SELF-ORGANIZING IN NEURAL NETWORKS BASED ON MEMORIZING

A. Pchelkin

Neural networks, self-organization, inductive automaton, emotional center.

1. Introduction

The goal of artificial intelligence is to develop a system that could deal with intelligent tasks in autonomous way. This goal is not achieved, but through this research many interesting algorithms have been investigated and developed, and each of them can provide some specific aspect of intelligence. For example, theorem proving [1], task solving [1], expert systems [1], pattern and speech recognition [1], clustering [1], decision tree learning [2], adaptive behavior in reinforcement learning [2], deterministic automata learning [3], error minimization in neural networks [4], and so on. However, if we would like to build an autonomous adaptive intelligent system (for example, adaptive autonomous robot), we have to combine many of developed approaches into a single learning algorithm. And usually we are not able to perform this combination in natural way, so a very sophisticated system is obtained, that could not be robust.

Artificial intelligence research has been inspired by biology because intelligence phenomena were mostly observed in human beings. However, all these approaches have a weak relation with biological systems. What does it mean? It means a potential danger – one day the selected direction of developing alternative, to natural, intelligence can find itself in impasse. One of the most biologically based approaches in artificial intelligence is Kohonen neural network [4], however it omits implementation of many important phenomena, for example, emotional control of neural network.

Our investigation is connected with a new artificial neural network, proposed by Emelyanov-Yaroslavsky in 1990 [5]. We will refer this neural network by a term "inductive automaton". The main idea was not to implement a specific functionality for each class of tasks but to define one basic task – energy consumption minimization and then to obtain all other functionalities as consequences of it, as by-products. The second criterion of developing that network was looking for a strong analogy between the artificial network and natural neural systems. Therefore, it is a good example of trying to solve two previously discussed problems.

2. The problem statement

Inductive automaton, like natural neural networks, has emotional center that controls level of activity in network and excitatory links formation. Our opinion is that the explanation of emotional center function, presented by Emelyanov-Yaroslavsky, is not complete. During his investigation the author is trying to understand biological logic of natural neural networks, he investigates different mechanisms observed in natural networks but he is not trying to answer a question: why all these complicated mechanisms are needed? The question arises: may be it is possible to use more simple and transparent mechanisms that provide the same intelligence with a little lost of functionality? These questions are not answered in his monograph about inductive automaton. Therefore, there is a need for additional investigation.

Our aim is to develop complete and transparent understanding of complex dynamic biological processes described in the monograph about inductive automaton. But in this paper we are going to *present our understanding only of emotional center function*. The goal is to show that not only biological neural networks need it but also abstract artificial neural networks need an artificial emotional center. We will be trying to find clear explanation of phenomena and mechanisms that provide intelligence as a consequence.

Researches in artificial intelligence can be divided into two groups: biological and informational approaches. The pure biological approach means studying phenomena in biological systems and trying to implement the same. In contrast, the pure informational approach means not studying biology but investigating intelligent system as abstract mathematical system from information processing point of view. The author of inductive automaton classifies his approach as a pure biological. In contrast, our position is that the most constructive way is to combine these two approaches. We need to study biology to obtain new ideas but we also need to study these ideas from information processing point of view, if we are going to implement them as precise learning algorithms. In any case, the natural neural network must have two basics of explanations – biological and informational.

3. Biological concept of inductive automaton

Here, we are going to present a short review of biological concept of inductive automaton [5]. The central attention we will pay to the implementation of an emotional center (EC).

The main output parameter of a neuron is spike frequency, therefore the transfer function of the neuron is gradual, but unlike popular sigmoid it has not two, but three stable states: a neuron can be inactive, half active and fully active. After full activation of a neuron occurs self-locking and forced deactivation of the neuron because full activation is only a temporal state. However, a neuron can maintain half active state for a longer period of time.
The activity of the neural network is controlled by a shared threshold D (in the original concept [5] this threshold was not distinctly referred – it is only a coefficient of the dynamic threshold of a neuron; this coefficient depends on the state of EC: is it good or bed). Greater values of the shared threshold D makes this intermediate half active state more stable: easier to reach this state but more difficult to come out by full activation or full deactivation. Smaller values of the threshold D make half active state less stable.

