AI Magazine Volume 4 Number 3 (1983) (© AAAI)



# Machine Learning: A Historical and Methodological Analysis

Jaime G. Carbonell
*Computer Science Department*
*Carnegie-Mellon University*
*Pittsburg, PA 15213*

Ryszard S. Michalski
*Department of Computer Science*
*University of Illinois at Urbana-Champaign*
*Champaign-Urbana, IL 61801*

Tom M. Mitchell
*Department of Computer Science*
*Rutgers University*
*New Brunswick, NJ 08903*

**Editors' Note:** Machine Learning has been a constant theme throughout AI's two decades of existence In this overview the authors analyze various aspects including the major methodological approaches advocated in Machine Learning research, and how they have related to major contemporary themes in "mainstream" AI Research. In a subsequent issue we plan to include a sequel to this rarticle which will give the authors' views on current research directions in Machine Learning.

In the meanwhile, we are *very* anxious to get readers reactions to *this* and all earlier contributions to this column — *Derek Sleeman and Jaime Carbonell*

Abstract

Machine learning has always been an integral part of artificial intelligence, and its methodology has evolved in concert with the major concerns of the field In response to the difficulties of encoding ever-increasing volumes of knowledge in modern AI systems, many researchers have recently turned their attention to machine learning as a means to overcome the knowledge acquisition bottleneck This article presents a taxonomic analysis of machine learning organized primarily by learning strategies and secondarily by knowledge representation and application areas A historical survey outlining the development of various approaches to machine learning is presented from early neural networks to present knowledge-intensive techniques.

This paper is a modified and extended version of the first chapter of *Machine Learning, An Artificial Intelligence Approach,* with permission of the publisher: Tioga Press (Palo Alto, CA) The research described here was sponsored in part by the Office of Naval Research (ONR) under grant number N00014-79-C-0661, and in part by the National Science Foundation grant MCS82-05166

LEARNING IS A MANY-FACETED PHENOMENON. Learning processes include the acquisition of new declarative knowledge, the development of motor and cognitive skills through instruction or practice, the organization of new knowledge into general, effective representations, and the discovery of new facts and theories through observation and experimentation. The study and computer modelling of learning processes in their multiple manifestations constitutes the subject matter of machine learning

Although machine learning has been a central concern in artificial intelligence since the early days when the idea of "self-organizing systems" was popular, the limitations inherent in the early neural network approaches led to a temporary decline in research volume. More recently, new symbolic methods and knowledge-intensive techniques have yielded promising results and these in turn have led to the current revival in machine learning research This article examines some basic methodological issues, proposes a classification of machine learning techniques, and provides a historical review of the major research directions

## The Objectives of Machine Learning

The field of machine learning can be organized around three primary research foci:

- **Task-Oriented Studies** the development and analysis of learning systems oriented toward solving a predetermined set of tasks (also known as the "engineering approach")

- **Cognitive Simulation**—the investigation and computer simulation of human learning processes (also known as the "cognitive modelling approach")
- **Theoretical Analysis**—the theoretical exploration of the space of possible learning methods and algorithms independent of application domain.

Although many research efforts strive primarily towards one of these objectives, progress in one objective often leads to progress in another For instance, in order to investigate the space of possible learning methods, a reasonable starting point may be to consider the only known example of robust learning behavior, namely humans (and perhaps other biological systems) Similarly, psychological investigations of human learning may be helped by theoretical analysis that may suggest various plausible learning models. The need to acquire a particular form of knowledge in some task-oriented study may itself spawn new theoretical analysis or pose the question: "How do *humans* acquire this specific skill (or knowledge)?" The existence of these mutually supportive objectives reflects the entire field of artificial intelligence, where expert systems research, cognitive simulation, and theoretical studies provide some cross-fertilization of problems and ideas

**Applied Learning Systems: A Practical Necessity.** At present, instructing a computer or a computer-controlled robot to perform a task requires one to define a complete and correct algorithm for that task, and then laboriously program the algorithm into a computer These activities typically involve a tedious and time-consuming effort by specially trained personnel.

Present-day computer systems cannot truly learn to perform a task through examples or by analogy to a similar, previously-solved task. Nor can they improve significantly on the basis of past mistakes, or acquire new abilities by observing and imitating experts Machine learning research strives to open the possibility of instructing computers in such new ways, and thereby promises to ease the burden of hand-programming growing volumes of increasingly complex information into the computers of tomorrow. The rapid expansion of applications and availability of computers today makes this possibility even more attractive and desirable

When approaching a task-oriented knowledge acquisition task, one must be aware that the resultant computer systems must interact with humans, and therefore should closely parallel human abilities. The traditional argument that an engineering approach need not reflect human or biological performance is not truly applicable to machine learning. Since airplanes, a successful result of an almost pure engineering approach, bear little resemblance to their biological counterparts, one may argue that applied knowledge acquisition systems could be equally divorced from any consideration of human capabilities This argument does not apply here because airplanes need not interact with or understand birds Learning machines, on the other hand, will have to interact with the people who make use of them, and consequently the concepts and skills they acquire – if not necessarily their internal mechanisms—must be understandable to humans.

