} \qquad (2)$$

where E_i is the input excitation of i-th neuron, y_i is an output of i-th neuron, and w_{ij} is a synaptic weight of connection between j-th and i-th neurons, Θ_i is a threshold of i-th neuron. If y_i = 1, we say that the i-th neuron is active. If y_i = 0, the i-th neuron is not active.

Each neuron has synaptic connections with all other neurons and one or more external inputs [14] (Fig. 3).

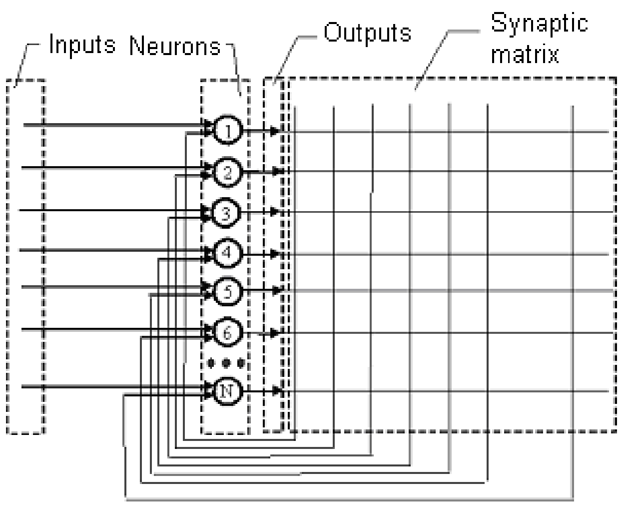

Fig. 3. Structure of ensemble neural network

External inputs are used to excite different subsets of neurons. When a subset is excited, the synaptic weights among its neurons increase. Initially all the synaptic weights are equal to 0 (Fig. 4).

We will use one of the Hebbian rules for synaptic weights change. If the neurons i and j are excited at the time t synaptic weights between them are changed in the following way:

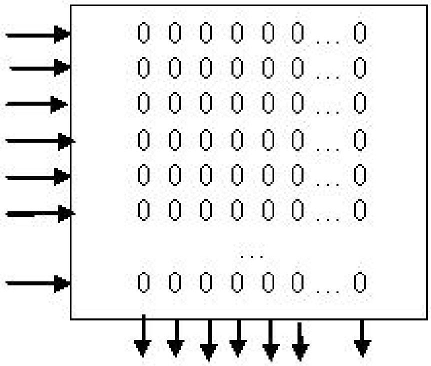

Fig. 4. Initial synaptic weight matrix

$$w_{ij}(t+1) = w_{ij}(t-1) + 1\,; \tag{3}$$
$$w_{ji}(t+1) = w_{ji}(t-1) + 1\,; \tag{4}$$

where $w_{ij}(t+1)$ and $w_{ji}(t+1)$ are synaptic weights between neurons *i* and *j* after their excitation, $w_{ij}(t-1)$ and $w_{ji}(t-1)$ are synaptic weights between neurons *i* and *j* before their excitation.

Let us excite the neurons 2, 3, 4, and 5 in the neural network presented in Fig.3. After the excitation synaptic weight matrix will be changed as it is shown in Fig. 5.

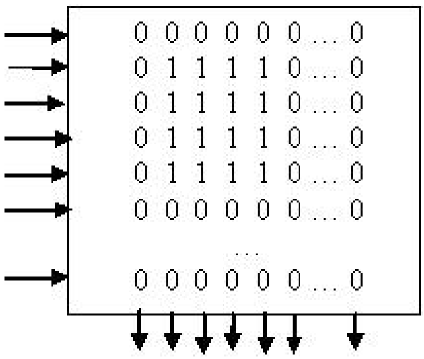

Fig. 5. Synaptic weight matrix after excitation of neurons 2, 3, 4, 5

Excited neurons 2, 3, 4 and 5 are connected to each other with excitation connections. The subset of neurons connected with excitation connections was termed neural ensemble**.** Now, if we will excite some of these neurons the connections between them will excite all other neurons of the neural ensemble. This property of neural ensemble is very important for applications, because neural ensemble composed from many different neurons functions as a single information unit. The theory of neural ensembles was proposed by D. Hebb [11].

Let us add new neural ensemble which contains the neurons 3, 4, 5, and 6. After excitation of these neurons we will have synaptic matrix presented in Fig. 6.

