



Available at [www.sciencedirect.com](http://www.sciencedirect.com)

SciVerse ScienceDirect

journal homepage: [www.elsevier.com/locate/bica](http://www.elsevier.com/locate/bica)



## INVITED ARTICLE

# Probabilistic dynamic logic of cognition

Evgenii E. Vityaev <sup>a,b,\*</sup>, Leonid I. Perlovsky <sup>c</sup>, Boris Ya. Kovalerchuk <sup>d</sup>,  
Stanislav O. Speransky <sup>b</sup>

<sup>a</sup> Sobolev Institute of Mathematics, 630090 Novosibirsk, Russia

<sup>b</sup> Novosibirsk State University, 630090 Novosibirsk, Russia

<sup>c</sup> Harvard University, Air Force Research Laboratory, USA

<sup>d</sup> Central Washington University, Ellensburg, WA 98926-7520, USA

Received 15 March 2013; received in revised form 25 June 2013; accepted 28 June 2013

### KEYWORDS

Artificial intelligence;  
Probabilistic dynamic logic;  
Probabilistic inference;  
Neuron model;  
Breast cancer

### Abstract

We developed an original approach to cognition, based on the previously developed theory of neural modeling fields and dynamic logic. This approach is based on the detailed analysis and solution of the problems of artificial intelligence – combinatorial complexity and logic and probability synthesis. In this paper we interpret the theory of neural modeling fields and dynamic logic in terms of logic and probability, and obtain a Probabilistic Dynamic Logic of Cognition (PDLC). We interpret the PDLC at the neural level. As application we considered the task of the expert decision-making model approximation for the breast cancer diagnosis. First we extracted this model from the expert, using original procedure, based on monotone Boolean functions. Then we applied PDLC for learning this model from data. Because of this model may be interpreted at the neural level, it may be considered as a result of the expert brain learning. In the last section we demonstrate, that the model extracted from the expert and the model obtained by the expert learning are in good correspondence. This demonstrate that PDLC may be considered as a model of learning cognitive process.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

Previously, an original approach was developed to the cognition simulation based on the theory of neural modeling

fields and dynamic logic (Kovalerchuk & Perlovsky, 2008; Perlovsky, 1998; Perlovsky, 2006; Perlovsky, 2007). On the one hand, this approach is based on the detailed analysis of the cognition problem for artificial intelligence – combinatorial complexity and logic and probability synthesis. On the other hand, it is based on the psychological, philosophical or cognitive science data for the basic mechanisms of cognition. The main idea behind success of NMF is matching the levels of uncertainty of the problem/model and the levels of uncertainty of the evaluation criterion used to identify the model. When a model becomes more certain then

\* Corresponding author at: Sobolev Institute of Mathematics, 630090 Novosibirsk, Russia. Tel.: +7 383 363 4562; fax: +7 383 333 2598.

E-mail addresses: [vityaev@math.nsc.ru](mailto:vityaev@math.nsc.ru) (E.E. Vityaev), [leonid@seas.harvard.edu](mailto:leonid@seas.harvard.edu) (L.I. Perlovsky), [borisk@cwu.edu](mailto:borisk@cwu.edu) (B.Ya. Kovalerchuk), [netid@ya.ru](mailto:netid@ya.ru) (S.O. Speransky).

the evaluation criterion is also adjusted dynamically to match the adjusted model. This process is called dynamic logic of model construction, which mimics processes of the mind and natural evolution.

Analysis of the cognition problems has, in fact, a broader meaning and overcoming these problems can lead to the other formal characterizations of the cognition process. With this purpose, in Kovalerchuk and Perlovsky (2008) a generalization of the theory of neural modeling fields and dynamic logic was obtained in the form of dynamic logic of cognition and cognitive dynamic logic. These logics are formulated in rather general terms: *relations of generality, uncertainty, simplicity; maximization the similarity with empirical content; training method*.

In this paper, we interpret these concepts in terms of logic and probability: the uncertainty we interpret as a probability, and the process of training as a semantic probabilistic inference (Smerdov & Vityaev, 2009; Vityaev, 2006a; Vityaev, 2006b; Vityaev & Smerdov, 2009). Obtained Probabilistic Dynamic Logic of Cognition (PDLC) belongs to the field of probabilistic models of cognition (Probabilistic Inductive Logic Programming, 2008; Probabilistic models of cognition, 2006; The Probabilistic Mind, 2008). We show that this logic also solves the cognition problems (combinatorial complexity and logic and probability synthesis). Thus, by the generalization obtained in Kovalerchuk and Perlovsky (2008), we extend the interpretation of the theory of neural modeling fields and dynamic logic on the probabilistic models of cognition. Probabilistic dynamic logic had been used for simulation of brain activity and cognitive processes in Demin and Vityaev (2008).

## 2. Cognition problem from the probabilistic point of view

Now we repeat and extend the description of cognition problem for artificial intelligence stated in Perlovsky (1998) and Kovalerchuk and Perlovsky (2008). The founders of artificial intelligence in the 1950s and 1960s believed that by reference to the rules of logic, they would soon create a computer which intelligence would be far superior to the human brain. But the application of logic to artificial intelligence didn't lead to the results expected. We need to clearly distinguish the theoretical and empirical attitudes. In theory using the idealized knowledge, for example, in physics, geometry, chemistry and other sciences, logic and logical inference are justified and work perfectly. But the intelligent systems are based on the empirical learning process, with the knowledge, obtained as a result, being inductive. For the inductively derived knowledge logical inference does not work well.

The brain is not a logical but a predictive one. However, a suitable definition of prediction for the inductively derived knowledge is a problem.

A generally accepted definition of prediction belongs to Karl Popper and is based on the fact that for the prediction of some fact it is necessary to infer it from the available facts and laws. But this definition does not work for the inductively derived knowledge with estimations of probabil-

ity, confirmation, etc. At the same time, in the logical inference of predictions it is necessary to deduce the estimations of probability, confirmation, etc. for the obtained prediction. For probability estimations there is a probabilistic logic (Gabbay, Johnson, Ohlbach, & Woods, 2002) and probabilistic inductive logic programming (Raedt et al., 2008) to deal with it. But it is well known that prediction estimations may fall during the logical inference and leading to zero prediction estimations. Predictions with zero estimation can not be regarded as predictions. This problem is now regarded as a problem of logic and probability synthesis (Cozman et al., 2009). There have already been five symposia between 2002 and 2011 under the common title Prolog (Probability + Logic).

We have introduced a new concept of prediction (Smerdov & Vityaev, 2009; Vityaev, 2006b; Vityaev & Smerdov, 2009) which is not use a logical inference and replace the "true" and "false" values by probability. Instead of logical inference we introduced semantic probabilistic inference. The new definition of the prediction is fundamentally different from the Karl Popper's one – the prediction of some fact occurs not as a logical inference of the particular fact from the existing ones but as a direct inductive inference of the rule that predicts the fact we are interested in. Estimates of probability strictly increase in the process of semantic probabilistic inference.

