Bio-Inspired NMF Object Perception

Leonid I. Perlovsky

Abstract – **The paper describes Neural Modeling Fields (NMF) for object perception, a bio-inspired paradigm. I discuss previous difficulties in object detection and recognition, and describe how NMF overcomes this difficulties. Neural and mathematical mechanisms are described and future research directions outlined.**

I. Past Difficulties, Complexity and Logic

Biological object perception involves signals from sensory organs and internal mind's representations (memories) of objects. During perception, the mind associates subsets of signals corresponding to objects with representations of object. This produces object recognition.

Mathematical descriptions of the very first *recognition* step in this seemingly simple association-recognition-understanding process met a number of difficulties during the past fifty years. These difficulties were summarized under the notion of combinatorial complexity (CC)[1]. CC refers to multiple combinations of various elements in a complex system; for example, recognition of a scene often requires concurrent recognition of its multiple elements that could be encountered in various combinations. CC is prohibitive because the number of combinations is very large: for example, consider 100 elements (not too large a number); the number of combinations of 100 elements is 100^{100}, exceeding the number of all elementary particle events in life of the Universe; no computer would ever be able to compute that many combinations.

Algorithmic complexity was first identified in pattern recognition and classification research in the 1960s and was named "the curse of dimensionality"[2]. It seemed that adaptive self-learning algorithms and neural networks could learn solutions to any problem 'on their own', if provided with a sufficient number of training examples. The following thirty years of developing adaptive statistical pattern recognition and neural network algorithms led to a conclusion that the required number of training examples often was combinatorially large. Thus, self-learning approaches encountered *CC of learning requirements*.

Rule-based systems were proposed in the 1970's to solve the problem of learning complexity [[3, 4]]. An initial idea was that rules would capture the required knowledge and eliminate a need for learning. However in presence of variability, the number of rules grew; rules became contingent on other rules; combinations of rules had to be considered; rule systems encountered *CC of rules*.

Model-based systems were proposed in the 1980s. They used models, which depended on adaptive parameters. The idea was to combine advantages of rules with learning-adaptivity by using adaptive models. The knowledge was encapsulated in models, whereas unknown aspects of particular situations were to be learned by fitting model parameters[5]. Fitting models to data required selecting data subsets corresponding to various models. The number of subsets, however, is combinatorially large. A general popular algorithm for fitting models to the data, multiple hypothesis testing[6], is known to face CC of computations. Model-based approaches encountered *computational CC* (N and NP complete algorithms).

Later research related CC the type of logic, underlying various algorithms and neural networks. Combinatorial complexity of algorithms based on logic is related to Gödel theory: it is a manifestation of the inconsistency of logic in finite systems[7]. Various manifestations of CC are all related to formal logic and Gödel theory[8]. Rule systems rely on formal logic in a most direct way. Self-learning algorithms and neural networks rely on logic in their training or learning procedures: every training example is treated as a separate logical statement. Fuzzy logic systems rely on logic for setting degrees of fuzziness. CC of mathematical approaches to the mind is related to the fundamental inconsistency of logic.

II. Mathematical Formulation

Biological object perception is modeled by Neural Modeling Fields (NMF) as follows. NMF is a multi-level, hetero-hierarchical system. It mathematically implements several mechanisms of the mind, and this paper describes the mechanisms of perception.

This section describes a basic mechanism of interaction between two adjacent hierarchical levels of bottom-up and top-down signals (fields of neural activation). At each hierarchical level, we enumerate neurons by index n = 1,… N. These neurons receive bottom-up input signals, $\mathbf{X}(n)$, from lower levels in the processing hierarchy. $\mathbf{X}(n)$ is a field of bottom-up neuronal synapse activations, coming from neurons at a lower level. Each neuron has a number of synapses; for generality, we describe each neuron activation as a set of numbers, $\mathbf{X}(n) = \{X_d(n), d = 1,... D\}$. Top-down, or priming signals to these neurons are sent by concept-models, $\mathbf{M}_m(\mathbf{S}_m,n)$; we enumerate models by index m = 1,… M. Each model is characterized by its parameters, $\mathbf{S}_m$; in the neuron structure of the brain they are encoded by strength of synaptic connections, mathematically, we describe them as a set of numbers, $\mathbf{S}_m = \{S_{a\ m}, a = 1... A\}$. Models represent signals in the following way. Say, signal $\mathbf{X}(n)$, is coming from sensory neurons activated by object m, characterized by parameters $\mathbf{S}_m$. These parameters may include position, orientation, or lighting of an object m. Model $\mathbf{M}_m(\mathbf{S}_m,n)$ 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 $\mathbf{M}_m(\mathbf{S}_m,n)$ from an object-concept-model m. A neuron n is