The half active state of a neuron corresponds to state "badly" because the energy consumption and deficit become greater. However, fully active state corresponds to state "well" because it allows the economy of energy in long-terms. The idea of self-organization is based on these two qualitative states "badly" and "well": inductive automaton builds new links between neurons in order to make the state of network better. The level of "badly" for the network is defined by the number of half active neurons. Emelyanov-Yaroslavsky introduces EC to make such self-organization more efficient motivating his decision this way:

> "To transform the unordered neuronal set into the automaton, one should have a sole index of state for this set and a certain device, capable of predicting this state. The simple way of resolving this problem is the introduction of a critical element into the neuronal set, so that common motor means of the set are controlled only by this critical element. Such "many into single" transformation one can compare to the sore tooth, when the perception of the person is defined only by a level of pain, and his behavior is aimed only to the decreasing of this level. In the neuroenergetic concept the control of associative memory is organized on the basis of such critical element, called the emotional center."

EC is a mechanism of neural network activity modulation. It is a subset of special neurons. Other neurons are called memory neurons. These special neurons control other neurons by two parameters: (1) the shared threshold D – shared between all neurons and (2) reinforcement coefficient that determine proportion of temporal links transformation to persistent links. The first parameter controls the activity of neural network but the second one – self-organizing.

EC needs a regular influence from memory neurons. EC and memory neurons are interacting as an oscillating system. In the state "well" of EC the value of the shared threshold D must increase. It makes more difficult for memory neurons to reach fully active state. After a very short period of time the state of EC must become worse. But in the state "badly" of EC the value of the shared threshold D must decrease making easier for memory neurons to reach fully active state. And it will make the state of the EC better. The value of reinforcement coefficient increases only during improving of the state of EC. Such interrelations between EC and memory neurons promote the setting of oscillatory process: state "well" in EC creates state "badly" in memory neurons, state "badly" in memory neurons creates state "badly" in EC, state "badly" in EC activates the network to reach state "well". During this oscillations the EC controls count of half active neurons.

4. Informational concept of inductive automaton

Since this paragraph we are going to present informational concept of inductive automaton. It is called so because it could be considered as a complementary to purely biological concept presented by Emelyanov-Yaroslavsky. During his investigation he is trying to understand biological logic of natural neural networks, investigating different mechanisms observed in natural neural networks but we are trying to answer a question: why these complicated mechanisms are needed for artificial neural networks.

In biological concept of inductive automaton energy consumption minimization is considered as a basic task (Emelyanov-Yaroslavsky, 1990). In contrast, to present an informational concept of inductive automaton we are going to consider *memory* as a general phenomenon. This phenomenon can be observed in different biological systems, for example, neural networks, immune systems, ant colonies and so on. We propose to consider memorizing as a mechanism of self-organization of inductive automata. It means that *inductive automaton must solve one basic task – memorizing of external and internal information flow*. Here, we understand information flow as a sequence of neural group activations. External flow is a perceptual sequence, which is generated by sensors, but internal flow - by neurons themselves.

Inductive automaton has a single long-term storage for memory – there are links between neurons: mainly weak excitatory, and not so often – strong inhibitory links. The configuration of all links we will call “structure“. This structure is memory storage and determines internal reactions of neural network to external stimuli. Inductive automaton memorizes information by adding to its structure new successful constructions, i.e. excitatory links that helps to memorize information flow.

We will say that external or internal information flow is memorized by inductive automaton if and only if it is able to reproduce this information flow, and there is no matter how it is recorded in links between neurons. This understanding makes possible to separate the logical layer from the physical layer of inductive automaton functioning. Physical layer consists of links that store information and neurons that interpret storing information. Logical layer is a set of different internal reactions (sequences of neuron activations in neural network) to external stimuli and previous internal reactions.