Another problem of cognition in artificial intelligence is the combinatorial complexity problem (CC), (Kovalerchuk & Perlovsky, 2008; Perlovsky, 1998). Perception associates a subset of signals corresponding to the external objects with representations of these objects. The process of association-recognition-understanding turned out to be not at all easy, and is connected with the notion of combinatorial complexity. Subsequent studies found a connection between CC and logic in a variety of algorithms. Logic considers a very small change in data or models as a new proposition. Attribution of truth values "true" and "false" does not allow comparing statements and this leads to CC. In (Hyafil & Rivest, 1976) it is proved that even the simplest task of finding the set of propositions describing the decision trees is NP-hard.

Follow the work (Kovalerchuk & Perlovsky, 2008) we introduced two order relations on propositions: relations of generality and comparison that are used in semantic probabilistic inference. This essentially reduces the search and, along with the use of statistical estimates, makes it acceptable and solves CC problem.

Now we recall and extend the basic definitions related to cognition (Kovalerchuk & Perlovsky, 2008; Perlovsky, 2006). We assume that the basic mechanisms of cognition include: instincts, concepts, emotions and behavior. Further we explain how semantic probabilistic inference may be used in formalization of these concepts.

Ray Jackendoff (2002) believes that the most appropriate term for mechanism of concepts is a model, or an internal model of cognition. Concepts are models in a literal sense. Within our cognitive process, they construct world objects and situations. Cognition involves the multi-levelled hierarchy of concept models: from the simplest elements of perception (line, point) to the concept models of objects, relations between objects and complex situations.

The fundamental role of emotions in cognition is that they are connected with the instinct for knowledge that maximizes a measure of similarity between the concept models and the world (Kovalerchuk & Perlovsky, 2008). This emotional mechanism turned out to be fundamental in “breaking the vicious circle” of the combinatorial complexity. In the process of learning and understanding of the input signals, the models are adapted to represent the input signals better and to increase the similarity between them. This increase of similarity satisfies the instinct for knowledge and feels like an aesthetic emotion.

Experiments confirming the relation between emotions and the instinct for knowledge can be found in the P.V. Simonov’s Information theory of emotions (Simonov, 1981): “Summing up the results of our experiments and literature material we came to the conclusion... that emotion is a reflection of human and animal brain of any actual need (its quality and quantity) and probability (possibility) of its satisfaction that the brain evaluates on the basis of genetic and acquired earlier individual experience...”.

The following experiment shows that the instinct for knowledge causes positive emotions (Simonov, 1981): “In our experiments we projected rows of five digits – ones and zeros – on the screen, which was placed in front of the test subject. The test subject was warned that some of the frames containing a common feature (e.g., two zeros in a row 00) would be accompanied by a beep. The task was to find this feature out... Until the first hypothesis about corroborated feature (usually erroneous frames, e.g. 01) neither new frames nor beeping called GSR (galvanic skin response, emotional detector) emerged... The emergence of hypothesis was accompanied by GSR... After the formation of hypotheses two situations are possible, we regard them as the experimental models of negative and positive emotional responses... The hypothesis is wrong, and the frame... containing corroborated feature (00 and, consequently, not confirming the hypothesis 01) does not cause GSR. When beeping indicates that the test subject made a mistake, GSR is registered as a result of a mismatch between the hypothesis and the present stimulus... The test subject changes their hypothesis several times and at some point it starts being accurate. Now, the very appearance of the corroborated frame causes GSR and its corroboration with a beep leads to even stronger galvanic skin shifts. How can we understand this effect? Indeed, in this case we see a complete match of the hypothesis... with the present stimulus. The absence of the mismatch should entail the absence of GSR... In fact, in the latter case we also encounter mismatch, but mismatch of another sort when testing the false hypothesis. Formed in the process of repeated combinations prediction contains not only the afferent model of the goal... but also the probability of achieving this goal. At the moment of the frame corroboration... predicted by a beep, *the probability of solving the problem (that the hypothesis is correct) has increased* – and this mismatch between the prediction and received information led to a strong GSR”.

Thus, confirmation of hypothesis, which causes a positive emotion, increases its credibility and, consequently, the closeness of concept model to our world (demonstration of the instinct for knowledge). The whole process of learning, when a person achieves more accurate and correct

actions in the real world, is supported by the emotions – positive emotions reinforce correct actions (and corresponding correct predictions, increasing their probability), and negative emotions correct mismatches between the model and the real world (and corresponding wrong predictions, reducing their probability).

In this case similarity of the concept models to our world, controlled by emotions, is measured by the probability of predictions. Semantic probabilistic inference, underlying the training operator, implements directed search of more and more probable rules by adding to the premise of the rules additional properties of the world that allow increasing the conditional probability of prediction and, therefore, provide greater value and the similarity to the world. This directed search eliminates the CC problem. The main definitions of PDLC presented in the next section.

In the third section we argue that semantic probabilistic inference may be also considered as a formalization of the Hebb’s rule (Caporale & Dan, 2008; Hebb, 1949) and a formal model of neuron. It gives us possibility to interpret the PDLC at the neuronal level. We have applied the PDLC for the approximation of expert decision-making model for the diagnosis of breast cancer (Kovalerchuk, Vityaev, & Ruiz, 2001).

### 3. The data, models, relation of generality, similarity between the model and the data

We define the basic concepts of PDLC in the propositional case. Expanded definitions in the language of first-order logic (in several different versions) can be found in (Smerdov & Vityaev, 2009; Vityaev, 2006b; Vityaev & Smerdov, 2009).

By data we mean the standard object-feature matrix, in which the set of features  $x_1(a), \dots, x_n(a)$  are given over the set of objects  $A = \{a_1, \dots, a_m\}$ , where  $a$  stands for the variable on objects. For each value of feature  $x_i$  we define an atomic proposition  $P_j^i(a) = (x_i(a) = x_{ij})$ ,  $j = 1, \dots, n_i$ , where  $x_{j1}, \dots, x_{jn_i}$  are all values of the feature  $x_j$ ,  $i = 1, \dots, n$ . We denote the set of all atomic propositions (atoms) as  $At$ . We shall denote these atoms as Boolean variables  $a, b, c, \dots$ . We call *literal* a set consisting of atomic propositions or their negations, which we shall also denote as Boolean variables  $a, b, c, \dots$  (possibly with indices). The set of all literals we denote by  $L$ .

We assume that *data* is presented by some *empirical system* (algebraic system)

$$Data = \langle A, \{P_1^1, \dots, P_{n_1}^1, P_1^2, \dots, P_{n_2}^2, \dots, P_1^k, \dots, P_{n_k}^k\} \rangle,$$

in which the truth values of all atomic propositions over the set of objects  $A$  are defined.