Harvard University, Cambridge, MA; and the US Air Force Research Laboratory, Sensors Directorate, Hanscom, MA; leonid@seas.harvard.edu

activated if both bottom-up signal from lower-level-input and 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. The most benign everyday visual perception uses many levels from retina to object perception. The NMF premise is that the same laws describe the basic interaction dynamics at each level. Perception of minute features, or everyday objects, or cognition of complex abstract concepts is due to the same mechanism described below. Perception and cognition involve models and learning. In perception, models correspond to objects; in cognition models correspond to relationships and situations.

In NMF, bottom-up signals are unstructured data $\{\mathbf{X}(n)\}$ and output signals are recognized or formed concepts $\{m\}$. Top-down, "priming" signals are models, $\mathbf{M}_m(\mathbf{S}_m,n)$, which upon recognition become bottom-up signals for the hext, higher level.

Learning is an essential part of perception and cognition. NMF learns driven by the knowledge instinct, an internal "desire" to improve correspondence between top-down and bottom-up signals. It increases a similarity measure between the sets of models and signals, L({X},{M})**Error! Bookmark not defined.**

$$L(\{\mathbf{X}\},\{\mathbf{M}\}) = \prod_{n\in N} l(\mathbf{X}(n)). \qquad (1)$$

This expression contains a product of partial similarities, l(X(n)). Before perception occurs, the mind does not know which retinal neuron corresponds to which object. Therefore a partial similarity measure is constructed so that it treats each model as an alternative (a sum over models) for each input neuron signal. Its constituent elements are conditional partial similarities between signal $\mathbf{X}(n)$ and model $\mathbf{M}_m$, $l(\mathbf{X}(n)|m)$. This measure is "conditional" on object m being present, therefore, when combining these quantities into the overall similarity measure, L, they are multiplied by r(m), which represents the measure of object m actually being present. Combining these elements, a similarity measure is constructed as follows**Error! Bookmark not defined.**:

$$L(\{\mathbf{X}\},\{\mathbf{M}\}) = \prod_{n\in N} \sum_{m\in M} r(m)\, l(\mathbf{X}(n)\mid m). \qquad (2)$$

The learning process consists in estimating model parameters **S** and associating signals with concepts by maximizing the similarity (2). Note, all possible combinations of signals and models are accounted for in expression (2). This can be seen by expanding a sum in (2), and multiplying all the terms; it would result in H^N items, a huge number. This is the number of combinations between all signals (N) and all models (M). Here is the source of combinatorial complexity (CC) of many algorithms used in the past. For example, multiple hypothesis testing algorithms[9] attempts to maximize similarity L over model parameters and associations between signals and models, in two steps. First it takes one of the H^N items, which is one particular association between signals and models; and maximizes it over model parameters; this is performed over all items. Second, the largest item is selected (that is the best association for the best set of parameters). Such a program inevitably faces a wall of CC, the number of computations on the order of H^N.

Neural modeling fields overcomes this fundamental difficulty of many learning algorithms and solves this problem without CC by using dynamic logic[,10,11]. An important aspect of dynamic logic is matching vagueness or fuzziness of similarity measures to the uncertainty of models. Initially, parameter values are not known, and uncertainty of models is high; so is the fuzziness of the similarity measures. In the process of learning, models become more accurate and the similarity measure more crisp, the value of the similarity increases. This is the mechanism of dynamic logic.

