Fuzzy Dynamic Logic

Leonid I. Perlovsky

Air Force Research Lab., 80 Scott Rd., Hanscom AFB, MA 01731

Tel. 781-377-1728; e-mail: Leonid.Perlovsky@hanscom.af.mil

ABSTRACT

Fuzzy Logic is extended toward dynamic adaptation of the degree of fuzziness. The motivation is to explain the process of learning as joint model improvement and fuzziness reduction. A learning system with fuzzy models is introduced. Initially, the system is in a highly fuzzy state of uncertain knowledge, and it dynamically evolves into a low-fuzzy state of certain knowledge. We present an image recognition example of patterns below clutter. The paper discusses relationships to formal logic, fuzzy logic, complexity and draws tentative connections to Aristotelian theory of forms and working of the mind.

KEYWORDS: *fuzzy logic, complexity, fuzzy dynamic logic, mind, image recognition.*

1. INTRODUCTION. LEARNING AND COMPLEXITY

Relating complexity and learning dates at least to the 1960s, when Bellman discussed "the curse of dimensionality" [1]. He considered learning in the context of statistical pattern recognition; learning required estimating probability distribution functions in feature spaces for various classes; and Bellman found that such estimation was often difficult in high-dimensional feature spaces. The following thirty years of developing adaptive statistical pattern recognition and neural network algorithms for learning led to a conclusion that these approaches often encountered *combinatorial complexity (CC) of learning requirements*: recognition of any object, it seemed, could be learned if "enough" training examples were used for an algorithm learning. The required examples had to account for all possible variations of "an object", in all possible geometric positions and in *combinations* with other objects, sources of light, etc. The number of combinations quickly grows with the number of objects and conditions; say, a medium complexity problem of recognizing 100 different objects in various combinations might lead to a need to learn 100! ($\sim 100^{100}$) combinations; this number is larger than the number of elementary particles in the Universe in its entire history. It would require a corresponding number of training examples [2].

The complexity of learning had to be resolved, and during the 1960s and 70s a new paradigm was invented, interests shifted from learning algorithms to artificial intelligence relying on rule-based systems. An initial idea was that rules would capture the required knowledge and eliminate a need for learning. Rules are attractive in complex situations, they can solve an important problem in intelligence: define complex courses of actions according to designer's understanding of the problem. Rule systems worked well, when system goals, objects, and environments did not change, and when all aspects of the problem could be predetermined. When situations vary, learning was needed. For example, the first Chomsky ideas concerning the mind mechanisms for learning language [3] were based on rules. Some founders of the rule-based artificial intelligence emphasized that rule systems cannot explain learning [4], but these warnings were not heeded. Attempts to add learning to rule systems dominated artificial intelligence for about thirty years. However, rule systems in presence of unexpected variability, encountered *CC of rules*: detailed sub-rules and sub-sub-rules, one contingent on another, had to be specified.

The complexity of learning based on rules had to be overcome. A new paradigm emerged in the 1980s, model-based systems were proposed to combine advantages of learning and rules by using adaptive models. Existing knowledge was to be encapsulated in models, and unknown aspects of concrete situations were to be described by adaptive parameters. Learning consisted in estimating model parameters from training data. Complicated models could describe complex situations with relatively few unknown parameters, so that limited amount of training data would suffice for learning [2]. In the process of learning, each model had to be associated with corresponding data or signals (we use "data" and "signals" interchangeably), so that the model parameters could be estimated. This however was not easy to accomplish: In complex situations objects or processes responsible for incoming signals, which are described by different models, are not clearly separable. For example, in a visual scene it might not be clear where different objects are, and where shades are; similarly when tracking in clutter, it is not clear which signal belongs to which track or clutter. In sensor fusion, it is not clear, which signals in different sensors come from the same objects. When searching information in multiple databases or

Internet, it is not clear which descriptions refer to the same situations. To find a solution, one has to define a measure of similarity between models and signals (say, mean square error), then consider many combinations between signal subsets and models, and estimate parameters by minimizing errors for each combination. These type algorithms are called Multiple Hypothesis Testing or Multiple Hypothesis Tracking (MHT) [5, 6]. The number of subsets and the number of combinations are combinatorially large. Therefore, model-based systems encountered *computational CC* (N and NP complete algorithms). The CC became a ubiquitous feature of intelligent algorithms and seemingly, a fundamental mathematical limitation.