By *the model* we shall mean the Boolean function  $\Phi$ , with atoms from  $At$  being substituted for Boolean variables. It is known that any Boolean function can be represented by a number of rules  $\{R\}$  of the form

$$(a_1 \& \dots \& a_k \Rightarrow b), \quad a_1, \dots, a_k, \quad b \in L, \quad (1)$$

Therefore by the model  $\Phi$  we shall mean a set of rules  $\{R\}$ . We denote the set of all type (1) rules by  $Pr$ . We say that *the rule (1) is applicable to the data* if the premise

$a_1, \dots, a_k$  of the rule is true in the Data, and that *the model*  $\Phi = \{R\}$  is applicable to the data if every rule in the model is applicable to the data.

For the models, defined as a set of rules, there emerges a combinatorial complexity problem. To avoid this problem, we shall define a partial order on the sets of rules and models, as well as a measure of similarity between the model and the data.

**Definition 1.** We shall call the rule  $R_1 = (a_1^1 \& a_2^1 \& \dots \& a_{k_1}^1 \Rightarrow c)$  more general than the rule  $R_2 = (b_1^2 \& b_2^2 \& \dots \& b_{k_2}^2 \Rightarrow c)$ , and denote it as  $R_1 \succ R_2$  if and only if  $\{a_1^1, a_2^1, \dots, a_{k_1}^1\} \subset \{b_1^2, b_2^2, \dots, b_{k_2}^2\}$ ,  $k_1 < k_2$ , and not less general  $R_1 \succsim R_2$  if  $k_1 \leq k_2$ .

**Corollary 1.**  $R_1 \succsim R_2 \Rightarrow R_1 \vdash R_2$  and  $R_1 \succ R_2 \Rightarrow R_1 \vdash R_2$  where  $\vdash$  is the logical inference in propositional calculus (atoms being perceived as propositional characters).

Thus, no less general (and more general) statements are logically stronger. Besides, more general rules are easier, because they contain a smaller number of letters in the premise of the rule, so the relation can be interpreted as the relation of simplicity, in sense of (Kovalerchuk & Perlovsky, 2008).

**Definition 2.** We shall call the model  $\Phi_1 = \{R_1^1\}$  no less general  $\Phi_1 \succsim \Phi_2$  than the model  $\Phi_2 = \{R_2^2\}$  iff for every rule  $R^2 \in \Phi_2$  there exists no less general rule  $R^1 \in \Phi_1$ ,  $R_1^1 \succsim R_2^2$  and more general  $\Phi_1 \succ \Phi_2$  if at least for one rule  $R^2 \in \Phi_2$  there exists more general rule  $R_1^1 \succ R_2^2$ ,  $R^1 \in \Phi_1 \setminus \Phi_2$ .

**Corollary 2.**

$$\Phi_1 \succ \Phi_2 \Rightarrow \Phi_1 \vdash \Phi_2.$$

Corollary 2 implies that the more general model is logically stronger and simultaneously easier.

We define the set of propositions  $F$  as a set of propositions obtained from the atoms  $At$  by the closure under logic operations  $\wedge, \vee, \neg$ .

**Definition 3.** We call the mapping  $\mu: F \mapsto [0, 1]$  as *probability* on the set of propositions  $F$ , if it does satisfy the following conditions (Halpern, 1990):

1. If  $\vdash \varphi$ , then  $\mu(\varphi) = 1$ ;
2. If  $\vdash \neg(\varphi \wedge \psi)$ , then  $\mu(\varphi \vee \psi) = \mu(\varphi) + \mu(\psi)$ .

We define the conditional probability of the rule  $R = (a_1 \& \dots \& a_k \Rightarrow c)$  as

$$\mu(R) = \mu(c/a_1 \& \dots \& a_k) = \frac{\mu(a_1 \& \dots \& a_k \& c)}{\mu(a_1 \& \dots \& a_k)}, \text{ if } \mu(a_1 \& \dots \& a_k) > 0.$$

For the case  $\mu(a_1 \& \dots \& a_k) = 0$  the conditional probability is not defined. We denote by  $Pr_0$  the set of all rules of  $Pr$ , for which the probability is defined. We assume that the probability  $\mu$  gives the probabilities of events, represented by propositions above the empirical system Data.

**Definition 4.** We shall call by a *probability law* a rule  $R \in Pr_0$  such that it cannot be generalized (logically increased) without reducing its conditional probability, i.e. for any  $R' \in Pr_0$  if  $R' \succ R$ , then  $\mu(R') < \mu(R)$ .

Probability laws are the most common, simple and logically strongest rules among the rules with lesser or equal conditional probability. We denote the set of all probability laws by PL (Probabilistic Laws). Any rule can be generalized (simplified and logically enforced) to a probability law without reducing its conditional probability.

**Lemma 1.** For every rule  $R \in Pr_0$ , it is either a probability law or there is a probability law  $R' \in PL$  such that  $R' \succ R$  and  $\mu(R') \geq \mu(R)$ .

**Definition 5.** By *probabilistic regularity model* we mean the model  $\Phi = \{R\}$ ,  $R \in PL$ .

**Lemma 2.** For any model  $\Phi$ , it is either a probabilistic regularity model or there is no less general model  $\Phi' \succsim \Phi$ , for some probabilistic regularity model  $\Phi'$ .

We define an order relation on the set of probabilistic laws PL.

**Definition 6.** By *relation of probabilistic inference*  $R_1 \sqsubseteq R_2$ ,  $R_1, R_2 \in PL$  we mean simultaneous fulfillment of two inequalities  $R_1 \succsim R_2$  and  $\mu(R_1) \leq \mu(R_2)$ . If both inequalities are strict, then the relation of probabilistic inference is called a *strict probabilistic inference*  $R_1 \sqsubset R_2 \iff R_1 \succ R_2 \& \mu(R_1) < \mu(R_2)$ .

**Definition 7.** By *semantic probabilistic inference* (Vityaev, 2006a; Vityaev, 2006b), we shall mean the maximal (that cannot be extended) sequence of probability laws in relation of strict probabilistic inference  $R_1 \sqsubset R_2 \sqsubset \dots \sqsubset R_k$ . We shall call *the most specific* the latter probability law  $R_k$  in this inference.

We now extend the definition of semantic probabilistic inference on the models and define the relation of similarity on the probabilistic regularity models.

**Definition 8.** Probabilistic regularity model  $\Phi_2 = \{R_2^2\}$  is *closer to the data* than the probabilistic regularity model  $\Phi_1 = \{R_1^1\}$  (we denote it as  $\Phi_1 \triangleleft \Phi_2$ ) if and only if  $\Phi_1 \succ \Phi_2$  and for any probabilistic law  $R^2 \in \Phi_2$  there exists probabilistic law  $R^1 \in \Phi_1$  such that  $R_1^1 \in R_2^2$  and for at least one probabilistic law  $R^2 \in \Phi_2$  there is a probabilistic law  $R^1 \in \Phi_1 \setminus \Phi_2$  with strict relation of probabilistic inference  $R_1^1 \sqsubset R_2^2$ .

