INDUCTIVE LEARNING: A REVIEW OF SOME RECENT WORK

by

Ryszard S. Michalski

Invited paper for the Third Yale Workshop on Applications of Adaptive Systems Theory, June 15-17, 1983.

Invited Paper for the Third Yale Workshop on Applications of Adaptive Systems Theory June 15-17, 1983

INDUCTIVE LEARNING: A Review of Some Recent Work

Ryszard S. Michalski

Department of Computer Science
University of Illinois
Urbana, Illinois 61801

EXTENDED ABSTRACT

Inductive learning is a process of constructing general descriptions of phenomena from scattered observations and specific facts about them. This form of learning represents a fundamental strategy by which people acquire knowledge about their environment.

Recently, inductive learning has received increasing attention from artificial intelligence researchers, because it is viewed as a key for improving methods of introducing knowledge into computers. Knowledge acquisition is the bottleneck in the development of modern knowledge-intensive AI systems. Such systems include expert systems that provide consultation and advice in various domains of application, vision systems, natural language understanding systems, speech recognition systems and vision-equiped robots.

The central distinctive feature of inductive learning systems is that they have the ability to create, modify, generalize or specialize complex symbolic descriptions. These descriptions may be represented differently in different systems. They may be sets of expressions in predicate or propositional logic, semantic networks, decisions trees or the so-called frames or scripts. The latter descriptions can be viewed as units of steoretypical knowledge about some aspect of our experience. For example, a restaurant "script" can be a sequence of typical events that happen when one visits a restaurant. A "room" frame may be a data structure specifying typical components of a room. Frames and scripts are usualy organized into a generalization/specialization hierarchy. Such a hierarchy makes it possible for specialized frames (those representing more specific concepts) to inherit properties of more general frames. An example of a frame-based inductive learning system is AM (Lenat, 1981). Given a set of initial concepts about numbers and sets, as well as various heuristic rules, the system generates new concepts and new relationships. Those which score high on "interestingness" attribute are retained, and those which score low are forgotten. Thus, the working of the system can be compared to a process of mathematical research.

An advantage of symbolic logic-based representations is that they have a well defined syntax and semantics, and have precise inference rules. Because of this, a number of researchers have used such logic-based descriptions. For a review of some methods using logic as an underlying formalism see [Dietterich and Michalski, 1981], and [Dietterich et al, 1981].

This research was supported in part by National Science Foundation under grant No. MCS 82-05166.

The inductive learning process can be characterized formally by the following paradigm:

Given:

(a) *observational statements (facts)*, F, that represent specific knowledge about some objects, situations, processes, etc.,

(b) *tentative inductive assertion* (which may be null),

(c) *background knowledge* that defines the assumptions and constraints imposed on the observational statements and generated candidate inductive assertions, and any relevant problem domain knowledge. The last includes the *preference criterion* characterizing the desirable properties of the sought inductive assertion.

Find:

An *inductive assertion (hypothesis)*, H, that tautologically or weakly implies the observational statements, and satisfies the background knowledge.

A hypothesis H tautologically implies facts F if F is a logical consequence of H, i.e., if the expression $H \Rightarrow F$ is true under all interpretations ('$\Rightarrow$' denotes logical implication). This is expressed as follows.

$$H \rhd F \quad \text{(read: H \textit{specializes} to F)}$$

or

$$F \mid\!< H \quad \text{(read: F \textit{generalizes} to H).}$$

Symbols $\rhd$ and $\mid\!<$ are called the *specialization* and *generalization* symbols, respectively. If $H \Rightarrow F$ is valid, and H is true, then by the law of detachment (modus ponens) F must be true. Deriving F from H (deductive inference), is, therefore, truth-preserving. In contrast, deriving H from F (inductive inference) is not truth-preserving, but falsity-preserving, i.e., if some facts falsify F then they also must falsify H.

The condition that H *weakly implies* F means that facts F are not certain but only plausible or partial consequences of H. By allowing weak implication, this paradigm includes methods for generating "soft" hypotheses, which hold only probabilistically, and partial hypotheses, which account for some but not all of the facts (e.g., hypotheses representing 'dominant patterns' or characterizing inconsistent data).

Two major types of inductive learning can be distinguished:

1. learning from examples (concept acquisition)
2. learning from observation (descriptive generalization)

In learning from examples there is a teacher (or "oracle") that generates examples and counter-examples of the concept to be learned. In learning from observation, there is no teacher: the system receives various facts and attempts to produce from them a consistent body of general rules.