Mathematical structure of many neural networks was shown to be equivalent to algorithmic concepts analyzed above [7]. Among zillions of publications aimed at overcoming the CC, several fundamental mathematical ideas could be identified: Fuzzy Logic (FL), Genetic Algorithms (GA), and Hierarchical Decomposition (HD). GA attempt to concur the CC by combinatorial accumulation of experience from generation to generation, similar to genetic evolution. However, it is not clear how to set parameters of GA so that the system does not evolve into an evolutionary dead end. There is no agreement among biological geneticists that simple rules of classical genetics forming a foundation of GA are sufficient to explain evolution. More and more complex mechanisms of fine-tuning evolution are discussed (gene-mutators, punctuated equilibrium, competitive cooperation Gaia, etc.) HD fights the CC by exponential reduction of complexity at multiple levels of the hierarchy. However, there are no general rules of how to define hierarchical levels, how many are needed, and if they have to be predetermined or evolving. Promising schemes are emerging from combining these ideas, in particular, granularity combines FL and HD [8, 9, 10].

2. COMPLEXITY AND LOGIC

Combinatorial complexity is related to the type of logic, underlying various algorithms [2]. Formal logic is based on the “law of excluded middle” (or “excluded third”), according to which every statement is either true or false and nothing in between. Therefore, algorithms based on formal logic have to evaluate every little variation in data or models as a separate logical statement; combinations of subset of signals have to be considered, a large number of combinations of these variations causes combinatorial complexity. In fact, combinatorial complexity of algorithms based on logic is related to Gödel theory: it is a finite system manifestation of the incompleteness of logic [11]. Multivalued logic and fuzzy logic were proposed to overcome limitations related to the law of excluded third [12]**.** Yet the mathematics of multivalued logic is no different in principle from formal logic; excluded third is replaced by excluded n+1. Fuzzy logic encountered a difficulty related to the degree of fuzziness: if too much fuzziness is specified, the solution does not achieve a needed accuracy, if too little, it becomes similar to formal logic. FL eliminates combinatorial complexity of subsets; this works however, at a fixed degree of fuzziness. In complex cases fuzziness has to be properly selected at each algorithmic step in various parts of the system; it leads to combinatorial complexity of fuzziness optimization.

This paper concentrates on extending FL to Fuzzy Dynamic Logic, for shorter, Dynamic Logic (DL), which adds dynamic evolution of fuzziness to FL. At each algorithmic step in various parts of the system, the degree of fuzziness evolves according to local similarity between models and signals, without CC.

3. MODELING FIELD THEORY (MFT)

Modeling field theory (MFT) briefly summarized below is a learning system framework based on parametric models [7]. Learning in MFT is equivalent to joint parameter estimation and model-data association. It is achieved by maximizing a measure of similarity, L, between models, $\{\mathbf{M}_h\}$, h = 1,… H, and data $\{\mathbf{X}(n)\}$, n = 1,… N. Models represent expected data in the following sense: each model depends on its parameters $\{\mathbf{S}_h\}$, and for certain values of these parameters, $\{\underline{\mathbf{S}}_h\}$, the model is the conditional expectation for the data:

$$\mathbf{M}_h(\underline{\mathbf{S}}_h, n) = E\{ \mathbf{X}(n) \mid h \}. \tag{1}$$

In the r.h.s. of this equation, h is a hypothesis (about the source of the data, say an object and its orientation, etc.) and E{.} is the statistical expectation (please note, for shortness, we use the same notation h on the l.h.s. to denote just a model number, but not specific values of the parameters). For the parameter values $\mathbf{S}_h = \underline{\mathbf{S}}_h$, the model, $\mathbf{M}_h(\underline{\mathbf{S}}_h, n)$, is the expected value of the signal sample n, conditioned on the hypothesis h. The similarity measure can be defined as a likelihood; in absence of a priori knowledge about signal-model association, it treats each model as an alternative for each subset of signals

