An extended and improved version of the invited paper published in the Proceedings of the First World Conference on the Fundamentals of Artificial Intelligence. Paris, July 1-5, 1991.

TOWARD A UNIFIED THEORY OF LEARNING:
Basic Ideas and a Classification of Learning Processes

Ryszard S. Michalski
Center for Artificial Intelligence
George Mason University
Fairfax, VA 22030

Abstract

The paper presents initial results toward developing a unifying conceptual framework for characterizing diverse learning strategies and paradigms. It outlines the *inferential learning theory* (*ILT*) that aims at understanding the *competence* aspects of learning processes. Thus, the theory has different goals than computational learning theory that is concerned with computational complexity of learning processes. The ILT views learning as a goal-oriented process of modifying learner's knowledge by exploring the learner's experience. Such a process is accomplished by various *knowledge transmutations*, such as selection, replication, reformulation, abstraction, similization, generalization, and/or their opposites, such as generation, deletion, randomization, concretion, dissimilization and specialization, respectively. These transmutations are characterized as specific forms of underlying types of inference, deduction, induction or analogy. Several fundamental concepts, like analytic vs. synthetic learning, induction, abduction, abstraction and generalization, are analyzed in a novel way. It is shown, for example, that inductive generalization, inductive specialization and abductive derivation can be viewed as different forms of induction, and deductive generalization and abstraction are related forms of deduction. The above concepts are used to develop a general classification of learning processes.

Key words: learning theory, machine learning, inferential learning theory, inference types, deduction, induction, abduction, classification of learning.

1. Introduction

The last several years have witnessed a great proliferation of methods and approaches to machine learning. Research in this field has spanned such subareas as empirical concept learning from examples, explanation-based learning, discovery systems, computational learning theory, neural net learning, genetic algorithm based learning, constructive induction, conceptual clustering, cognitive models of learning, reinforcement learning, multistrategy learning, and applications of machine learning to varius practical domains. Among many sources that have reported the progress in various subareas of this field, the reader may consult Laird (1988), Haussler and Pitt (1988), Touretzky, Hinton and Sejnowski (1988), Goldberg (1989), Schafer (1989), Segre (1989), Fulk and Case (1990), Porter and Mooney (1990), Kodratoff and Michalski (1990), and Birnbaum and Collins (1991). In view of such an expansion and diversification of research in machine learning, there is an intense need for developing a conceptual framework that would clarify the interrelationships among different subareas and approaches, and determine the conditions for their most effective applicability.

The purpose of this paper is to report an effort toward the development of such a general framework. It outlines the *inferential learning theory,* which analyzes and characterizes learning processes in terms of operators that transform the initial learner's knowledge to the knowledge desired. These operators are called *knowledge transmutations*, and are characterized in terms of the types of transformation they perform, and of the underlying type of inference involved in this transformation. The main aims of the theory are to investigate the properties of different knowledge transmutations, their relation to different types of inference, and their role in various learning methods and paradigms. The theory strives to provide a basis for developing effective techniques and tools for analyzing diverse learning processes from the viewpoint of their logical capability.

Learning has been traditionally characterized as a behavior change due to experience. While such a view is generally appealing, it does not give many clues into how to actually build learning systems. To build a learning system, one needs to understand, in computational terms, what types of knowledge changes occur in learning, and how they are accomplished in response to different kinds of experience.

To provide answers to such questions, the inferential learning theory assumes that learning is a process of creating or improving knowledge representations by exploring the learner's experience. Such a process can be characterized by the types of knowledge transmutations that transform the initial knowledge (the learner's prior knowledge plus learner's experience) into the knowledge needed for accomplishing the learning goal. These knowledge transmutations include such processes as selection, replication, reformulation, abstraction, similization, generalization, and their opposites.

These operations can be done by a learner explicitly, by well-defined rules of inference, or implicitly, as results of specific mechanisms involved in information processing. Since these operations can be applied in a great variety of ways, learning processes have to be guided by the learning goal(s). The goals can also be expressed explicitly or implicitly. The learner's experience (an input information to the learning process) can be in the form of sensory observations, facts or knowledge communicated by a source of information, e.g., a teacher.

The underlying goal of the Inferential Theory is to understand the *competence* aspects of learning processes, in contrast to the computational learning theory (e.g., Fulk and Case, 1990) that is concerned with the *computational complexity* of the processes. These competence aspects address such questions as what types of knowledge the learner is able to learn from what kinds of inputs, given certain prior knowledge; what is the logical relationship between the learned knowledge, the input information and the learner's prior knowledge; what types of inference and knowledge transformations underlie such processes, etc. The presented work draws upon or is an extension of previous ideas described by Michalski (1983 & 1990a) and Michalski and Kodratoff (1990). The next section presents basic tenets of the inferential learning theory. To explain simply the underlying ideas and research aims, the presentation relies primarily on conceptual explanations and examples, rather than on precise definitions and formal elaborations, which at this stage of the theory could obscure the presentation.

2. Basic tenets of the theory

Any learning process aims at improving the learner's knowledge or skill by interacting with some information source. A key idea of the inferential learning theory is that this improvement is done by performing various *knowledge transmutations* of the learner's prior knowledge and/or external inputs to the learning process. Consequently, the inferential learning theory analyzes learning processes in terms of knowledge transmutations involved in them.

The underlying tenet of the theory is that learning can be usefully viewed as a process of creating or modifying certain knowledge structures to achieve a given learner's goal. Such a process involves an interaction between the learner's prior knowledge, the inputs from an information source, and the goal. These three components, the input, the learner's prior knowledge and the learner's goal, define what we call the *learning task.*

According to the theory, the interactions among the components of a learning task can be characterized, at a conceptual level, in terms of knowledge transmutations. These transmutations affect the relationship between the knowledge and the world the knowledge describes. For example, one of the most important knowledge transmutations is a generalization operation, which transforms knowledge that characterizes a set of entities in the world into knowledge that characterizes a superset of these entities. Knowledge transmutations are bidirectional processes, and are characterized by pairs of opposite operations. Current theory distinguishes among the following knowledge transmutations: selection vs. generation, replication vs. deletion, reformulation vs. randomization, abstraction vs. concretion, similization vs. dissimilization, and generalization vs. specialization. (See Section 4 for more details on knowledge transmutations.) These knowledge transmutations represent different specific forms of underlying types inference, such as deduction, induction or analogy. (See section 3.)

In symbolic learning systems, knowledge transmutations are performed in a more or less explicit way, and in conceptually comprehensible steps. For example, a generalization operation may be done according to some defined rules of generalization (Michalski, 1983). In subsymbolic systems (e.g., neural networks), the transmutations are performed implicitly, in steps solely dictated by the underlying computational mechanism. For example, a neural network may generalize an input example by performing a sequence of small modifications of the weights of internode connections. Although these weight modifications do not directly correspond to any explicit inference steps, they, nevertheless, can be characterized as certain knowledge transformations. For example, Wnek et al. (1990) described a simple method for visualizing generalization operations performed by a neural network, genetic algorithm, and symbolic learning systems (see also Figure 2).

Any learning process needs to be always guided by some underlying goal, otherwise the proliferation of choices of what to learn would quickly overwhelm any realistic system. Such a learning goal can be explicitly defined, or only implicitly defined, by the way the learner processes the input information, by what it pays attention to, etc. The input information (input) can be observations, stated facts, concept instances, previously formed generalizations, conceptual hierarchies, information about the validity of some pieces of knowledge, or some combinations of such types of knowledge. At the beginning of a learning process, the input activates segments of the learner's prior knowledge that are relevant to the learning goal. Such a goal-relevant part of learner's prior knowledge is called *background knowledge* (*BK*).

The BK can be in different forms, e.g., in a declarative form which is most useful for explicit reasoning (conceptual knowledge), or in procedural from, as sequences of instructions for executing specific tasks (control knowledge, skills).

Figure 1 illustrates major components and the information flow in a general learning process, according to the theory. In each learning cycle, the learner analyzes the input information in terms of its BK and its goals, and generates new knowledge and/or a better form of knowledge (depending on the learning goal) through various knowledge transmutations. The results are fed back to the learner's "knowledge base," and may be used in subsequent learning processes.

The central aspect of any knowledge transmutation is the basic type of underlying inference, which may be deduction or induction. As mentioned ealier, transmutations change the relationship between knowledge and the world they describe. The underlying type of inference involved in a transmutation characterizes such a change along the truth-falsity dimension. A deductive transmutation is truth-preserving, and an inductive transmutation is falsity-preserving (see Sec. 3).

