Conundrum of Combinatorial Complexity

Leonid I. Perlovsky, *Senior Member, IEEE*

Abstract—This paper examines fundamental problems underlying difficulties encountered by pattern recognition algorithms, neural networks, and rule systems. These problems are manifested as combinatorial complexity of algorithms, of their computational or training requirements. The paper relates particular types of complexity problems to the roles of a priori knowledge and adaptive learning. Paradigms based on adaptive learning lead to the complexity of training procedures, while nonadaptive rule-based paradigms lead to complexity of rule systems. Model-based approaches to combining adaptivity with a priori knowledge lead to computational complexity. Arguments are presented for the Aristotelian logic being culpable for the difficulty of combining adaptivity and a priority. The potential role of the fuzzy logic in overcoming current difficulties is discussed. Current mathematical difficulties are related to philosophical debates of the past.

Index Terms—Pattern recognition, neural networks, rule systems, complexity, training, learning, a priori knowledge, fuzzy logic, Aristotelian logic

——————————— ✦ ———————————

1 INTRODUCTION

RECOGNITION of complex patterns has met with difficulties that are often expressed in terms of the complexity of a recognition process. Various recognition paradigms have their own sets of difficulties, but it seems that there always is a step in the recognition process that is exponentially or combinatorially complex. A well-known term used in this regard is "the curse of dimensionality" [4]. This designates a phenomenon of exponential (or combinatorial) increase in the required number of training samples with the increase of the dimensionality of a pattern recognition problem. The curse of dimensionality is characteristic of adaptive algorithms and neural networks.

Another set of difficulties is encountered by those approaches to the problem of recognition that utilize systems of a priori rules. In the case of rule systems, the difficulty is in a fast (combinatorial) growth of the number of rules with the complexity of the problem [48]. Model-based approaches that utilize object models in the recognition process encounter difficulties manifested as combinatorial complexity of required computations [7], [15], [32]. The difficulties of various pattern recognition paradigms have been summarized in recent reviews as follows. "Much of our current models and methodologies do not seem to scale out of limited 'toy' domains" [31]. The key issue is the "combinatorial explosion inherent in the problem" [16].

The seemingly inexorable combinatorial explosion that reincarnates in every pattern recognition paradigm is related in this paper to a fundamental issue of the roles of a priori knowledge vs. adaptive learning. This relationship has been discussed recently for geometric patterns and for function approximation [14], [36]. The issue of the roles of a priori knowledge vs. adaptive learning has been of an overriding concern in the research of mathematics of intelligence since its inception. In fact, the controversy about a priori knowledge and learning can be traced throughout the entire history of the concepts of mind throughout the Middle Ages to Aristotle and Plato. The philosophical thoughts of the past turn

————————————————

• *The author is with the Nichols Research Corporation, 70 Westview Street, Lexington, MA 02173. E-mail: perlovsl@nichols.com.*

out to be directly relevant to the development of mathematical concepts of intellect today. Section 2 discusses mathematical difficulties of relying exclusively on adaptive learning or on a priori knowledge. Then, Section 2 turns to difficulties of combining a priority with adaptivity. Section 3 presents a brief discussion of interrelationships between philosophical concepts of intellect and mathematical concepts of pattern recognition. Possible future directions are discussed in Section 4.

2 COMPUTATIONAL CONCEPTS

A contemporary direction in the theory of intellect based on modeling neural structures of the brain was founded by McCulloch and his coworkers [24]. In search of a mathematical theory unifying neural and cognitive processes, they combined an empirical analysis of biological neurons with the theory of information and mathematically formulated the main properties of neurons. McCulloch believed that the material basis of the mind is in complicated neural structures of a priori origin. Specialized, genetically inherited a priori structures have to provide for specific types of learning and adaptation abilities. An example of such a structure investigated by McCulloch was a group-averaging structure providing for scale-independent recognition of objects, which McCulloch believed serves as a material basis for *concepts* or *ideas* of objects independent of their apparent size [38].

