Neural Networks 22 (2009) 247–257

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

2009 Special Issue

Language and cognition

Leonid Perlovsky *

Harvard University, SEAS, Cambridge, USA
The AF Research Laboratory, Sensors Directorate, Hanscom AFB, USA

ARTICLE INFO

Article history:
Received 5 January 2009
Received in revised form 9 March 2009
Accepted 16 March 2009

Keywords:
Language
Cognition
Emotions
Knowledge instinct
Dynamic logic
Mind
Heterarchy
Hierarchy

ABSTRACT

What is the role of language in cognition? Do we think with words, or do we use words to communicate made-up decisions? The paper briefly reviews ideas in this area since 1950s. Then we discuss mechanisms of cognition, recent neuroscience experiments, and corresponding mathematical models. These models are interpreted in terms of a biological drive for cognition. Based on the Grossberg–Levine theory of drives and emotions, we identify specific emotions associated with the need for cognition. We demonstrate an engineering application of the developed technique, which significantly improves detection of patterns in noise over the previous state-of-the-art. The developed mathematical models are extended toward language. Then we consider possible brain–mind mechanisms of interaction between language and cognition. A mathematical analysis imposes restrictions on possible mechanisms. The proposed model resolves some long-standing language–cognition issues: how the mind learns correct associations between words and objects among an astronomical number of possible associations; why kids can talk about almost everything, but cannot act like adults, what exactly are the brain–mind differences; why animals do not talk and think like people. Recent brain imaging experiments indicate support for the proposed model. We discuss future theoretical and experimental research.

Published by Elsevier Ltd

1. Nativism, cognitivism, evolutionism

Complex innate mechanisms of the mind were not appreciated in the first half of the last century. Thinking of mathematicians and intuitions of psychologists and linguists were dominated by logic. Considered mechanisms of logic were not much different for language or cognition; both were based on logical statements and rules. Even fundamental Gödelian theory (Gödel, 1931/1994) establishing the deficiency of logic did not move thinking about the mind away from logic.

Contemporary linguistic interests in the mind mechanisms of language were initiated in the 1950s by Chomsky (1965). He identified the first mysteries about language that science had to resolve. "Poverty of stimulus" addressed the fact that the tremendous amount of knowledge needed to speak and understand language is learned by every child around the world even in the absence of formal training. It has seemed obvious to Chomsky that surrounding language cultures do not carry enough information for a child to learn language, unless specific language learning mechanisms are inborn in the mind of every human being. This inborn mechanism should be specific enough for learning complex language grammars and still flexible enough so that a child of any ethnicity from any part of the world would learn whichever language is spoken around, even if he or she is raised on the other side of the globe. Chomsky called this inborn learning mechanism Universal Grammar and set out to discover its mechanisms. He emphasized the importance of syntax and thought that language learning is independent of cognition. This approach to language based on innate mechanisms, is called *nativism*.

Chomsky and his school initially used available mathematics of logical rules, similar to rule systems of artificial intelligence. In 1981, Chomsky (Chomsky, 1981) proposed a new mathematical paradigm in linguistics, *rules and parameters*. This was similar to model-based systems emerging in mathematical studies of cognition. Universal properties of language grammars were supposed to be modeled by parametric rules or models, and specific characteristics of grammar of a particular language were fixed by parameters, which every kid could learn from a limited exposure to the surrounding language. Another fundamental change of Chomsky's ideas (Chomsky, 1995) was called *the minimalist program*. It aimed at simplifying the rule structure of the mind mechanism of language. Language was modeled in closer interactions to other mind mechanisms, closer to the meaning, but stopped at an interface between language and meaning. Chomsky's linguistics still assumes that meanings appear independently from language. Logic is the main mathematical modeling mechanism.

Many linguists disagreed with separation between language and cognition in Chomsky's theories. Cognitive linguistics emerged in the 1970s to unify language and cognition, and explain

* Tel.: +1 781 377 1728; fax: +1 781 377 8984.
E-mail address: leonid@deas.harvard.edu.

0893-6080/$ – see front matter. Published by Elsevier Ltd
doi:10.1016/j.neunet.2009.03.007

creation of meanings. Cognitive linguistics rejected Chomsky's idea about a special module in the mind devoted to language. The knowledge of language is no different from the rest of cognition, and is based on conceptual mechanisms. It is embodied and situated in the environment. Related research on construction grammar argues that language is not compositional, not all phrases are constructed from words using the same syntax rules and maintaining the same meanings; metaphors are good examples (Croft & Cruse, 2004; Evans & Green, 2006; Ungerer & Schmid, 2006). Cognitive linguistics so far has not led to computational linguistic theory explaining how meanings are created. Formal apparatus of cognitive linguistics is dominated by logic.