It is well known, that neural networks can realize an associative memory function. Usually in neural network it is assumed that a network stores static images. However, we assume that network primarily stores memorized sequences and only then – images. Associative memory function could be understood as follows: if in situation S a neural network has generated reaction R (sequence of activation) then next time in the same situation

S the probability of the same reaction R must increase. We assume that learning rule provides such a property. However, it is possible to obtain a conflict situation – memorizing a new sequence it is possible to damage memory about previously memorized sequences. Therefore, there is a need for a mechanism that protects previously memorized sequences.

The idea of proposed memory protection mechanism has been inspired by the deep analysis of emotional center implementation [5] *in inductive automaton* and seems as follows:

1. *Weak memorizing.* Inductive automaton records the current fragment of information flow by minimal increasing of corresponding inter-neuron excitatory links.
2. *Reproduction test.* Neural network tries to restore previously recorded fragment of information flow, *i.e.* it tries to activate neurons in the same logical sequence without presents of factors (*e.g.* sensor activation sequence) produced such a sequence.
3. *Strong memorizing.* If the performed reproduction test was successful then inductive automaton records the repeated fragment of information flow by normal increasing of corresponding inter-neuron excitatory links, and it records by maximal increasing fragments where reproduction was more difficult but possible.

Thus, inductive automaton records sequences that it is able to reproduce. We consider this property as the key mechanism for formation of more and more abstract model of external information flow. On the other hand, there is no sense to records in memory anything without ability to reproduce it, and the reproduction test guarantee needed ability to retrieve storing in neural network sequences. Described principle is similar to reinforcement learning [2] where positive reward depends on the ability of neural network to restore recorded activations in the same logical order.

Usually in computer sciences memory is understood as something that can store information and memory architecture guarantees precise retrieving of stored information. Such approach makes implementation simpler but potential abilities of memory architecture - more bounded. Therefore, we propose to analyze another point of view on memory – *memory is something that can restore information*. This can allow defining of abstractions formation process as a consequence of memorizing unlimited information flow by physically fixed memory storage. Consequently, we can define abstraction as something that helps to restore information without recording it (or by minimal recording) but sensors in combination with motor neurons could be understood as an external memory storage.

Self-organizing of the structure of inductive automaton we will understand as a consequence of information flow memorizing. All the time the neural network must perform in parallel recording of the current sequence and reproduction test for restoring the previous sequence. Inductive automaton must record only such sequences that could be successfully attached into the structure, formed till the current moment, as complements that don't damage its integrity.

To illustrate the described idea, let us consider an example. Let us assume, that stimulus 1 activates neuron group A, stimulus 2 – group B and stimulus 3 – group C. Stimulus sequence 1-2-3 could activate sequence A-B-C and imply minimal increasing of excitatory links A→B and B→C. Then, a question arises: what could prevent to recover this sequence? It could be another previously created links, *e.g.* excitatory link A→C or inhibitory link A→B.

Thus, the formation of a structure must imply memorizing, and growth of logical layer based on fixed physical layer. This process is possible only when inductive automaton is being successful at searching of more abstract forms of storing memorized information. Thus, the formation of a structure implies formation of more abstract model of the external world.

5. Inductive automaton: states and transitions

In this section, we are going to analyze the inductive automaton analogy – finite automaton. We define the term "state" as a set of neurons that are active at the current time moment. Since different states can include the same neurons, it is possible to define a similarity measure between different states, *e.g.* using a distance measure between the corresponding sets of neurons. Using these definitions we are going to explain the described above memorizing principle (see Fig. 1). States of inductive automaton are shown as cycles, but transitions between the states - as arrows. Information flow could be considered as a sequence of states ($s^{(1)}, s^{(2)}, s^{(3)}, s^{(4)}, \ldots, s^{(t-1)}, s^{(t)}, \ldots$).