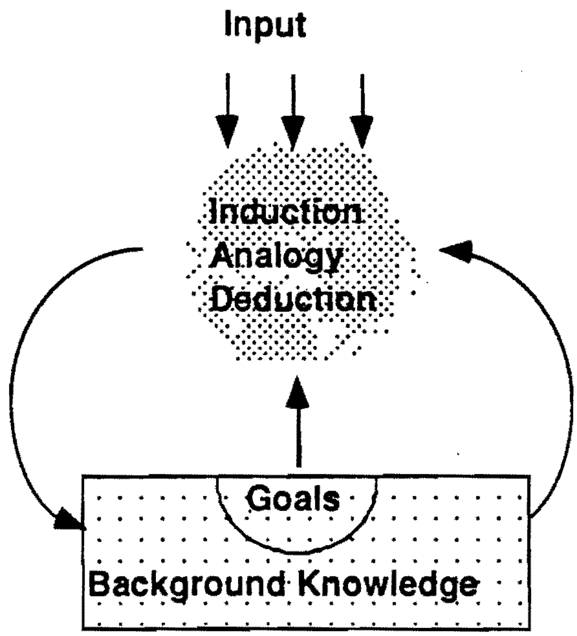


Figure 1. An illustration of a general learning process.

A learning strategy is defined by the overall type of knowledge transmutation that turns the input into the stored knowledge. The lowest strategy is *rote learning* (*direct knowledge implantation*) in which the information from a source is essentially copied from a source into the learner's knowledge base. This strategy thus involves a replication transmutation. Such a process requires a proper arrangement of the new knowledge within the learner's current knowledge structure.

The next level of strategy, *learning from instruction*, involves selecting relevant parts from the knowledge supplied by a source (a selection transmutation), and performing various desirable transformations to fit the learner's conceptual structure (reformulation transmutations).

The above two strategies, rote learning and learning from instruction, are not supposed to change the meaning of the knowledge obtained from the source, therefore they involve only truth-preserving transmutations. In these strategies, the learning process engages primarily the learner's memory, rather than its reasoning capabilities.

Although these two learning strategies are relatively simple in comparison to other strategies, they should not be viewed as uninteresting or unimportant. Their implementation poses a number of significant problems, such as those concerning the choice of knowledge representation, knowledge organization, and access to the stored knowledge. In the case of learning from instruction, there are also problems of determining what parts of the source knowledge are relevant to the learner's goals, and what transmutations are required to assimilate knowledge into the learner's knowledge structures.

These two strategies are also quite important because they are very widely used in human learning, as well as in computer systems. For example, buidling a database can be characterized, from the viewpoint of a computer, basically as a form of direct knowledge implantation. (Modern databases exhibit also elements of learning from instruction, as they are capable of performing some deductive knowledge transformations by themselves.) Most of the current methods of knowledge acquisition for knowledge-based systems can be viewed as combinations of the above two learning strategies plus various knowledge elicitation techniques.

Higher learning strategies require a learner to perform correspondingly more advanced forms of knowledge transmutations. For example, in the explanation based-learning strategy, the overall transmutation can be characterized as deductive generalization. In learning by analogy or case-based learning, the overall transmutation is similization. In learning causal explanations, it is abductive derivation. In learning from examples and learning from observation strategies (the later is also called learning by discovery) it is inductive generalization.

As mentioned earlier, the learning strategy to be applied depends on the learning task (defined by the available input information, learner's prior knowledge, and the learning goal). The input information comes from an information source, which may be the learner's environment, a teacher, or a learner's own internal process.

The learning goal can be implicit or explicit, but is necessary for determining what parts of prior knowledge are relevant, what knowledge is to be acquired, and how to evaluate the learned knowledge. There can be many different types of learning goals, e.g., to solve a problem, to perform an action, to "understand" observed facts, to concisely describe given data, to discover a regularity in a collection of observations, to explain or express a regularity in terms of high level concepts, to confirm a given piece of knowledge, etc. A learner may have more than one goal, and the goals may be conflicting. In such a situation, their relative importance affects the decision about the amount of effort the learner extends in pursuing any of them. A weakness of some machine learning research is that it considers a learning process separately from the learning goal(s), and as a result many developed systems are method-oriented rather than problem-oriented. Studying the role of goals in learning is an important research topic for machine learning.

In summation, the inferential learning theory states that in order to learn, an agent has to be able to perform *inference*, and has to posses *memory* that supplies the BK needed for performing the inference, and records the results of the inference for future use. Without either of the two components—the ability to reason and the ability to store and retrieve information from memory—no learning can be accomplished. Thus, one can write an "equation":

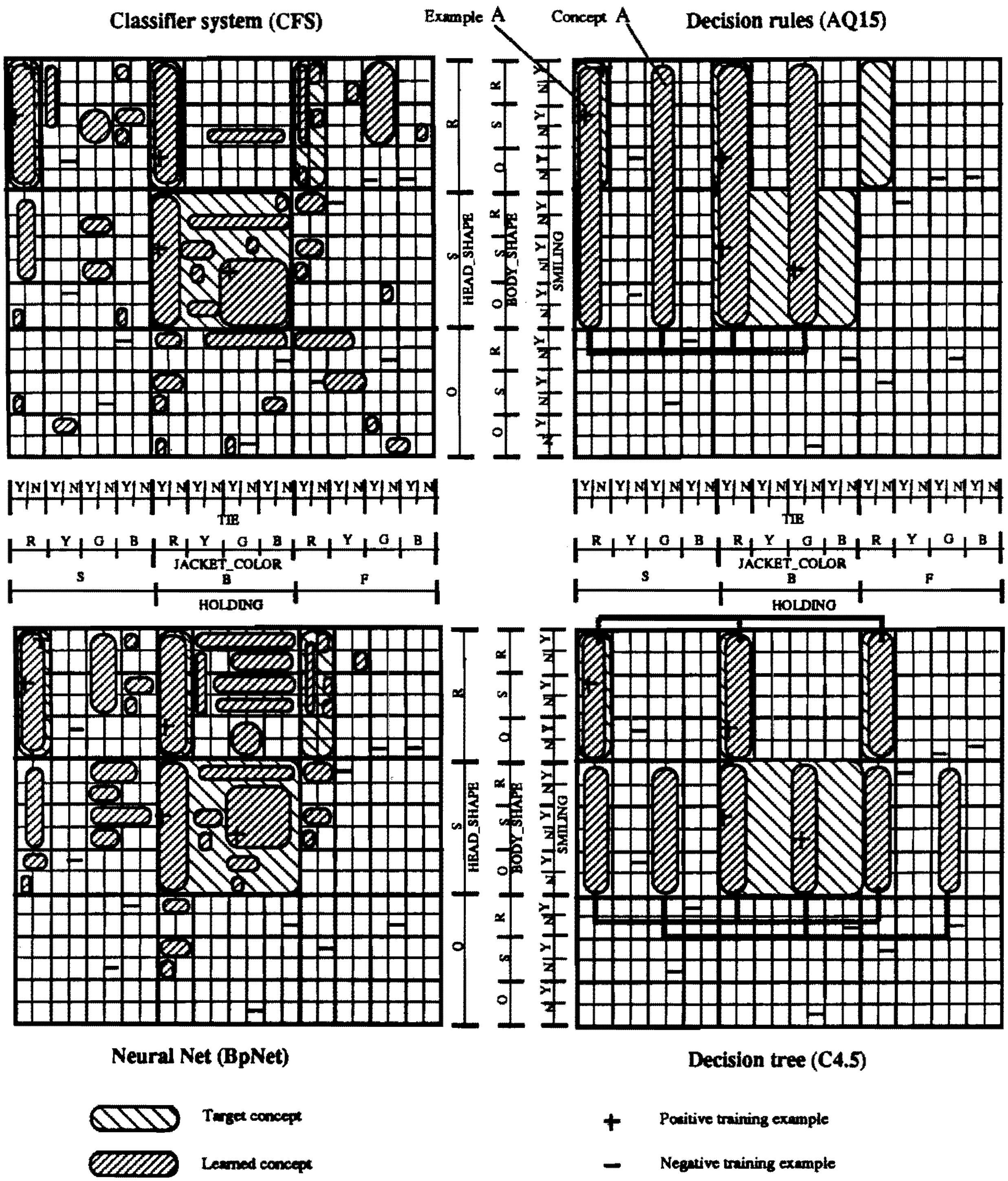
Learning = Inference + Memory

It should be noted that the term "inference" is used here in a very general sense, meaning any type of inference or knowledge transmutation or manipulation, including syntactic or semantic transformations, as well as random searching for a specified entity.