$$L(\{\mathbf{X}\},\{\mathbf{M}\}) = \prod_{n\in N} \sum_{h\in H} r(h)\, l(\mathbf{X}(n) \mid \mathbf{M}_h(n)); \tag{2}$$

here, $l(\mathbf{X}(n)|\mathbf{M}_h(n))$ (or simply $l(n|h)$) is a conditional partial similarity between one signal sample $\mathbf{X}(n)$ and one model $\mathbf{M}_h(n)$. Parameters $r(h)$ are proportional to the number of signals $\{n\}$ associated with the model h and normalized as

$$\sum_{h' \in H} r(h') = 1. \tag{3}$$

In probability theory, $r(h)$ are called priors, we call them class rates, or rates for short; they are usually unknown, and have to be estimated as well as other model parameters. For L to be interpreted as a likelihood for the entire set of signals $\{\mathbf{X}(n)\}$, $l(n|h)$ must be defined as conditional probability distribution functions, pdf. This probabilistic interpretation is valid, if we use correct values for model parameters; otherwise, these similarity measures can be considered as fuzzy similarities, parametrically dependent on model parameters.

In probability and statistics, a product of probabilities or pdf is used for independent random numbers. Expression (2) is different in this regard: although it contains a product over signal samples, n, signal samples are not assumed independent. Dependencies among $\{ \mathbf{X}(n) \}$ are due to models, $\mathbf{M}_h(n)$, each model determines conditional expected values of many signals. Note, all possible combinations of signals and models are accounted for in this expression. This can be seen by expanding a sum in (2), and multiplying all the terms; it would result in H^M items. This is the number of combinations between all signals (N) and all models (H). Here is the source of CC of many algorithms used in the past, related to the idea of multiple hypothesis testing (MHT). MHT attempted to maximize L over model parameters and associations between signals and models, in two steps, first by maximizing each of the H^M items over model parameters, and second by selecting the largest item. Such a program inevitably faces a wall of CC.

Likelihood interpretation of similarity measure (2) has an undesirable aspect: it requires that for some values of model parameters, models are exact conditional expectations of the data in the sense of (1). In many applications however, models are in principle approximate representation of reality. It is therefore more satisfactory to use another type of similarity measure, inspired by the concept of mutual information. Maximization of mutual information leads to extraction of maximum information from data, given available models. Such a similarity measure can be defined as follows (to preserve a notional correspondence between information and a logarithm of likelihood, we define log-similarity measure, LL_{in}, related to information),

$$LL\,(\{\mathbf{X}\},\{\mathbf{M}\}) = \sum_{n} \text{abs}(X(n))\ [\ \ln \sum_{h' \in H} r(h)\ l(\mathbf{X}(n) \mid \mathbf{M}_h(n))\ ]. \qquad (4)$$

Formally, this expression can be considered a logarithm of (2) with the following modification: every data point n is accounted for with a weight proportional to the strength of signal at this point. Relationship of this expression to mutual information in the models about data is considered in [7].

3. FUZZY DYNAMIC LOGIC (DL)

Fuzzy dynamic logic [13] maximizes similarity (2) or (4) without combinatorial complexity as follows. First, we define conditional similarities l(n|h) in such a way, that their degrees of fuzziness correspond to accuracies of models. This could be done in several ways, for example by using Gaussian parameterization for conditional similarities,

$$l(n|h) = (2\pi)^{-d/2} (\det\mathbf{C}_h)^{-1/2} \exp\{- 0.5(\mathbf{X}(n) - \mathbf{M}_h(n))^T \mathbf{C}_h^{-1} (\mathbf{X}(n) - \mathbf{M}_h(n)) \}. \tag{5}$$

Here, d is the dimensionality of the vectors **X** and **M**, and $\mathbf{C}_h$ is the covariance matrix. If model $\mathbf{M}_h(n)$ matches signals $\mathbf{X}(n)$ well, then covariance $\mathbf{C}_h$ is small, and expression (5) is close to the delta-function (low fuzziness). If the model does not match signals, then covariance $\mathbf{C}_h$ is large, and expression (5) is a wide distribution (highly fuzzy). The value of the covariance is to be estimated along with other parameter values. We must note, that this parameterization is very different from standard 'Gaussian assumption'. We do not assume that signals are Gaussian. Expression (5) assumes only that deviations between signals and models ($\mathbf{X}(n) - \mathbf{M}_h(n)$) are Gaussian. As will be clear from the following, this is only needed for a single model, which is the correct model for the signal n. And even this assumption is not necessary, later we will consider modifications for the general case.

