

Information Conveyed by the Neural Network Systems and its Applied Significance

Algis GARLIAUSKAS

*Institute of Mathematics and Informatics
Akademijos 4, 08663 Vilnius, Lithuania
e-mail: galgis_1@ktl.mii.lt*

Received: January 2007

Abstract. In this paper, the information theory interpreted as the neural network systems of the brain is considered for information conveying and storing. Using the probability theory and specific properties of the neural systems, some foundations are presented. The neural network model proposed and computational experiments allow us to draw a conclusion that such an approach can be applied in storing, coding, and transmission of information.

Key words: information retrieval, probability theory, neural networks, conveying information.

1. Introduction

In the neural network (NN) as a main brain structure, the information system provides the means for quantifying a moving amount of information from one neuron or their ensemble to an other neuron or, respectively, other ensemble of neurons. There are fundamental questions about information processing in the brain. The amount of information can be carried by different codes characterizing the description of neuron activities. Another important characteristic is the speed of the transferring information which defines the information rates from neuronal responses.

The information theory is based on the probabilistic approach because of randomness of many factors of sources, channels through which the information is conveyed by complex information nets to distribute knowledge among demanders. Beside the probabilistic approach, the statistical uncertainty plays a very important role. In connection with this, the entropy notation has been introduced by Hartley in 1928. The theoretical background of information theory was originated by Claude Shannon in 1948. More complete treatments of the information theory and coding can be found in many books such as Abramson (1990), Hamming (1990), and Rieke *et al.* (1993), where the information transmitting by neuronal firing is considered. The conceptions and modeling of information evaluation by neural networks have been proposed in the recent paper of Garliauskas (2004).

In Sections 2 and 3, the main conceptions of information theory based on the probability theory are considered. The conveying of information in a single neuron and retrieval of information from an associative memory are described in Sections 4, 5. The neural

network model, modeling strategy of the transmitted information, and computational experiment are discussed in Section 6.

2. Probabilities and Information

The statement of possibilities is connected with the information measure. More possibilities mean more information. The main measure in the information theory is the amount of information. Hartley (1928) was the first to try to determine the measure of information on an example of symbols from a given alphabet. He introduced an uncertainty measure as a logarithm of the number of possibilities in the statistical experiment. The probability notion was introduced by Shannon (1948) in defining the amount of information. He was one of those who, on the basis of the integrated mathematical theory, created the information theory introducing a simple information determination measure based on the logarithm of probability of possibilities with a minus sign. Thus, if we have k possibilities, then, according to Hartley $I_N = \log_2 k$ (\log_2 means the logarithm of a binary base) and according to Shannon (1948), $I_{sh} = -\log_2 P(k)$, and if $P(k) \cong 1/k$ is an equiprobable distributio

n, then Hartley's and Shannon's measures of information are identical.

Another notion of the information theory is entropy which was first introduced by Boltzmann in 1868 explaining the behaviour of gas in the thermodynamical equilibrium. The link of entropy with information was first examined by Maxwell. Boltzman established a bond between the amount of information and the considered microstates in the gas medium with equiprobable distribution, the Boltzmann constant being equal to a unit. Then entropy as a measure of uncertainty is expressed as

$$S = - \sum_{i=1}^N \sigma_i \log_2 \sigma_i, \quad (1)$$

where σ_i is the probability of the i th microstate, N is the number of microstates.

Since N is large and microstates are equiprobable, S becomes

$$S = \log_2 N, \quad (2)$$

the same as I_N or I_{sh} above.

It is important to note that the resolution with the microstate determines the amount of information provided. This occurs in many situations when is analyzing neural codings. Here the resolution could be made progressively finer that coressponds to an increase in information proportional to the logarithm of the number of possibilities.