Let's assume, that at some time moment a neural network is in state X, and also it is known that in the nearest past the network was in a similar state X_0, and after this the network proceeded into state Y_0. And now, it is necessary to test the network: is it possible to reach state Y similar to state Y_0? The positive answer means that the restored transition of the last sequence has been successfully memorized and doesn't damage integrity of self-organized structure. In general case the principle of states sequence memorizing could be defined by following cycle of neural network control by the shared threshold D:

1. *Stabilizing of the current state* (state X). The shared threshold D takes the maximal value. It makes the half active state more stable: easier to reach this state but more difficult to come out by full activation or full deactivation. Therefore, it stabilizes the current state by fixing the group of half active neurons. This process is similar to image recovery in Hopfield networks [4].
2. *Transition to the next state* (state Y). The shared threshold D takes the minimal value. It makes easier for neurons to become deactivated through fully activation or without it. In the case of full activation a neuron may activate other inactive neuron groups. Thus, this phase activates the process of the current state modifying - searching for a new state.
3. *Saving temporal links.* Inductive automaton checks this transition: is it equivalent to the transition in the nearest past? (Is state Y similar to state Y_0, or not?) The positive answer implies saving of temporal links that were created in order to record the proceeded transition.

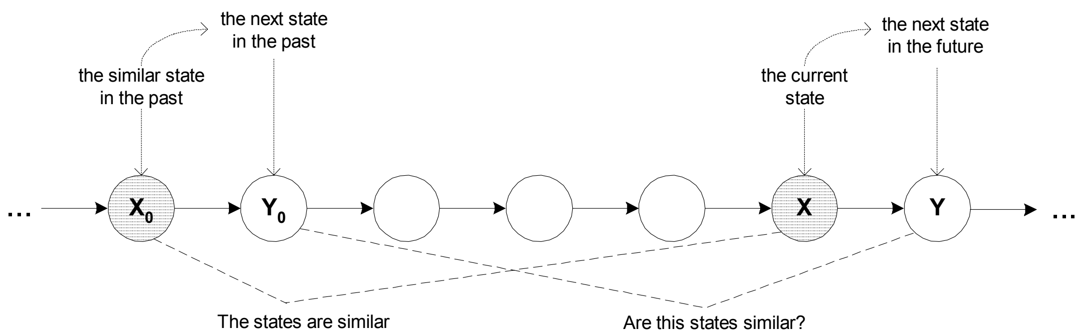

Figure 1. Test for recovery of the transition between states performed in the nearest past

The described cycle implies emotional center oscillations in order to produce states (the recovery of the current state consistency) and transitions (searching for the next state) within the neural network. Thus, the neural network step by step is trying to repeat the previous activation sequence, and in the case of success the structure is being formed. Memory neurons try to memorize sequence of states by forming temporal additional structures, but emotional center stores only the last sequence in order to evaluate the ability of memory neurons to recover this sequence and to allow temporal structures transformation to persistent structures in the case of successful recovering.

The neural network can't store memorized sequences directly because it can store only connections between distinct neurons. However, questions arise: how to recover the last sequence in the case when network has memorized more then one sequence? How to know which sequence to choose? The analysis of the biological concept shows that the neural network can use special hints to recover the last sequence. It is the synchronization of activity of simultaneously fully activated neurons in the last sequence. Using this partial information and connections between distinct neurons the neural network must recover the logical order of the last activation.

6. A simulation

To illustrate the described principle of the control of neural network activity, a simulation was performed. A partly formed neural network was selected for this simulation (see. fig. 2). Neurons are drawn by cycles and links – by arrows (thin arrow – weak excitatory link; thick arrow – strong inhibitory link). For better understanding there are dashed ellipses around the groups of neurons (A,B,C,K,E,F,H,G) and neurons are numerated. The marked neurons have lower ability to activate than other neurons, *e.g.* neuron `3` could become active only if neuron `1` is fully active.