This definition means that when passing from the probabilistic regularity model  $\Phi_1$  to the model  $\Phi_2$  there is such a build-up of premises of rules, which (strictly) increases the conditional probability of these rules. The increase of the conditional probabilities of model rules means the increase of the model predictive ability and its similarity to the data.

As mentioned in the introduction, the instinct for knowledge is to "maximize a measure of similarity between the concept models and the world". In our definition the

measure of similarity is defined by the set of conditional probabilities of the rules of the model, i.e. by total accuracy of the predictions of the model.

**Definition 9.** We call *the measure of similarity between the probabilistic regularity model  $\Phi = \{R\}$  and the data* – the set  $\{\mu(R), R \in \Phi\}$  of conditional probabilities of the model rules.

**Corollary 3.** If  $\Phi_1 \triangleleft \Phi_2$ ,  $\Phi_1 = \{R_1^1\}$ ,  $\Phi_2 = \{R_2^2\}$ , then the measure of similarity  $\{\mu(R), R \in \Phi_2\}$  of the model  $\Phi_2$  approximates the measure of similarity  $\{\mu(R), R \in \Phi_1\}$  of the model  $\Phi_1$  in the sense that for any probability  $\mu(R_2) \in \{\mu(R_2), R_2 \in \Phi_2\}$  there is the same  $\mu(R_1) \leq \mu(R_2)$  (or strictly less probability  $\mu(R_1) < \mu(R_2)$ ),  $\mu(R_1) \in \{\mu(R_1), R_1 \in \Phi_1\}$ .

*The instinct for knowledge is a process that develops dynamically by successive approximation to the data.*

**Definition 10.** The *training operator*  $L: \Phi_1 \rightarrow \Phi_2$  is the transformation of model  $\Phi_1$  to model  $\Phi_2$  such that both models are applicable to the data and the similarity of the model to the data becomes higher  $\Phi_1 \triangleleft \Phi_2$ , and also all the maximally specific laws of model  $\Phi_1$  transfer to the model  $\Phi_2$ .

We have developed a software system named Discovery (Kovalerchuk & Vityaev, 2000; Vityaev, 2006a), which precisely implements this training operator by the following steps:

- (1) the initial set of rules  $\Phi_0 = \{R\}$  consists of the type (1) rules with the number of predicates in the premise of  $k$  that is not more than some number  $N$ , defined by the user;
- (2) only those rules that are probabilistic laws  $\Phi_1 = \{R \mid R \in \Phi_0, R \in PL\}$  are selected from the set  $\Phi_0$ ;
- (3) hereafter the training operator  $L: \Phi_i \rightarrow \Phi_{i+1}$ ,  $i = 1, 2, \dots$  is applied as long as possible. If reinforcement of the model is no longer possible, the learning process stops. The strict increase of the probability during probabilistic inference is verified by Fisher's exact test for contingency tables.

In the book (Kovalerchuk & Vityaev, 2000) there is a pseudo code for this algorithm. This program has been successfully used for a wide range of applications in various domains (Kovalerchuk & Vityaev, 1998; Kovalerchuk & Vityaev, 2000; Kovalerchuk et al., 2001; Scientific Discovery website, xxxx; Vityaev, 2006a).

#### 4. Neural organization of probabilistic dynamic logic

Semantic probabilistic inference may be regarded as the formal model of neuron (Vityaev, 2006a). Here we briefly present this formal model. So, we can illustrate definitions, introduced in the previous section at the neuronal level.

By the information coming to the "entrance" of the brain, we shall mean all afferentation perceived by the brain. We define afferentation transmitted through the neuron dendrites, by monadic predicates  $P_j^i(a) = (x_i(a) = x_{ij})$ ,  $j = 1, \dots, n_i$ , where  $x_i(a)$  is some characteristic, and  $x_{ij}$  is a value of this characteristic in the situation (on object)  $a$ . If this afferentation passes on the excitatory synapse, neuron perceives it as information about the truth of the predicate  $P_j^i(a)$ ; if it passes on the inhibitory synapse, neuron perceives it as a negation of the predicate  $\neg P_j^i(a)$ .

We shall define the excitation of neuron in the situation (on object)  $a$  and transfer of this excitation on its axon by monadic predicate  $P_0(a)$ . If neuron is inhibited in the situation  $a$  and does not transmit excitation by its axon, we define this situation as a negation of the predicate  $\neg P_0(a)$ . It is known that each neuron has a receptive field, stimulation of which excites it unconditionally. The initial (before training) semantics of the predicate  $P_0$  is the information that it receive from the receptive field. In the process of learning this information enriches.

We assume that the formation of conditional rules (relations) at the neuronal level originates according to the Hebb's rule (Caporale & Dan, 2008; Gerstner & Kistler, 2002; Hebb, 1949) and can be formalized accurately enough by semantic probabilistic inference.

Predicates and their negations are literals (predicates or its negations), and, as in the previous section, we will denote them by Boolean variables  $a, b, c, \dots \in L$ .

We shall define neuron as a set  $\{R\}$  of type (1) rules:

$$(a_1 \& \dots \& a_k \Rightarrow b), \quad a_1, \dots, a_k, \quad b \in L,$$

where  $a_1, \dots, a_k$  are values (excitatory/inhibitory) of the predicates coming at the input of the neuron, and  $b$  is value of the predicate  $P_0(a)$  that denotes the output of neuron.

We now define a method for calculating the conditional probabilities of the neuron rules  $(a_1 \& \dots \& a_k \Rightarrow b)$ . We calculate the number of cases  $n(a_1, \dots, a_k, b)$  where an event  $\langle a_1, \dots, a_k, b \rangle$  has occurred as a simultaneous excitation/inhibition of the neuron inputs  $\langle a_1, \dots, a_k \rangle$  and the neuron itself just before the reinforcement. Reinforcement can be both positive or negative.

Among the cases  $n(a_1, \dots, a_k, b)$  we calculate the number  $n^+(a_1, \dots, a_k, b)$  of cases when the reinforcement was positive and the number  $n^-(a_1, \dots, a_k, b)$  of cases when the reinforcement was negative. Then the conditional (empirical) probability of the neuron rule  $(a_1 \& \dots \& a_k \Rightarrow b)$  can be computed as follows:

$$\mu(b/a_1, \dots, a_k) = \frac{n^+(a_1, \dots, a_k, b) - n^-(a_1, \dots, a_k, b)}{n(a_1, \dots, a_k, b)}.$$

In the process of elaborating classical conditioning the conditional signals associate with the result. At the neuronal level classical conditioning appears in the form of neuronal plasticity (Caporale & Dan, 2008; Pfister, Barber, & Gerstner, 2003).