3. Bayes Approach to the Information Measure

In general, when uncertainty takes place in the case where there is more than one possible *a posteriori* state, the average information values to be defined for each state correspond

to the *a posteriori* probability measure with regard to appropriate *a priori* state probabilities. Let us have s events with an *a priori* probability $P(s)$ and *a posteriori* probabilities are nonzero. Then the correct expression involves summing over of all events with the corresponding probabilities

$$I(s) = \sum_s P^{post}(s) \log_2 P^{post}(s)/P^{prior}(s). \quad (3)$$

We consider a discrete set independent of events, using the metaphor of a set of symbols in the alphabet S . The event now is the corresponding symbol omitted by an information source. The entropy of the information source will be as follows

$$H(S) = - \sum_{s \in S} P(s) \log_2 P(s). \quad (4)$$

If we take into account that the information source has a channel for transforming information of symbols (s) from the alphabet S to receive input codes and (s') from the alphabet S' to emit output codes, the joint probability of the channel is given by the product

$$P(s, s') = P(s)P(s') \quad (5)$$

for any pair (s, s'). When the input and output codes of symbols are independent, the channel transmits zero information, the connection of symbols will be expressed by condition probabilities $P(s'/s)$, and if two symbols are independent, $P(s'/s)$ becomes an unconditional probability $P(s')$. Then such a case has two ways of defining conditional probabilities

$$P(s, s') = P(s'/s)P(s) = P(s/s')P(s'). \quad (6)$$

Relation (6) yields

$$P(s'/s) = P(s/s')P(s')/P(s), \quad (7)$$

which is called a Bayes theorem.

According to (6) the posterior probability corresponds to the conditional probability $P(s'/s)$, and the prior probability corresponds to the unconditional probability $P(s')$. Then the information transmitted in the channel that receives the symbol s will be as follows:

$$I(s) = \sum_{s'} P(s'/s) \log_2 P(s'/s)/P(s'). \quad (8)$$

Symmetrically, the transformation under the conditional probability $P(s/s')$ and emitted symbol s' is described by

$$I(s) = \sum_s P(s/s') \log_2 P(s/s')/P(s). \quad (9)$$

Finally, the amount of mutual information can be expressed in quite symmetrical form

$$I = \sum_s \sum_{s'} P(s'/s) \log_2 P(s'/s)/P(s')$$

$$= \sum_{s,s'} P(s, s') \log_2 P(s, s')/[P(s)P(s')]. \quad (10)$$

The mutual information can also be characterized by the entropy of the source using the alphabet S minus the conditional entropy equivocation of S with respect to the new alphabet S' used by the channel, and it is written as follows

$$I = H(S) - H(S, S'). \quad (11)$$

In the transition of information, an important characteristic is the capacity of the channel. It can be defined as the maximal mutual information via all possible sets of input probabilities $P(s)$. The information transmitted via a channel can range from zero in the case of input and output independence to the lower bound of the two upper ones: the entropy of the source, and the capacity of the channel.

4. Conveying Information in a Single Neuron and Neural Networks

A single neuron can have some input data given to the synapses, which provide the firing of the neuron-soma (Fig. 1). All signals are summarized in the soma and if the integrated signal becomes higher than the neuron threshold, the output signal is generated at the next discrete time moment or not in other cases. Such a complex neuron can be presented as an information channel that conveys input data to the output of the neuron.

Let us have some learning rule with the aid of which a neuron either produces an output signal +1 as a response to all inputs, or 0, if all input signals do not achieve the neuron threshold value. Depending on the input data, the synapse weights allow us to distinguish between just two sorts of input data sets. The maximum information that the learning rule could transmit on the data is one bit. This amount of information is achieved if two sorts of data sets are equiprobable.

The complex neuron is presented with N inputs of signals $\mathbf{S}(t)$ from the set A , the signals which are delivered by N axons to the same number of synapses with the weight vector \mathbf{W} and neuron-soma with an axon, having the activation signals $\mathbf{S}(t + 1)$ at the $t + 1$ moment (Fig. 1). The sender is given a complex neuron with N inputs and uses an adaptive rule to try to find \mathbf{W} that can reproduce the input information exactly. We assume that such a rule finds \mathbf{W} if it exists. The receiver evaluates the information on N

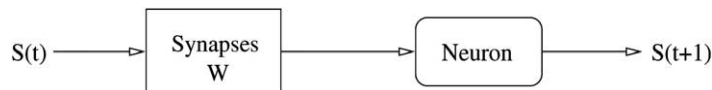


Fig. 1. The neuron as a channel of information.

input signals. The input information will be correctly reproduced only with the certain probability.