Evolutionary linguistics emphasized that language evolved together with meanings. A fundamental property of language is that it is transferred from generation to generation, and language mechanisms are shaped by this process. (Christiansen & Kirby, 2003; Hurford, 2008). Evolutionary linguistics by simulation of societies of communicating agents (Brighton, Smith, & Kirby, 2005) demonstrated the emergence of a compositional language.

2. Cognition, dynamic logic, and the knowledge instinct

Consider a seemingly simple experiment. Close your eyes and imagine an object in front of you. The imagined image is vague, not as crisp and clear as with opened eyes. As we open eyes; the object becomes crisp and clear. It seems to occur momentarily, but actually it takes 1/5th of a second. This is a very long time for neural brain mechanisms – hundreds of thousands of neural interactions. Let us also note: with opened eyes we are not conscious about initially vague imagination, we are not conscious about the entire 1/5th of a second, we are conscious only about the end of this process: crisp, clear object in front of our eyes. The explanation of this experiment has become simple after many years of research that have found out what goes on in the brain during these 1/5th of a second.

2.1. Instincts, emotions, concepts

Explaining this experiment requires us to consider mechanisms of concepts, instincts, and emotions. We perceive and understand the world around due to the mechanism of concepts. Concepts are like internal models of objects and situations; this analogy is quite literal, e.g., during visual perception of an object, a concept-model of the object stored in memory projects an image (top-down signals) onto the visual cortex, which is matched there to an image projected from the retina (bottom-up signal; this simplified description will be refined later; see Grossberg (1988)).

The mechanism of concepts evolved for instinct satisfaction. The word instinct is not used currently in the psychological literature; the reason is that the notion of instinct was mixed up with instinctual behavior and other not very useful ideas. We use the word instinct to denote a simple inborn, non-adaptive mechanism described in Grossberg and Levine (1987). Instinct is a mechanism of the internal "sensor", which measures vital body parameters, such as blood pressure, and indicate to the brain when these parameters are out of safe range. This simplified description will be sufficient for our purposes, more details could be found in Gnadt and Grossberg (2008) and Grossberg and Seidman (2006) and the references therein. We have dozens of such sensors, measuring sugar level in blood, body temperature, pressure at various parts, etc.

According to instinctual–emotional theory (Grossberg & Levine, 1987), communicating satisfaction or dissatisfaction of instinctual needs from instinctual parts of the brain to decision making parts of the brain is performed by emotional neural signals. The word emotion refers to several neural mechanisms in the brain (Juslin & Västfjäll, 2008); in this paper we always refer to the mechanism connecting conceptual and instinctual brain regions. Perception and understanding of concept-models corresponding to objects or situations that can potentially satisfy an instinctual need receive preferential attention and processing resources in the mind.

Projection of top-down signals from a model to the visual cortex primes or makes visual neurons to be more receptive to matching bottom-up signals. This projection produces imagination that we perceive with closed eyes, as in the closed–open eye experiment. Conscious perception occurs, as mentioned, after top-down and bottom-up signals match. The process of matching for a while presented difficulties to mathematical modeling, as discussed below.

2.2. Combinatorial complexity, logic, and dynamic logic

Perception and cognition abilities of computers still cannot compete with those of kids and animals. Most algorithms and neural networks suggested since 1950s for modeling perception and cognition, as discussed in Perlovsky (2006a), faced difficulty of combinatorial complexity (CC). Rule systems of artificial intelligence in the presence of variability has grown in complexity: rules have become contingent on other rules, and rule systems faced CC. Algorithms and neural networks designed for learning have to be trained to understand not only individual objects, but also combinations of objects, and thus faced CC of training. Fuzzy systems required a fuzziness level to be set appropriately in different parts of systems, also degrees of fuzziness vary in time, an attempt to select efficient levels of fuzziness would lead to CC.

These CC difficulties were related to Gödelian limitations of logic, they were manifestations of logic inconsistency in finite systems (Perlovsky, 2000). Even approaches designed specifically to overcome logic limitations, such as fuzzy logic and neural networks, encountered logical steps in their operations: neural networks are trained using logical procedures (e.g. "this is a chair"), and fuzzy systems required logical selection of the degree of fuzziness.

To overcome limitations of logic, dynamic logic was proposed (Perlovsky, 2000, 2006a; Perlovsky & McManus, 1991). In the next section we summarize the mathematical description of dynamic logic, here we describe it conceptually. Whereas logic works with statements (e.g. "this is a chair"), dynamic logic is a process from vague to crisp, from vague statement, decision, plan, to crisp ones. It could be viewed as fuzzy logic that automatically sets a degree of fuzziness corresponding to the accuracy of learning models.