Furthermore, if there emerge new stimuli allowing to predict neuron excitation even better (with higher probability), they attached to this conditional relation. The formal connection of new stimuli to the existing relation is determined by probabilistic inference (Definition 6), which actually means that new stimuli are added to the conditional

relation if they increase the conditional probability of prediction of neuron excitation.

Formalization of the process of conditional connections at the neuronal level is given by semantic probabilistic inference (Definition 7). Fig. 1 schematically shows a few semantic probabilistic inferences implemented by neuron:

1.  $(b \Leftarrow a_1^1 \& a_2^1) \sqsubset (b \Leftarrow a_1^1 \& a_2^1 \& a_3^1 \& a_4^1)$   
 $\sqsubset (b \Leftarrow a_1^1 \& a_2^1 \& a_3^1 \& a_4^1 \& a_5^1 \& \dots \& a_k^1);$
2.  $(b \Leftarrow a_1^2 \& a_2^2 \& a_3^2) \sqsubset (b \Leftarrow a_1^2 \& a_2^2 \& a_3^2 \& a_4^2 \& \dots \& a_m^2);$
3.  $(b \Leftarrow a_1^3) \sqsubset (b \Leftarrow a_1^3 \& a_2^3 \& a_3^3) \sqsubset (b \Leftarrow a_1^3 \& a_2^3 \& a_3^3 \& a_4^3 \& \dots \& a_n^3).$

According to the definition of semantic probabilistic inference, the rules must be probabilistic laws (Definition 4). This means that rules do not include stimuli that are not a signal ones, i.e. each stimulus must increase the probability of neuron excitation, which is needed for the conditional relation.

Another feature of semantic probabilistic inference is the requirement of increasing the probability of rules during probabilistic inference  $R_1 \sqsubset R_2$  (Definition 6), i.e.  $(b \Leftarrow a_1^1 \& a_2^1) \sqsubset (b \Leftarrow a_1^1 \& a_2^1 \& a_3^1 \& a_4^1)$ . This means that the conditional relation  $(b \Leftarrow a_1^1 \& a_2^1)$  is enhanced by new stimuli  $a_3^1 \& a_4^1$  to the relation  $(b \Leftarrow a_1^1 \& a_2^1 \& a_3^1 \& a_4^1)$ , if they increase the conditional probability of neuron  $b$  excitation.

The set of semantic probabilistic inferences, which neuron detects in the process of learning, forms its probabilistic regularity model (Definition 5), predicting neuron excitation  $P_0(a)$ .

The probabilistic regularity model of the neuron becomes closer to the data (Definition 8) if at least one of its conditional relations (one of the probability laws) becomes stronger – new stimuli were added, which increase the conditional probability (probabilistic inference was realized – Definition 6). The data in this case represents a record of learning experience.

Similarity to data means that the neuron responds more accurately to conditioned stimuli, i.e. with higher conditional probability. Measure of similarity between the proba-

bilistic regularity model and data (Definition 9) is a set of conditional probabilities of neuron excitation over all its conditional relations.

Neuron training (by the training operator – Definition 10) is a transformation of regularity neuron model, in which at least one of its conditional relations is strictly increasing. Thus, the training operator is constantly increasing neuron conditional relations in virtue of experience.

It is known that in the process of elaborating conditional relations the speed of conduction impulse from the conditional stimulus to the neuron's axon became higher, i.e. the higher the probability of conditional relation is, the higher the speed of neuron response to the conditional signal. This confirms that the brain responds primarily to the high probability predictions and neurons excite to the strongest patterns with the highest conditional probabilities.

A model of neurons group, that express the expert intuition, as in the problem of approximation the expert decision-making model, considered in the following section, is the set of neuron models.

If we consider the task of modeling a Boolean function of the expert decision-making model, presented in the next section, it is necessary to consider all regularity models of neurons, predicting Boolean variables of these functions.

The theory of neural modeling fields and dynamic logic can also be represented at the neuronal level. To determine the strength of the conditional relation at the neuronal level we use weights attributed to the conditional stimulus and determining the strength of its impact on the neuron excitation. During training, these weights – as probabilities – are modified.

The main similarity between dynamic logic and probabilistic dynamic logic is in their dynamics:

- in the theory of neural modeling fields the fact that the measure of similarity changes gradually and, at first, the approximate solutions and models are found, and then the measure of similarity is specified, and more accurate models are found;
- in the case of semantic probabilistic inference – at first simple rules are used for prediction, with one or two conditional stimuli in the premise, which do not give very good conditional probability (a measure of similarity), and then the rules are built up by adding new specifying conditions to strengthen this conditional probability.

## 5. Extraction of expert decision-making models for the diagnosis of breast cancer

We have applied the developed training operator for the approximation of expert decision-making models for the diagnosis of breast cancer (Kovalerchuk et al., 2001). First, we extracted this model from the expert radiologist J.Ruiz, using special procedure for extraction of expert knowledge (Kovalerchuk, Triantaphyllou, Despande, & Vityaev, 1996; Kovalerchuk et al., 2001), based on monotone Boolean functions. Then we applied the 'Discovery' system (Kovalerchuk & Vityaev, 2000; Vityaev, 2006a), that implements the training operator, for the approximation of this model.

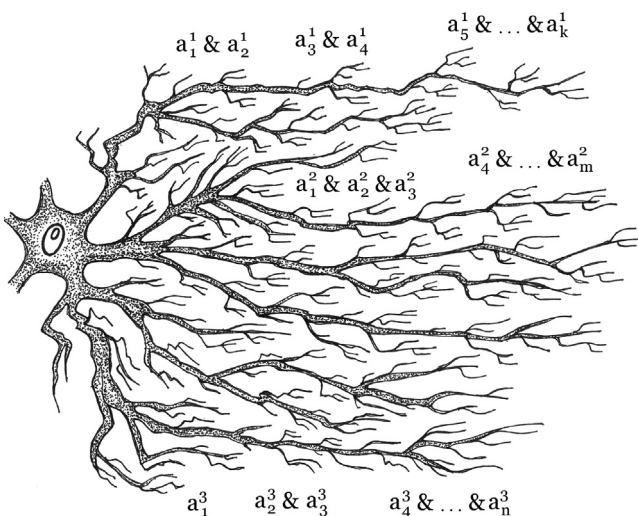


Fig. 1 Semantic probabilistic inference with neuron.