Thus, the process in the single neuron and in the neural network can be viewed as a data transmission process, in which the sender examines the data and creates a message \mathbf{W} that depends on those data. The receiver the \mathbf{W} uses to try to restore what the data of A was. This means conveying of information about the training data to a future user of that neural network.

5. Information Retrieval from an Associative Memory

The neural network is able to store information and to retrieve it in the sense of patterns in the *a posteriori* period. Such a process is similar to the determination of the capacity of channels. Let there be a vector \mathbf{r}^μ with components $\{r_i^\mu\}$, where μ is the number of a pattern, and i is the index of a neuron. The neural network reproduces a distinct firing pattern from the pattern, which was introduced into the network like as a channel. Let us denote it simply as vector \mathbf{r} . The quality of retrieval patterns can be measured by the average mutual information

$$\begin{aligned} \langle I(\mathbf{r}^\mu, \mathbf{r}) \rangle &= \sum_{\mathbf{r}^\mu, \mathbf{r}} P(\mathbf{r}^\mu, \mathbf{r}) \log_2 P(\mathbf{r}^\mu, \mathbf{r}) / P(\mathbf{r}^\mu) P(\mathbf{r}) \\ &\cong \sum_i \sum_{r^\mu, r} P(r^\mu, r) \log_2 P(r^\mu, r) / P(r^\mu) P(r). \end{aligned} \quad (12)$$

If r_i^μ and r_i are continuous distribution values, it means that, instead of the number of spikes, the firing rate of neurons will be computed by the firing train with a smoothing kernel, then one has to deal with probability densities, which we denote as $P(r)dr$. Substituting $P(r)dr$ for $P(r)$ and $P(r^\mu, r)dr$ for $P(r^\mu, r)$, the average mutual information will be as follows:

$$\langle I(r^\mu, r) \rangle_i = \int dr^\mu dr P(r^\mu, r) \log_2 P(r^\mu, r) / P(r^\mu) P(r). \quad (13)$$

If r^μ exactly determines r , the probability density will be

$$P(r^\mu, r) dr^\mu dr = P(r^\mu) \delta(r - r(r^\mu)) dr^\mu dr = P(r^\mu) dr^\mu \quad (14)$$

after losing one differential on the way.

It is necessary to distinguish the conceptual difference between the information of a fired pattern in the output of network and that stored in the network. There two cases are defined as: (a) the information selecting the correct memory pattern and (b) the information of the fired pattern under the influence of something in the outside world, for example, the sources of noises. Case (a) is well-defined where a formal model is considered, and it is incorrect where a real system is analyzed.

6. Neural Network Model and Computational Experiment

6.1. The Walsh Patterns as Stimuli

A neuron or, in general, neurons and their nets can be represented as information channels with two sets of symbols from two alphabets, as mentioned above. One set is a stimulus set and the other a response set.

Such consideration is linked with processing of information in the brain when the amount of information of the response of a neuron, or of an ensemble of neurons characterize an event from the outside world, for example, a stimulus which is shown to a testing being. As a rule, the experiment of extracellular recordings was carried out in the primary visual cortex of rhesus monkeys (Kjaer *et al.*, 1989). The responses of each neuron to several hundred or several thousand presentations were recorded, i.e., from 3 to 34 presentations of each stimulus. The Walsh patterns provide a set of stimuli of spatial structure. The Walsh pattern set is constructed by 144 patterns, 64 are basic ones and 64 are black and white elements reversed plus 8 bar-like stimuli with reversed elements (Kjaer *et al.*, 1989). The examples of such stimuli formed by NWorks software (Klimashauskas *et al.*, 1989) the adaptive resonance theory proposed by Carpenter and Grossberg (1988) are presented in Fig. 2. On the right-hand side of Fig. 2 the top pattern is one of the base stimuli, the middle one is the reversed the basic stimulus, and the bottom one is the reversed bar-like stimulus.