However, this investigation into the a priori aspect of the intellect was not continued during the neural network research in 1950s and 1960s, and neural networks developed at that time utilized simple structures. These neural networks were based on the concept of general, nonspecific adaptive learning using concrete empirical data.[1] By emphasizing the adaptive aspect of intellect and neglecting its a priori aspect, this approach deviated from the program outlined by McCulloch. Simple structures of early neural networks and learning based entirely on the concrete empirical data were in agreement with behaviorist psychology dominant at the time. When the fundamental, mathematical character of limited capabilities of perceptrons was analyzed by Minsky and Papert [30], interest in the field of neural networks fell sharply.

Concurrent with early neural networks, adaptive algorithms for pattern recognition have been developed based on statistical techniques and the concept of classification space [12], [13], [33], [44]. In order to recognize objects (patterns) using these methods, the objects are characterized by a set of classification features that are designed based on a preliminary analysis of a problem and thus contain a priori information needed for a solution of this type of problem. Application of statistical pattern recognition methods has been limited by the fact that general mathematical methods for the design of classification features have not been developed. Design of classification features is based on a priori knowledge of specific problems and remains an art requiring human participation. When a problem complexity is not reduced to a few classification features in a preliminary analysis, these approaches lead to difficulties related to exorbitant training requirements.

The exorbitant training requirements of statistical pattern recognition algorithms can be understood due to the geometry of high-dimensional classification spaces. There are three basic approaches used to partition a classification space among classes: discriminating surfaces, nearest neighbors, and model-based parametric estimation [36]. Early parametric methods based on relatively simple probabilistic models such as Gaussian distributions were limited to simple classifier shapes (such as the quadratic classifier). Nonparametric paradigms (discriminating surfaces and nearest neighbors) have been used to surpass the limitations of simple parametric methods. However, due to the fact that the volume of a classification space grows exponentially with the dimensionality (number of features), training requirements for nonparametric paradigms are often exponential in terms of the problem complexity [36]. This is essentially the same problem that was encountered earlier in the field of adaptive control and was named "the curse of dimensionality" [4]. The father of cybernetics, Wiener, also saw this problem. He pointed out that using higher-order predictive models, or combining many simple models, is inadequate for the description of complex nonstationary systems, because of insufficient data for learning [47].

Facing exorbitant training requirements of statistical pattern recognition algorithms and being dissatisfied with limited capabilities of mathematical methods of modeling neural networks, which existed at the time, Minsky suggested a different concept of artificial intelligence based on the principle of a priority. He argued that intelligence could only be understood on the basis of extensive systems of a priori rules [28]. This was the next attempt (after McCulloch) to understand the intellect from the principle of a priority. The main advantage of this method is that it requires no training, because it explicitly incorporates detailed, high-level, a priori knowledge into the decision making. This knowledge is represented in a symbolic form similar to high-level cognitive concepts utilized by a human in conscious decision-making processes.

The main drawback of this method is the difficulty of combining rule systems with adaptive learning; while modeling the a priori aspect of the intellect, rule-based systems were lacking in adaptivity. Minsky emphasized that his method does not solve the problem of learning [29], notwithstanding attempts to add learning to rule-based artificial intelligence that continued in various fields of modeling the mind, including linguistics and pattern recognition [5], [6], [20], [21], [48]. In linguistics, Chomsky has proposed to build a self-learning system that could learn a language similarly to a human, using a symbolic mathematics of rule systems [10]. In Chomsky's approach, the learning of a language is based on a language faculty, which is a genetically inherited component of the mind, containing an a priori knowledge of language. This direction in linguistics, named the Chomskyan Revolution, was about recognizing the two questions about the intellect: first, how is it possible? and second, how is learning possible? as the center of a linguistic inquiry and of a mathematical theory of mind [6]. However, combining adaptive learning with a priori knowledge proved difficult: Variabilities and uncertainties in data required more and more detailed rules, leading to combinatorial complexity of logical inference [48].