**Machine Learning as a Science.** The question of what are the genetically-endowed abilities in a biological system (versus environmentally-acquired skills or knowledge) has fascinated biologists, psychologists, philosophers and artificial intelligence researchers alike. A clear candidate for a cognitive invariant in humans is the learning mechanism the innate ability to acquire facts, skills and more abstract concepts. Therefore, understanding human learning well enough to reproduce aspects of that learning behavior in a computer system is, in itself, a worthy scientific goal. Moreover, the computer can render substantial assistance to cognitive psychology, in that it may be used to test the consistency and completeness of learning theories, and enforce a commitment to fine-structure process-level detail that precludes meaningless, tautological or untestable theories (Sloman 1978, Carbonell 1981).

The study of human learning processes is also of considerable practical significance. Gaining insights into the principles underlying human learning abilities is likely to lead to more effective educational techniques Thus, it is not surprising that research into intelligent computer-assisted instruction, which attempts to develop computer-based tutoring systems, shares many of the goals and perspectives with machine learning research One particularly interesting development is that computer tutoring systems are starting to incorporate abilities to infer models of student competence from observed performance. Inferring the scope of a student's knowledge and skills in a particular area allows much more effective and individualized tutoring of the student (Sleeman 1983).

An equally basic scientific objective of machine learning is the exploration of possible learning mechanisms, including the discovery of different induction algorithms, the scope and theoretical limitations of certain methods, the information that must be available to the learner, the issue of coping with imperfect training data, and the creation of general techniques applicable in many task domains. There is no reason to believe that human learning methods are the only possible means of acquiring knowledge and skills In fact, common sense suggests that human learning represents just one point in an uncharted space of possible learning methods—a point that through the evolutionary process is particularly well suited to cope with the general physical environment in which we exist Most theoretical work in machine learning has centered on the creation, characterization and analysis of general learning methods, with the major emphasis on analyzing generality and performance rather than psychological plausibility

Whereas theoretical analysis provides a means of exploring the space of possible learning methods, the task-oriented approach provides a vehicle to test and improve the performance of functional learning systems By constructing and testing applied learning systems, one can determine the cost-effectiveness trade-offs and limitations of particular approaches to learning. In this way, individual data points in the space of possible learning systems are explored, and the space itself becomes better understood

**Knowledge Acquisition versus Skill Refinement.** There are two basic forms of learning: *knowledge acquisition* and *skill refinement* When we say that someone learned physics, we mean that this person acquired concepts of physics, understood their meaning, and their relationship to each other as well as to the physical world. The essence of learning in this case is the acquisition of knowledge, including descriptions and models of physical systems and their behaviors, incorporating a variety of representations—from simple intuitive mental models, examples and images, to completely tested mathematical equations and physical laws. A person is said to have learned more if his knowledge explains a broader scope of situations, is more accurate, and is better able to predict the behavior of the physical world (Popper 1968) This form of learning is typical to a large variety of situations and is generally termed *knowledge acquisition* Hence, knowledge acquisition is defined as learning new symbolic information coupled with the ability to apply that information in an effective manner.

A second kind of learning is the gradual improvement of motor and cognitive skills through practice, such as learning to ride a bicycle or to play the piano. Acquiring textbook knowledge on how to perform these activities represents only the initial, and not necessarily critical, phase in developing the requisite skills. The bulk of the learning process consists of refining the acquired skill, and improving the mental or motor coordination by repeated practice and a correction of deviations from desired behavior This form of learning, often called *skill refinement,* differs in many ways from knowledge acquisition. Whereas the essence of knowledge acquisition may be a conscious process whose result is the creation of new symbolic knowledge structures and mental models, skill refinement occurs by virtue of repeated practice without concerted conscious effort. Most human learning appears to be a mixture of both activities, with intellectual endeavors favoring the former, and motor coordination tasks favoring the latter.

Present machine learning research focuses on the knowledge acquisition aspect, although some investigations, specifically those concerned with learning in problem-solving and transforming declarative instructions into effective actions, touch on aspects of both types of learning. Whereas knowledge acquisition clearly belongs in the realm of artificial intelligence research, a case could be made that skill refinement comes closer to non-symbolic processes, such as those studied in adaptive control systems It may indeed be the case that skill acquisition is inherently non-symbolic in biological systems, but an interesting symbolic model capable of simulating gradual skill improvement through practice has been proposed by Newell and Rosenbloom (1981). Hence, perhaps both forms of learning can be captured in artificial intelligence models.

## A Taxonomy of Machine Learning Research

This section presents a taxonomic road map to the field of machine learning with a view towards presenting useful criteria for classifying and comparing most artificial intelligence-based machine learning investigations Later, the main directions actually taken by researchers in this area over the past twenty years are surveyed.

One may classify machine learning systems along many different dimensions. We have chosen three dimensions as particularly meaningful:

- Classification on the basis of the *underlying learning strategy* used. The strategies are ordered by the amount of inference the learning system performs on the information provided to the system
- Classification on the basis of the type of *representation of knowledge (or skill)* acquired by the learner
- Classification in terms of the *application domain* of the performance system for which knowledge is acquired

Each point in the space defined by the above dimensions corresponds to a system employing a particular learning strategy, a particular knowledge representation, and applied to a particular domain. Since many existing learning systems employ multiple strategies and knowledge representations, and some have been applied to more than one domain, such learning systems are characterized by a collection of points in the space

The subsections below describe explored values along each of these dimensions Future research may well reveal new values on these dimensions as well as new dimensions. Indeed, the larger space of all possible learning systems is still only sparsely explored and partially understood Existing learning systems correspond to only a small portion of the space because they represent only a small number of possible combinations of the values