The authors in (Kjaer *et al.*, 1998) have additionally included the Walsh index which indicates the number of intensity of the pattern shown. The patterns with a small index

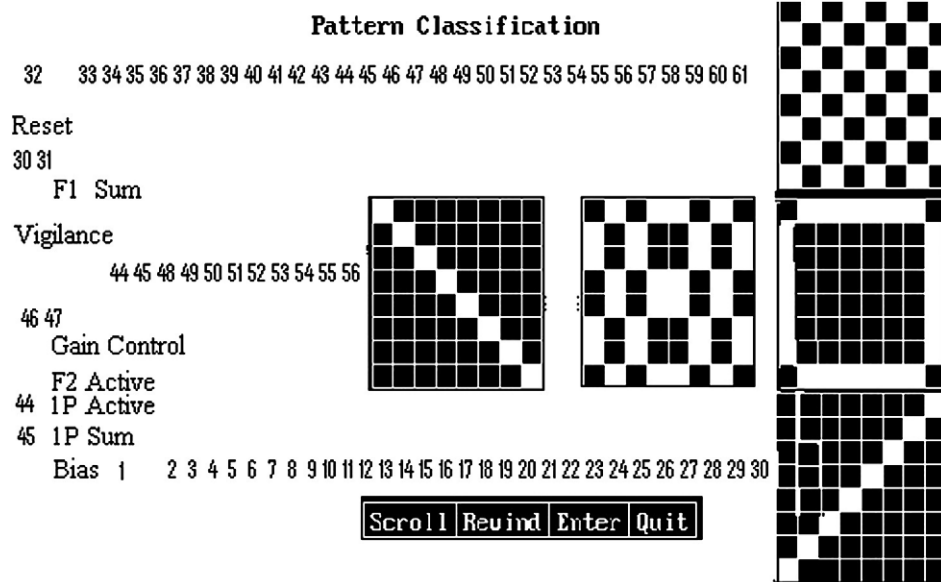


Fig. 2. Examples of the Walsh patterns.

have a coarse structure, and those with a big index have a fine structure. The Walsh index can be changed from 0 to 7.

The data of stimuli are based on the spike sequences recorded in a time window between 20 and 340 ms after a showing of stimulus. The sequences are represented by 320-bit binary strings in which gets 1 if the spike appears at some ms or 0 in other cases. In 5-ms intervals the spikes are convolved with a Gaussian kernel of standard deviation 3 ms. As a result, the 64-dimensional response vector \mathbf{r} with r_i components was defined.

6.2. Modeling Strategy of Transmitted Information by NN

We denote that a stimulus s with *a priori* probability $P(s)$ is shown from a given set. The conditional probability $P(s/r)$ represents *a posteriori* probability that is obtained by updating the responses r . The amount of information of responses is expressed like in (8):

$$I(r) = \sum_s P(s/r) \log_2 P(s/r)/P(s). \quad (15)$$

It accepts the maximum information value with $I(r) \approx -\log_2 P(s(r))$ if r unequivocally determines $s(r)$, that is, $P(s/r) = 1$ (for one stimulus, and 0 for all others) and no information because of $I(r) = \sum_s P(s) \log_2 P(s)/P(s) = 0$, and if s and r are independent, r tells us nothing about s .

The amount of information conveyed by each response according to the (first part of (10)) can be averaged over all possible responses r ,

$$\langle I(r) \rangle_r = \sum_r P(r) \sum_s P(s/r) \log_2 P(s/r)/P(s). \quad (16)$$

Denoting the joint probability as $P(s, r)$ for the pair of events s and r , and using the Bayes theorem (7), expression (16) reduces to the symmetric form (10) for the mutual information

$$\langle I(r) \rangle_r = \sum_{s,r} P(s, r) \log_2 P(s, r)/P(s)P(r), \quad (17)$$

which means that responses tell us about stimuli as much as stimuli tell us about responses. This is the general feature, which emphasizes the independence of the two variables: stimuli and neuronal responses. Beside the mutual information of (15), (16), the information conveyed about each stimulus is also interesting

$$I(s) = \sum_r P(r/s) \log_2 P(r/s)/P(r). \quad (18)$$

Since $P(r)$ is the probability distribution of responses averaged by stimuli, it means that the stimulus information measure of (18) depends not only on the stimulus s , but also

on all the other stimuli used. This emphasizes a relative nature of all information measures. More clearly, it emphasizes the relevance of measuring the information, conveyed by a given neuronal population, stimuli that are either representatives of real-life stimulus statistics, or of particular interest for the properties of the population being examined.