To simulate this neural network we have applied a simplified activation procedure that includes only impulse generating, potentials calculating and temporal links creating. Full description of this procedure and complete discussion on the current experiment are out of the scope of this paper, and we are going to present them in future papers.

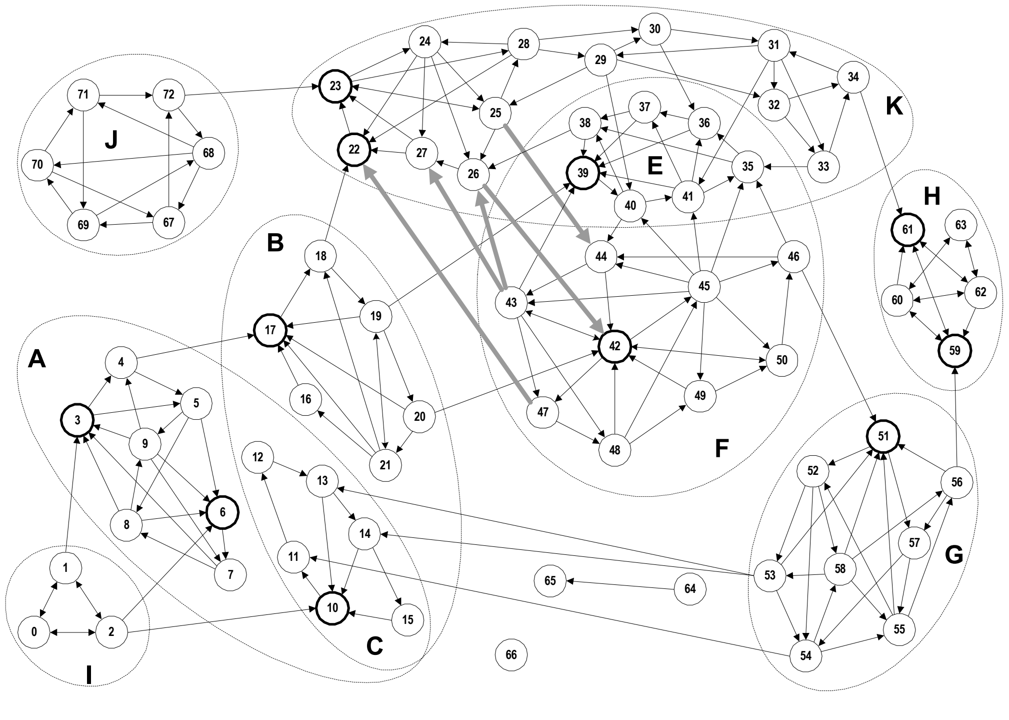

Figure 2. A neural network selected for the simulation

In the beginning of the experiment neuron 0 was made half active. The value of shared threshold has set manually for each time moment. In this experiment neurons have been activated as follows: I→A,C→C,B→C,F,E→C,G→H. Fig. 3 shows activity (average impulse generating frequency) dynamics in each group of neurons. It should be noted, that all the activity maximums correspond to the minimal value of shared threshold *D*. Before each activity maximum half activation occurs (state consistency recovery), but after each maximum the deactivation of active neurons occurs, and activity is transferred to other groups of neurons. However, some neurons can maintain half active state for a longer period of time, *e.g.* group C. The union of groups K,E,F could be considered as a single neural assembly, that was split by inhibitory links on two competing assemblies. For example, if in the beginning of the experiment neurons 0 and 69 would be made half active, we could obtain a different sequence of activation: I,J→A,C,K,E→C,B,H→ C.