**Classification Based on the Underlying Learning Strategy.** Since we distinguish learning strategies by the amount of inference the learner performs on the information provided, we first consider the two extremes: performing no inference, and performing a substantial amount of inference If a computer system is programmed directly, its knowledge increases, but it performs no inference whatsoever on the new information; all cognitive effort is on the part of the programmer. Conversely, if a system independently discovers new theories or invents new concepts, it must perform a very substantial amount of inference; it is deriving organized knowledge from experiments and observations An intermediate point in the spectrum would be a student determining how to solve a mathematics problem by analogy to worked-out examples in the textbook—a process that requires inference, but much less than discovering a new branch of mathematics without guidance from teacher or textbook

As the amount of inference that the learner is capable of performing increases, the burden placed on the teacher or external environment decreases. It is much more difficult to teach a person by explaining each step in a complex task than by showing that person the way that similar tasks are usually handled. It is more difficult yet to program a computer to perform a complex task than to instruct a person to perform the task; as programming requires explicit specification of all requisite detail, whereas a person receiving instruction can use prior knowledge and common sense to fill in most mundane details The taxonomy below captures this notion of trade-offs in the amount of effort required of the learner and of the teacher

**Rote Learning and Direct Implanting of New Knowledge** In rote learning no inference or other transformation of the knowledge is performed by the learner. Variants of this strategy of knowledge acquisition method include:

- *Learning by being programmed, constructed, or modified* by an external entity, (for example, the usual style of computer programming)
- Learning by memorization of given facts and data with no inferences drawn from the incoming information (for example, as performed by existing database systems) The term "rote learning" is used primarily in this context

**Learning from Instruction**. Acquiring knowledge from a teacher or other organized source, such as a textbook, requires that the learner transform the knowledge from the input language to an internally-usable representation, and that the new information be integrated with prior knowledge for effective use Hence, the learner is required to perform some inference, but a large fraction of the burden remains with the teacher, who must present and organize knowledge in a way that incrementally augments the student's existing knowledge. Learning from instruction, also termed "learning by being told," parallels most formal education methods. Therefore, the machine learning task is one of building a system that can accept instruction or advice and can store and apply this learned knowledge effectively

**Learning by Analogy**. Learning by analogy is the process of transforming and augmenting existing knowledge (or skills) applicable in one domain to perform a similar task in a related domain For instance, a person who has never driven a small truck, but drives automobiles, may well transform his existing skill (perhaps imperfectly) to the new task. Similarly, a learning-by-analogy system might be applied to convert an existing computer program into one that performs a closely-related function for which it was not originally designed Learning by analogy requires more inference on the part of the learner than does rote learning or learning from instruction. A fact or skill analogous in relevant parameters must be retrieved from memory; then the retrieved knowledge must be appropriately transformed, applied to the new situation, and stored for future use.

**Learning from examples**. Learning from examples is a special case of inductive learning. Given a set of examples and counterexamples of a concept, the learner induces a general concept description that describes all of the positive examples and none of the counterexamples Learning from examples is a method that has been heavily investigated in artificial intelligence. The amount of inference performed by the learner is much greater than in learning from instruction, as no general concepts are provided by a teacher, and is somewhat greater than in learning by analogy, as no similar concepts are provided as "seed" from which the new concept may be grown Learning from examples can be subcategorized according to the *source* of the examples:

- The source is a *teacher* who knows the concept and generates examples of the concept that are meant to be as helpful as possible If the teacher also knows (or, more typically, infers) the knowledge state of the learner, the examples can be generated to optimize convergence on the desired concept (as in Winston's (1975) near-miss analysis)
- The source is the *learner itself* The learner typically knows its own knowledge state, but clearly does not know the concept to be acquired. Therefore, the learner can generate instances (and have an external entity such as the environment or a teacher classify them as positive or negative examples) on the basis of the information it believes necessary to discriminate among contending concept descriptions For instance, a learner trying to acquire the concept of "ferromagnetic substance," may generate as a possible candidate "all metals." Upon testing copper and other metals with a magnet, the learner will then discover that copper is a counterexample, and therefore the concept of ferromagnetic substance should not be generalized to include all metals Mitchell's LEX system (1983) and Carbonell's plan generalization method (1983) illustrate the process of internal instance generation.

  The source is the *external environment* In this case the example generation process is operationally random, as the learner must rely on relatively uncontrolled observations For example, an astronomer attempting to infer precursors to supernovas must rely mainly upon unstructured data presentation Although the astronomer knows the concept of a supernova, he cannot know *a priori* where and when a supernova will occur, nor can he cause one to exist Michalski's STAR methodology (1983) exemplifies this type of learning.

One can also classify learning from examples by the *type* of examples available to the learner:

- *Only positive examples available* Whereas positive examples provide instances of the concept to be acquired, they do not provide information for preventing overgeneralization of the inferred concept In this kind of learning situation, overgeneralization might be avoided by considering only the minimal generalizations necessary, or by relying on *a priori* domain knowledge to constrain the concept to be inferred

- *Positive and negative examples available* In this kind of situation, positive examples force generalization whereas negative examples prevent overgeneralization (the induced concept should never be so general as to include any of the negative examples) This is the most typical form of learning from examples

Learning from examples may be one-trial or incremental In the former case, all examples are presented at once. In the latter case, the system must form one or more hypotheses of the concept (or range of concepts) consistent with the available data, and subsequently refine the hypotheses after considering additional examples. The incremental approach more closely parallels human learning, allows the learner to use partially learned concepts (for performance, or to guide the example generation process), and enables a teacher to focus on the basic aspects of a new concept before attempting to impart less central details. On the other hand, the one-step approach is less apt to lead one down garden paths by an injudicious choice of initial examples in formulating the kernel of the new concept.