6.3. Neural Network Structure and the Learning Algorithm

To evaluate the information transmitted in the neural networks of the brain, the artificial neural network (ANN) learning and retrieval method was used. Despite that there exist many kinds of ANN architectures, we have chosen a simple multi-layered network and the learning algorithm based on the backpropagation procedure.

To estimate the amount of information according to (16), it is necessary first to estimate $P(s/\mathbf{r})$ corresponding to the experimental data. Suppose that $Q_s(\mathbf{r})$ is the estimate of $P(s/\mathbf{r})$. Then a log-likelihood measure can be used as a criterion. The cost function

$$E = - \sum_{s,\mu} T_s^\mu \log_2 Q_s(\mathbf{r}^\mu) \quad (19)$$

provides an appropriate fitting criterion. T_s^μ is the target function and it is equal to 1 if the response pattern μ was evoked by stimulus s , and vice versa.

The structure of ANN is shown in Fig. 3. Going from bottom to upper units, the update of the response data begins with preprocessing of data.

In this stage, the 64-dimensional response vector components have been transformed by the Karhunen-Loeve method, expressing the responses by their principal components.

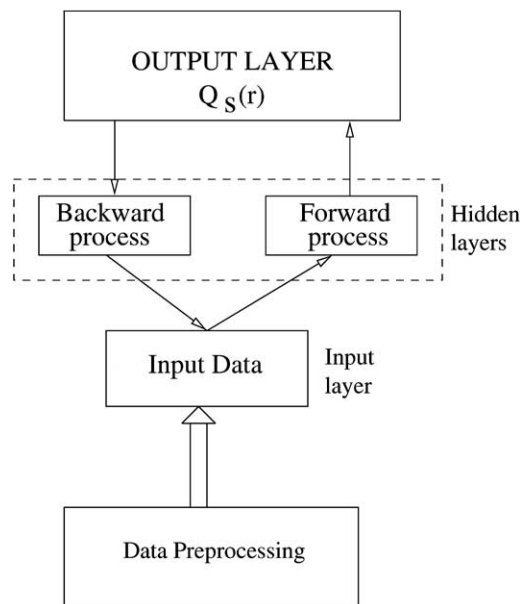


Fig. 3. The structure of ANN.

For this, the covariance matrix $\mathbf{C} = cov(\mathbf{r}_t, \mathbf{r}_{t'})$ is formed. Here \mathbf{r}_t and $\mathbf{r}_{t'}$ are the vectors with components r_t and $r_{t'}$ (t and t' equal from 1 to n), $cov(r_t, r_{t'})$ is the element of the covariance matrix and it is expressed by a known formula

$$cov(r_t, r_{t'}) = M\{(r_t - M\{r_t\})(r_{t'} - M\{r_{t'}\})\}, \quad (20)$$

where $M\{\cdot\}$ is the mathematical expectation of responses.

The eigenvalues are defined according to the matrix \mathbf{C} . The principal components are ordered according to the magnitude of their eigenvalues. The higher principal component is defined by a higher eigenvalue for the least amount of variance. Thus, now it is possible to use a smaller number of components when the modeling of responses, i.e., instead n the number N with r_j ($j = 1, 2, \dots, N$) is used, where $N < n$.

The next unit of ANN is that of input data representing the input layer. The input layer only gives the chosen principal responses without transforming into the hidden layers through then influence of synaptic weights. The neurons of layers are interconnected among the layers. There are no connections within a layer. The neurons of hidden layers and the output layer are fed by the bias neuron with a constant response $r_0 = 1$. The hidden unit activation H_i is expressed as follows:

$$H_i(\mathbf{r}) = \tanh\left(\sum_i W_{ij}r_j + b_i\right), \quad (21)$$

where W_{ij} are the weights between the pair of neurons i and j , r_j is the principal component of responses, and b_i is the bias term. The hidden units perform the two functions in the learning procedure: updating of weights in the feed-forward process, and in the feed-backward one.