Second, we introduce fuzzy class memberships, f(h|n), which define certainty of fuzzy logical statement that signal n originated from an object represented by model h,

$$f(h|n) = r(h)\, l(n|h) / \sum_{h' \in H} r(h')\, l(n|h'). \tag{6}$$

These variables give a measure of correspondence between signal $\mathbf{X}(n)$ and model $\mathbf{M}_h$ normalized within [0,1]. The definition (6) formally follows the Bayesian expression for a posteriori probabilities. We should remember, however, that the probabilistic interpretation is valid, if we use correct values for model parameters; otherwise fuzzy class memberships, f(h|n) can be called estimated probabilities.

Third, we define a dynamic logic process, leading to estimation of class membership functions and model parameters, which maximize similarity (2) without combinatorial complexity. First, consider a straightforward maximization of similarity (or, equivalently, its logarithm) over the parameter values, $\mathbf{S}_h$ and $\mathbf{C}_h^{-1}$, by gradient ascent (for now, we consider class rates r(h) known),

$$d\mathbf{S}_h/dt = d\ \mathrm{lnL}/d\mathbf{S}_h = \sum_n \ f(h|n)[\partial \mathrm{lnl}(n|h)/\partial \mathbf{M}_h]\ (\partial \mathbf{M}_h/\partial \mathbf{S}_h). \tag{7}$$

$$d\mathbf{C}_h^{-1}/dt = d\ \mathrm{lnL}/d\ \mathbf{C}_h^{-1} = \sum_n \ f(h|n)[\partial \mathrm{lnl}(n|h)/\partial \mathbf{C}_h^{-1}]\ . \tag{8}$$

If this internal process operates on a given set of signals { $\mathbf{X}$(n) } much faster than external real-time of the system operation, we can consider the dynamic logic process independent from the external environment. In this case the convergence of process (7) and (8) is assured as long as similarity is finite and its maximum is attained within the bounded region of parameter values. The first condition holds according to definitions (2) and (5), as long as none of the covariances turn to 0. In fact, $\mathbf{C}_h = 0$, is a maximum of similarity measures (2) and (5). It is a degenerate non-fuzzy (crisp) solution in which just a few data points belong to class h (if $\mathbf{M}_h$depends on p(h) parameters, $\mathbf{S}_h$, a degenerate solution might contain exactly p(h) data points). Such solutions are usually undesirable and must be prevented by limiting the minimal values of covariances. The second condition can be imposed by limiting the maximal values of parameters. Therefore we impose the following conditions:

$$\mathbf{C}_h > \mathbf{C}_h^{min}; \quad abs(\mathbf{S}_h) < \mathbf{S}_h^{max}. \tag{9}$$

Here $\mathbf{C}_h^{min}$ and $\mathbf{S}_h^{max}$ values are based on physical meanings of the parameters. For example, $\mathbf{C}_h^{min}$ might be determined by errors of sensor measurements responsible for data $\mathbf{X}(n)$. The procedure described above guarantees convergence to a local maximum of the similarity measure.

For the following analysis and for numerical integration of the dynamic process, it is convenient to substitute equations (7) and (8), with the following,

$$f(h|n) = r(h)\, l(n|h) / \sum_{h' \in H} r(h')\, l(n|h'). \tag{10}$$

$$r(h) = N_h / N; \quad N_h = \sum_n f(h|n); \tag{11}$$

$$\mathbf{C}_h = \sum_n f(h|n)\, (\mathbf{X}(n) - \mathbf{M}_h(n))(\mathbf{X}(n) - \mathbf{M}_h(n))^T / N_h. \tag{12}$$

$$\mathbf{S}_h = \mathbf{S}_h + dt * \sum_n f(h|n)[\partial \ln l(n|h)/\partial \mathbf{M}_h](\partial \mathbf{M}_h/\partial \mathbf{S}_h), \tag{13}$$

The l.h.s. of these equations contain new parameter values and the r.h.s. depend on the parameter values at the previous iteration; parameter dt determines the iteration step and speed of convergence. Eq.(11) estimates class rates r(h); N_h is an estimated number of signals $\mathbf{X}(n)$ associated with or coming from object h. Whereas convergence of the process (7) and (8) is a consequence of gradient ascent, we have to prove that the above equations also define a process converging to a local maximum of the similarity.