**Learning from Observation and Discovery**. This "unsupervised learning" approach is a very general form of inductive learning that includes discovery systems, theory-formation tasks, the creation of classification criteria to form taxonomic hierarchies, and similar tasks to be performed without benefit of an external teacher. Unsupervised learning requires the learner to perform more inference than any approach thus far discussed The learner is not provided with a set of instances of a particular concept, nor is it given access to an oracle that can classify internally-generated instances as positive or negative examples of any given concept. Moreover, rather than focusing on a single concept at a time, the observations may span several concepts that need to be acquired, thus introducing a severe focus-of-attention problem One may subclassify learning from observation according to the *degree of interaction* with an external environment. The extreme points in this dimension are:

- *Passive observation*, where the learner classifies and taxonomizes observations of multiple aspects of the environment (as in Michalski and Stepp's (1983) conceptual clustering )
- *Active experimentation*, where the learner perturbs the environment to observe the results of its perturbations Experimentation may be random, dynamically focused according to general criteria of interestingness, or strongly guided by theoretical constraints As a system acquires knowledge, and hypothesizes theories it may be driven to confirm or disconfirm its theories, and hence explore its environment applying different observation and experimentation strategies as the need arises. Often this form of learning involves the generation of examples to test hypothesized or partially acquired concepts. This type of learning is exemplified in Lenat's AM and EURISKO systems (Lenat 1976, 1983) book-lena82) )

An Intermediate point in this dimension is the BACON system (Langley, Simon & Bradshaw, 1983), attention but does not design new experiments

The above classification of learning strategies should help one to compare various learning systems in terms of their underlying mechanisms, in terms of the available external source of information, and in terms of the degree to which they rely on pre-organized knowledge.

**Classification According to Type of Knowledge Acquired.** A learning system may acquire rules of behavior, descriptions of physical objects, problem-solving heuristics, classification taxonomies over a sample space, and many other types of knowledge useful in the performance of a wide variety of tasks. The list below spans types of knowledge acquired, primarily as a function of the representation of that knowledge

1 **Parameters in algebraic expressions**– Learning in this context consists of adjusting numerical parameters or coefficients in algebraic expressions of a fixed functional form so as to obtain desired performance. For instance, perceptrons adjust weighting coefficients for threshold logic elements when learning to recognize two-dimensional patterns (Rosenblatt 1958, Minsky & Papert 1969).
2 **Decision trees**—Some systems acquire decision trees to discriminate among classes of objects The nodes in a decision tree correspond to selected object attributes, and the edges correspond to predetermined alternative values for these attributes. Leaves of the tree correspond to sets of objects with an identical classification Feigenbaum's EPAM exemplifies this discrimination-based learning approach (Feigenbaum, 1963).
3 **Formal grammars**—In learning to recognize a particular (usually artificial) language, formal grammars are induced from sequences of expressions in the language. These grammars are typically represented as regular expressions, finite-state automata, context-free grammar rules, or transformation rules
4. **Production rules**—A production rule is a condition-action pair $C \rightarrow A$, where $C$ is a set of conditions and $A$ is a sequence of actions If all the conditions in a production rule are satisfied, then the sequence of actions is executed. Due to their simplicity and ease of interpretation, production rules are a widely-used knowledge representation in learning systems The four basic operations whereby production rules may be acquired and refined are:
   1 *Creation:* A new rule is constructed by the system or acquired from an external entity
   2. *Generalization:* Conditions are dropped or made less restrictive, so that the rule applies in a larger number of situations.
   3 *Specialization* Additional conditions are added to the condition set, or existing conditions made more restrictive, so that the rule applies to a smaller number of specific situations
   4 *Composition:* Two or more rules that were applied in sequence are composed into a single larger rule,

thus forming a "compiled" process and eliminating any redundant conditions or actions

5 **Formal logic-based expressions and related formalisms**—These general-purpose representations have been used to formulate descriptions of individual objects (that are input to a learning system) and to formulate resultant concept descriptions (that are output from a learning system). They take the form of formal logic expressions whose components are propositions, arbitrary predicates, finite-valued variables, statements restricting ranges of variables (such as "a number between 1 and 9"), or embedded logical expressions.

6. **Graphs and Networks**—In many domains graphs and networks provide a more convenient and efficient representation than logical expressions, although the expressive power of network representations is comparable to that of formal logic expressions. Some learning techniques exploit graph-matching and graph-transformation schemes to compare and index knowledge efficiently

7 **Frames and schemas**—These provide larger organizational units than single logical expressions or production rules. Frames and schemas can be viewed as collections of labeled entities ("slots"), each slot playing a certain prescribed role in the representation. They have proven quite useful in many artificial intelligence applications For instance, a system that acquires generalized plans must be able to represent and manipulate such plans as units, although their internal structure may be arbitrarily complex Moreover, in experiential learning, past successes, untested alternatives, causes of failure, and other information must be recorded and compared in inducing and refining various rules of behavior (or entire plans) Schema representations provide an appropriate formalism.