The output activation is presented by the exponential normalized function

$$Q_s(\mathbf{r}) = \frac{\exp[\sum_i (W'_{si}H_i + B_s)]}{\sum_{s'} \exp[\sum_i (W'_{s'i}H_i + B_{s'})]}. \quad (22)$$

The learning iteration process is like this. The responses are given from the input layer to the hidden layers and the output layer according to the feed-forward procedure. The obtained output activation (22) value is compared with the target function (19), and if the error function value is significant or the convergence value is not achieved, the feedback procedure is performed under the gradient descent algorithm. The growth of weights

$$\Delta W'_{si} = \eta \sum_{\mu} (T_s^{\mu} - K_s(\mathbf{r}_{\mu}))H_i(\mathbf{r}_{\mu}) \quad (23)$$

allows a correction of new weights

$$W'_{si}{}^{new} = W'_{si}{}^{old} + \Delta W'_{si},$$

where η is a learning rate.

The changes of weights in the hidden layer are presented as follows

$$\Delta W_{ij} = \eta \left\{ \sum_{\mu s} [T_s^\mu - K_s(\mathbf{r}_\mu)] W_{si} \right\} (1 - H_i^2(\mathbf{r}_\mu)) r_j^\mu, \quad (24)$$

where $1 - H_i^2(\mathbf{r}_\mu)$ is a derivative of the activation function (21). Afterwards the new weight values for the input data are corrected according to

$$W_{si}^{new} = W_{si}^{old} + \Delta W_{si} + \alpha W_{si}^{prev},$$

where an additional member of the moment α and the previous weight value were included.

6.4. Results of Modeling

We have chosen the reversed task, that is when the information on each stimulus is expressed by mutual information (18). Here the probability distribution of responses is averaged by stimuli. It means, that the information measure depends on all stimuli. This variant of modeling is connected with the coding information. For modeling, we have taken 64-element Walsh patterns and 8 responses. The scheme of the ANN is shown in Fig. 4.

We tried different numbers of hidden layers and neurons in each layer and estimated the mean square errors and the amount of information after some hundred epochs of training and testing processes. We did especially not investigate the methods of division of the Walsh patterns set for training and testing parts because of limit experimental data.

First, the Walsh patterns were modified as the binary input data. The black elements of the stimulus pattern (Fig. 2) are coded as 1, and the white ones as -1. Besides the input data, there were chosen the output values, where the linear independent codes of responses were used. Analogically the test data were coded. For our example of the computational experiment, 12 training patterns and 9 testing ones were formed.

Second, we have realized many variants of modeling in search of the optimal structure of the ANN, parameters of learning, activation functions from linear to Gaussian or exponential functions in hidden and output layers, learning by the cross-validation method and without it, with different fault tolerance levels, and so on. The better results were obtained in the case of sigmoid activation functions for hidden and output layers, and by the cross-validation technique

The changes of mean square errors dependent on the number of training and testing trials are shown in Fig. 5a, b.

The training error (1) decays quicker than the testing one (2). The training error reaches the minimum fault tolerance at 175 epochs and the testing one at about 300 epochs (Fig. 5a). The training (1), testing (2), and output value (3) errors under changed learning parameters from the learning rate $\eta = 0.1$ and the moment $\alpha = 0.9$ to $\eta = 0.08$ and