**1. Hierarchical approach.** First, we asked the expert to describe specific cases using binary features. Then, we asked the expert to make a decision on each case. A typical request for the expert had the following form: “If feature 1 has value  $v_1$ , feature 2 – value  $v_2$ , ..., feature  $n$  – value  $v_n$ , is this case suspicious for cancer or not.” Each set of values  $v_1, v_2, \dots, v_n$  is a possible clinical case. It is almost impossible to ask a radiologist to provide diagnoses for thousands of possible cases. We used a hierarchical approach together with the property of monotonicity to reduce the number of questions for the expert to several tens. First, we constructed a hierarchy of clinical features. At the bottom of the hierarchy 11 clinical binary features  $w_1, w_2, w_3, y_1, y_2, y_3, y_4, y_5, x_3, x_4, x_5$  were set (with values: 1 – “suspicious for cancer”, 0 – “non-cancerous case”). The expert found these 11 features could be organized into a hierarchy by introducing two generalized features:  $x_1$  – “the number and amount of calcification”, depending on the features  $w_1, w_2, w_3$ :

- $w_1$  – the number of calcifications/cm<sup>3</sup>,
- $w_2$  – the amount of calcification/cm<sup>3</sup>,
- $w_3$  – total number of calcifications. and  $x_2$  – “the shape and density of calcification”, depending on the features  $y_1, y_2, y_3, y_4, y_5$ :
- $y_1$  – ‘irregularity in the form of individual calcifications’
- $y_2$  – ‘variations in the form of calcification’
- $y_3$  – ‘variations in the size of calcification’
- $y_4$  – ‘variations in the density of calcification’
- $y_5$  – ‘the density of calcification’.

We will consider the feature  $x_1$  as function  $g(w_1, w_2, w_3)$  and the feature  $x_2$  as function  $x_2 = h(y_1, y_2, y_3, y_4, y_5)$  that are to be found. The result is a decomposition of the problem in the form of the following function:  $f(x_1, x_2, x_3, x_4, x_5) = f(g(w_1, w_2, w_3), h(y_1, y_2, y_3, y_4, y_5), x_3, x_4, x_5)$ .

**2. Monotonicity.** In terms of introduced features and functions we can present clinical cases in terms of vectors with five generalized variables ( $x_1, x_2, x_3, x_4, x_5$ ). We consider two clinical cases represented with vectors: (10110) and (10100). If the radiologist correctly diagnoses the case (10100) as suspicious for cancer, then using the property of monotonicity we can conclude that the case (10110) should also be suspicious for cancer, as there are more features with the value 1 conducting cancer. The expert agrees with the assumption of monotonicity of the functions  $f(x_1, x_2, x_3, x_4, x_5)$  and  $h(y_1, y_2, y_3, y_4, y_5)$ .

## 6. Extraction of expert decision-making rules

We describe an interview with an expert using a minimal sequences of questions to fully restore functions  $f$  and  $h$ . These sequences are based on the fundamental Hansel’s lemma (Hansel, 1966; Kovalerchuk et al., 1996). We shall omit the description of mathematical details which can be found in (Kovalerchuk et al., 2001).

We consider Table 1. Columns 2 and 3 represent the values of functions  $f$  and  $h$  that need to be restored. We omit

the restoring of function  $g(w_1, w_2, w_3)$ , because it takes only a few questions to restore it. All 32 possible cases for five Boolean variables  $\langle x_1, x_2, x_3, x_4, x_5 \rangle$  are presented in column 1. These cases were grouped in a Hansel chains (Hansel, 1966; Kovalerchuk et al., 1996). The sequence of chains begins with a short chain #1 – (01100) < (11100). The largest chain #10 includes 6 ordered cases: (00000) < (00001) < (00011) < (00111) < (01111) < (11111). The chains are numbered from 1 to 10, and each case has its own number in the chain, for example, 1.2 is the second case in the first chain. Asterisks in columns 2 and 3 represent the responses from the expert, for example, 1\* for the case (01100) in column 3 means that the expert said ‘Yes’ (‘suspicious for cancer’) about this case. Responses for some other cases in columns 2 and 3 are automatically derived in virtue of the property of monotonicity. The value  $f(01100) = 1$  for the case 1.1 can be extended to the cases of 1.2, 6.3., 7.3 due to monotonicity. Values of monotone Boolean function  $h$  are computed similarly, only the attributes, for example, in the sequence (10010) are interpreted as features  $y_1, y_2, y_3, y_4, y_5$ . Hansel chains remain the same if the number of variables does not change.

Columns 4 and 5 list the cases on which you can expand the values of the functions  $f$  and  $h$  without consulting the expert. Column 4 is the expansion of value 1 of the functions and column 5 is for expansion of value 0 of the functions. If the expert gives the answer  $f(01100) = 0$  for the case (01100), this value must be expanded in column 2 to the cases 7.1 (00100) and 8.1 (01000), written in column 5. So we should not ask the expert about the cases 7.1 and 8.1, as they follow from the monotonicity. The answer  $f(01100) = 0$  cannot be extended to the case 1.2 with the value  $f(11100)$  so the expert should give an answer about the case  $f(11100)$ . If their answer is  $f(11100) = 0$ , this value should be extended to the cases 5.1. and 3.1 being written in column 5.

The total number of questions marked with an asterisk (\*) in columns 2 and 3 is 13 and 12 respectively. This table shows that 13 questions are needed to restore function  $f(x_1, x_2, x_3, x_4, x_5)$  and 12 questions – to restore function  $h(y_1, y_2, y_3, y_4, y_5)$ . This is only 37.5% of the 32 possible questions. The total number of questions required to restore these functions without the monotonicity condition and hierarchy is  $2 * 2^{11} = 4096$ .

## 7. Decision rules (model) derived from the expert

We can find Boolean functions  $f(x_1, x_2, x_3, x_4, x_5)$  and  $h(y_1, y_2, y_3, y_4, y_5)$  according to Table 1 as follows:

1. We need to find all lowest ones for all chains and present them in the form of conjunctions.
2. We need to take the disjunction of obtained conjunctions.
3. We need to eliminate unnecessary (inferring from the others) conjunctions.

By column 1 and 3 we find

$$x_2 = h(y_1, y_2, y_3, y_4, y_5) = y_2 y_3 \vee y_2 y_4 \vee y_1 y_2 \vee y_1 y_4 \vee y_1 y_3 \vee y_2 y_3 y_5 \vee y_2 \vee y_1 \vee y_3 y_4 y_5 \equiv y_2 \vee y_1 \vee y_3 y_4 y_5.$$