The double role of memory, as a supplier of background knowledge, and as a storer of the results, is often reflected in the organization of a learning system. For example, in a neural net, background knowledge resides in the structure of the network (in the type of units used and in the way they are interconnected), and in the initial weights of the connections. The learned knowledge usually resides only in the new values of the weights. In a decision tree learning system, the BK includes an attribute evaluation procedure and knowledge about the domains of the attributes. The knowledge created is in the form of a decision tree. In a "self-contained" rule learning system, all background knowledge and the learned knowledge would be in the form of rules. A learning process would involve modifying prior rules and/or creating new ones. The ultimate learning capabilities of a learning system are determined by what it can or cannot change in its knowledge base during a learning process.

Because inferential theory views learning as an inference process, it may appear that it only applies to symbolic methods, and does not apply to subsymbolic or hybrid forms of learning, such as neural net learning, reinforcement learning or genetic algorithm-based learning. It is argued that it also applies to them because these methods can also be analyzed from the viewpoint of the types of knowledge transmutations performed by them. They can generalize, specialize, similize, reformulate, select, etc. information, as any other systems.

To illustrate this point, Figure 2 presents "images" of concepts learned by a neural network, a classifier system using a genetic algorithm, a decision tree learning program (C4.5), and a rule learning program (AQ15). Each cell of a diagram represents a single combination of attribute values, i.e., an instance in the description space .



Example A: Head_Shape = R & Body_Shape= S & Smiling= y & Holding=S & Jacket_color =R & Tie=y

Concept A: Head_Shape = S or R & Holding = S or B & Jacket_color = R or G

Figure 2. Images of the *target* concept and the *learned concept* (the concept was learned from 6% positive and 3% negative examples). [from Wnek et al., 1990].

To determine the combination of attribute values corresponding to a given cell (e.g., see Example A in Fig. 2), one projects the cell on the ranges of attributes values associated with the scales aside of the diagram. The area "target concept" includes all possible instances of the concept to be learned. The area (set of cells) marked "target concept" represents all instances that belong to the given concept. The area "learned concept" denotes all instances that a given learning system would classify as belonging to the concept, after the learning process has been completed.

The four diagrams presented in Figure 2 display the "learned concepts" for each of the learning system compared: a classifier system (CFS), and rule learning system (AQ15), and neural net (BpNet) and a decision tree learning system (C4.5).

The set-theoretic difference between the "target concept" and the "learned concept" thus represents errors--an "error image." Each instance in this area will be incorrectly classified by the learned concept. By analyzing the images of the concepts learned by different paradigms, one can determine the degree to which they generalized the original examples, can "see" the differences between these generalizations and can determine the classification of new or hypothetical examples according to the learned concept, etc. (For more details, see Wnek et al., 1990.)

Thus, from the viewpoint of the inferential learning theory, the difference between symbolic and subsymbolic systems is that the latter perform knowledge transformations implicitly, e.g., by modifying weights of connections rather than explicitly, as in the former. The prior knowledge in subsymbolic systems is also represented in an implicit way, e.g., by the structure of the neural net and the initial settings of the weights of the connections. This prior knowledge could also be re-represented, at least conceptually, in the form of logical expressions or rules, and then dealt with as with any other knowledge.

The subsymbolic approaches, obviously, also have the ability to memorize results of their learning. For example, in a neural net, the acquired knowledge is manifested in the new weights of the connections among the net's units.

3. Types of inference

As stated earlier, the inferential theory postulates that a learner learns by conducting inferences to derive the desirable knowledge representation from the input and current BK, and then stores the results for future use. Such a process may involve any type of inference. Therefore, from this viewpoint, a complete learning theory has to include a complete theory of inference.

Such a theory of inference should account for and explain all possible types of knowledge transformations. Figure 3 presents an attempt to schematically illustrate all basic types of inference.

The first major classification is to divide inference types into deductive and inductive. The difference can be explained by considering an entailment:

$$P \cup BK \models C \tag{1}$$

where P denotes a set of statements, called *premise*, BK represents the reasoner's BK (including rules of inference), and C denotes a set of statements, called *consequent*. Deductive inference is deriving consequent C, given premise P and BK. Inductive inference is hypothesizing premise P, given consequent C and BK. If $\models$ is the formal logic entailment (i.e., (1) is a valid formula), then deductive inference can be viewed as "tracing forward" the relationship (1) and induction as "tracing backward" such relationship.

In a general view of deduction and induction, which also captures their approximate or commonsense forms, $\models$ may denote a "weak" entailment, i.e., plausible, probabilistic or partial. The difference between the "strong" (valid) and "weak" entailment leads to another major classification of inference types. Specifically, inferences can be divided into those based on the *universal* or *domain-independent* dependencies, and those based on *contingent* or *domain-dependent* dependencies. A universal dependency between individual statements or sets of statements represents a necessarily true relationship, i.e., a relationship that must be true in all possible worlds. For example, valid rules of inference represent universal dependencies.

To illustrate a universal dependency, consider the statement "All elements of the set X have the property q." If this statement is true, then the statement "x, an element of X, has the property q" must also be necessarily true. This relationship between the statements is true independently of the domain of discourse, i.e., of the nature of elements in the set X.

If a reasoning process involves only statements that are assumed to be true, such as axioms, "true" observations, "true" implications, etc., and/or universal dependencies, then deriving C, given P, is the *universal* (or *crisp*) *deduction*, and hypothesizing P, given C, is *universal* (or *crisp*) *induction*. For example, suppose that BK is "All elements of the set X have the property q" and the input (premise P) is "x is an element of the set X." Deriving a statement "x has the property q" is universal deduction. On the other hand, suppose that BK is "x is an element of the set X" and the input (the observed consequent C) is "x has the property q. Hypothesizing the premise P that "All elements of the set X have the property q" is universal induction.

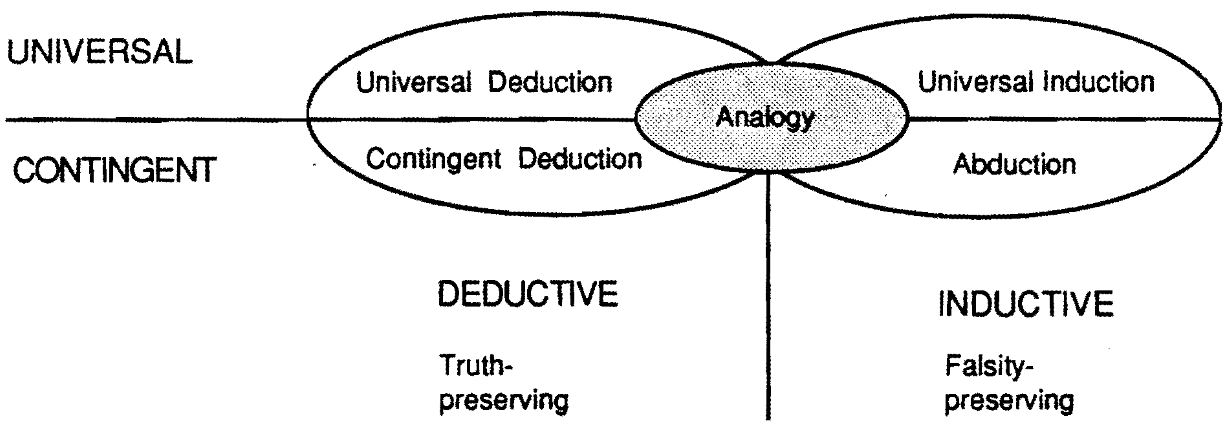


Figure 3. A classification of different kinds of inference.

Contingent dependencies are relationships that are domain-dependent in the sense that they represent some world knowledge that is not totally certain, or that there may be worlds in which they are not true. They can be in the form of probabilistic dependencies, plausible implications, partial dependencies, etc. The contingency of these relationships is usually due to the fact that they represent incomplete, or not totally precise or correct information about all the factors in the world that enter a dependency. These relationships may hold with different "degrees of strength." The conclusions from inferences based on contingent dependencies (even using valid rules of inference) are therefore uncertain, and may be characterized by different "degrees of belief" (probabilities, degrees of truth, likelihoods, etc.). They also usually hold in both directions, although not with the same strength in each direction (Collins and Michalski, 1989).