Lemma. Given $p_k > 0$, $q_k > 0$, $\Sigma_k p_k = 1$, $\Sigma_k q_k = 1$. The $\max_{pk} [\Sigma_k q_k \ln(q_k/p_k)] = 0$. It is proved by maximization of the bracketed expression, using the method of Lagrange multipliers.

Theorem. Equations (2), (3) and (9) through (13) define a convergent dynamic process with stationary states given by $\max_{\{\mathbf{S}_h\}} L$.

Proof. We will use notations ll(n|h,t) = ln(r(h)l(n|h)). Note that due to the definition (6),

$$\sum_{h' \in H} f(h'|n) = 1. \tag{14}$$

Consider a change in log-similarity from iteration to iteration:

$$LL(t+dt) - LL(t) = \sum_n \ln l(n|t+dt) - \ln l(n|t)$$

$$= \sum_n \sum_{h' \in H} f(n|h', t)*[ll(n|h',t+dt) - \ln f(h'|n,t+dt) - ll(n|h',t) + \ln f(h'|n,t)]$$

$$= \sum_n \sum_{h' \in H} f(n|h', t)*\{ [\ln f(h'|n,t) - \ln f(h'|n,t+dt)] + [ll(n|h',t+dt) - ll(n|h',t)] \} \tag{15}$$

Here, in the second line we used definition (6) and its logarithm (inside brackets). In the third line, the first square bracketed expression is non-negative due to the lemma. Therefore, to achieve a nonnegative change in log-similarity from iteration to iteration, it is sufficient to maintain nonnegative the second square bracketed expression. It is therefore sufficient to maintain nondecreasing ll(n| h,t+dt) on each iteration. Eqs.(11) through (13) achieve this as follows: eqs.(11) and (12) maximize ll(n|h,t+dt) over r(h) and $\mathbf{C}_h$ correspondingly; and eq.(13) is a step along the gradient ascent. This proves the local convergence of the dynamic process (10) through (13) to a local maximum of similarity LL. In this respect, the iterative dynamic process defined by iterations (10) through (13) is equivalent to the continuous dynamic process defined by differential equations (7) and (8). The above proof is easily extended to the information-type similarity measure (5) with all Σ_n (.) modified to Σ_n abs(X(n)) (.).

Qualitatively, the dynamic logic process can be characterized as follows. It is initiated using inaccurate models, large covariance matrices, and highly fuzzy membership functions. The process dynamics leads to improved parameter estimation, improved models, small covariances and low-fuzzy membership functions (nearly crisp logical statements about the signal class memberships). We would like to emphasize again that in the dynamic logic process, fuzziness of class memberships, f(h|n), is matched to uncertainties of model parameters, $\mathbf{S}_h$. This is due to the fact that covariances in eq.(12) are defined by mismatches between the data and the models. This correspondence is maintained at every moment (every iteration step) in the dynamic process, and locally for each model (if model h matches its corresponding data, then its covariance is small, and the membership functions for its data are close to 1).

4. GLOBAL VS. LOCAL CONVERGENCE

The global convergence for highly nonlinear functions, like similarity LL, might require unrealizable combinatorially complex computations. We would like to emphasize however that the global convergence is not necessarily attained in human perception and cognition, and is not necessary for practical engineering problems. In visual perception, for example, correct recognition of objects in a scene might require the second or third look from a different viewpoint. Therefore object recognition in human perception does not require the best match (or global maximization of some measure) between the signals and object models, stored in the mind, to be necessarily attained from the first look. In real life, working of the mind proceeds through continuous inflow of sensory signals, which are continuously matched to the previously learned models. Similarly in many engineering problems, say tracking multiple objects in clutter, it is not necessary that each initiated track converge to a valid moving object trajectory. New tracks are continuously initiated, and it is sufficient that upon track convergence, an engineering system can differentiate real tracks from spurious ones, with sufficient reliability.