8 **Computer programs and other procedural encodings**—The objective of several learning systems is to acquire an ability to carry out a specific process efficiently, rather than to reason about the internal structure of the process Most automatic programming systems fall in this general category. In addition to computer programs, procedural encodings include human motor skills (such as knowing how to ride a bicycle), instruction sequences to robot manipulators, and other "compiled" human or machine skills Unlike logical descriptions, networks or frames, the detailed internal structure of the resultant procedural encodings need not be comprehensible to humans, or to automated reasoning systems. Only the external behavior of acquired procedural skills become directly available to the reasoning system.

9. **Taxonomies**—Learning from observation may result in global structuring of domain objects into a hierarchy or taxonomy Clustering object descriptions into newly-proposed categories and forming hierarchical classifications require that the system formulate relevant criteria for classification.

10 **Multiple representations**—Some knowledge acquisition systems use several representation schemes for the newly-acquired knowledge Most notably, some discovery and theory-formation systems acquire concepts, operations on those concepts, and heuristic rules for new domains. These learning systems must select appropriate combinations of representation schemes applicable to the different forms of knowledge acquired.

**Classification by Domain of Application** Another useful dimension for classifying learning systems is their area of application The list below specifies application areas to which various existing learning systems have been applied Application areas are presented in alphabetical order, not reflecting the relative effort or significance of the resultant machine learning system.

1. Agriculture
2. Chemistry
3. Cognitive Modeling (simulating human learning processes)
4. Computer Programming
5. Education
6. Expert Systems (high-performance, domain-specific AI programs)
7. Game Playing (chess, checkers, poker, and so on)
8. General Methods (no specific domain)
9. Image Recognition
10. Mathematics
11. Medical Diagnosis
12. Music
13. Natural Language Processing
14. Physical Object Characterizations
15. Physics
16. Planning and Problem-solving
17. Robotics
18. Sequence Extrapolation
19. Speech Recognition

Now that we have a basis for classifying and comparing learning systems, we turn to a brief historical outline of machine learning.

## A Historical Sketch of Machine Learning

Over the years, research in machine learning has been pursued with varying degrees of intensity, using different approaches and placing emphasis on different aspects and goals. Within the relatively short history of this discipline, one may distinguish three major periods, each centered around a different paradigm:

- neural modeling and decision-theoretic techniques
- symbolic concept-oriented learning
- knowledge-intensive approaches combining various learning strategies

**The Neural Modelling Paradigm.** The distinguishing feature of the first paradigm was the interest in building general purpose learning systems that start with little or no initial structure or task-oriented knowledge. The major thrust of research based on this *tabula rasa* approach involved constructing a variety of neural model-based machines, with random or partially random initial structure. These systems were generally referred to as *neural nets* or *self-organizing systems.* Learning in such systems consisted of incremental changes in the probabilities that neuron-like elements (typically threshold logic units) would transmit a signal.

Due to the primitive nature of computer technology at that time, most of the research under this paradigm was either theoretical or involved the construction of special purpose experimental hardware systems, such as perceptrons (Rosenblatt 1958), pandemonium (Selfridge 1959) and adelaine (Widrow 1962). The groundwork for this paradigm was laid in the forties by Rashevsky and his followers working in the area of mathematical biophysics (Rashevsky 1948), and by McCulloch and Pitts (1943), who discovered the applicability of symbolic logic to modeling nervous system activities. Among the large number of research efforts in this area, one may mention many works such as Ashby (1960), Rosenblatt (1958, 1962), Minsky & Papert (1969), Block (1961), Yovits (1962), Widrow (1962), Culberson (1963), Kazmierczak (1963). Related research involved the simulation of evolutionary processes, that through random mutation and "natural" selection might create a system capable of some intelligent behavior (for example, Friedberg 1958, 1959; Holland 1980)

Experience in the above areas spawned the new discipline of pattern recognition and led to the development of a decision-theoretic approach to machine learning. In this approach, learning is equated with the acquisition of linear, polynomial, or related discriminant functions from a given set of training examples Example include Nilsson (1965), Koford (1966), Uhr (1966), and Highleyman (1967). One of the best known successful learning systems utilizing such techniques (as well as some original new ideas involving non-linear transformations) was Samuel's checkers program (Samuel, 1959, 1963). Through repeated training, this program acquired master-level performance Somewhat different, but closely related, techniques utilized methods of statistical decision theory for learning pattern recognition rules (for example, Sebestyen 1962, Fu 1968, Watanabe 1960, Arkadev 1971, Fukananga 1972, Duda & Hart 1973, Kanal 1974).

In parallel to research on neural modeling and decision-theoretic techniques, researchers in control theory developed adaptive control systems able to adjust automatically their parameters in order to maintain stable performance in the presence of various disturbances, for example, Truxal (1955); Davies (1970); Mendel (1970); Tsypkin (1968, 1971, 1973); and Fu (1971, 1974).

Practical results sought by the neural modeling and decision theoretic approaches met with limited success. High expectations articulated in various early works were not realized, and research under this paradigm began to decline Theoretical studies have revealed strong limitations of the "knowledge-free" perceptron-type learning systems.