For example, "If there is fire, then there is smoke" is a (bidirectional) contingent dependency, because there could be a situation or a world in which it is false. If one sees fire, then one may derive (deductively) a conclusion that there may be smoke. This conclusion, however, is not certain. In a reverse direction of reasoning ("tracing backward" the above dependency), observing smoke, one may hypothesize (abductively) that there is fire. This is also an uncertain inference. Therefore, it may appear that there is no principal difference between contingent deduction and abduction.

These two types of inferences are different if one assumes that $\models$ in (1) indicates a causal ordering (i.e., P is viewed as a cause, and C as a consequence). Contingent deduction derives a plausible consequent, C, of the causes represented by P. Abduction derives plausible causes, P, of the consequent C. Contingent deduction can thus be viewed as "tracing forward," and abduction as "tracing backward" such contingent, causally ordered, dependencies.

In sum, both contingent deduction and abduction are based on contingent domain-dependent relationships. Contingent deduction produces likely consequences of given causes, and abduction produces likely causes of given consequences. If a dependency is truly symmetrical (e.g., A <=> B), then the difference between contignent deduction and abduction ceases to exist.

Universal deductive inference is strictly truth-preserving, and universal induction is strictly falsity-preserving (if C is not true, then the hypothesis P cannot be true either). A universal deduction thus produces a provably correct (valid) consequent from a given premise. A universal induction produces a hypothesis that logically entails the given consequent (though the hypothesis itself may be false). Contingent deduction is truth-preserving, and abduction is falsity-preserving only to the extent to which the contingent dependencies involved in reasoning are true.

The intersection of the deduction and induction (i.e., an inference that is both truth-preserving and falsity-preserving for universal or true dependencies), represents an equivalence-based inference. Analogy can be viewed as an extension of such equivalence-based inference, namely, as a similarity-based inference. Every analogical inference can be characterized as a combination of deduction and induction. Induction is involved in detecting an analogical match (i.e., in determination of the properties and/or relations that are similar between the analogs), whereas deduction uses the analogical match to derive unknown properties of the target analog. Therefore, in the diagram, analogy occupies the central area.

As mentioned above, universal induction produces a premise that (together with BK) tautologically implies a given consequent. The tautological implication stems from the set-superset relationship. There are two types of universal induction: *inductive generalization* and *inductive specialization.* Inductive generalization is a widely known form of induction.

For example, given that "*bean 1*, *bean 2*, and *bean 3* from a bag B are white" one may hypothesize that "All beans in bag B are white." Clearly, if the hypothesized premise "All beans in bag B are white," is true, then the given consequent (i.e., *bean 1*, *bean 2*, and *bean 3* from bag B are white) must necessarily be true.

Inductive specialization is a less known form of induction. To illustrate this form, suppose, for example, that we are told that

"There is a University in Virginia designed by Jefferson." (2)

Suppose that knowing (2), and having BK that Charlottesville is a town in Virginia, an agent hypothesizes that

"There is a University in Charlottesville designed by Jefferson." (3)

This would be an example of inductive specialization. To see that this is a form of induction, notice that if (3) is true, then (2) must also be true (assuming the background knowledge is true).

In sum, induction is a process opposite of deduction, that has as its aim to produce justifiable premises that entail consequents, or justifiable explanations for the given facts. These explanations can be in the form of generalizations (theories, rules, laws, etc.), causal explanations, or both. The term "justifiable" is important here because induction is an underconstrained problem, and just "reversing" deduction could lead to an unlimited number of alternatives. For this reason, the "symmetry" between deduction and induction is only partial.

Taking into consideration the above property, the previously given description of inductive inference based on (1) can be further elaborated. Namely, given a consequent C (observations, facts, rules, etc.), and BK BK, the reasoner searches for a hypothetical premise P, consistent with BK, such that

$$P \cup BK \models C \tag{4}$$

and which satisfies the *hypothesis selection criteria.*

In different contexts, the selection criteria are called a *bias* (e.g., Utgoff, 1986), a *comparator* (Poole, 1989), or *preference criteria* (Michalski, 1983). These criteria are necessary for any act of induction because for any given consequent and a non-trivial hypothesis description language there could be a very large set of distinct hypotheses that can be expressed in that language, and that satisfy the relation (4). The selection criteria specify how to choose among them. Ideally, these criteria should reflect the properties of a hypothesis that are desirable from the viewpoint of the reasoner's (or learner's) goals. Often, these criteria (or bias) are partially hidden in the description language used. Specifically, the description language may be limited to only conjunctive statements involving a given set of attributes, or determined by the mechanism performing induction (e.g., a method that generates decision trees is automatically limited to using only operations of conjunction and disjunction in the hypothesis representation).

Generally, these criteria reflect three basic desirable characteristics of a hypothesis: *accuracy*, *utility*, and *generality*. The accuracy expresses a desire to find a "true" hypothesis. Because the problem is logically underconstrained, the "truth" of a hypothesis cannot be guaranteed in principle. To satisfy the entailment (4), a hypothesis has to be *complete* and *consistent* with regard to the input facts (Michalski, 1983). In some situations, however, an inconsistent and/or incomplete hypothesis may give a better overall predictive performance than a complete and consistent one (e.g., Quinlan, 1989; Bergadano et al., 1990). The utility requires a hypothesis to be simple, and easily implementable or applicable to performing an expected set of tasks. The generality criterion expresses the desire to have the hypothesis useful for predicting new cases.

While the above described view of induction is by no means universally accepted, it is consistent with some long-standing scientific thoughts on this subject going back to Aristotle (e.g., Adler and Gorman, 1987; see also the reference under Aristotle). Aristotle, and many subsequent thinkers, e.g., Bacon (1620), Whewell (1857), Cohen (1970) and others, viewed induction as a fundamental inference type that underlies all processes of creating new knowledge. They did not assume that knowledge is created only from low-level observations, without use of prior knowledge, and based only on universal dependencies.

Based on the role and amount of background knowledge, induction, as defined above, can be divided into *empirical induction* and *constructive induction*. In empirical induction there is little background knowledge, and the generated hypothesis is typically expressed using the same terms (attributes, relations, etc.) as the statements in the input (a consequent to be explained). For example, the hypothesis may use the attributes selected from among those that are used in describing the instances in the input to induction. For this reason, such induction is sometimes called *selective* (Michalski, 1983).

In contrast, a constructive induction would use background knowledge and/or experiments to generate additional, more problem-oriented terms or concepts, and use them in the formulation of the hypothesis. To illustrate different kinds of induction, below are a few examples. To test if the inferences are inductive, one needs to see if, given BK and the hypothesis, the input is a logical consequence of them.

Empirical inductive generalization (Background knowledge limited)

Input: The "A girl's face" is a beautiful painting. The "Lvow cathedral" is a beautiful painting.
BK: "A girl's face" and "Lvow cathedral" are paintings by Dawski.

Hypothesis: All paintings by Dawski are beautiful.

Constructive inductive generalization (Backgound knowledge intensive)

Input: The "A girl's face" is a beautiful painting. The "Lvow cathedral" is a beautiful painting.
BK: "A girl's face" and "Lvow cathedral" are paintings by Dawski. Dawski is a known painter. Paintings are pieces of art. Beautiful pieces of art by a known painter are expensive.

Hypothesis: All paintings by Dawski are expensive.

Inductive specialization
Input: John lives is Virginia.
BK: Fairfax is a town in Virginia.

Hypothesis: John lives in Fairfax.

Abductive derivation
Input: There is smoke in the house.
BK: Fire causes smoke.

Hypothesis: There is a fire in the house.

General (constructive) induction: (e.g., generalization plus abduction)
Input: Smoke is coming from John's apartment.
BK: Fire causes smoke. John's apartment is in the Golden Key building.

Hypothesis: The Golden Key building is on fire.
As mentioned earlier, in the most general formulation of induction, the union of BK and a hypothesis may only weakly entail the consequent. In such cases, the hypothesis could be logically inconsistent and/or incomplete in relation to the given the input.