NMF-dynamic logic is described mathematically as follows. First, assign any values to unknown parameters, $\{\mathbf{S}_m\}$. Then, compute association variables f(m|n),

$$f(m|n) = r(m)\, l(\mathbf{X}(n)|m) \,/ \sum_{m'\in M} r(m')\, l(\mathbf{X}(n)|m'). \qquad (3)$$

Next, iteratively improve the association variables and parameters, according to:

$$df(m|n)/dt = f(m|n) \sum_{m'\in M} \{[\delta_{mm'} - f(m'|n)] \bullet \qquad (1)$$

$$[\partial \ln l(n|m')/\partial \mathbf{M}_{m'}\ ]\ )/\partial \mathbf{M}_{m'}/\partial \mathbf{S}_{m'} \bullet d\mathbf{S}_{m'}/dt,$$

$$d\mathbf{S}_m/dt = \sum_{n\in N} f(m|n)[\partial \ln l(n|m)/\partial \mathbf{M}_m]\partial \mathbf{M}_m/\partial \mathbf{S}_m \qquad (4)$$

here $\delta_{mm'}$ is 1 if m=m', 0 otherwise.

Parameter t is the time of the internal dynamics of the NMF system (like a number of internal iterations).

Convergence of these equations was proven in**Error! Bookmark not defined.**. Engineering application of this theory to multiple target tracking in strong clutter was presented in[12]. It resulted in significant savings in complexity and in about two orders of magnitude improvement in signal-to-clutter ratio.

III. Example of NMF Object Perception (2)

Finding patterns below noise can be an exceedingly complex problem. If an exact pattern shape is not known and depends on unknown parameters, these parameters should be found by fitting the pattern model to the data. However, when the locations and orientations of patterns are not known, it is not clear which subset of the data points should be selected for fitting. A standard approach for solving this kind of problem, which has already been discussed, is multiple hypothesis testing []. Here, since all combinations of subsets and models are exhaustively searched, it faces the problem of combinatorial complexity. In the current example, we are looking for 'smile' and 'frown' patterns in noise shown in Fig.1a without noise, and in Fig.1b with noise, as actually measured. Each pattern is characterized by a 3-parameter parabolic shape. The image size in this example is 100x100 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. Using a multiple hypothesis testing brute-force approach will take about $M^N = 10^{6,000}$ operations. Alternatively, fitting 4x3=12 parameters to 100x100 grid by a brute-force testing of all parameter values would take about 10^{32} to 10^{40} operations, a prohibitive computational complexity in both cases.

To apply NMF and dynamic logic to this problem one needs to develop parametric adaptive models of expected patterns. The models and conditional partial similarities for this case are described in details in []: a uniform model for noise, Gaussian blobs for highly-fuzzy, poorly resolved patterns, and parabolic models for 'smiles' and 'frowns'. The number of computer operations in this example was about 10^{10}. Thus, a problem that was not solvable due to CC becomes solvable using dynamic logic.

During an adaptation process, initial fuzzy and uncertain models are associated with structures in the input signals, and fuzzy models become more definite and crisp with successive iterations. The type, shape, and number, of models are selected so that the internal representation within the system is similar to input signals: the MF concept-models represent structure-objects in the signals. The figure below illustrates operations of dynamic logic. In Fig. 1(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 –2dB and –0.7dB); (c) an initial fuzzy model, a large fuzziness corresponds to uncertainty of knowledge; (d) through (h) show improved models at various iteration stages (total of 22 iterations). Every five iterations the algorithm tried to increase or decrease the number of pattern-models. Between iterations (d) and (e) the algorithm decided, that it needs three Gaussian models for the 'best' fit. 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 used 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 stopped increasing.