**The Symbolic Concept-Acquisition Paradigm.** A second major paradigm started to emerge in the early sixties stemming from the work of psychologists and early AI researchers on models of human learning (Hunt *et al*, 1963, 1966). The paradigm utilized logic or graph structure representations rather than numerical or statistical methods Systems learned symbolic descriptions representing higher level knowledge and made strong structural assumptions about the concepts to be acquired. Examples of work in this paradigm include research on human concept acquisition (Hunt & Hovland 1963, Feigenbaum 1963, Hunt *et al* 1966, Hilgard 1966, Simon & Lea 1974) and various applied pattern recognition systems (Bongard 1970, Uhr 1966, Karpinski & Michalski 1966)

Some researchers constructed task-oriented specialized systems that would acquire knowledge in the context of a practical problem. For instance, the META-DENDRAL program (Buchanan & Feigenbaum 1978) generates rules explaining mass spectrometry data for use in the DENDRAL system (Buchanan *et al* 1971).

Winston's (1975) structuraly learning system was an influential development in this paradigm. In parallel with Winston's work, different approaches to learning structural concepts from examples emerged, including a family of logic-based inductive learning programs, AQVAL (Michalski 1972, 1973, 1978), and related work (Hayes-Roth 1974, Hayes-Roth & McDermott 1978, Vere 1975, Mitchell 1978). See Dietterich and Michalski (1983) and Michie (1982) for additional discussion of this paradigm )

**The Modern Knowledge-Intensive Paradigm** The third paradigm represents the most recent period of research starting in the mid-seventies. Researchers have broadened their interest beyond learning isolated concepts from examples, and have begun investigating a wide spectrum of learning methods, most based upon knowledge-rich systems Specifically, this paradigm can be characterized by several new trends, including:

1 **Knowledge-Intensive Approaches:** Researchers are strongly emphasizing the use of task-oriented knowledge and the constraints it provides in guiding the learning process One lesson from the failures of earlier *tabula rasa* and knowledge-poor learning systems is that to acquire new knowledge a system must already possess a great deal of initial knowledge

2 **Exploration of alternative methods of learning:** In addition to the earlier research emphasis on learning from examples, researchers are now investigating a wider variety of learning methods such as learning from instruction (e g , Mostow 1983, Haas & Hendrix 1983, Rychener 1983), learning by analogy (e g , Winston 1979, Carbonell 1983, Anderson 1983), and discovery of concepts and classifications (e g ,

Lenat 1976, Langley, *et al* 1983, Michalski 1983, Michalski & Stepp 1983, Hayes-Roth 1983, Quinlan 1983)

3 **Incorporating abilities to generate and select learning tasks:** In contrast to previous efforts, a number of current systems incorporate heuristics to control their focus of attention by generating learning tasks, proposing experiments to gather training data, and choosing concepts to acquire (e g., Lenat 1976, Mitchell *et al* 1983, Carbonell 1983).

In contrast with the knowledge-free parametric learning methods used in the neural networks, and in contrast with the early symbolic methods that learned isolated, "disembodied" concepts, the current approaches use a wealth of general and domain-specific knowledge However, the availability of large volumes of knowledge does not mean that the inductive inference processes are themselves domain dependent and non-generalizable The generality lies in the inductive inference methods and the power is derived from their ability to use domain knowledge to focus attention and structure new concepts. The current methodological assumption is that machine learning systems, much like humans, must learn incrementally, slowly expanding a highly-organized knowledge base, rather than by some gestalt self-organization process. A recently published book on machine learning (Michalski, Carbonell & Mitchell, 1983) presents some of the major research directions in this general approach.

## References


Anderson, J. A. (1983) Acquisition of proof skills in geometry. In R. S. Michalski, J. G. Carbonell & T M Mitchell (Eds ), *Machine Learning, An Artificial Intelligence Approach* Palo Alto, CA: Tioga Press 191-219

Arkadev, A. G & Braverman, E. M. (1971) Learning in pattern classification machines. Nauka, Moscow

Ashby, W Ross (1960) *Design for a brain, the origin of adaptive behavior* New York: John Wiley and Sons, Inc

Block, H. D. (1961) The perceptron: A model of brain functioning. *I. Rev Math Physics* 34(1):123-135

Bongard, N. (1970) *Pattern recognition.* New York: Spartan Books Translation from Russian original, published in 1967.

Buchanan, B G & Mitchell, T. M. (1978) Model-directed learning of production rules In Waterman, D A & Hayes-Roth, F. (Eds.), *Pattern-directed inference systems* New York: Academic Press.

Buchanan, B G , Feigenbaum, E A & Lederberg, J. (1971) *A heuristic programming study of theory formation in sciences* In *Proceedings of the Second International Joint Conference on Artificial Intelligence,* 40-48 International Joint Conferences on Artificial Intelligence, London

Carbonell, J G. (1981) Issues in computer modeling of cognitive phenomena: An AI perspective (Commentary on K M Colby's 'Modeling a paranoid mind') *The behavioral and brain sciences* 4(4), December

Carbonell, J. G (1983) Learning by analogy: formulating and generalizing plans from past experience In R. S. Michalski, J G Carbonell & T M. Mitchell (Eds ), *Machine Learning, an Artificial Intelligence Approach* Palo Alto, CA: Tioga Press 137-160

Culberson, J T (1963) *The minds of robots* Urbana, Illinois: University of Illinois Press

Davies, W. D. T. (1970) *System identification for self-adaptive control* Part I Machine Learning. New York: Wiley-Interscience, Wiley & Sons, Ltd

Dietterich, T G & Michalski, R. S. (1983) A comparative review of selected methods for learning from examples In J G Carbonell & T M. Mitchell (Eds ), *Machine Learning, an Artificial Intelligence Approach.* Palo Alto, Ca : Tioga Press, 41-82

Duda, R O & Hart, P E (1973) *Pattern classification and scene analysis* New York: John Wiley and Sons

Feigenbaum, E. A. (1963) The simulation of verbal learning behavior In Feigenbaum, E A & Feldman, J. (Eds ), *Computers and Thought* New York: McGraw-Hill

Friedberg, R. M (1958) "A learning machine: Part 1." *IBM Journal* (3) (2):2-13

Friedberg, R., Dunham, B. & North, T. (1959) "A learning machine: Part 2 " *IBM Journal of Research and Development* 3(3):282-287

Fu, K. S. (1968) *Sequential methods in pattern recognition and machine learning.* New York: Academic Press.