4. Knowledge transmutations

As mentioned before, the process of deriving desirable knowledge from a given input may be characterized in terms of various operations on the input information, and/or the learner's background knowledge. Such elementary operations may be replication and/or deleting some parts of the input, changing the measurement units, changing the quantization of the attributes, or generally reducing the amount of information conveyed by the input.

More complex operations may involve a translation of the input information from one representation system to another, an abstraction, a generalization or specialization, etc. Some operations do not change the inherent meaning of the information, i.e., are truth-preserving, but some may generate hypothetical knowledge. All such operations are *knowledge transmutations*. The transmutations are typically bidirectional operations, and can be characterized by pairs of opposite processes. We distinguish the following basic knowledge transmutations:

1. Selection vs. generation

The selection transmutation selects a subset of knowledge from a given source of knowledge that satisfies some goal. For example, choosing a subset of relevant attributes from a set of candidates, or determining the most plausible hypothesis among a set of candidates is a selection of operation. The opposite operation is generation, that generates additional components of some knowledge structure. For example, generating a new attribute that is not present in the initial set of attributes, or creating a new decision rule is a generation operation.

2. Replication vs. deletion

The replication transmutation involves identifying a relevant knowledge segement in the given source (the input or prior knowledge), and directly reproducing it within the learner's knowledge (e.g., *rote learning*). There is no change in the basic form of knowledge. The learner or teacher must determine what parts of the source are relevant to the learning goal. The opposite operation to replication is *deletion* of some parts of knowledge (e.g., forgetting).

3. Reformulation vs. randomization

The reformulation operation transforms a segment of knowledge about some entity into another segment of knowledge according to well-defined truth-preserving rules of transformation. For example, mapping a subspace represented in a right-angled coordinate system into a radial coordinate system without any other change is a reformulation. A translation of a segment of knowledge from one formal language to another is also a form of reformulation. The opposite operation to reformulation is *randomization*, which transforms a knowledge segment to another one by making random changes. For example, mutation in a genetic algorithm represents a randomization operation. There is a whole range of intermediate transmutations between reformulation and randomization, called in this context *derivations*. For example, a derivation obtained by tracing backward a causal relationship is an *abductive derivation* (see Section 3). Another example of an intermediate derivation is *crossover* in a genetic algorithm.

4. Abstraction vs. concretion

Abstraction reduces the amount of detail in a description of an entity (an object, or a class of objects). To do so, it often transfers a description from one language to another that is more suitable for expressing the properties of the entity that are relevant to the reasoner's goal. The purpose of abstraction is to reduce the amount of information about an entity in such a way that information relevant to the learner's goal is preserved, and other information is discarded. An opposite operation to abstraction is *concretion*, which generates additional details about a given entity.

5. Generalization vs. specialization

The *generalization* operation extends the set of entities to which certain properties are assigned. Generalization is typically inductive, which means that the extended set is inductively hypothesized. Generalization can also be deductive, when a more general assertion is a logical consequence of the more specific one, or is deduced from other knowledge. For example, transforming a statement "Mary went to France" into "Mary went to Europe" is a deductive generalization. The deductive generalization can be viewed as a form of abstraction. The opposite operation to generalization is *specialization*, which narrows the set under consideration. A typical form of specialization is deductive, but, as shown in section 3, there can also be an inductive specialization.

5.Abstraction & Generalization

Since abstraction and generalization are very important operations, and sometimes confused with each other, we give them special attention here. To further elaborate the above description, abstraction is defined as a process of creating a less detailed representation of a given entity from a more detailed representation of this entity, using truth-preserving operations. The latter means that the set of inferences that can be drawn from an abstract description of the entity is a subset of the inferences that can be drawn from the original description of that entity (given the same background knowledge). In other words, details that are preserved should not suggest any new meaning that is not implied by the original description.

To illustrate an abstraction operation, consider a transformation of the statement "My workstation has a Motorola 25-MHz 68030 processor" to "My workstation is quite fast." To make such an operation, the system needs domain-dependent background knowledge that "a processor with the 25-MHz clock speed can be viewed as quite fast," and a rule "If a processor is fast then the computer with that processor can be viewed as fast." Note that the more abstract description is a logical consequence of the original description, and carries less information.

The abstraction process often involves a change in the representation language, from one that uses more specific terms to one that uses more general terms, with a proviso that the statements in the a second language are logically implied by the statements in the first language. A very simple form of abstraction is to replace in a description of an entity a specific attribute value (e.g., the length in a centimeter) by a less specific value (e.g., the length stated in linguistic terms, such as short, medium and long). A more complex abstraction would involve a significant change of the description language, e.g., taking a description of a computer in terms of electronic circuits and connections, and changing it into a description in terms of the functions of the individual modules.

The term abstraction is sometimes confused with generalization. As mentioned before, generalization extends the set under consideration. To illustrate the difference between the two, consider a statement d(S,v), which says that attribute (descriptor) d takes value v for the set of entities S. Let us write such a statement in the form:

$$d(S) = v \tag{5}$$

Changing (5) to the statement d(S) = v', in which v' represents a more general concept (e.g., a parent node in a generalization hierarchy of values of the attribute d), is an abstraction operation. Changing (5) to a statement d(S') = v, in which S' is a superset of S, is a generalization operation. For example, transferring the statement "color(my-pencil) = light-blue" into "color(my-pencil)=blue" is an abstraction operation. Transforming the original statement into "color(all-my-pencils) = light-blue" is a generalization operation. Finally, transferring the original statement into "color(all-my-pencils)=blue" is both generalization and abstraction. In other words, associating the same information with a larger set is a generalization operation; associating a smaller amount of information with the same set is an abstraction operation.

An abstraction process is usually done to serve a certain purpose, namely, to express only the information about some entity that is relevant to a given goal. An abstraction then can be viewed as a process of transforming knowledge from one form to another form, so that information relevant to a given goal is preserved, and irrelevant information is removed. Thus, formally, an abstraction is a transformation:

$$D_1(S) \quad \text{--->} \quad D_2(S) \tag{6}$$

such that

$$INF_G(D_1, BK) \supseteq INF_G(D_2, BK) \tag{6'}$$

where $D_1(S)$ and $D_2(S)$ are descriptions of the set S (in the same or different languages), and $INF_G(D_1)$ and $INF_G(D_2)$ are sets of all deductive inferences, relevant to the goal G, that can be drawn about S from D_1 and D_2, respectively, using BK. If the goal G does not require to remove any parts from the descriptions D_1 and D_2, then (6') is equivalent to saying that D_1 implies D_2 (meaning that if an entity has properties stated by D_1, then it has the properties stated by D_2). The goal defines what parts of the description are relevant and cannot be removed, and what parts of the description can be ignored. Often, the goal of an abstraction process is only implicit. Since an abstraction is a truth-preserving process (from the viewpoint of the goal), it can be considered a form of deduction.

In contrast to the above, generalization is a transmutation that changes the reference set of a given description. Such an operation can be formally characterized as a transformation:

$$D(S_1) \quad \text{--->} \quad D(S_2) \tag{7}$$

where D is a description applied to sets of entities, S_1 and S2, and

$$S_2 \supseteq S1 \tag{7'}$$

When generalizing a set of descriptions, generalization usually involves also a removal of information that is not shared by individual descriptions, and this is a form of abstraction. Thus, a general formulation of generalization includes also an abstraction operation:

$$D_1(S_1) \quad \text{--->} \quad D_2(S_2) \tag{8}$$

where D_1 and D_2 are descriptions of, S_1 and S2, respectively; description D_1 implies D_2, and

$$S_2 \supseteq S1 \tag{8'}$$

As mentioned earlier, a generalization transmutaion can be deductive or inductive depending on the form of the description.

To illustrate these ideas, let us take a *source statement* "John is 6 feet tall, weighs 190 pounds, has blue eyes, and lives in Fairfax." A transformation of this statement into a *target statement:*" John is a big man who lives in Virginia" is an abstraction of the source statement. To make this abstraction one needs to utilize BK that "Being 6 feet tall and weighing 190 pounds classifies one to be called big," and that "Fairfax is a town in Virginia." The implied goal here is that information about the height, weight and the place where a person lives is relevant to the reasoner's goal, while the eye color is not. The abstracted statement clearly tells us less about John, but whatever can be inferred from it about John, can also be inferred from the original statement (given the same BK). The target statement does not introduce or hypothesize any more information about John. The goal is an important component in a general formulation of abstraction, because an abstraction process may introduce information that is incidental, and should not be taken into consideration while making inferences about the entity under consideration. For example, from an abstract drawing of a person one should not infer that the person is made out of paper.