The figure shows that after excitation of two overlapping neural ensembles new ensemble is formed. The internal structure of the new ensemble is more complex. It contains the nucleus and the fringe. The nucleus includes the neurons that have connections with large synaptic weights, and the fringe contains the neurons with connections of smaller synaptic weights.

If the neural network contains a large number of overlapping ensembles almost all connections may have nonzero synaptic weights. In this case the problem of excitation of the whole network appears. Different authors proposed different versions of the System for Activity Control (SAC) to avoid this problem. For example, P. Milner proposed decentralized SAC based on inhibitory connections between the neurons [12], [13].

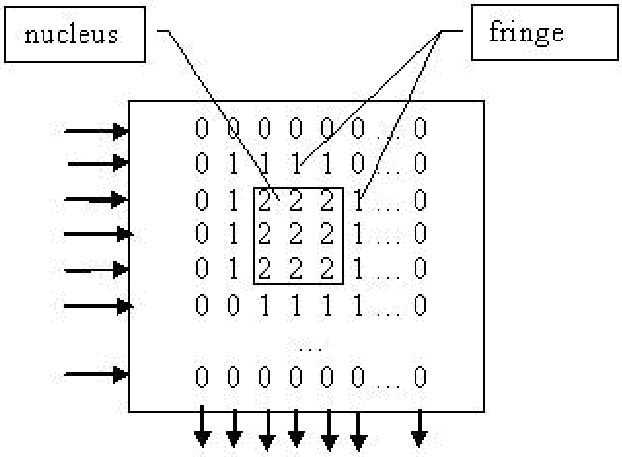

Fig. 6. Synaptic weight matrix after excitation of two ensembles

N. Amosov proposed centralized SAC for semantic networks [7] that was adapted for ensemble neural network by E. Kussul [2], A. Goltsev [15] and T. Baidyk [16]. V. Breitenberg described centralized SAC that exists in real neural structures of the brain [17].

We will use the following scheme of the SAC: when the input excitations E_i are calculated using equation (1), we select the threshold Θ, common for all the neurons, in such a way that:

$$p(\Theta) > m \tag{5}$$
$$p(\Theta + 1) \le m\,, \tag{6}$$

where $p(\Theta)$ is the number of active neurons that correspond to selected threshold Θ, $p(\Theta+1)$ is the number of active neurons that corresponds to the threshold $\Theta+1$, and m is the number of the neurons in the neuron ensemble. In our paradigm all the ensembles have approximately the same size (number of active neurons).

We will denote the state of the neuron outputs with the binary vector $Y = \{y_1, y_2, \dots y_N\}$. We will consider the vector that contains approximately *m* components that equal to "1" as the code of an ensemble. We will represent different situations in the form of ensembles. For elementary objects we will create the ensemble codes as binary masks containing m components equal to “1”, located at random positions. Complex objects will be created from elementary objects using the rules presented below. The situations will be created from the objects and their relations. For example, situation “the car is near the man” is a relation between two objects, “car” and “man”.

Let us create the object that can be described as “car moves rapidly with acceleration”. For each elementary object “car”, “moves rapidly”, and “with acceleration” we create the random binary mask: Y_{car} for “car”, $Y_{rapidly}$ for “moves rapidly” and $Y_{accelerate}$ for “with acceleration”. After this we create a binary vector Y that will correspond to “car moves rapidly with acceleration”. To create this vector we use the following procedure:

1. Create intermediate vector X:

$$X = (Y_{car}) \cup (Y_{rapidly}) \cup (Y_{accelerate}), \quad (7)$$

where $\cup$ stands for bitwise logic operation “OR”. The vector X corresponds to all the neurons included in the ensembles “car”, “moves rapidly”, and “with acceleration”, but it cannot represent the ensemble “car moves rapidly with acceleration” because it contains the number of neurons much larger than m. To make the number of neurons approximately equal to m we use the procedure termed “normalization”. There are several requirements to perform this procedure. First of all, the representation of every vector Y_i in expression (7) must be proportional to the number of ones in initial vector Y_i. It is important to have possibility to restore Y_i from normalized vector X in hierarchical structures.

Secondly, if there are two vectors X_1 and X_2 that have many common elements (that is there is a large overlap) then the normalized vectors X_1 and X_2 also must have many common elements. It is important to preserve similarity of characteristics presented by these vectors during the process of training of ensemble neural network. These demands will be fulfilled if the procedure of normalization will be realized with the following steps.

2. Create vector Z_0 :

$$Z_0 = \neg X, \quad (8)$$

where symbol “$\neg$” means bitwise inversion of vector X.