Dynamic logic corresponds to the open–close eye experiment: initial states of models are vague. This experiment was recently performed with much more details using brain imaging. Bar et al. (2006) used functional Magnetic Resonance Imaging (fMRI) to obtain high spatial resolution of processes in the brain, which they combined with magneto-encephalography (MEG), measurements of the magnetic field next to the head, which provided high temporal resolution of the brain activity. Combining these two techniques the experimenters were able to receive high resolution of cognitive processes in space and time. Bar et al. concentrated on three brain areas: early visual cortex, object recognition area (fusiform gyrus), and object information semantic processing area (OFC). They demonstrated that OFC is activated 130 ms after the visual cortex, but 50 ms before object recognition area. This suggests that OFC represents the cortical source of top-down facilitation in visual object recognition. This top-down facilitation was unconscious. In addition they demonstrated that the imagined image generated by top-down signals facilitated from OFC to cortex is *vague*, similar to the closed–open eye experiment. Conscious perception of an object occurs when vague projections become crisp and match the crisp and clear image from the retina, and an object recognition area is activated.

2.3. The knowledge instinct and neural modeling field theory

The process of matching concept-models in memory to bottom-up signals coming from sensory organs is necessary for perception; otherwise an organism will not be able to perceive the surroundings and will not be able to survive. Therefore humans and higher animals have an inborn drive to fit top-down and bottom-up signals. We call this mechanism the instinct for knowledge (1991 (Perlovsky, 2006a)). This mechanism is similar to other instincts in that our mind has a sensor-like mechanism that measures a similarity between top-down and bottom-up signals, between concept-models and sensory percepts. Brain areas participating in the knowledge instinct were discussed in Levin and Perlovsky (2008). As discussed in that publication, biologists considered similar mechanisms since 1950s; without a mathematical formulation, however, its fundamental role in cognition was difficult to discern. All learning algorithms have some models of this instinct, maximizing correspondence between sensory input and an algorithm internal structure (knowledge in a wide sense). According to Grossberg and Levine (1987) instinct–emotion theory, satisfaction or dissatisfaction of every instinct is communicated to other brain areas by emotional neural signals. We feel these emotional signals as harmony or disharmony between our knowledge-models and the world. At lower layers of everyday object recognition these emotions are usually below the level of consciousness; at higher layers of abstract and general concepts this feeling of harmony or disharmony could be strong, as discussed in Perlovsky (2006b) it is a foundation of our higher mental abilities. We summarize now a mathematical theory combining the discussed mechanisms of cognition as interaction between top-down and bottom-up signals at a single layer in multi-layer heterarchical system following Perlovsky (2006a).

Neurons are enumerate by index $n = 1, \ldots, N$. These neurons receive bottom-up input signals, $\mathbf{X}(n)$, from lower layers in the processing heterarchy. The word heterarchy is used by many neural and cognitive scientists to designate that the mind is organized in an approximate hierarchy; this hierarchy is not exact, cross-layer interactions are abundant (Grossberg, 1988); we would use *hierarchy* for simplicity. $\mathbf{X}(n)$ is a field of bottom-up neuronal synapse activations, coming from neurons at a lower layer. Top-down, or priming signals to these neurons are sent by concept-models, $\mathbf{M}_h$ $(\mathbf{S}_h, n)$; we enumerate models by index $h = 1, \ldots, H$. Each model is characterized by its parameters, $\mathbf{S}_h$. Models *represent* signals in the following sense. Say, signal $\mathbf{X}(n)$, is coming from sensory neurons activated by object h, characterized by parameters $\mathbf{S}_h$. These parameters may include position, orientation, or lighting of an object h. Model $\mathbf{M}_h$ $(\mathbf{S}_h, n)$ predicts a value $\mathbf{X}(n)$ of a signal at neuron n. For example, during visual perception, a neuron n in the visual cortex receives a signal $\mathbf{X}(n)$ from retina and a priming signal $\mathbf{M}_h$ $(\mathbf{S}_h, n)$ from an object concept-model h. A neuron n is activated if both a bottom-up signal from lower layer input and a top-down priming signal are strong. Various models compete for evidence in the bottom-up signals, while adapting their parameters for better match as described below. This is a simplified description of perception. Models $\mathbf{M}_h$ specify a field of primed neurons {n}, hence the name for this modeling architecture, *modeling fields*.