In summary, generalization transforms descriptions along the set-superset dimension, and is typically falsity-preserving (Michalski and Zemankowa, 1991). In contrast, abstraction transforms descriptions along the level-of-detail dimension, and is typically truth-preserving. Generalization often uses the same description space (or language), abstraction often involves a change in the representation space (or language). The reason why generalization and abstraction are frequently confused may be attributed to the fact that many reasoning acts involve both processes. The opposite process to abstraction is called *concretion* (Webster's dictionary defines it as being a process of concretizing something). Given a description of an entity, concretion creates a description that has more details about this entity. Concretion is usually a form of inductive specialization.

As a parallel concept to constructive induction, discussed before, one may introduce the concept of *constructive deduction.* Similarly, to constructive induction, constructive deduction is a process of deductively transforming a source description into a target description, which uses new, more relevant terms and concepts than the source description. As in constructive induction, the process uses background knowledge for that purpose.

Looking at abstraction from such a viewpoint, one may classify it as a form of constructive deduction. The latter is a more general concept than abstraction, however, as it also includes any other possible deductive knowledge transformations resulting in descriptions that contain concepts that were not present in the original description. For example, changing the problem representation space may be a form of constructive deduction, but not an abstraction.

In a general sense, constructive deduction may also involve various forms of probabilistic reasoning (e.g., Schum, 1986; Pearl, 1988), or plausible reasoning (e.g., Collins and Michalski, 1989). In such cases, the distinction between constructive induction and constructive deduction becomes just a matter of the degree to which different forms of reasoning are stressed.

6. A classification of learning processes

Learning processes can be classified according to many criteria, such as the type of learning strategy used, the type of knowledge representation employed, the way information is supplied to a learning system, the application area, etc. Classifications based on such criteria have been discussed, for example, in Carbonell, Michalski and Mitchell (1983) and Michalski (1986).

The inferential learning theory offers a new way of looking at learning processes, and suggests some additional classification criteria. The theory considers learning as a knowledge transformation process whose primary goal is typically either to increase the amount of learner's knowledge or its effectiveness. Therefore, the primary learning goal (or purpose) can be used as a major criterion for classifying learning processes.

Based on this criterion, learning processes are divided into two categories—synthetic and analytic. The main goal of synthetic learning is to acquire new knowledge that goes beyond the knowledge already possessed, or beyond its deductive closure. The primary inference types involved in synthetic processes are induction and/or analogy. The word "primary" is important, because every inductive or analogical inference also involves deductive inference. The latter form is used, for example, to test whether a generated hypothesis entails the observations, to perform an analogical knowledge transfer based the hypothesized analogical match and to generate new terms using background knowledge, etc.

The main goal of analytic learning processes is to transform knowledge that the learner already possesses into the form that is most desirable according to the given learning goal. For example, one may know how to type on the typewriter, and through practice learns how to do it more rapidly. Likewise, one may have a complete knowledge of how an automobile works, and therefore can in principle diagnose the problems with it. By analytic learning, one can learn simple tests for more efficient diagnosis. From the viewpoint of the inferential theory, the primary inference type used in analytic learning is deduction.

Other important criteria include the type of input information, the type of primary inference employed and finally, the role of the learner's background knowledge in the learning process.

Figure 4 presents a classification of learning processes according to the above criteria. This classification shows the basic characteristics of all major machine learning approaches and paradigms. The categories presented are not to be viewed as having precisely delineated borderlines, but rather as labels of central tendencies that can transform from one to another by differently emphasizing various principal components.

If the input to a synthetic learning method are examples classified by an independent source of knowledge, e.g., a teacher, then we have *learning from examples*. When the input includes facts that need to be described or organized into a knowledge structure by the learner itself, then we have *learning from observation*. The latter is exemplified by learning by discovery, conceptual clustering and theory formation.

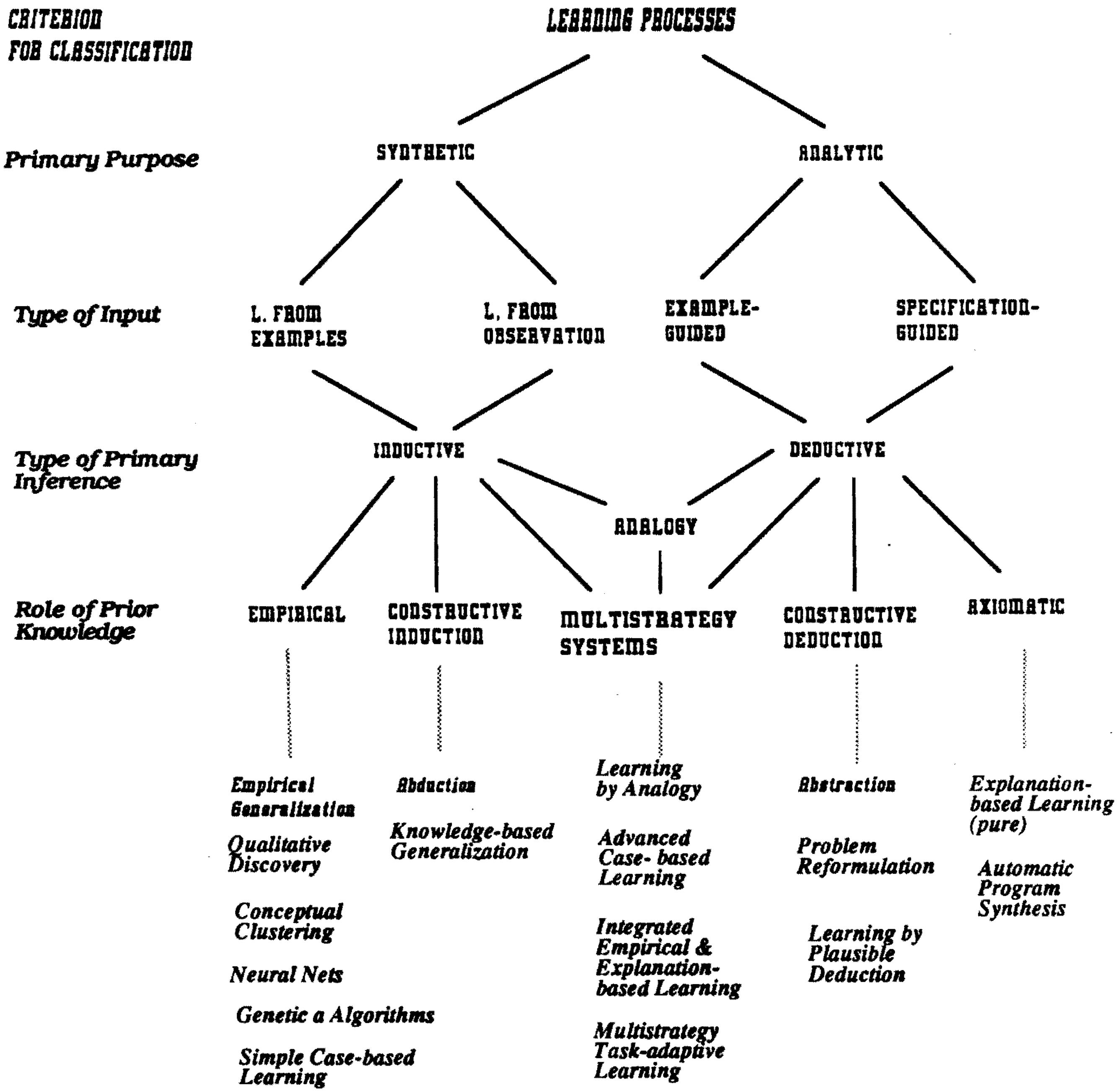


Figure 4. A general classification of learning processes.

The primary type of inference used in synthetic learning is induction. As described earlier, inductive learning can be empirical (background knowledge-limited) or constructive (background knowledge-intensive). Most work in empirical induction has been concerned with empirical generalization of concept examples using attributes selected from among those present in the descriptions of the examples. Another form of empirical learning includes quantitative discovery, in which learner constructs a set of equations characterizing given data.

Learning methods employed in neural nets or genetic algorithms can also be viewed as forms of empirical inductive learning. They typically rely on relatively small amounts of background knowledge, and their primary inference type is inductive. This inference, however, is not executed in an explicit way, like in typical symbolic methods, but in an implicit way.