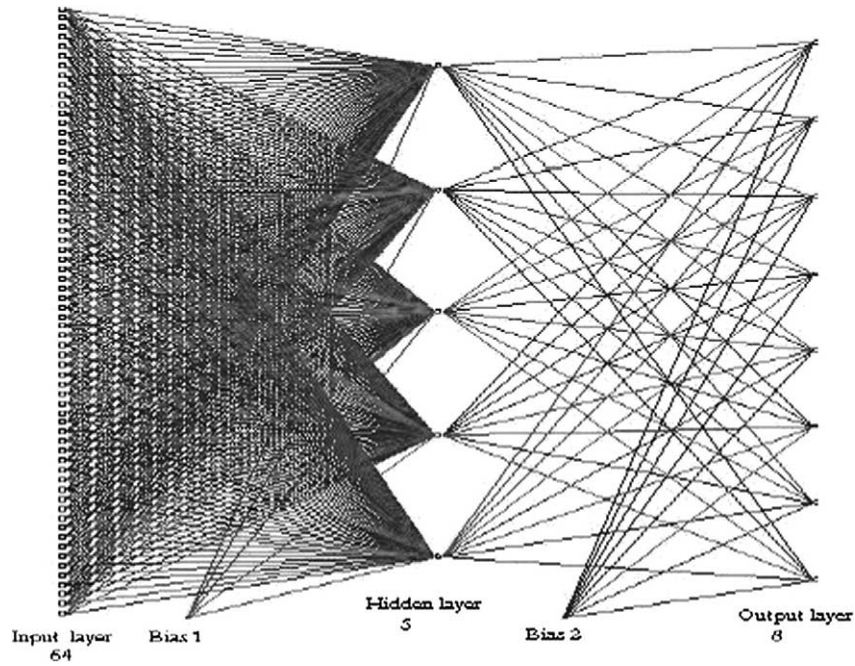


Fig. 4. The scheme of ANN.

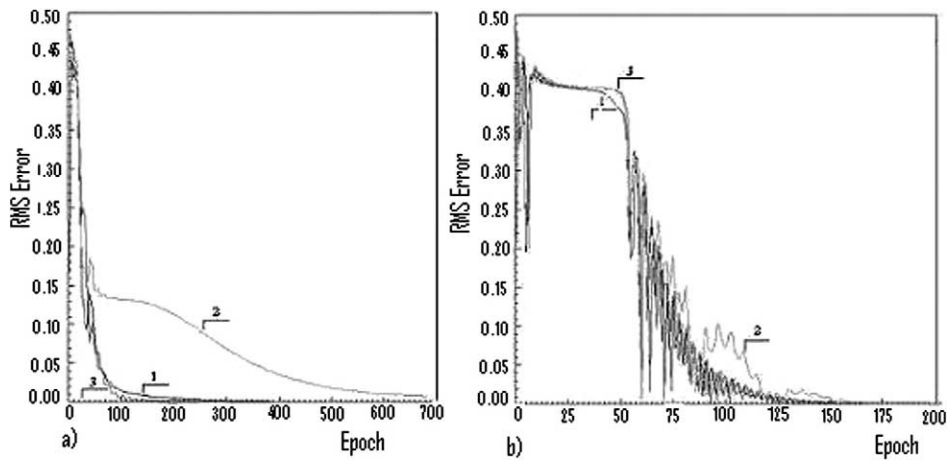


Fig. 5. The error changes versus epochs of training: (a) Training (1), testing (2), and output (3) errors dependent on epochs, (b) The errors with change of learning parameters.

the same α tend quicker (about 180 epochs) to the minimum value, but in a more complicated way, as it is shown in Fig. 5b, and minimum error reached the minimum rather early.

In Fig. 5a, the correct training patterns were 10 out of 12, that is about 83%, and the

correct testing patterns were 8 out of 9, i.e., about 89%. In Fig. 5b, almost the same result was reached, however the training and testing time was significantly shorter.

We also tried different output activation functions as mentioned above. The linear function of type $x_{out} = kx$ at $x \geq 0$ and $x_{out} = 0$ at $x < 0$ was included. The results of errors are shown in Fig. 6.

The training error reaches the minimum value, but the testing and output activation values decay up to 90 epochs and after that stabilize. The correct training patterns were those levels as above, while the correct testing patterns were about 44.5%.

Finally, the number of hidden layers as well as of neurons were evaluated from the point of view of the maximal amount of information. Some results are shown in Table 1.