The knowledge instinct maximizes a similarity measure between top-down and bottom-up signals,

$$L(\{\mathbf{X}\}, \{\mathbf{M}\}) = \prod_{n\in N} \sum_{h\in H} r(h) l(n|h). \tag{1}$$

Here $l(n|h)$ is a partial similarity of a bottom-up signal in pixel n given that it originated from concept-model h; functional shape of $l(n|h)$ often can be taken as a Gaussian function of $\mathbf{X}$(n) with the mean $\mathbf{M}_h$ $(\mathbf{S}_h, n)$. Partial similarities are normalized on objects (or concepts) h being definitely present, and coefficient $r(h)$ estimate a probability of them actually being present. Similarity L accounts for all combinations of signals n coming from any model h, hence the huge number of items H^N in Eq. (1); this is a basic reason for combinatorial complexity of most algorithms. From time to time a system forms a new concept-model, while retaining an old one as well; alternatively, old concepts are sometimes merged or eliminated. This requires a modification of the similarity measure (1); the reason is that more models always result in a better fit between the models and data. Therefore similarity (1) has to be reduced using a "skeptic penalty function", $p(N, M)$ that grows with the number of models M, and this growth is steeper for a smaller amount of data N.

The learning instinct demands maximizing the similarity L over model parameters $\mathbf{S}$. Dynamic logic maximizes similarity L while matching vagueness or fuzziness of similarity measures to the uncertainty of models. It starts with any unknown values of parameters $\mathbf{S}$ and defines association variables $f(h|n)$,

$$f(h|n) = r(h) l(n|h) \Big/ \sum_{h'\in H} r(h') l(n|h'). \tag{2}$$

Dynamic logic determining the Modeling Field (MF) dynamics is given by

$$\mathrm{d}f(h|n)/\mathrm{d}t = f(h|n) \sum_{h'\in H} [\delta_{hh'} - f(h'|n)] \cdot$$
$$[\partial lnl(n|h')/\partial \mathbf{M}_{h'}] \partial \mathbf{M}_{h'}/\partial \mathbf{S}_{h'} \cdot d\mathbf{S}_{h'}/\mathrm{d}t, \tag{3}$$
$$\mathrm{d}\mathbf{S}_h/\mathrm{d}t = \sum_{n\in N} f(h|n) [\partial lnl(n|h)/\partial \mathbf{M}_h] \partial \mathbf{M}_h/\partial \mathbf{S}_h, \tag{4}$$

here

$$\delta_{hh'} \text{ is 1 if } h = h', \text{ 0 otherwise.} \tag{5}$$

Initially, parameter values are not known, and uncertainty of partial similarities is high (e.g., if $l(n|h)$ is modeled by Gaussian functions, variances are high). So the fuzziness of the association variables is high. In the process of learning, models become more accurate, and association variables more crisp, as the value of the similarity increases. The number of models is determined in the learning process. The system always keeps a store of dormant models, which are vague, have low $r(h)$, and do not participate in the parameter fitting; only their parameters $r(h)$ are updated. When $r(h)$ exceeds a threshold, a model is activated; correspondingly, an active model is deactivated when its $r(h)$ falls below the threshold. MF organization is similar to ART (Carpenter & Grossberg, 1987) in that it models interaction between bottom-up and top-down signals. It is different in that it fits all models in parallel.

Dynamic logic process always converges (Perlovsky, 2000); it is proven by demonstrating that at each time step in Eqs. (3) and (4) (as long as the bottom-up signals remain constant), the knowledge instinct (1) increases; thus dynamic logic and the knowledge instinct are mathematically equivalent.

2.4. Perception example

Here we illustrate the developed technique with an example described in Perlovsky (2006a), which demonstrates that the described theory can find patterns below noise at about 100 times better in terms of signal-to-noise ratio, than previous state-of-the-art algorithms. The reason for choosing such an example is to demonstrate, in a relatively simple way, that engineering algorithms based on the mind cognitive mechanisms significantly exceed capabilities of ad hoc algorithms (Fig. 1).

As exact pattern shapes are not known and depend on unknown parameters, these parameters should be found by fitting the

Fig. 1. Finding 'smile' and 'frown' patterns in noise, an example of dynamic logic operation: (a) true 'smile' and 'frown' patterns are shown without noise; (b) actual image available for recognition (signal is below noise, signal-to-noise ratio is between 1/2 and 1/4); (c) an initial fuzzy blob-model, the fuzziness corresponds to uncertainty of knowledge; (d) through (h) show improved models at various iteration stages (total of 22 iterations). Between stages (d) and (e) the algorithm tried to fit the data with more than one model and decided, that it needs three blob-models to 'understand' the content of the data. There are several types of models: one uniform model describing noise (it is not shown) and a variable number of blob-models and parabolic models, which number, location, and curvature are estimated from the data. Until about stage (g) the algorithm 'thought' in terms of simple blob-models, at (g) and beyond, the algorithm decided that it needs more complex parabolic models to describe the data. Iterations stopped at (h), when similarity (2) stopped increasing. This example is discussed in more detail in (Linnehan et al., 2003).