In contrast to empirical inductive learning, constructive inductive learning is knowledge-intensive, as it uses background knowledge to create new and/or high-level characterizations of the input information . Such characterizations may use terms (attributes, relations, etc.) not present in the input. As described before, abduction can be viewed as a form of constructive induction, which "traces backward" domain-dependent rules.

To be more complete, we will discuss some other classifications of inductive methods, not shown in this classification.

One classification is based on the way facts or examples are presented to the learner. If examples are presented all at once, then we have one-step or batch (non-incremental) inductive learning. If they are presented one by one, or in portions, and the system may have to modify the hypothesis after each input, we have an incremental inductive learning. Incremental learning may be without memory, with partial memory, or with complete memory of past facts.

Another classification is based on whether the input can be assumed to be totally correct, or that it can have errors and/or noise.

The third classification characterizes methods based on the types of matching instances with concept descriptions. Such matching can be done in a direct or an indirect way. The latter employs a substantial amount of background knowledge. For example, case-based learning methods employ matching procedures that allow the system to recognize new examples that do not directly match any past example (e.g., Bareiss, Porter and Wier, 1990). Learning methods based on the *two-tiered concept representation* (Bergadano et al., 1990) also use sophisticated matching procedures.

Analytic methods can be divided into those that are guided by an example in the process of knowledge reformulation (example-guided), and those that start with a specification (specification-guided). The former category includes *explanation-based learning* (e.g., DeJong et al., 1986), *explanation-based-generalization* (Mitchell et al., 1986), and *explanation-based specialization* (Minton et al., 1987; Minton, 1988). If deduction employed in the method is based on axioms, then it is called *axiomatic*. A "pure" explanation-based generalization can be viewed as an example of an axiomatic method because it is based on a deductive process that utilizes complete and consistent domain knowledge. This domain knowledge plays the role analogous to the axioms in formal theories. Synthesizing a computer program from its formal specification is another form of this class of learning processes. Analytic methods that involve deductive transformations of description spaces are classified as methods of "constructive deduction." A major member of this class are methods that utilize abstraction as the primary knowledge transformation operation. Other members of this class include methods of transforming problem representation spaces, or utilizing contingent (e.g., plausible) deduction.

Multistrategy learning systems integrate two or more learning strategies or paradigms. Among the most widely known systems that can be classified into this category are Unimem (Lebowitz, 1986), Odysseus (Wilkins, Clancey, and Buchanan, 1986), Prodigy (Minton et al., 1987), DISCIPLE-1 (Kodratoff and Tecuci, 1987), GEMINI (Danyluk, 1987 and 1989), OCCAM (Pazzani, 1988), IOE (Dietterich and Flann, 1988) and ENIGMA (Bergadano et al., 1990). Other examples are in Segre (1989) and Porter and Mooney (1990). With few exceptions, existing multistrategy systems are concerned with integrating an empirical method with an explanation-based method. Some, like DISCIPLE, also include an analogical learning. The integration of these methods is typically done in a predefined, problem-independent way.

An approach to building a *multistrategy task-adaptive learning* (MTL) is outlined in Michalski, 1990a. An MTL system is supposed to determine by itself which strategy or a combination thereof is most suitable for a given learning task.

The type of knowledge representation employed in a learning system can be used as another dimension for classifying learning systems (not shown in Figure 4). That is, learning systems can be classified on the basis of the knowledge representation employed, e.g., a logic-style representation, production rules, frames, semantic network, grammar, decision tree, neural network, classifier system, PROLOG program, etc., or a combination of different representations. The knowledge representation used in a learning system is often dictated by the application domain. It also depends on the type of learning strategy employed, as not every knowledge representation is suitable for every type of learning strategy.

Thus, in parallel to multistrategy systems that combine several strategies, one can also distinguish multirepresentation learning systems that apply different knowledge representations in the process of learning. Such systems might employ various forms of constructive deduction or constructive induction to create and use representations at different levels of abstraction. The latter systems would thus be capable of changing the representation of the original problem statements. The importance of this area has been acknowledged by pattern recognition researchers (e.g., Bongard, 1970), as well as by AI researchers (Amarel, 1986; Mozetic, 1989).

Summarizing, reasoning/learning processes can be described in terms of three major dimensions characterizing the relationship between the input to the output:

A. the type of logical relationship: induction vs. deduction vs. analogy.

B. the direction and the degree of the change in the reference set: generalization vs. specialization.

C. the direction and the degree of change in the level-of-detail dimension: abstraction vs. concretion.

Each dimension corresponds to a different mechanism of knowledge transformation that may occur in a learning process, and may involve two opposite operations. The operations involved in the first two mechanisms, induction vs. deduction, and generalization vs. specialization, have been relatively well-explored in machine learning. The operations involved in the third mechanism, abstraction vs. concretion, have been relatively less studied. Because these three mechanisms are interdependent, not all combinations of operations can occur in a single learning process (Michalski and Zemankova, 1991). The problems of how to properly and effectively measure the amount of change in the reference set and in the level-of-detail of descriptions are important topics for future research.

The "grand" classification above appears to be the first attempt to characterize and relate to each other all major methods and subareas of machine learning. As such it can be criticized on various grounds. As any classification, this classification is useful only to the degree to which it illustrates important distinctions and relations among various categories. The ultimate goal of this classification effort is to show that diverse learning mechanisms and paradigms can be viewed as parts of one general structure, rather than as a collection of unclearly related components and research efforts.

7. Summary

The goals of this research are to develop a theoretical framework and an effective methodology for characterizing and unifying diverse learning strategies and approaches. The proposed Inferential Theory looks at learning as a process of making knowledge transformations. Consequently, it proposes to analyze any learning method or strategy in terms of the types of knowledge transformations that occur in the learning process. Basic knolwedge transformations have been classified according to three interrelated dimensions, defined by the type of the logical relationship between an input and output (induction vs. deduction vs. analogy), by the change in the reference set (generalization vs. specialization), and by the change in the level-of-detail dimension (abstraction vs. concretion).

The classification of types of inferences proposed here relates to each other such basic types of inference as deduction, induction, abduction and analogy. It has been shown that in addition to widely known inductive generalization, one can also distinguish inductive specialization. Is has been also shown that abduction can be viewed as a form of general induction, and abstraction as a form of deduction. These concepts have been used to develop a general classification of learning processes. The proposed inferential learning theory can serve as a basis for the development of multistrategy learning systems that combine diferent learning strategies and paradigms. Early results in this direction have led to the formulation of the *multistrategy task-adaptive learning methodology*, that dynamically chooses the learning strategy, or a combination of them, according to the learning task (Michalski, 1990a).

Many of the ideas discussed are at a very early state of development, and many issues have not been resolved. For example, future research should develop more precise characterization of various concepts discussed, and effective methods for characterizing different knowledge transformations, and for measuring their "degrees." Another important research topic is to determine how basic operations in various learning algorithms and paradigms map into the described knowledge transformations.

In conclusion, the inferential learning theory provides a new viewpoint for analyzing and characterizing learning processes. By addressing their logical capabilities and limitations, it strives to analyze and understand the competence aspects of diverse learning processes. Among its major goals are to develop effective methods for determining what kind of knowledge a learner can acquire from what kind of inputs, and for formally characterizing the knowledge transformations occuring in diverse learning systems. Related goals are to develop a clear understanding of the areas of the most effective applicability of different learning methods and paradigms, and to gain new insights into how to build more advanced learning systems.

Acknowledgments

The author expresses his gratitude to Hugo De Garis, Ken DeJong, Bob Giansiracusa, Mike Hieb, Heedong Ko, Yves Kodratoff, David Littman, Elizabeth Marchut, David A. Schum, Gheorge Tecuci, Brad Utz, Janusz Wnek and Jianping Zhang for insightful comments on various topics reported in this paper. Thanks also go to Janet Holmes and Susan Lyons for stylistic suggestions and proofreading.

This research was supported in part by the Defense Advanced Research Projects Agency under the grants administered by the Office of Naval Research No. N00014-K-85-0878 and N00014-91-J-1854, and in part by the Office of Naval Research under grants No. N00014-88-K-0397, No. N00014-88-K-0226 and No. N00014-91-J-1351.

References

Adler, M. J., Gorman (Eds..) The Great Ideas: A Synoptic of Great Books of the Western World, Vol. 1, Ch. 39, *Encyclopedia Britannica*, 1987.