Therefore, local convergence discussed in the previous section might be sufficient in many cases. In practical applications of the dynamic logic process, new sensory signals are continuously fed into the system, and new highly fuzzy models are continuously initiated. Models, which attained convergence (that is, their parameters stopped change significantly between iterations), are evaluated for their correspondence to real or spurious objects (this procedure depends on an application, it is not a basic part of the dynamic logic process considered here). Spurious models are removed from the dynamic process, as well as real-object models are removed after corresponding objects have left the field of view.

5. IMAGE RECOGNITION EXAMPLE

For this example we use similarity measure (5). Models should be defined corresponding to objects of interest expected to be present in images. Usually, it is neither possible nor feasible to model exactly all types of signals present. Still, all signals must be accounted for; this is simply a consequence of definition (6). A simple and efficient way to accomplish this is to utilize a simple clutter model specified as a uniform distribution across the image, we will denote it model h=1,

$$l(n|h=1) = 1/A. \tag{16}$$

Here, A is the area of the image (in the same units that pixel size is measured, meters or pixel numbers).

Fig. 1 illustrates the considered example, recognition of 'smiles' and 'frowns' in the background of a strong noise. Fig. 1A shows these objects without noise. Position and curvature of objects is unknown. We model each 'smile' or 'frown' pattern by a sum of Gaussian shapes along a parametric parabolic curve, as follows, for h > 1,

$$\mathbf{M}_h(n,k) = \mathbf{n}_h + a_h*(k^2, k). \tag{17}$$

$$l(n|h) = \sum_{k=-K/2}^{K/2} (2\pi)^{-1} (\sigma_{0h})^{-2}$$

$$\exp\{- 0.5(\mathbf{n} - \mathbf{M}_h(n,k))^T \mathbf{C}_h^{-1} (\mathbf{n} - \mathbf{M}_h(n,k)) \} * \exp(-k^2/2\sigma_h^2). \tag{18}$$

Here, $\mathbf{n}_h$ is a center position of the model h, it is a 2-D vector in image pixel coordinate space; a_h is a curvature parameter, k is a nonadaptive parameter of the model, parameterizing the parabolic shape in image pixel coordinates; σ_{0h} determines the width of pattern models, $\mathbf{C}_h$ is a diagonal matrix with σ_{0h}^2 along the diagonal; and σ_h determines the extent of pattern h. Parameters $\mathbf{n}_h$, a_h, σ_{0h}, and σ_h are adaptive and estimated from the image data; K was set to $4\sigma_h$. The minimal value for all standard deviations was set to sensor resolution, 1 pixel.

The dynamic logic convergence process is illustrated in Fig.1C through 1H. In this case we did not know how many objects are present. The algorithm started with a single 'blob' model described by a Gaussian distribution. After every five iterations the algorithm checked if more models would better fit the data. This procedure requires a modification to similarity (5). The reason is that more models always ‘better’ fit any data. Similarity must be penalized for additional free parameters. Many types of penalty functions were studied in literature [7, 14]. We used here Akaike Information Criterion, which compensates for the bias in estimated pdf due to free parameters; for N_{par} parameters, the penalty (for log likelihood or information) is $– N_{par}/2$.

For comparative illustration, the complexity of dynamic logic algorithm in this example was about 10^8 operations; in MHT type algorithm, if all combinations of three models were fit to this data, it would take approximately 10^{30} operations. Complexity of dynamic logic process is linear in the number of models.