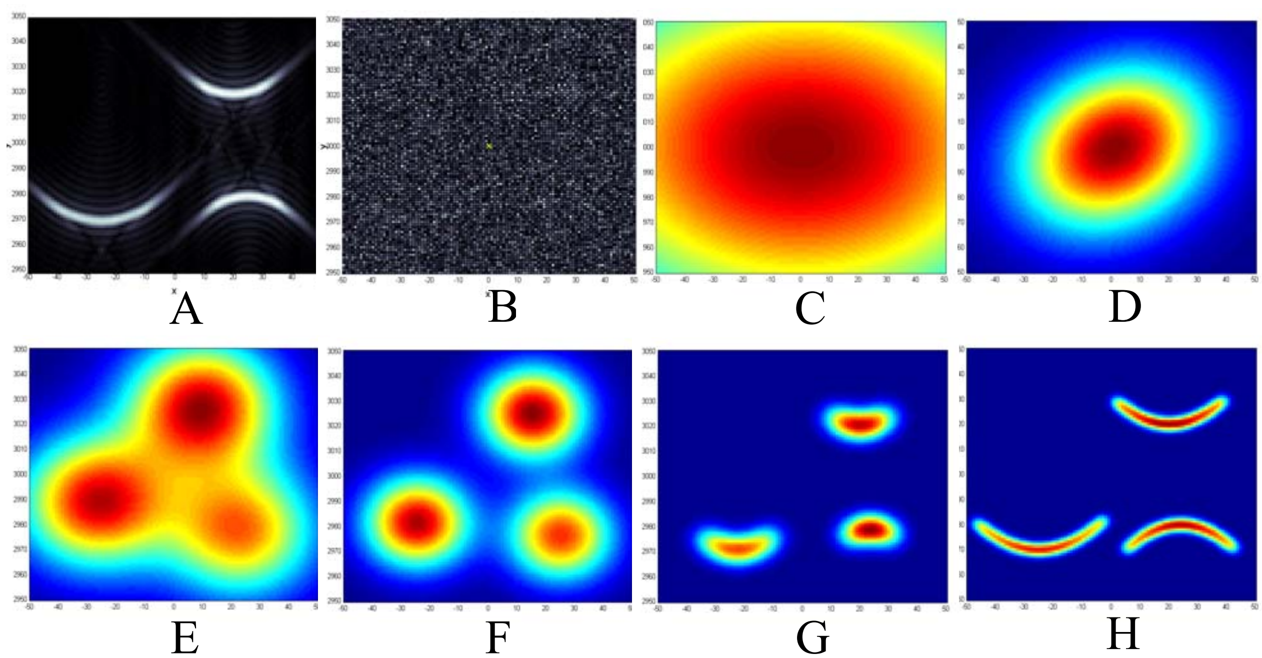

Fig.1. Finding 'smile' and 'frown' patterns in noise, an example of dynamic logic operation: (a) true 'smile' and 'frown' patterns are shown without noise; (b) actual image available for recognition (signal is below noise, signal-to-noise ratio is between –2dB and –0.7dB); (c) an initial fuzzy blob-model, 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 details in [[13]].

VI. Biological validation of Dynamic Logic

Biological experimental validation of dynamic logic can be Obtained by everyone in few seconds. Just close you eyes and imagine a familiar object that you observed in front of you just a second ago. Your imagination is vague, not as crisp as perception of the object with open eyes. Complex neural processes involved in this experiment were understood only recently**Error! Bookmark not defined.**: imagination is produced in the visual cortex by top-down signals from models in our memory. This proves that in the initial stages of perception memories-models are vague, as in dynamic logic.

Recently, detailed neurological and fMRI neuroimaging studies[14,15,16] confirmed that conscious perceptions are preceded by activation of cortex areas, where top-down signals originate from memories-models. Also, initial top-down signals (models) are driven by "low-spatial frequency," that is *vague* contents of bottom-up signals. Thus, dynamic logic processes "from vague to crisp" were confirmed by direct observations.

VII. Future Directions

The next step is to extended NMF and dynamic logic to perception of complex objects. Initial vague models can be obtained from exact models by convolving the model with a low resolution kernel. The kernel uncertainty, like in the demonstrated example, should be wide enough to account for model parameters. Parameterization of the model should allow for a reasonably good match with sensor data. If constructing an analytic expression for the image shape is unfeasible, an alternative could be to construct a weighted combination of possible views, with weights being the model parameters.

Longer-term research should concentrate first on implementing multi-layer NMF, which would be able to model higher level cognition, including complex situations. Second, it should be combined with language-learning NMF.[17],[18]. Third, this integrated cognitive-language hierarchical NMF system should be used for acquiring, in interaction with human users, language along with high-level cognition.

In parallel multi-agent systems with NMF mind can be developed for several applications.