Amarel, S., "Program Synthesis as a Theory Formation Task: Problem Representations and Solution Methods," in *Machine Learning: An Artificial Intelligence Approach Vol.II*, Morgan Kaufmann, Los Altos, CA, R. S. Michalski, J. G. Carbonell and T. M. Mitchell (Eds.), 1986.

Aristotle, Posterior Analytics, in *The Works of Aristotle*, Volume 1, R. M. Hutchins (Ed.), Encyclopedia Britannica, Inc., 1987.

Bacon, F., *Novum Organum*, 1620.

Bareiss, E. R., Porter, B. and Wier, C.C., PROTOS, An Exemplar-based Learning Apprentice, in *Machine Learning: AN Artificial Intelligence Approach vol. III*, Morgan Kaufmann, 1990.

Bergadano, F., Matwin, S., Michalski, R.S. and Zhang, J., Learning Two-tiered Descriptions of Flexible Concepts: The POSEIDON System, *Machine Learning and Inference Reports, No. MLI-3,* Center for Artificial Intelligence, George Mason University, 1990.

Birnbaum, L. and Collins, G., *Proceedings of the 8th International Conference on Machine Learning*, Chicago, June 1991.

Bongard, N., *Pattern Recognition,* Spartan Books, New York, 1970 (translation from Russian).

Carbonell, J. G., Michalski R.S. and Mitchell, T.M., An Overview of Machine Learning, in *Machine Learning: AN Artificial Intelligence Approach,* Michalski, R.S., Carbonell, J.G., and Mitchell , T. M. (Eds.), Morgan Kaufmann Publishers, 1983.

Cohen, L.J., *The Implications of Induction*, London, 1970.

Danyluk, A.P., "The Use of Explanations for Similarity-Based Learning," *Proceedings of IJCAI-87*, pp. 274-276, Milan, Italy, 1987.

Danyluk. A. P., "Recent Results in the Use of Context for Learning New Rules," *Technical Report No.* TR-98-066, Philips Laboratories, 1989.

DeJong, G. and Mooney, R., "Explanation-Based Learning: An Alternative View," *Machine Learning Journal,* Vol 1, No. 2, 1986.

Dietterich, T.G., and Flann, N.S., "An Inductive Approach to Solving the Imperfect Theory Problem," *Proceedings of 1988 Symposium on Explanation-Based Learning,* pp. 42-46, Stanford University, 1988.

Fulk, M. and Case, J. *Proceedings of the 3rd Annual Workshop on Computational Learning Theory*, University of Rochester, N.Y., August 6-8, 1990.

Goldberg, D.E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.

Haussler D. and Pitt, L. (Eds.), *Proceedings of the 1988 Workshop on the Computational Learning Theory* (COLT 88), Morgan Kaufmann Publishers, San Mateo, CA, 1988.

Kodratoff, Y., and Tecuci, G., "DISCIPLE-1: Interactive Apprentice System in Weak Theory Fields," *Proceedings of IJCAI-87*, pp. 271-273, Milan, Italy, 1987.

Kodratoff, Y. and Michalski, R.S., (eds.) *Machine Learning: An Artificial Intelligence Approach Volume III*, Morgan Kaufmann Publishers, Inc. 1990.

Laird, J.E., (Ed.), *Proceedings of the Fifth International Conference on Machine Learning*, University of Michigan, Ann Arbor, June 12-14, 1988.

Lebowitz, M., "Integrated Learning: Controlling Explanation," *Cognitive Science*, Vol. 10, No. 2, pp. 219-240, 1986.

Michalski, R. S., "Theory and Methodology of Inductive Learning," *Machine Learning: An Artificial Intelligence Approach*, R. S. Michalski, J. G. Carbonell, T. M. Mitchell (Eds.), Tioga Publishing Co., 1983.

Michalski, R.S., Understanding the Nature of Learning: Issues and Research Directions, in *Machine Learning: An Artificial Intelligence Approach Vol. II,* Michalski, R.S., Carbonell, J.G., and Mitchell , T. M. (Eds.), Morgan Kaufmann Publishers, 1986.

Michalski, R.S., Toward a Unified Theory of Learning: Multistrategy Task-adaptive Learning, *Reports of Machine Learning and Inference Laboratory MLI-90-1,* January 1990a.

Michalski, R.S., LEARNING FLEXIBLE CONCEPTS: Fundamental Ideas and a Method Based on Two-tiered Representation, in *Machine Learning: An Artificial Intelligence Approach vol. III*, Kodratoff, Y. and Michalski, R.S. (eds.), Morgan Kaufmann Publishers, Inc., 1990b.

Michalski, R.S.and Kodratoff, Y. "Research in Machine Learning: Recent Progress, Classification of Methods and Future Directions," in *Machine Learning: An Artificial Intelligence Approach vol. III*, Kodratoff, Y. and Michalski, R.S. (eds.), Morgan Kaufmann Publishers, Inc., 1990.

Michalski, R. S. and Zemankova, M., "What is Generalization: An Inquiry into the Concept of Generalization and its Types," to appear in *Reports of Machine Learning and Inference Laboratory*, Center for Artificial Intelligence, George Mason University, 1991.

Minton, S., "Quantitative Results Concerning the Utility of Explanation-Based Learning," *Proceedings of AAAI-88*, pp. 564-569, Saint Paul, MN, 1988.

Minton, S., Carbonell, J.G., Etzioni, O., et al., "Acquiring Effective Search Control Rules: Explanation-Based Learning in the PRODIGY System," *Proceedings of the 4th International Machine Learning Workshop*, pp. 122-133, University of California, Irvine, 1987.

Mitchell, T.M., Keller,T., Kedar-Cabelli,S., "Explanation-Based Generalization: A Unifying View," *Machine Learning Journal*, Vol. 1, January 1986.

Mozetic, I., Hierarchical Model-based Diagnosis, *Reports of Machine Learning and Inference Laboratory*, No. MLI89-1, 1989.

Pazzani, M.J., "Integrating Explanation-Based and Empirical Learning Methods in OCCAM," *Proceedings of EWSL-88*, pp. 147-166, Glasgow, Scotland, 1988.

Pearl J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, 1988.

Poole, D., Explanation and Prediction: An Architecture for Default and Abductive Reasoning, *Computational Intelligence*, No. 5, pp. 97-110, 1989.

Porter, B. W. and Mooney, R. J. (eds.), *Proceedings of the 7th International Machine Learning Conference*, Austin, TX, 1990.

Quinlan, J. R., "Probabilistic Decision Trees," chapter in *Machine Learning: An Artificial Intelligence Approach, Vol. III*, Y. Kodratoff and R. S. Michalski (eds.), Morgan Kaufmann, Los Altos, CA, 1989.

Schafer, D., (Ed.), *Proceedings of the 3rd International Conference on Genetic Algorithms*, George Mason University, June 4-7, 1989.

Schum, D.,A.,"Probability and the Processes of Discovery, Proof, and Choice," *Boston University Law Review*, Vol. 66, No 3 and 4, May/July 1986.

Segre, A. M. (Ed.), *Proceedings of the Sixth International Workshop on Machine Learning*, Cornell University, Ithaca, New York, June 26-27, 1989.

Touretzky, D., Hinton, G., and Sejnowski, T. (Eds.), *Proceedings of the 1988 Connectionist Models*, Summer School, Carnegie Mellon University, June 17-26, 1988.

Utgoff, P. Shift of Bias for Inductive Concept Learning, in *Machine Learning: An Artificial Intelligence Approach Vol. II,* Michalski, R.S., Carbonell, J.G., and Mitchell , T. M. (Eds.), Morgan Kaufmann Publishers, 1986.

Whewell, W., *History of the Inductive Sciences*, 3 vols., Third edition, London, 1857.

Wilkins, D.C., Clancey, W.J., and Buchanan, B.G., *An Overview of the Odysseus Learning Apprentice*, Kluwer Academic Press, New York, NY, 1986.

Wnek, J., Sarma, J., Wahab, A. A. and Michalski,R.S., COMPARING LEARNING PARADIGMS VIA DIAGRAMMATIC VISUALIZATION: A Case Study in Concept Learning Using Symbolic, Neural Net and Genetic Algorithm Methods, *Proceedings of the 5th International Symposium on Methodologies for Intelligent Systems*, University of Tennessee, Knoxville, TN, North-Holland, October 24-27, 1990.