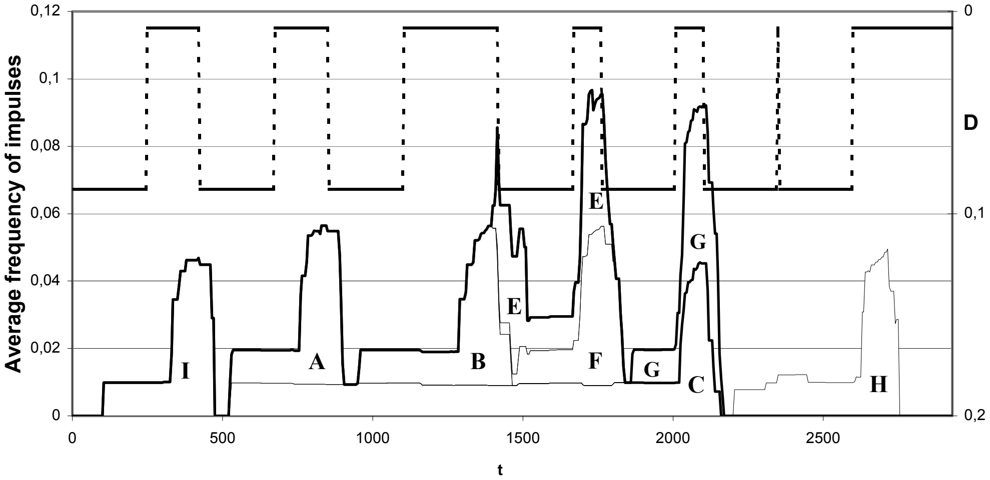

Figure 3. Neural network activation dynamic

Conclusions

In this paper we have presented informational concept of inductive automaton. In biological concept energy consumption minimization is considered as a basic task. In contrast, in informational concept inductive automaton must solve one basic task – memorizing of external and internal information flow. Self-organizing of the structure and abstractions formation are considered as a consequence of attempting to memorize unlimited information flow by physically fixed memory storage. During these attempts, it is possible to obtain conflicts – memorizing a new sequence it is possible to damage memory about previously memorized sequences. Therefore, there is a need for emotional center - a mechanism that protects previously memorized information. Emotional center controls recording of the current sequence and reproduction test for restoring the previous sequence, and in case of success allows structure forming. Thus, the performed analysis has discussed a need for emotional center not only for natural neural network, but also for artificial neural networks.

Acknowledgement. This research was supported in part by the Latvian Science Foundation under grant No.02-86d.

References

1. Хант Э. Искусственный интеллект // Мир, Москва, 1978 - 558 p.
2. Mitchel, T.H., Machine learning // The McGraw-Hill Companies. Inc., 1997 - 432 p.
3. Angluin D. Learning regular sets from queries and counterexamples // Information and Computation, Nr. 75, 1987 - pp 87.-106.
4. Fausett, L. V. Fundamentals of neural networks: architectures, algorithms and applications // Prentice-Hall, Inc. New Jersey, 1994 - 461 p.

5. Emelyanov-Yaroslavsky L.B. Intelligent quasi-biological system (Inductive automaton). // M.: "Nauka", 1990 - 112 p.

Artūrs Pčolkins, Mg. sc. dat., RTU, Datorzinātnes un informacījas tehnoloģijas fakultāte, Kalķu 1, Rīga, E-mail: arturp@balticom.lv