An example of a simple inductive learning program is AQ11 (Michalski & Larson, 1978). Given a set of vectors (sequences of attribute-value pairs) characterizing "positive" and "negative" examples of a concept, the system infers a general description of the concept. For example, here is a decision rule for diagnosing a soybean disease "Charcoal Rot" that was inductively learned from several specific cases of the disease:

$$[\text{leaves} = n] \;\&\; [\text{stem} = \text{abn}] \;\&\; [\text{internal discoloration} = \text{black}]$$
$$::> [\text{Diagnosis} = \text{Charcoal Rot}]$$

For contrast, the rule below describes the same disease, but was formulated from a definition of the disease provided by a plant pathologist:

$$Q_s([\text{time} = \text{Jul..Aug}] \;\&\; [\text{precip.} <= n] \;\&\; [\text{temp.} >= n] \;\&\;$$
$$[\text{growth} = \text{abn}] \;\&\; \underline{[\text{leaves} = n]} \;\&\; \underline{[\text{stem} = \text{abn}]} \;\&\; [\text{sclerotia} = p] \;\&\;$$
$$[\text{roots} = \text{rotted}] \;\&\; \underline{[\text{internal discoloration} = \text{black}]})$$

$$+$$

$$Q_c([\text{damaged area} = \text{upland areas}] \;\&\; [\text{severity} = \text{severe}] \;\&\;$$
$$[\text{seed size} = \text{small v medium}] \;\&\; [\#\ \text{years crop repeated: } {}^{\wedge}\text{ER2}])$$
$$::> [\text{Diagnosis} = \text{Charcoal Rot}]$$

where Q_s and Q_c are coefficients representing "significant" and "confirmatory" weights and ^ER2 is a function specifying the dependence of the weight coefficient on the value of the variable "#years crop repeated":

$$\text{ER2} = \begin{cases} 1.0, \text{ if } \#\ \text{years crop repeated} >= 2 \\ 0.6, \text{ if } \#\ \text{years crop repeated} = 1 \\ 0.2, \text{ if crop not repeated} \end{cases}$$

Operator "&" is evaluated as the avarage of the operands. Underlined statements indicate the corespondence between the two rules. Details on these results can be found in [Michalski and Chilausky, 1980]. Examples needed for inductive learning do not neccessarily have to be given by a teacher but may be generated as results of experiments performed by a learning system (e.g., Mitchell, 1983).

A simple problem of descriptive generalization is to construct a classification of a given set of observations. Creating a classification is often the first step in understanding and formulating a theory about a collection of facts. Research on this topic has been done under the headings of cluster analysis and numerical taxonomy. This research relied solely on simple measures of similarity of objects defined over a given set of attributes. Classes were defined as sets of objects whose intra-class similarity is high and inter-class similarity is low. The major disadvantage of these methods is that they are unable to take into consideration background knowledge about the objects and attributes. Nor can they generate new attributes or describe the classes (clusters) that they are creating.

An interesting new method that does not suffer from these disadvantages is based on the so-called "conjuctive conceptual clustering." Given a set of observations, the method structures them into a hierarchy of classes and produces a description of each class. The description is in the form of a conjunctive statment involving relations over selected object attributes. The method and a program CLUSTER implementing it are described by Michalski and Stepp [1983].

The current research in the area of inductive learning explores issues of building complex structured descriptions, automatically generating new attributes, concepts and heuristics, learning from noisy data, applying analogy in learning, and a host of related problems. A review of contemporary research on inductive and other forms of learning from artificial intelligence perspective is in [Michalski, Carbonell, Mitchell, etc., 1983].

References

Davis, R. and Lenat, B. D., "Knowledge-based systems in artificial intelligence," McGraw-Hill, 1982.

Dietterich, T. G., London, R., Clarkson, K., and Dromey, G., "Learning and inductive inference," Chapter XIV in Vol. 3 of The Handbook of Artificial Intelligence, Cohen, P. R., and Feigenbaum, E.A., (eds.). 1982. Also available as Rep. No. STAN-CS-82-913, Department of Computer Science, Stanford University, May, 1982.

Dietterich, T. G. and Michalski, R. S., "Inductive Learning of Structural Descriptions", Artificial Intelligence, Vol. 16, 1981.

Michalski, R.S., Carbonell J.G., Mitchell T.M., MACHINE LEARNING: An Artificial Intelligence Approach, Tioga, Palo Alto, CA, 1983.

Michalski, R. S. and Chilausky, R. L., "Learning by Being Told and Learning from Examples: an Experimental Comparison of the Two Methods of Knowledge Acquisition in the Context of Developing an Expert System for Soybean Disease Diagnosis", Policy Analysis and Information Systems, Vol. 4, No. 2, June 1980, (Special issue on knowledge acquisition and induction).

Michalski, R. S. and Larson, J. B., "Selection of Most Representative Training Examples and Incremental Generation of VL1 Hypotheses: The Underlying Methodology and Description of Programs ESEL and AQ11", Report 867, University of Illinois, 1978.

Michalski, R. S. and Stepp, R.E., "Learning from Observation: Conceptual Clustering",in MACHINE LEARNING, Michalski, R. S., Carbonell, J. G. and Mitchell, T.M. (Eds.), Tioga, Palo Alto, CA, 1983.

Mitchell,T. M., Utgoff, P. E. and Banerji, R. B., "Learning by Experimentation: Acquiring and Refining Problem-Solving Heuristics", in MACHINE LEARNING, Michalski, R. S. Carbonelli, J. G. and Mitchell, T. M. (Eds.), Tioga, Palo Alto, CA, 1983.