Model-based approaches in machine vision have been used to extend the rule-based concept to 2D and 3D sensory data. Use of physically based models permits utilization of detailed a priori information on objects' properties and shape in algorithms of image recognition and understanding [5], [7], [8], [9], [15], [20], [22], [27], [31], [41], [48]. Models used in machine vision typically are complicated geometrical 3D models that require no adaptation. These models are useful in applications where variabilities are limited and types of objects and other parameters of the recognition problem are constrained. When unforeseen variabilities are a constant factor in the recognition problem, utilization of such models faces difficulties that are common to rule-based systems. More and more detailed models are required, potentially leading to a combinatorial explosion.

Parametric model-based approaches have been proposed to overcome the difficulties of previously used methods and to combine the adaptivity of parameters with a priority of models. In these approaches, adaptive parameters are used to adapt models to variabilities and uncertainties in data. Parametric adaptive

---

1. This statement should be further qualified. Widrow's (1959) Adaline neural network was based on the a priori cybernetic signal model (Wiener filter). For the problems of linear signal filtering, Adaline adaptation is efficient, that is, it learns as fast as theoretically possible. However, Adaline's internal model was relatively simple, and it seemed that Adaline was in line with the dominant philosophy at the time that emphasized self-learning and ignored complicated a priori knowledge. See further discussions in Sections 3 and 4.

methods date back to Widrow's Adaline and linear classifiers. These early parametric methods can be efficiently trained using few samples, however, they are limited to simple decision regions and are not suitable for many complicated problems. Complicated problems, such as image recognition, require utilization of multiple flexible models. In the process of recognition, an algorithm has to decide which subset of data corresponds to which model. This step is called segmentation, or association, and it requires a consideration of *multiple combinations* of subsets of the data. Because of this, complicated adaptive models often lead to combinatorial explosion of the complexity of the recognition process.

A mathematical analysis of existing approaches to the design of systems and algorithms of mathematical intelligence leads to the conclusion that computational concepts of most of today's neural networks originate in pattern recognition algorithms, and that there are four basic concepts forming the foundation for all the multiplicity of existing algorithms and neural networks [35], [36], [37]. These are

1) the concept of rule-based systems [28], defined by the factor of a priority;
2) the concept of nearest neighbors and
3) the concept of discriminating surfaces, both defined by the factor of adaptivity [12], [33]; and
4) the concept of parametric models [41], [48], that attempts to combine a priority and adaptivity.

While methods based on adaptivity face combinatorial explosion of *the training process*, those based on a priority face combinatorial explosion of *the complexity of rule systems,* and attempts to combine the two face combinatorial explosion of *the computational complexity.* Factors of a priority and adaptivity ought to be combined by physically acceptable concepts of the intellect. Therefore, approaches to combining both factors are of paramount interest. However, existing approaches to this problem have not resolved the conundrum of combinatorial complexity. To repeat, "Much of our current models and methodologies do not seem to scale out of limited 'toy' domains" [31]; "The key issues (are) ... the inherent uncertainty of data measurements" and "combinatorial explosion inherent in the problem" [16].

3 MATHEMATICS AND PHILOSOPHY

The problem of combining adaptivity and a priority is fundamental to computational intelligence as well as to understanding human intelligence. There is an interrelationship among concepts of mind in mathematics, psychology, and philosophy, which is much closer than currently thought among scientists and philosophers of today. From the contemporary point of view, the questions about mind posed by ancient philosophers are astonishingly scientific. A central question to the work of Plato, Aristotle, Avicenna, Maimonides, Aquinas, Occam, and Kant was the question of the origins of universal concepts. Are we born with a priori knowledge of concepts or do we acquire this knowledge adaptively by learning from experience? This question was central to the work of ancient philosophers and medieval theologists, and it was equally important to theories of Freud, Jung, and Skinner. The different answers they gave to this question are very similar to the answers given by McCulloch, Minsky, Chomsky, and Grossberg [10], [17], [18], [29], [26].