Further experimental studies should extend results of,, to high cognitive functions, to cognition of abstract concepts, and to mechanisms of language and cognition interaction.

Acknowledgments

I am thankful to R. Brockett, R. Deming, D. Levine, R. Kozma, and B. Weijers, for discussions, help and advice, and to AFOSR for supporting part of this research under the Lab. Task 05SN02COR, PM Dr. Jon Sjogren.

References

[1] Perlovsky, L.I. (1998). *Conundrum of Combinatorial Complexity.* IEEE Trans. PAMI, **20**(6) p.666-70.

[2] Bellman, R.E. (1961). *Adaptive Control Processes.* Princeton University Press, Princeton, NJ.

[3] Minsky, M.L. (1975). A Framework for Representing Knowlege. In The Psychology of Computer Vision, ed. P. H. Whinston, McGraw-Hill Book, New York.

[4] Winston, P.H. (1984). Artificial Intelligence. 2nd edition, Addison-Wesley. Reading, MA.

[5] Singer, R.A., Sea, R.G. and Housewright, R.B. (1974). Derivation and Evaluation of Improved Tracking Filters for Use in Dense Multitarget Environments, IEEE Transactions on Information Theory, IT-20, pp. 423-432.

[6] Perlovsky, L.I., Webb, V.H., Bradley, S.R. & Hansen, C.A. (1998). *Improved ROTHR Detection and Tracking Using MLANS* . AGU Radio Science, **33**(4), pp.1034-44.

[7] Perlovsky, L.I. (1996). *Gödel Theorem and Semiotics.* Proceedings of the Conference on Intelligent Systems and Semiotics '96. Gaithersburg, MD, v.2, pp. 14-18.

[8] Perlovsky, L.I. (2001). Neural Networks and Intellect: using model based concepts. New York: Oxford University Press.

[9] Singer, R.A., Sea, R.G. and Housewright, R.B. (1974). Derivation and Evaluation of Improved Tracking Filters for Use in Dense Multitarget Environments, *IEEE Transactions on Information Theory*, **IT-20**, 423-32.

[10] Perlovsky, L.I. & McManus, M.M. (1991). Maximum Likelihood Neural Networks for Sensor Fusion and Adaptive Classification. *Neural Networks* **4 (1)**, pp. 89-102.

[11] Kovalerchuk, B. & Perlovsky, L. (2008). Dynamic Logic of Phenomena and Cognition, The World Congress on Computational Intelligence, WCCI'08, Hong Kong, China.

[12] Perlovsky, L.I. and Deming, R.W. (2007). *Neural Networks for Improved Tracking*, IEEE Trans. Neural Networks, **18(6)**, pp. 1854-1857.

[13] Linnehan, R., Mutz, Perlovsky, L.I., C., Weijers, B., Schindler, J., Brockett, R. (2003). *Detection of Patterns Below Clutter in Images*. Int. Conf. On Integration of Knowledge Intensive Multi-Agent Systems, Cambridge, MA Oct.1-3, 2003.

[14] Schacter, D.L., Dobbins, I.G., & Schnyer, D.M. (2004). Specificity of priming: A cognitive neuroscience perspective. Nature Reviews Neuroscience, **5**, 853-862.

[15] Bar, M., Kassam, K.S., Ghuman, A.S., Boshyan, J., Schmid, A.M., Dale, A.M., Hamalainen, M.S., Marinkovic, K., Schacter, D.L., Rosen, B.R., & Halgren, E. (2006). Top-down facilitation of visual recognition. Proceedings of the National Academy of Sciences USA, **103**, 449-54.

[16] Schacter, D.L. & Addis, D.R. (2007). The ghosts of past and future. Nature, 445, 27.

[17] Perlovsky, L.I. (2004). Integrating Language and Cognition. *IEEE Connections*, Feature Article, **2**(2), pp. 8-12.

[18] Perlovsky, L.I. (2006). *Modeling Field Theory of Higher Cognitive Functions*. Chapter in Artificial Cognition Systems, Eds. A. Loula, R. Gudwin, J. Queiroz. Idea Group, Hershey, PA, pp.64-105.