**Table 1** The dynamic sequence of questions to the expert.

Case	$f$	$h$	Expansion on the monotonicity		Chain	Case
			4: $1 \mapsto 1$	5: $0 \mapsto 0$		
1	2	3	4: $1 \mapsto 1$	5: $0 \mapsto 0$	6	7
(01100)	1*	1*	1.2, 6.3, 7.3	7.1, 8.1	Chain 1	1.1
(11100)	1	1	6.4, 7.4	5.1, 3.1		1.2
(01010)	0*	1*	2.2, 6.3, 8.3	6.1, 8.1	Chain 2	2.1
(11010)	1*	1	6.4, 8.4	3.1, 6.1		2.2
(11000)	1*	1*	3.2	8.1, 9.1	Chain 3	3.1
(11001)	1	1	7.4, 8.4	8.2, 9.2		3.2
(10010)	0*	1*	4.2, 9.3	6.1, 9.1	Chain 4	4.1
(10110)	1*	1	6.4, 9.4	6.2, 5.1		4.2
(10100)	1*	1*	5.2	7.1, 9.1	Chain 5	5.1
(10101)	1	1	7.4, 9.4	7.2, 9.2		5.2
(00010)	0	0*	6.2, 10.3	10.1	Chain 6	6.1
(00110)	1*	0*	6.3, 10.4	7.1		6.2
(01110)	1	1	6.4, 10.5			6.3
(11110)	1	1	10.6			6.4
(00100)	1*	0*	7.2, 10.4	10.1	Chain 7	7.1
(00101)	1	0*	7.3, 10.4	10.2		7.2
(01101)	1	1*	7.4, 10.5	8.2, 10.2		7.3
(11101)	1	1	5.6			7.4
(01000)	0	1*	8.2	10.1	Chain 8	8.1
(01001)	1*	1	8.3	10.2		8.2
(01011)	1	1	8.4	10.3		8.3
(11011)	1	1	10.6	9.3		8.4
(10000)	0	1*	9.2	10.1	Chain 9	9.1
(10001)	1*	1	9.3	10.2		9.2
(10011)	1	1	9.4	10.3		9.3
(10111)	1	1	10.6	10.4		9.4
(00000)	0	0	10.2		Chain 10	10.1
(00001)	0*	0	10.3			10.2
(00011)	1*	0	10.4			10.3
(00111)	1	1*	10.5			10.4
(01111)	1	1	10.6			10.5
(11111)	1	1				10.6
Questions	13	12				

Function  $g(w_1, w_2, w_3) = w_2 \vee w_1 w_3$  can be obtained by asking  $2^3 = 8$  questions to the expert. By columns 1 and 2 we find

$$f(\bar{x}) = x_2 x_3 \vee x_1 x_2 x_4 \vee x_1 x_2 \vee x_1 x_3 x_4 \vee x_1 x_3 \vee x_3 x_4 \vee x_3 \\ \vee x_2 x_5 \vee x_1 x_5 \vee x_4 x_5 \equiv x_1 x_2 \vee x_3 \vee (x_2 x_1 x_4) x_5 \equiv \\ (w_2 \vee w_1 w_3)(y_1 \vee y_2 \vee y_3 y_4 y_5) \vee x_3 \vee (y_1 \vee y_2 \vee y_3 y_4 y_5)(w_2 \vee w_1 w_3) x_4 x_5.$$

## 8. Approximation of the expert model with the training operator

To approximate the expert decision-making model we used the software system 'Discovery', which implements the training operator. Several tens of diagnostic rules that approximate this expert model were discovered. They were statistically significant with respect to the statistical criterion for the selection of the rules for levels 0.01, 0.05, 0.1. The rules were found for 156 cases (73 malignant, 77 benign, 2 suspicious for cancer and 4 with a mixed diagnosis) (Kovalerchuk et al., 2001).

The set of rules was tested by the cross-validation method. The diagnosis was obtained for 134 cases (in 22 cases the diagnosis was not made). Accuracy of the diagnosis was 86%. Incorrect diagnosis was obtained in 19 cases (14% of all diagnostic cases). Type I error made 5.2% (7 malignant cases were diagnosed as benign) and type II error made 8.9% (12 benign cases were diagnosed as malignant). Some of these rules are given in Table 2. The following table gives examples of rules along with their statistical significance by Fisher-test. In this table: 'NUM' is the number of calcification in  $\text{cm}^3$ ; 'VOL' is the amount in  $\text{cm}^3$ ; 'TOT' is the total number of calcification; 'DEN' is the density of calcification; 'VAR' is variations in the form of calcification; 'SIZE' is variations in the size of calcification; 'IRR' is the irregularity in the form of calcification; 'SHAPE' is the form of calcification.

We considered three levels of similarity measure for the training operator 0.75, 0.85 and 0.95, which meant that the conditional probabilities of all probability laws of probabilistic regularity model are higher than or equal to these values.

Higher levels of conditional probability reduce the number of rules and diagnosed cases, but increase the accuracy



**Table 2** Examples of detected diagnostic rules.

Diagnostic rule	<i>f</i> -test	<i>f</i> -test value			Cross-validation accuracy (%)
		0.01	0.05	0.1	
If 10 < NUM < 20 and VOL > 5 then <i>malignant</i>	NUM	0.0029	+	+	93.3
	VOL	0.0040	+	+	
If TOT > 30 and VOL > 5 and DEN is <i>medium</i> then <i>malignant</i>	TOT	0.0229	–	+	100
	VOL	0.0124	–	+	
	DEN	0.0325	–	+	
If VAR is <i>detectable</i> and 10 < NUM < 20 and IRR is <i>medium</i> then <i>malignant</i>	VAR	0.0044	+	+	100
	NUM	0.0039	+	+	
	IRR	0.0254	–	+	
If SIZE is <i>medium</i> and SHAPE is <i>weak</i> and IRR is <i>weak</i> then <i>benign</i>	SIZE	0.0150	–	+	92.86
	SHAPE	0.0114	–	+	
	IRR	0.0878	–	–	

of diagnosis and similarity to the data. Forty-four statistically significant diagnostic rules were found with the level of criterion (*f*-test) 0.05 and the conditional probability of not less than 0.75; 30 rules with the conditional probability of not less than 0.85; 18 rules with the conditional probability of not less than 0.95. Of these, 30 rules gave accuracy of 90% and 18 rules gave the accuracy of 96.6% with only 3 cases of type II errors (3.4%).

As is clear from these results, the required similarity to the data 0.75, 0.85 and 0.95 is less than 86%, 90% and 96.6% accuracy obtained with the cross-validation method. This is due to the Fisher’s exact test that is used to examine the statistical significance of the increase of the conditional probabilities in semantic probabilistic inference, used by the Discovery system. This prevents retraining of the Discovery system. Other experiments (Kovalerchuk & Vityaev, 1998; Kovalerchuk & Vityaev, 2000) show rather high Discovery retrain resistance.

As a result the training operator managed to approximate reported data accurately enough. This result turned out to be better than using neural networks (Brainmaker) which gave 100% accuracy on the training, but only 66% on cross-validation. Decision trees (software SIPINA) gave 76–82% accuracy on the training.

### 9. Comparison of the expert model and its approximation by the training operator

To compare the model (rule), obtained with Discovery system, and the model (rules) extracted from an expert, we asked the expert to evaluate the first of the rules. Here are some of the rules detected by the Discovery system and radiologist’s commentaries on their compliance with its decision-making model.

*IF* “the total number of calcification” is more than 30 and “volume” is greater than 5 cm<sup>3</sup> and “density” of calcification is medium, *THEN* “suspicious for cancer.”

The significance of *f*-test is 0.05. Cross-validation accuracy is 100%. Radiologist commentary – “this rule is promising but might be risky.”