Plato faced the very first question about the intellect 2,300 years ago: How is it possible at all? He came to a conclusion that our ability to think is founded on the fact that *concepts* or abstract ideas (Eidos) are known to us a priori, through a mystic connection with a world of Ideas [39]. This conception of mind based on *realism* of ideas did not bode well with a new way of thinking that had emerged from the medieval scholasticism and was formulated in the 16th century as a scientific method. A forerunner of the scientific method, Occam, who is considered one of the last great medieval scholastic thinkers, held nominalistic views [34] that

are opposite to realism. Following Antisthenes, the founder of the Cynic school of philosophy, *nominalism* considers ideas to be just names (nomina) for classes or collections of similar empirical facts.

Time has obscured the influence of Occam on the development of the scientific method, and his name is hidden behind the figures of great philosophers and scientists that came after him. However, despite the realism of Descartes, Leibnitz, and Newton, nominalism of the forerunner of contemporary scientific thinking continues to pervade scientific attitudes of today. One of the reasons for the influence of nominalism is the unbreakable tie between the scientific method and objectivization of the subject of inquiry. In physics, the theoretical tradition of Newton's realism counterbalanced the influence of nominalism, but in the area of empirical sciences, such as psychology in the last century, the reality of facts seemed more significant than the reality of ideas that have not been clad in a mathematical form.

Near the end of 19th century, the success of the mathematical method in physics had advanced a requirement of objectivization and, in the empirical sciences, where the only criterion of objectivity was seen in reproducible experiments, the theoretical possibility of a priori concepts was questioned. A priori concepts started losing ground, became lowered to the level of (at best) unproved hypothesis, and it could even be argued that in some areas of science, the desire for objectivity inhibited deep theoretical scientific thinking. Concepts dressed not in the strict language of mathematical computations seemed compromised. In this atmosphere, to resolve the dilemma between the objectivity and depth of investigation, there was born behaviorism, a new scientific direction redefining psychology as a science of human behavior [45] and an accompanying intellectual and philosophical movement [42].

Emergence of cybernetics proceeded under the influence of the dominating psychological concept of behaviorism, which can be seen from the cybernetics' program paper [40]. The mutual influence of behaviorism, nominalistic philosophy, and cybernetics was enhanced by the fact that available cybernetic models were relatively simple linear Wiener filters, suitable for utilization of only simple a priori knowledge. It was truly revolutionary that despite the prevailing nominalistic orientation, McCulloch came to a conclusion that understanding of mind required the realistic philosophy. He wrote: "under the influence of nominalistic concepts since Occam, the realistic logic decayed, which caused problems for scientific understanding of mind" [25], [26]. A realistic philosophy, created by the school of Plato and Aristotle, McCulloch saw as the foundation for the search of the material structures of mind. However, early neural network research in 1950s and 1960s did not follow this direction and pursued the nominalistic concept of learning from examples, without using complicated a priori knowledge, until the demise of behaviorism in 1960s.

Whichever the reasons for the influence of the nominalistic concept, today it still forms the basis for nonparametric algorithms and neural networks, which do not utilize complicated a priori information in the process of learning and adaptation. However, a concerted research effort toward combining a priori knowledge and learning is emerging. And today, tracing the relationships between philosophical and mathematical theories of the intellect and outlining future research directions, we move away from Occam, who stands near the roots of scientific objectivization, toward the idealistic realism of Plato and Aristotle, explaining the possibility of mind by combining a priority and adaptivity on a realistic basis.

The rule-based systems approach to utilization of extensive a priori knowledge represents the next (after McCulloch) attempt to understand intelligence on the basis of realistic philosophy. The similarity between rule-based systems and Plato's conception of mind based on a priori *ideas* [39] has been discussed by Chomsky [10]. He has directly related the principle of a priority in algorithm design to the philosophy of realism. He has also hoped that the problem of learning can be solved using a rule-based approach to

intelligence. However, the mathematics of rule-based systems is inadequate for adaptation and learning, which has been emphasized by the founder of this approach to computational intelligence, Minsky [29], and has been later confirmed in multiple attempts to solve the problem of learning on the basis of rule-based systems [48]. Chomsky came to a similar conclusion, and later he proposed a different approach to the problem of learning based on a priori principles and adaptive parameters [11], which is similar to parametric model-based approaches to combining a priority and adaptivity in pattern recognition. As discussed in the previous section, existing mathematical methods used for this purpose face combinatorial computational complexity.