pattern model to the data. At the same time it is not clear which subset of the data points should be selected for fitting. A previous state-of-the-art algorithm, multiple hypothesis testing (Singer, Sea, & Housewright, 1974) tries various subsets. In difficult cases, all combinations of subsets and models are exhaustively searched, leading to combinatorial complexity. In the current example the searched patterns are shown in Fig. 2(a) without noise, and in Fig. 2(b) with noise, as actually measured. Direct search through all combinations of models and data leads to complexity of $M^N = 10^{5000}$. A search in parameter space yields less complexity. Each pattern is characterized by a 3-parameter parabolic shape plus the 4th parameter for the average signal strength. The image size is 100×100 points, and the true number of patterns is 3, which is not known. Therefore, at least 4 patterns should be fit to the data, to decide that 3 patterns fit best. Fitting $4 \times 4 = 16$ parameters to 100×100 grid by a brute-force testing would take about 10^{40} to 10^{42} operations, still a prohibitive computational complexity.

The models and conditional similarities for this case are described in detail in Linnehan et al. (2003): a uniform model for noise, Gaussian blobs for highly fuzzy, poorly resolved patterns, and parabolic models for the patterns of interest. The number of computer operations in this example was about 10^9. Thus, a problem that was not solvable due to CC becomes solvable using dynamic logic.

In this example dynamic logic performs better than the human visual system. This is understood due to the fact that the human visual system is optimized for different type of images, not for parabolic shapes in noise.

An ability of dynamic logic to extract signals from strong noise and clutter was used in many applications; we would mention here an application to EEG signals (Kozma, Deming, Perlovsky, Levine, & Perlovsky, 2007). Potentially, EEG signals contain information about brain cognitive events; detecting these signals and estimating their parameters could be utilized to allow quadriplegics to move a computer cursor or steer their wheelchairs with their thoughts; or those playing computer games could control actions on the screen with their thoughts. The difficulty is that EEG signals are notoriously noisy. The referenced article describes a dynamic logic algorithm for extracting cognitively related events from EEG.

3. Extension to language

All linguistic theories, as reviewed at the beginning of the paper, are formulated as logical systems, and face combinatorial complexity. This is possibly why computers do not understand human language, and in particular, Google, Yahoo, and other search engines, while being immensely useful, cause so much frustrations to their users. Extension of dynamic logic to language promises to remedy the situation. Here we briefly summarize this extension following Perlovsky (2006c). The challenge in extending dynamic logic to language has been in substituting derivatives in Eqs. (3) and (4) with equivalent procedures suitable for linguistic constructs that are essentially discrete, non-differentiable structures. For example, consider a phrase "Leonid sits in a chair". A language learning procedure should be able to figure out that the gist of this phrase, its most essential part is {sit, chair}; "Leonid" can be substituted by many other nouns, and "in", "a" are even more dispensable. The main idea of dynamic logic is learning sentence structures not by trying all possible combinations of words, but by taking a "derivative" of a phrase with respect to constituent words. But of course standard mathematical definition of a derivative is not applicable to this situation in principle. Language constructs are essentially discrete and non-differentiable.

A suitable derivative-like procedure was described in Perlovsky (2006c). Here we summarize it for a word–phrase layer; where bottom-up signals are comprised of words, top-down models are phrases, and these phrase-models are learned without combinatorial complexity. The bottom-up input data, $\mathbf{X}$(n), in this "phrase-layer" MF system, are word strings, for simplicity, of a fixed length, S, $\mathbf{X}(n) = \{w_{n+1}, w_{n+2} \ldots w_{n+S}\}$. Here w_n are words from a given dictionary of size K, $W = \{w_1, w_2 \ldots w_K\}$, and n is the word position in a body of texts. A simple phrase-model often used in computational linguistics is "a bag of words", that is, a model is a subset of words from a dictionary, without any order or rules of grammar,

$$\mathbf{M}^L_h(\mathbf{S}_h, n) = \{w_{h,1}, w_{h,2} \ldots w_{h,S}\}. \tag{6}$$

A superscript L here denotes a language model, the parameters of this model are its words, $\mathbf{M}^L_h$ $(\mathbf{S}_h, n) = \mathbf{S}_h = \{w_{h,1},$