*IF* “variations in the form of calcification” are detectable and “the number of calcification” is between 10 and 20 and “the irregularity in the form of calcification” is medium *THEN* “suspicious for cancer.”

The significance of *f*-test is 0.05. Cross-validation accuracy is 100%. Radiologist commentary – “I would trust this rule”.

*IF* “variations in the size of calcification” are medium and “variations in the form of calcification” are weak and “the irregularity in the form of calcification” is weak *THEN* “benign.”

The significance of *f*-test is 0.05. Cross-validation accuracy is 92.86%. Radiologist commentary – “I would trust this rule”.

Thus, the training operator found the rules corresponding well enough with expert’s intuition. More detailed comparison of discovered rules and expert rules is given in (Kovalerchuk et al., 2001).

### 10. Conclusion

Thus, we interpreted the theory of neural modeling fields and dynamic logic in logical and probabilistic terms and proposed a PDLC for modeling cognitive processes. We demonstrated that the PDLC also solve artificial intelligence problems – combinatorial complexity and logic and probability synthesis. We also interpret the corresponding cognitive processes at the neural level.

We applied the PDLC for the approximation of expert decision-making model for the breast cancer diagnosis. First, we extracted this model from the expert, using

special procedure, based on monotone Boolean functions. Then we applied PDLC, training operator and Discovery system for learning the decision-making model from data. Because of the training operator may be interpreted at the neural level, this model may be considered as a result of the expert brain learning. In the last section we demonstrated that the model extracted from the expert, using Table 1 list, and the model obtained by the expert brain learning are in good correspondence. This demonstrates that the presented PDLC in good correspondence with the learning cognitive process.

## Acknowledgements

This work is supported by a RFBR Grant 11-07-00560-a; integration projects of the Siberian Branch of the Russian Academy of Sciences 3, 87, 136, the work was also supported by the Council of the President of the Russian Federation grants and state support of leading scientific schools SS-3606.2010.1.

## References

- Caporale, N., & Dan, Y. (2008). Spike timing-dependent plasticity: A Hebbian learning rule. *Annual Review of Neuroscience*, 31, 25–46.
- Cozman, F. G., Haenni, R., Romeijn Jan-Willem Russo, F., Wheeler, G. R., et al (2009). Combining probability and logic. *Journal of Applied Logic*, 7, 131–135.
- Demin, A. V., & Vityaev, E. E. (2008). Logical model of adaptive control system. *Neuroinformatics*, 3(1), 79–107, in Russian.
- Gabbay, D., Johnson, R., Ohlbach, H. J., & Woods, J. (Eds.). (2002). *Handbook of the logic of inference and argument: The turn toward the practical. Studies in logic and practical reasoning* (vol. 1). Elsevier.
- Gerstner, W., & Kistler, W. M. (2002). Mathematical formulations of Hebbian learning. *Biological Cybernetics*, 87, 404–415.
- Halpern, J. Y. (1990). An analysis of first-order logics of probability. *Artificial Intelligence*, 46, 311–350.
- Hansel, G. (1966). Sur le nombre des fonctions Booleennes monotones de n variables. *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences*, 262(20), 1088–1090, in French.
- Hebb, D. O. (1949). The organization of behavior. A neurophysiological theory, NY.
- Hyafil, L., & Rivest, R. L. (1976). Constructing optimal binary decision trees is NP-complete. *Information Processing Letters*, 5(1), 15–17.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. New York: Oxford Univ. Press.
- Kovalerchuk, B., Triantaphyllou, E., Despande, A., & Vityaev, E. (1996). Interactive learning of monotone boolean function. *Information Sciences*, 94(1–4), 87–118.
- Kovalerchuk, B., & Vityaev, E. (1998). Discovering lawlike regularities in financial time series. *Journal of Computational Intelligence in Finance*, 6(3), 12–26.
- Kovalerchuk, B. Y., & Vityaev, E. E. (2000). *Data mining in finance: Advances in relational and hybrid methods*. Kluwer Academic Publisher.
- Kovalerchuk, B. Y., Vityaev, E. E., & Ruiz, J. F. (2001). Consistent and complete data and "expert" mining in medicine. In *Medical data mining and knowledge discovery* (pp. 238–280). Springer.
- Kovalerchuk, B. Y., & Perlovsky, L. I. (2008). Dynamic logic of phenomena and cognition. *IJCNN*, 3530–3537.
- Perlovsky, L. I. (1998). Conundrum of combinatorial complexity. *IEEE Transactions on PAMI*, 20(6), 666–670.
- Perlovsky, L. I. (2006). Toward physics of the mind: Concepts, emotions, consciousness, and symbols. *Physics of Life Reviews*, 3, 23–55.
- Perlovsky, L. I. (2007). Neural networks, fuzzy models and dynamic logic. In R. Kohler & A. Mehler (Eds.), *Aspects of automatic text analysis (Festschrift in Honor of Burghard Rieger)* (pp. 363–386). Germany: Springer.
- Pfister, J. P., Barber, D., & Gerstner, W. (2003). Optimal Hebbian learning: A probabilistic point of view. In Kaynak et al. (Eds.), *ICANN/ICONIP. LNCS* (vol. 2714, pp. 92–98). Berlin Heidelberg: Springer-Verlag.
- Probabilistic models of cognition (2006). *Special Issue of the Journal: Trends in Cognitive Science*, 10(7), 287–344.
- Raedt, L., Frascioni, P., Kersting, K., & Muggleton, S. H. (Eds.). (2008). *Probabilistic inductive logic programming. LNAI* (vol. 4911). Berlin: Springer-Verlag.
- Scientific Discovery website <<http://math.nsc.ru/AP/ScientificDiscovery>>.
- Simonov, P. V. (1981). *Emotional brain*. Moscow: Science, in Russian.
- Smerdov, S. O., & Vityaev, E. E. (2009). Logic, probability and learning synthesis: Formalization of prediction. *Siberian mathematical electronic reports* (vol. 6, pp. 340–365). Novosibirsk: Sobolev institute of mathematics, in Russian.
- The Probabilistic Mind (2008). Prospects for Bayesian cognitive science. In N. Chater, M. Oaksford (Eds.), Oxford University Press.
- Vityaev, E. E. (2006a). Knowledge discovery. *Computational cognition. Cognitive process models*. Novosibirsk: Novosibirsk State University Press, in Russian.
- Vityaev, E. E. (2006b). The logic of prediction. In S. S. Goncharov, R. Downey, & H. Ono (Eds.), *Mathematical logic in Asia 2005* (pp. 263–276). World Scientific.
- Vityaev, E. E., & Smerdov, S. O. (2009). New definition of prediction without logical inference. In B. Kovalerchuk (Ed.), *Proceedings of the IASTED international conference on computational intelligence (CI 2009)* (pp. 48–54). Honolulu, Hawaii.