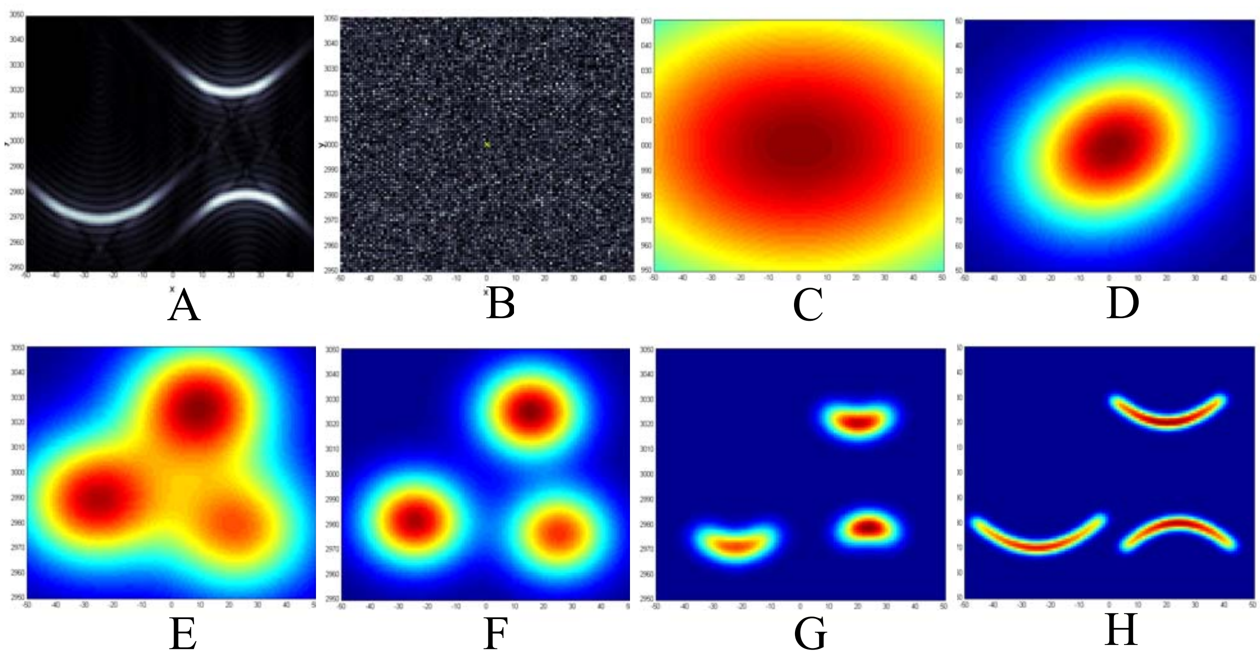

Fig.1. Finding ‘smile’ and ‘frown’ patterns in noise, an example of dynamic logic operation: (a) true ‘smile’ and ‘frown’ patterns shown without noise; (b) actual image available for recognition (signal is below noise, signal-to-noise ratio is between –2dB and –0.7dB); (c) an initial fuzzy model, the fuzziness corresponds to uncertainty of knowledge; (d) through (h) show improved models at various iteration stages (total of 22 iterations). At stage (d) the algorithm tried to fit the data with more than one model and decided, that it needs three models for the ‘best’ fit. There are three 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 (5) stopped increasing.

6. ARISTOTLE, ZADEH, AND FUZZY DYNAMIC LOGIC

Aristotle, I heard you are writing books now. Are you going to make our secret knowledge public?

from a letter by Alexander

Alexander, do not worry: nobody will understand.

from a reply letter by Aristotle

In a seminal paper in 1965 Lotfi Zadeh introduced fuzzy logic [15] to mirror the pervasive imprecision of the real world by providing a model for human reasoning in which everything - including truth - is a matter of degree. This is widely believed to be a sharp break with traditions of classical Aristotelian logic. It is interesting therefore to note that the original Aristotelian thinking might have been closer to fuzzy logic of Zadeh than usually appreciated. Aristotle closely tied logic to language, he emphasized that logical statements should not be formulated too specifically, otherwise meaning might be lost, "language contains necessary means for appropriate formulation of logical statements" and "common sense must be used to do it" [16]. However, Aristotle also formulated the "law of excluded middle", which contradicted uncertainty of language. This contradiction was noted in the 19th century by Boole, who decided to exclude any uncertainty from logic. A great school of logic formalization emerged, promising in the eyes of many to completely and forever formalize scientific discourse [17]. Soon, however Russell discovered disturbing inconsistencies and thirty years later Gödel proved that the entire enterprise was basically flawed [18, 19].