Fu, K S (1971) *Pattern recognition and machine learning* New York: Plenum Press.

Fu, K S & Tou, J T. (1974) *Learning systems and intelligent robots.* New York: Plenum Press.

Fukanaga, K (1972) *Introduction to statistical pattern recognition* New York: Academic Press

Haas, N & Hendrix, G G (1983) Learning by being told: acquiring knowledge for information management. In R. S. Michalski, J. G Carbonell & T. M. Mitchell (Eds ), *Machine Learning, an Artificial Intelligence Approach.* Palo Alto, CA: Tioga Press, 405-428.

Hayes-Roth, F (1974) Schematic classification problems and their solution *Pattern Recognition* 6:105-113.

Hayes-Roth, F (1983) Using Proofs and Refutations to Learn from experience. In R. S. Michalski, J G Carbonell & T. M Mitchell (Eds.), *Machine Learning, an Artificial Intelligence Approach* Palo Alto, CA: Tioga Press, 211-240

Hayes-Roth, F & McDermott, J. (1978) An interference matching technique for inducing abstractions *Communications of the ACM* 21(5):401-410

Highleyman, W H. (1967) Linear decision functions, with applications to pattern recognition. *Proceedings of IRE* (50):1501-1504

Hilgard, E. R. & Bower, G. H. (1966) *Theories of learning* - Third Edition New York: Appleton-Century-Groft.

Holland, J. H (1980) Adaptive Algorithms for Discovering and Using General Patterns in Growing Knowledge Bases *Policy Analysis and Information Systems* 4(3).

Hunt, E. B & Hovland, C. I (1966) Programming a Model of Human Concept Formation In Feigenbaum, E. A & Feldman, J (Eds.), *Computers and Thought* New York: McGraw-Hill, 310-325.

Hunt, E B., Marin, J. & Stone, P T (1966) *Experiments in Induction* New York: Academic Press

Kanal, L (1974) *Patterns in Pattern Recognition 1968-1974* IEEE *Transactions on Information Theory* IT-20(6):697-722

Karpinski, J. & Michalski, R. S (1966) A System that Learns to Recognize Hand-written Alphanumeric Characters. Technical Report 35, Proce Institute Automatyki, Polish Academy of Sciences *Machine Learning* Part I 30.

Kazmierczak, H & Steinbuch, K (1963) Adaptive Systems in Pattern Recognition *IEEE Transactions of Electronic Computers* EC-12(5):822-835

Koford, T S. & Groner, G F. (1966) The Use of an Adaptive Threshold Element to Design a Linear Optimal Pattern Classifier *IEEE Transactions-Information Theory* 1T-12:42-50

Langley, P. W., Simon, H. A & Bradshaw, G L. (1983) Rediscovering chemistry with the BACON system. In R. S. Michalski, J G Carbonell & T. M. Mitchell C. (Eds ), *Machine Learning, An Artificial Intelligence Approach* Palo Alto, CA: Tioga Press, 307-330.

Lenat, D B (1976) AM: an artificial intelligence approach to discovery in mathematics as heuristic search. PhD thesis, Stanford University.

Lenat, D B (1983) The role of heuristics in learning by discovery: three case studies In R S Michalski, J G. Carbonell & T M Mitchell (Eds ), *Machine learning, an artificial intelligence approach* Palo Alto, CA: Tioga Press, 243-306.

McCulloch, W. S. & Pitts, W (1943) A logical calculus of ideas imminent in nervous activity Bull Math. Biophysics 5:115-133.

Mendel, T. & Fu, K S (1970) *Adaptive Learning and Pattern Recognition Theory and Applications* New York: Spartan Books

Michalski, R. S. (1972) A variable-valued logic system as applied to picture description and recognition In Nake, F & Rosenfeld, A (Eds ), *Graphic languages* Netherlands: North-Holland

Michalski, R S & Larson (1978) *Selection of most representative training examples and incremental generation of* VLI *hypotheses· The underlying methodology and description of programs* ESEL *and* AQ*11*. Report 867, Universsity of Illinois.

Michalski, R S (1973) AQVAL/1 - Computer implementation of a variable valued logic system VL1 and examples of its application to pattern recognition *In proceedings of the First International Joint Conference on Pattern Recognition.* Washington, D C , 3-17

Michalski, R S (1983) A theory and methodology of learning from examples. In R. S. Michalski, J. G. Carbonell & T. M. Mitchell (Eds.), *Machine Learning, an Artificial Intelligence Approach* Palo Alto, CA: Tioga Press, 83-134

Michalski, R. S., & Stepp, R. E. (1983) Learning from observation: conceptual clustering In R S. Michalski, J. G. Carbonell & T M Mitchell (Eds ), *Machine Learning, an Artificial Intelligence Approach.* Palo Alto, CA: Tioga Press

Michalski, R S , Carbonell, J G , & Mitchell, T. M. (1983) (Eds) *Machine Learning, an Artificial Intelligence Approach* Palo Alto, CA: Tioga Press

Michie D (1982) "The State of the Art in Machine Learning." In D Michie (Ed ), *Introductory Readings in Expert Systems.*UK: Gordon and Breach.