Pčolkins A. Pašorganizācija neironu tīklos balstīta uz iegaumēšanu
Šis pētījums ir saistīts ar induktīvu automātu – jauno neironu tīklu, kuru piedāvāja Emeļjanov-Jaroslavski 1990. gadā. Galvenā ideja bija nevis realizēt specifisku funkcionalitāti katrai uzdevumu klasei, bet nodefinēt vienu pamatuzdevumu – enerģijas patēriņa minimizēšanu, un pēc tam iegūt pārējo funkcionalitātes kā blakus parādības. Induktīvam automātam, kā dabiskiem neironu tīkliem, ir emocionālais centrs. Tomēr, tā funkcijas paskaidrojums nav pilnīgs. Autors pēta dažādus sarežģītus bioloģiskus mehānismus, bet viņš necenšas atbildēt uz jautājumu: priekš kam visi šie sarežģīti mehānismi ir vajadzīgi mākslīgiem tikliem? Tāpēc mūsu mērķis bija izstrādāt pilnīgāku sapratni par tiem sarežģītiem procesiem, bet šajā rakstā mēs piedāvājam mūsu sapratni tikai par emocionālu centru.
Kontrastā ar tīri bioloģisko pieeju, mēs apskatam atmiņu kā pamatuzdevumu. Struktūras pašorganizācija un abstrakciju veidošana tiek apskatītas kā blakus parādības no mēģinājumiem iegaumēt neierobežotu informācijas plūsmu, izmantojot fiziski fiksētu atmiņas glabātuvi . Šo mēģinājumu laikā var dabūt konflikta situāciju – iegaumējot jaunu informāciju, ir iespējams bojāt atmiņu par agrāk iegaumēto informāciju. Tas noved pie secinājuma, ka ne tikai bioloģiskiem neironu tīkliem ir vajadzīgs emocionālais centrs, bet arī abstraktiem mākslīgiem neironu tīkliem arī ir vajadzīgs mākslīgais emocionālais centrs kā mehānisms, kurš aizsargātu izveidotas atmiņas integritāti.

Pchelkin A. Self-organizing in neural networks based on memorizing
Our investigation is connected with inductive automaton - a new artificial neural network proposed by Emelyanov-Yaroslavsky in 1990. The main idea is not to implement a specific functionality for each class of tasks but to define one basic task – energy consumption minimization and then to obtain all other functionalities as by-products. Inductive automaton, like natural neural networks, has emotional center. However, the explanation of its function is not complete. The author investigates different complex biological mechanisms, but he is not trying to answer a question: why all these complex mechanisms are needed for artificial networks? That's why, our aim is to develop a more complete understanding of those complex biological processes, but in this paper we present our understanding only of emotional center function.
In contrast to the purely biological concept, we consider memory as a basic task. Self-organizing of the structure and abstractions formation are considered as by-products of attempting to memorize unlimited information flow by physically fixed memory storage. During these attempts, it is possible to obtain a conflict situation – memorizing new information it is possible to damage memory about previously memorized information. It leads to conclusion that not only biological neural networks need emotional center, but also abstract artificial neural networks also need an artificial emotional center as a mechanism that protects formed memory integrity.

Пчелкин А. Самоорганизация в нейронных сетях, основанная на запоминании
Данное исследование связано с индуктивным автоматом – новая нейронная сеть, предложенная Емельяным-Ярославским в 1990 году. Главная идея состоит в том, чтобы не реализовывать специфичной функциональность для каждого класса задач, а определить одну основную задачу – минимизацию потребления энергии - и потом получить все остальные функциональности как побочные продукты. Индуктивный автомат, как и естественные нейронные сети, имеет эмоциональный центр. Однако, объяснение его функции не является достаточным. Автор исследует разные сложные биологические механизмы, но он не старается ответить на вопрос: для чего все эти сложные механизмы нужны искусственным сетям? Вот почему, наша цель состоит в поиске более полного понимания этих сложных биологических процессов, но в данной статье мы предлагаем наше понимание только функции эмоционального центра.
В отличие от чисто биологической концепции, мы рассматриваем запоминание как основную задачу. Самоорганизация структуры и формирование абстракций рассматриваются, как побочные продукты от попыток запомнить неограниченный поток информации с помощью физически ограниченного хранилища памяти. В течение этих попыток возможно получить конфликтную ситуацию – запоминая новую информацию, можно испортить память о ранее запомненной информации. Это приводит к выводу, что не только биологическим нейронным сетям нужен эмоциональный центр, но абстрактным искусственным нейронным сетям тоже нужен искусственный эмоциональный центр как механизм, который защищает целостность сформированной памяти.