The most striking fact is that the first one who pointed out that learning cannot be achieved in Plato's theory of mind was Aristotle. Aristotle recognized that in Plato's formulation, there could be no learning, since Eide (ideas, or concepts) are given a priori in their final form. Thus, learning is not needed and is impossible, and the world of ideas is completely separated from the world of experience. Seeking to unite the two worlds and to understand learning, Aristotle developed a concept of Form having a universal and higher reality and being a formative principle in an individual experience [1]. In the Aristotelian theory of Form, the adaptivity of the mind was due to a meeting between the a priori Form and matter, forming an individual experience. This theory was further developed by Avicenna [2], Maimonides [23], Aquinas [3], and Kant [19] among many other philosophers during the last 2,300 years.

4   WHITHER WE GO FROM HERE?

Let us summarize the results of our analysis. The fundamental issue in pattern recognition and computational intelligence is the relative role of a priori knowledge and adaptive learning. The computational intelligence techniques that utilized only one of these two factors have been limited in their suitability for complicated pattern recognition problems. These limitations have been manifested in two different ways for the two types of algorithms. For algorithms based on the factor of adaptivity alone, the limitations have been manifested by the exponential training requirements. And, for algorithms based on the factor of a priority alone, the limitations have been manifested by the exponential explosion of the complexity of rule systems. In order to surpass these limitations, model-based techniques have been developed for combining adaptivity and a priority, but they often lead to a combinatorial explosion of the computational complexity.

The human intellect combines the two factors of a priority and adaptivity. According to the philosophical analysis dating to Aristotle, adaptive learning is based on a priori Forms. In today's mathematical language, adaptive parametric models come closest to the Aristotelian Forms. A meeting between the a priori Form and matter can be understood as adaptive estimation of model parameters from the data. Thus, a successful approach to pattern recognition, following human intellect, ought to combine adaptivity and a priority in a model-based paradigm. However, algorithms that have been used in the past to combine adaptivity and a priority in a model-based paradigm lead to exponential computational complexity and are not suited for this purpose. The answer to the conundrum of combinatorial complexity requires understanding of this difficulty.

It seems that Aristotle provided us with the riddle and with the key to its answer. The major point of Aristotelian criticism of Plato's Ideas was that before a Form meets matter, it should be *not in its final form of a concept.* But the Aristotelian logic that underlies our algorithms, the Boolean calculus, and calculus of predicates that are based on Aristotelian logic operate with final forms of concepts. The same is true about geometrical models of model-based paradigms. It is the need to consider multiple combinations or associations between the concepts and the material world (signals, images) in the

process of recognition that leads to the combinatorial explosion. The answer to the conundrum of combinatorial complexity should be sought in overcoming the Aristotelian logic that underlies our algorithms. Fuzzy and probabilistic logics [49] may hold keys to the answer. Intermediate computational steps (before Forms meet matter) should employ fuzzy representations of concepts. Neural networks with their inherent capability for fuzzy logic at the intermediate computational steps emerge as a vehicle for this new computational concept. Existing neural networks, however, lack the capability for representing complicated a priori knowledge. Current approaches to combining neural and symbolic processing do so by eclectic means of combining the old computational concepts in hybrid systems, while "new computational concepts are needed" [43]. Thus, pattern recognition research should be looking for a nexus of model-based and neural network concepts. How could this Aristotelian mathematics of mind be achieved? It seems that the future mathematics of mind will utilize complicated a priori Forms represented as fuzzy spatiotemporal models in a neural architecture. A dynamical spatiotemporal model combining spatial information representation and temporal processing can be called a neural modeling field. Such a system will evolve by learning from external stimuli based on a priori models, becoming a paradigm of the Aristotelian mathematics of mind.

ACKNOWLEDGMENT

REFERENCES

[1]   Aristotle, *Metaphysics.* Translated H.G. Apostle, Bloomington, Ind: Indiana Univ. Press, 1966, IV BC.