Minsky, M & Papert, S (1969) *Perceptrons* Cambridge, Mass : MIT Press

Mitchell, T. M. (1978) Version spaces: an approach to concept learning PhD thesis, Stanford University Also: Stanford CS report STAN-CS-78-711, HPP-79-2

Mitchell, T M., Utgoff, P E & Banerji, R B (1983) Learning by experimentation: acquiring and refining problem-solving heuristics In R S Michalski, J. G. Carbonell & T. M Mitchell (Eds.), *Machine Learning, an Artificial Intelligence Approach* Palo Alto, CA: Tioga Press 163-190

Mostow, D. J. (1983) Transforming declarative advice into effective procedures: a heuristic search example In R. S. Michalski, J G. Carbonell & T. M. Mitchell (Eds ), *Machine Learning, an Artificial Intelligence Approach* Palo Alto, CA: Tioga Press, 367-404

Newell, A & Rosenbloom, P (1981) Mechanisms of skill acquisition and the law of practice. In Anderson, J. R. (Ed ), *Cognitive Skills and Their Acquisition.* Hillsdale, New Jersey: Erlbaum Associates

Nilsson, N. J. (1965) *Learning machines* New York: McGraw-Hill

Popper, K. (1968) *The logic of scientific discovery* New York: Harper and Row, 2nd edition

Quinlan, J. R (1983) Learning efficient classification procedures and their application to chess end games In R. S Michalski, J. G Carbonell & T M Mitchell (Eds.), *Machine Learning, an Artificial Intelligence Approach* Palo Alto, CA: Tioga Press, 463-480

Rashevsky, N (1948) *Mathematical biophysics.* Chicago, IL: University of Chicago Press

Rosenblatt, F. (1958) "The perceptron: a probabilistic model for information storage and organization in the brain " *Psychological Review* 65:386-407

Rosenblatt, F. (1962) *Principles of neurodynamics and the theory of brain mechanisms* Washington, D. C.: Spartan Books

Rychener, M. D (1983) The instructible production system: a retrospective analysis In R S Michalski, J. G. Carbonell & T. M. Mitchell (Eds.), *Machine Learning, an Artificial Intelligence Approach.* Palo Alto, CA: Tioga Press, 429-460

Samuel, A L (1959) Some studies in machine learning using the game of checkers *IBM Journal of Research and Development* (3):211-229

Samuel, A L. (1963) Some studies in machine learning using the game of checkers In Feigenbaum, E. A. & Feldman, J. (Eds.), *Computers and Thought* New York: McGraw-Hill, 71-105

Sebestyen, G. S. (1962) *Decision-making processes in pattern recognition.* New York: Macmillan Publishers.

Selfridge, O G (1959) Pandemonium: a paradigm for learning In Blake, D & Uttley, A (Eds ), *Proceedings of the symposium on mechanization of thought processes,* London: HMSO, 511-529

Simon, H. A. & Lea, G (1974) Problem solving and rule induction: a unified view In Gregg, L. W. (Ed.), *Knowledge and Cognition* Potomac, Maryland: Lawrence Erlbaum Associates, 105-127

Sleeman, D H (983) Inferring student models for intelligent computer-aided instruction In R S Michalski, J. G Carbonell & T. M. Mitchell (Eds.), *Machine Learning, An Artificial Intelligence Approach* Palo Alto, CA: Tioga Press, 483-510

Sloman, A (1978) *The Computer Revolution in Philosophy: Philosophy, Science And Models of the Mind* New York: Harvester Press

Truxal, T G (1955) *Automatic Feedback Control System Synthesis* New York: McGraw-Hill.

Tsypkin, Y Z (1968)"Self Learning - What is it?" IEEE *Transactions on Automatic Control.* AC-18(2):109-117.

Tsypkin, Y. Z (1971) *Adaptation and Learning in Automatic Systems* New York: Academic Press

Tsypkin, Y Z (1973) *Foundations of the Theory of Learning Systems.* New York: Academic Press.

Uhr, L. (1966) *Pattern Recognition.* New York: John Wiley and Sons

Vere, S A (1975) Induction of concepts in the predicate calculus *In Proceedings of the Fourth International Joint Conference on Artificial Intelligence,* 281-287

Watanabe, S (1960) Information-Theoretic Aspects of Inductive and Deductive Inference. IBM Journal of Research and Development 4(2):208-231.

Widrow, B. (1962) Generalization and Information Storage in Networks of Adelaine 'Neurons,'. Washington, D C.: Spartan Books. 435-461

Winston, P. H. (1975) Learning structural descriptions from examples In Winston, P. H. (Ed ), *The Psychology of Computer Vision,* chapter 5. New York: McGraw Hill. Original version published as a Ph D dissertaition, at MIT AI Lab, September, 1970

Winston, P H. (1979) Learning and Reasoning by Analogy CACM 23(12):689-703

Yovits, M. C., Jacobi, G T. & Goldstein, G. D. (1962) *Self-Organizing Systems* Washington, D C : Spartan Books.