Returning to Aristotle, we note that he saw logic as an instrument of public discourse, a way to correctly argue for conclusions, which have been already obtained by other means. This is clearly seen, for example, from [20], where he lists logical arguments that should be used in public speeches as needed for or against any of dozens of political issues. Never he left any impression that logic was a mechanism of obtaining truth. Logic is a tool of politics not of science. When Aristotle was seeking for explanation of human thinking, he developed a theory of Forms [16]. The main tenets of this theory are that perception is a process in which "a priori Forms meet matter." This process is the foundation for all our experience, this process creates concepts with which our mind thinks and perceives individual objects. A priori Forms exist in our minds in a different form than the 'final' form of concepts used for thinking and perception.

Aristotle called the initial states of a priori Forms also "potential states" or "potentialities", and their final forms (of concepts of perception and cognition) after they met matter, he called "actual states" or "actualities". He emphasized that whereas actualities obey the rule of excluded middle, potentialities do not [[21]]. It is my opinion that if Aristotle new fuzzy logic, he might have said that the a priori Forms are fuzzy, the 'final' concepts consciously used by the mind are crisp (or low-fuzzy), and the process by which "a priori Forms meet matter" is given by fuzzy dynamic logic.

Returning from guesses about Aristotle to summarizing results of the current paper, fuzzy dynamic logic was formulated. It is a process in which a priori fuzzy and uncertain knowledge (models) "meets" signals about the real world. The dynamic logic interaction between knowledge and signals leads to better knowledge without combinatorial complexity. An example was shown were patterns below noise were recognized in image data.

Acknowledgments. The author is grateful to R. Brockett, R. Linnehan, C. Mutz, J. Schindler, and B. Weijers for multiple discussions and for help with the illustrating example.

REFERENCES

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

[2] Perlovsky, L.I. (1998). Conundrum of Combinatorial Complexity. IEEE Trans. PAMI, **20**(6) pp. 666-670.

[3] Chomsky, N. (1972). *Language and Mind.* Harcourt Brace Javanovich, New York, NY.

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

[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] Blackman, S.S. (1986). Multiple Target Tracking with Radar Applications. Artech House, Norwood, MA.

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

[8] Zadeh, L.A. (2002). Toward a Perception-Based Theory of Probabilistic Reasoning with F-Granular Probability Distributions. BISC seminar, UC Berkeley.
[9] Pedrycz, W. (2001). Granular Computing: An Emerging Paradigm, Physica-Verlag.
[10] Caulfield, H.J. et al. (2002). Proc. 6th Joint Conf. Info. Sciences, Assoc. Intell. Mach., Durham, NC.
[11] 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.
[12] Pedrycz, W., Ruspini, E., Bonnisone, P. (1998). Handbook of Fuzzy Computing, Oxford University Press.
[13] Perlovsky, L.I. Mutz, C., Weijers, B., Linnehan, R., Schindler, J., Brockett, R. (2004). Synthesis of Formal And Fuzzy Logic To Detect Patterns In Clutter. IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, CIMSA'04, Boston, MA.
[14] Cherkassky, V., Mulier, F. (1998). Learning from Data: Concepts, Theory, and Methods. Wiley-Interscience, New York, NY.
[15] Zadeh, L.A. (1965). Fuzzy sets, Information and Control, vol. 8, pp. 338 - 353, 1965.
[16] Aristotle, (IV BCE). Metaphysics, In the Complete Works of Aristotle, Ed.J.Barnes, Bollingen Series, 1995, Princeton, NJ.
[17] Hilbert, D. (1900). See in The Honors Class: Hilbert's Problems and Their Solvers, Ben Yandell, AK Peters Ltd, 2003, Natick, MA.
[18] Russel, B. & Whitehead, A.N. (1908). *Principia Mathematica.* Cambridge Univ. Press, 1962, Oxford, England
[19] Gödel, K. (1986). Kurt Gödel collected works, I. (Ed. S.Feferman at al). Oxford University Press.
[20] Aristotle, (IV BCE). Rhetoric to Alexander, In the Complete Works of Aristotle, Ed.J.Barnes, Bollingen Series, 1995, Princeton, NJ.
[21] Aristotle, (IV BCE). Topics, In the Complete Works of Aristotle, Ed.J.Barnes, Bollingen Series, 1995, Princeton, NJ.