[2]   Avicenna, *Kitab al-Shifa.* Translated in Avicenna, S.M. Afnan, London, Great Britain: George Allen & Unwin, LTD, 1958, XI AD.

[3]   T. Aquinas. *Summa Contra Gentiles.* Tr. A.C. Pegis. Univ. of Notre Dame Press, 1997, 1324.

[4]   R.E. Bellman, *Adaptive Control Processes.* Princeton, NJ: Princeton Univ. Press, 1961.

[5]   P.P. Bonnisone, M. Henrion, L.N. Kanal, and J.F. Lemmer, *Uncertainty in Artificial Intelligence 6.* Amsterdam, The Netherlands: North Holland, 1991.

[6]   R.P. Botha, *Challenging Chomsky. The Generative Garden Game.* Oxford, UK: Basil Blackwell, 1991.

[7]   R.A. Brooks, "Model-Based Three-Dimensional Interpretation of Two-Dimensional Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 5, no. 2, pp. 140-150, 1983.

[8]   A. Califano and R. Mohan, "Multidimensional Indexing for Recognizing Visual Shapes," *IEEE Trans. Pattern Analysis Machine Intelligence,* vol. 16, no. 4, pp. 373-392, 1994.

[9]   R.T. Chen and C.R. Dyer, "Model-Based Recognition in Robotic Vision," *ACM Computing Surveys,* vol. 18, pp. 67-108, 1986.

[10] N. Chomsky, *Language and Mind.* New York, NY: Harcourt Brace Javanovich, 1972.

[11] N. Chomsky, *Principles and Parameters in Syntactic Theory.* N. Hornstein and D. Lightfoot eds., *Explanation in Linguistics. The Logical Problem of Language Acquisition,* London: Longman, 1981.

[12] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis.* New York, NY: J. Wiley & Sons, 1973.

[13] K. Fukunaga, *Introduction to Statistical Pattern Recognition.* New York, NY: Academic Press, 1972.

[14] F. Girosi, M. Jones, and T. Poggio, "Regularization Theory and Neural Networks Architectures," *Neural Computation,* vol. 7, no. 2, pp. 219-269, 1995.

[15] W.E.L. Grimson and T. Lozano-Perez, "Model-Based Recognition and Localization From Sparse Range or Tactile Data," *Int'l J. Robotics Research,* vol. 3, no. 3, pp. 3-35, 1984.

[16] W.E.L. Grimson and D.P. Huttenlocher, "Introduction to the Special Issue on Interpretation of 3-D Scenes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 10, pp. 969-970, 1991; vol. 14, no. 2, pp. 97-98, 1992.

[17] S. Grossberg, "How Does a Brain Build a Cognitive Code?" *Psychological Review*, vol. 87, pp. 1-51, 1980.

[18] S. Grossberg, "Nonlinear Neural Networks: Principles, Mechanisms, and Architectures," *Neural Networks,* vol. 1, no. 1, pp. 17-61, 1988.

[19] I. Kant, *Critique of Pure Reason*. Translated J.M.D. Meiklejohn, New York, NY: Wiley Book, 1943, 1781.

[20] H.R. Keshavan, J. Barnett, D. Geiger, and T. Verma, "Introduction to the Special Section on Probabilisitc Reasoning," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 3, pp. 193-195, 1993.

[21] J. Koster and R. May, *Levels of Syntactic Representation*. Dordrecht: Foris Publications, 1981.

[22] Y. Lamdan and H.J. Wolfson, "Geometric Hashing: A General and Efficient Recognition Scheme," *Proc. Second Int'l Conf. Computer Vision*, 1988.

[23] M. Maimonides, *The Guide for the Perplexed*, 2nd edition. Transl. M. Friedlander, New York, NY: Dover, 1956, 1190.

[24] W.S. McCulloch and W. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bull. Mathematical Biophysics*, vol. 7, pp. 115-133, 1943.

[25] W.S. McCulloch, "What Is a Number that a Man May Know It, and a Man, that He May Know a Number?" Ninth Alfred Korzybski Memorial Lecture, *General Semantics Bull.*, vol. 26, no. 27, pp. 17-18, 1961. Also in McCulloch, 1965.