Here, when only two hidden layers were taken, the number of neurons in the first layer was accepted 10. For 3 hidden layers in the first two layers, the number of neurons was taken 10 in each layer, as well as for 4 and 5 hidden layers. The trend of amount information decay (after maximum) has been observed in the increasing number of neurons

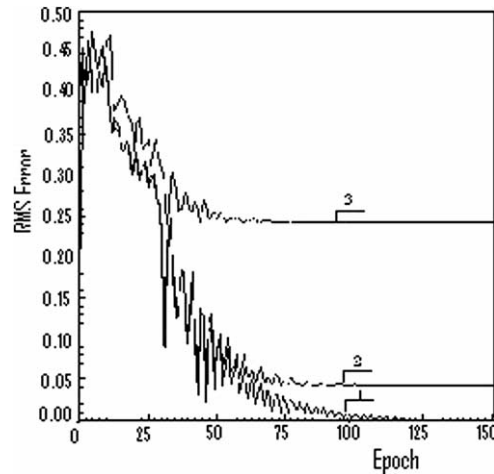


Fig. 6. The error changes versus epochs of training errors: (1) is the training, (2) is testing, and (3) is the output value errors.

Table 1
The amount of information (bits) in the output of ANN

Number of hidden layers	Number of neurons				
	1	5	10	15	20
1	2.8	2.74	0.14	0.2	0.11
2	3.4	0.23	0.17	0.12	0.15
3	5.1	0.15	0.21	0.17	0.15
4	7.0	0.22	0.13	0.14	0.11
5	6.1	0.34	0.15	0.16	0.14

in the hidden layers, but not so significant. On the basis of these limited experiments, the optimal number of neurons in hidden layers is being 1 for 64 input neurons.

Thus, the information may be coded and decoded by its storing by means of synaptic weights in a neural network as the main structure of communication in the brain.

7. Conclusions

The analysis of information theory applied to the neural network systems of the brain allows us to create some theoretical foundations, based on the probability theory for the convey of information in similar artificial systems. The neural network model proposed provides a kind of its realization, using the Walsh patterns as stimuli of the testing being. The computational experiments have showed the possibilities of coding and decoding information by implementing the artificial neural network.

Naturally, we have to take into consideration that investigations in this progressive area should be continued and under circumstances used in the practical sphere of communication systems.

References

- Abramson, N. (1963). *Information Theory and Coding*. McGraw-Hill, New York.
- Boltzmann, L. (1868). Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten. Akademie der Wissenschaften in Wien, *Mathematisch – Naturwissenschaftliche Klasse IIa*, **58**, 517–560.
- Carpenter, G. A. and S. Grossberg (1988). The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network. *Computer*, **3**, 77–88.
- Hamming, R.W. (1990). *Coding and Information Theory*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Hartley, R.V. (1928), Transmission of Information. *Bell System Tech. J.*, **7**, 535–564.
- Garliauskas, A. (2004). Conceptions and Modeling for Transmitted Information Evaluation by ANN. *Informatika*, **15**(4).
- Kjaer, T.W., J.A. Hertz and B.J. Richmond (1994). *Journal of Computational Neuroscience*, **1**, 109–139.
- Klimashauskas, C., J. Guiver and G. Pelton (1989). *Neural Computing*, **1**.
- Ricke, F., D. Warland and W. Bialek (1993). Coding efficiency and information rates in sensory neurons. *Europhysics Letters*, **22**, 151–156.
- Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Tech. J.*, 379–423, 623–656.

A. Garliauskas received his habil. dr. degree of technical sciences from the Computer Center, the Department of the USSR Academy of Sciences, Novosibirsk, USSR, in 1977. He is a head of the Laboratory of Neuroinformatics, Institute of Mathematics and Informatics and a professor in Vilnius Gediminas Technical University. His research interest includes neuroinformatics methodology, control problems and development of neural networks algorithms, chaos processes.

Informacijos perdavimas neuroninių tinklų sistemomis

Algis GARLIAUSKAS

Informacijos teorija interpretuojama kaip smegenų neuroninių tinklų sistema išsiminant ir perduodant informaciją. Teorijos pagrindimas remiasi tikimybių teorija ir specifinėmis neuroninių tinklų savybėmis. Pasiūlytas matematinis modelis ir atlikti skaitiniai eksperimentai, leidžiantys daryti išvadą apie galimą informacijos išsiminimą, kodavimą ir perdavimą.