3. Create the sequence of vectors:

$$\begin{aligned} X_1 &= X \,\&\, Z_1, \\ X_2 &= X_1 \,\&\, Z_2, \\ &\dots \\ X_k &= X_{k-1} \,\&\, Z_k, \\ X_{k+1} &= X_k \,\&\, Z_{k+1}. \end{aligned} \quad (9)$$

where $\&$ stands for bitwise logic operation “AND”.

Z_{i+1} is defined using the following equation:

$$Z_{i+1} = right \quad shift(Z_i), \quad (10)$$

where the function “right shift” is cyclic shift to the right by one bit.

Let the number of active neurons in vector U equal $n(U)$ and let:

$$n(X_k) > m \quad (11)$$

$$n(X_{k+1}) \le m \quad (12)$$

In this case vector that corresponds to the ensemble “car moves rapidly with acceleration” will be:

$$Y = X_k, \quad (13)$$

where k is an index that terminates the process of normalization.

Very important property of neuron ensemble is a measure of their similarity. Let us consider four objects:

1) “car moves rapidly with acceleration”,
2) “truck moves rapidly with acceleration”,
3) “truck moves slowly with acceleration”,
4) “truck moves slowly with deceleration”.

The intersections between first object and other objects are presented in Fig. 7.

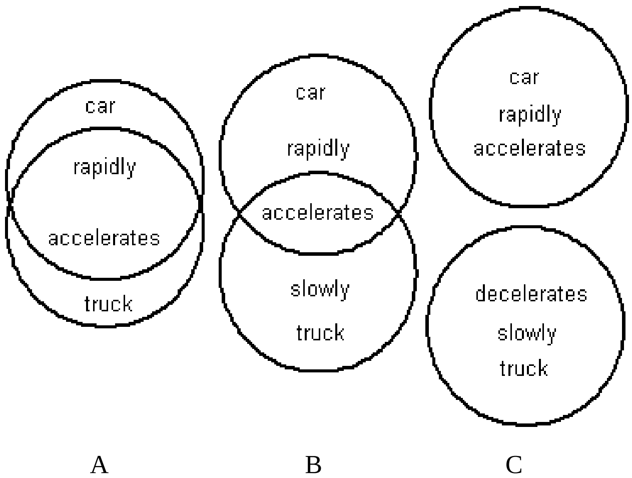

Fig. 7. Intersections between neuron ensembles

The ensembles “car moves rapidly with acceleration” and “truck moves rapidly with acceleration” have close resemblance – they have two common properties: “moves rapidly” and “with acceleration”. Figure 7A shows that corresponding ensembles have large intersection. The ensembles “car moves rapidly with acceleration” and “truck moves slowly with acceleration” have only one common property, and the Fig. 7B shows that they have smaller intersection. The ensembles “car moves rapidly with acceleration” and “truck moves slowly with deceleration” have no common properties, and the Fig. 7C shows that there is no intersection between corresponding ensembles. So it is easy to encounter the similarity of the objects by checking the intersection of corresponding ensembles.

V. STORAGE CAPACITY INVESTIGATION

To estimate the storage capacity of ensemble neural networks a simulation program was created. In this simulation the independent random ensembles that have m neurons each are created in the neural network that has N neurons. The ability of ensemble restoration was validated in the following way: in each ensemble 50% of neurons were eliminated and substituted with the neurons randomly selected from the network therefore introducing noise corresponding to 50% of the ensemble. After that the ensembles were input to the network. If in the output of the network we have the ensemble that contains 90% of neurons contained in the initial ensemble, we say that the network restored the ensemble correctly.

The results of simulation experiments are presented in Tables I and II. The maximum number of ensembles was obtained in cases where the number of retrieval errors was less than 1%.