[26] W.S. McCulloch, *Embodiments of Mind*, 2nd edition. Cambridge, Mass.: MIT Press, 1988, 1965.

[27] R.S. Michalski, J.G. Carbonell, and T.M. Mitchell, *Machine Learning: An Artificial Intelligence Approach*, vol 2. Los Altos, Calif.: Morgan Kaufmann, 1986.

[28] M.L. Minsky, *Semantic Information Processing*. Cambridge, Mass.: The MIT Press, 1968.

[29] M.L. Minsky, "A Framework for Representing Knowledge," *In The Psychology of Computer Vision*, P.H. Whinston, ed., New York, NY: McGraw-Hill Book, 1975.

[30] M.L. Minsky and S.A. Papert, *Perceptrons*. Cambridge, Mass.: The MIT Press, 1969, 1988.

[31] S. Negahdaripour and A.K. Jain, "Final Report of the NSF Workshop on the Challenges in Computer Vision Research," *Future Directions of Research*, U.S. Nat'l Science Foundation, 1991.

[32] R. Nevatia and T.O. Binford, "Description and Recognition of Curved Objects," *Artificial Intelligence*, vol. 8, no. 1, pp. 77-98, 1977.

[33] N.J. Nilsson, *Learning Machines*. New York, NY: McGraw-Hill, 1965.

[34] W. Occam, "Summa Logicae," Translated M.J. Loux, *Occam's Theory of Terms*, 1974, and Translated A.J. Freddoso and H. Schuurman, *Occam's Theory of Propositions*. Notre Dame, Ind.: Univ. of Notre Dame Press, 1980, XIV.

[35] L.I. Perlovsky, "Computational Concepts in ATR: Neural Networks, Statistical Pattern Recognition, and Model Based Vision," *ATR Working Group Meeting*, Seattle, Wash., 1991.

[36] L.I. Perlovsky, "Computational Concepts in Classification: Neural Networks, Statistical Pattern Recognition, and Model Based Vision," *J. Math. Imaging and Vision*, vol. 4, no. 1, pp. 81-110, 1994.

[37] L.I. Perlovsky, "Fuzzy Logic of Aristotelian Forms," *Proc. Conf. Intelligent Systems and Semiotics '96*, vol. 1, pp. 43-48, Gaithersburg, Md., 1996.

[38] W. Pitts and W.S. McCulloch, "How We Know Universals: The Perception of Auditory and Visual Forms," *Bull. Math. Biophysics*, vol. 9, pp. 127-147, 1947.

[39] Plato, *Phaedrus*. Translated in Plato, L. Cooper. New York, NY: Oxford Univ. Press, IV BC.

[40] A. Rosenblueth, N. Wiener, and J. Bigelow, "Behavior, Purpose and Teleology," *Philosophy of Science*, vol. 10, no. 1, pp. 18-24, 1943.

[41] A.M. Segre, "Applications of Machine Learning," *IEEE Expert*, vol. 7, no. 3, pp. 31-34, 1992.

[42] B.F. Skinner, *About Behaviorism*. New York, NY: Alfred A. Knopf, 1974.

[43] R. Sun and L.A. Bookman, *Computational Architectures Integrating Neural and Symbolic Processing*. Boston, Mass.: Kluwer Academic Publishers, 1995.

[44] S. Watanabe, *Pattern Recognition: Human and Mechanical*. New York, NY: John Wiley & Sons, 1985.

[45] J.B. Watson, "Psychology as the Behaviorist Views It," *Psychological Rev.*, vol. 20, pp. 158-177, 1913.

[46] B. Widrow, "Adaptive Sample-Data System—A Statistical Theory of Adaptation," *1959 WESCON Convention Record*, Part 4, pp. 74-85, 1959.

[47] N. Wiener, *Cybernetics*. New York, NY: Wiley, 1948.

[48] P.H. Winston, *Artificial Intelligence*, 2nd edition. Reading, Mass.: Addison-Wesley, 1984.

[49] L.A. Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8, pp. 338-352, 1965.