TABLE I
RESULTS OF EXPERIMENTS FOR SIZE OF NEURAL NETWORK N=28 000 NEURONS

Number of Ensembles	Ensemble Size (m)	Retrieval Errors
12 000	15	72
16 500	24	161
150 000	32	1472
190 000	64	1228

TABLE II
RESULTS OF EXPERIMENTS FOR SIZE OF NEURAL NETWORK N=40 000 NEURONS

Number of Ensembles	Ensemble Size (m)	Retrieval Errors
290 000	96	2797
410 000	64	4125

The table shows that the number of ensembles in the network can be much larger than the number of neurons. The number of ensembles in the neural network depends on the size of each ensemble. G.Palm and A.Knoblauch [18], [19] made theoretical estimations of optimal size of the neural ensembles and obtained the following asymptotical formula:

$$m = \ln N / \ln 2 . \qquad (14)$$

Using (14) for the case of N = 28 000 (Table I) gives us m = 14.8. The closest integer is 15. From the Table I we see that for m = 15 the maximum number of correctly restored ensembles is about 12 000. At the same time for m = 64 the number of correctly restored ensembles can go to over 188 000. G. Palm did not analyze the presence of noise in the input code. Insertion of noise requires an increase in the size of the ensemble to obtain optimal storage capacity.

VI. CONCLUSIONS

Analysis of street scenarios demands large memory storage for recognition of known situations. Ensemble neural networks can offer a potential implementation of such storage. Ensemble structure makes it possible to extract situations that are similar to the new situation encountered in the street and evaluate the new situation comparing it with the similar situations stored in memory.

ACKNOWLEDGMENT

We thank Dr. Neil Bruce for help in the article preparation.

REFERENCES

[1] E. Kussul, T. Baidyk, D. Wunsch, O. Makeyev, A. Martín, "Permutation coding technique for image recognition systems", *IEEE Transactions on Neural Networks*, vol. 17/6, pp. 1566-1579, November 2006.

[2] E.M. Kussul, *Associative neuron structures*. Kiev, Naukova Dumka, 1992, p. 144 (in Russian).

[3] T. Kohonen, *Self-organizing maps*, Berlin, Springer, 2001, p. 460.

[4] A.L. Mikaelyan, "Neuro-holographic processing methods and availability of neural nanostructure development", *Information Technology and Computer Systems*, no.1, 2004, pp. 9-19.

[5] B.V. Kryzhanovsky, L.B. Litinskii, "The vector models of associative memory", in *Proc. .Neuroinformatics-2003, MIFI Session*, Moscow, Russia, 2003, pp.72-85.

[6] K. Nakano, "Associatron – a model of associative memory", *IEEE Transactions on Systems, Man, and Cybernetics*, SMC–2, no. 3, pp. 380–388, July 1972.

[7] N.M. Amosov, T.N. Baidyk, A.D. Goltsev et al., *Neurocomputers and intelligent robots*, Kiev, Nauk. Dumka,1991, p. 272 (in Russian).

[8] E. Kussul, T. Baidyk, "Structure of neural ensemble", in *Proc. The RNNS/IEEE Symposium on Neuroinformatics and Neurocomputers*, Rostov-on-Don, Russia, 1992, pp. 423–434.

[9] A. Frolov, D. Husek, I . Muraviev, "Informational efficiency of sparsely encoded Hopfield-like associative memory", in *Proc. Neuroinformatics-2003, MIFI Session*, Moscow, Russia, 2003, pp.28-70.

[10] M.V. Tsodyks, "Associative memory in asymmetric diluted network with low level of activity", *Europhys. Lett*. 7 (3), 1988, pp. 203–208.

[11] D.O. Hebb, *The organization of behaviour*, New York: Wiley, 1949, p. 319.

[12] P.M. Milner, "The cell ensemble: Mark 2", *Psychol. Rev*., vol. 64, no. 4, 1957, pp. 242–252.

[13] P. Milner, *The autonomous brain: a neural theory of attention and learning*. Lawrence Erlbaum Associates, Inc., Publishers, 1999, pp. 153.

[14] T. Baidyk, E. Kussul, "Ensemble Neural Networks", *Optical Memory & Neural Networks (Information Optics)*, vol. 18, no. 4, pp.295-303, December 2009.

[15] A.D. Goltsev, *Neural networks with the assembly organization*, Kiev, Naukova Dumka, 2005, p. 200 (In Russian)

[16] T.N. Baidyk, *Neural networks and artificial intelligence problems*. Kiev, Naukova Dumka, 2001, p. 264 (In Russian)

[17] V. Breitenberg, "Cell ensembles in the cerebral cortex". *Lect. Notes Biomath*, vol. 21, 1978, pp. 171-178.

[18] G. Palm and F.T. Sommer, "Information capacity in recurrent McCulloch-Pitts networks with sparsely coded memory states", *Network* 3, 1992, pp.177-186.

[19] A. Knoblauch, "Neural associative memory for brain modeling and information retrieval", *Information Processing Letters*, vol.95, no. 6, pp. 537-544, September 2005.