A METHODOLOGICAL FRAMEWORK FOR MULTISTRATEGY COOPERATIVE LEARNING

by

Ryszard S. Michalski

Proceedings of the 5th International Symposium on Methodologies for Intelligent Systems - ISMIS'90, Knoxville, TN. pp. 404-411, Oct. 1990.

A METHODOLOGICAL FRAMEWORK FOR MULTISTRATEGY TASK-ADAPTIVE LEARNING

R.S. Michalski
Center for Artificial Intelligence
George Mason University
Fairfax, VA 22030
michalski@aic.gmu.edu

ABSTRACT

This paper outlines basic assumptions and a theoretical basis for *multistrategy task-adaptive learning* (MTL) methodology, which aims at ultimately integrating a spectrum of learning strategies, such as empirical learning, constructive induction, abduction, analytic learning, learning by analogy, and reinforcement learning. In MTL, in response to an input, a learner detemines the strategy, or a combination of strategies, that is most appropriate for the learning task. This detemination is based on the relationship between the input, the learner's background knowledge and the learner's task. By means of a simple example we show how an MTL learner can employ, depending on the above relationship, emprical learning, constructive inductive generalization, abduction, explanation-based learning and abstraction.

1. INTRODUCTION

In view of an immense diversification of research in machine learning, there is a need for developing frameworks that would clarify relationships among different learning paradigms and strategies, and provide a theoretical basis for their integration. In the last few years, several integrated systems have been developed, for example, Unimem (Lebowitz, 1986), Odysseus (Wilkins, Clancey, and Buchanan, 1986), Prodigy (Minton et al., 1987), DISCIPLE-1 (Kodratoff and Tecuci, 1987), GEMINI (Danyluk, 1987), OCCAM (Pazzani, 1988), IOE (Dietterich and Flann, 1988), and ENIGMA (Bergadano et al., 1990). With some exceptions, these systems typically integrate some simple empirical approach with an analytic approach, and do it in a predefined way. An open problem is how to develop a system that would integrate the whole spectrum of learning strategies, and would decide by itself which strategy(ies) is most suitable in a given situation.

This paper gives a brief account of our first efforts toward such a goal, and outlines a *multistrategy task-adaptive learning* (MTL) methodology, that aims at ultimately integrating all major learning strategies, such as empirical learning, constructive induction, explanation-based learning, constructive deduction, learning by analogy, and reinforcement learning. For any given learning situation, the system analizes the relationship between the input, the system's prior knowledge, and the current task, and then determines which strategy, or a combination of strategies, to apply.

Due to the space limitations, many underlying concepts and technical details have been omitted. For those the reader can consult (Michalski, 1989 & 1990, Michalski and Ko, 1990). The presented work is an extension of the ideas presented earlier in (Michalski and Ko, 1988; and Michalski and Watanabe, 1988).

2. UNDERLYING THEORETICAL ASSUMPTIONS

Among the currently two most active methodologies for building symbolic learning programs are empirical learning, which primarily exploits data, and analytical learning, which primarily exploits the learner's prior knowledge. These methodologies are complementary, and their usefulness depends on the specific domain of application. Most of the learning problems in the real world, however, do not seem to fit well either the empirical or the analytic mold. This is because most practical learning problems require an intricate and mutually interdependent interaction between the new facts available to the learner, and the learner's prior knowledge. The latter is rarely complete, directly relevant to the task and/or totally correct, and a learning process may have to simultaneously create new knowledge, debugg prior knowledge, or reformulate it into a better form.

© 1990 by Elsevier Science Publishing Co., Inc.
Methodologies for Intelligent Systems, 5
Z.W. Ras, M. Zemankova, and M.L. Emrich, Editors

The objective of research on MTL is to develop a methodology, which for any learning task can recognize what learning strategy, or combination thereof, is likely to be the most effective for solving it (hence, the term "task-adaptive"). The key idea is that any learning process can be viewed as a derivation of desired knowledge from the input information according to the principle of computational economy. What is "desired" knowledge depends on the task the learner wants to perform. How the learner proceeds to obtain this knowledge (learning strategy) depends on what is the most effective way to utilize the available information and the learner's prior knowledge. Recent experiments in cognitive science show that when people have to reason in order to answer a question, they utilize knowledge that is most easily available to them. For example, when they have a choice of the knowledge source, they typically rely on their personal knowledge rather than on the knowledge supplied to them externally (Michalski, Boehm-Davis and Dontas, 1989).

The MTL methodology postulates that a learner should be able to learn something from any input information, even from the facts that it already "knows." Depending on the interrelationship among the input information, the learner's prior knowledge and the given task, the learner constructs either new knowledge and/or better knowledge (by modifying the prior knowledge). The task may be defined by a specification of the desired knowledge, a performance measure, or may be derived from the general goal(s) of the learner. Another postulate is that whatever was learned in the previous act of learning should be reusable in subsequent learning. This implies that any knowledge learned must be expressed in the form compatible with the form in which any other knowledge is stored. Thus, a segment of knowledge in BK can itself be an input to a learning process. This aspect is called "closed-loop" learning. An interesting consequence of the above assumptions is that to satisfy them, a new form of knowledge representation has to be employed (Michalski, 1989, 1990).

Given some input, the learner first analizes how it relates to the background knowledge (BK) in the context of the task to be performed. If the input represents "novel" information that is relevant to the task, and is not inconsistent with BK, it is assimilated within the BK. This process may involve a generalization of appropriate segments of BK, as well as recording the input ("replication"). If the input is recognized to be already a part of the learner's BK, is implied by it, or implies it, the relevant segments are reinforced or restructured to facilitate their future use ("reformulation" or "tuning"). If the input is inconsistent with BK, but is believed to be correct, the appropriate parts of BK undergo modification. This process may involve specialization, storing exceptions, or weakening certainty parameters associated with relevant knowledge segments ("tuning"). If the input is similar to some component of KB, this component may be used to created desired knowledge. The results of any such a transformation are evaluated form the viewpoint of the learner's task, and stored if they pass the test.

To perform above knowledge transformations different inference types need to be employed. To determine if the input is a special case of what the learner already knows or to reformulate a segment of BK to make it more efficient or directly applicable to a given task, the system performs deductive inference. To synthesize new knowledge that entails the input facts, but is not a logical consequence of them, the system preforms inductive inference. To modify some exisiting knowledge so that it is useful for some new purpose, analogical inference is employed. The results of any inference are tested for their usefulness, and if they pass the test, are assimilated into BK. The next act of learning should be able to take advantage of the modified BK. A process in which results of one act of learning can be used in some subsequent act of learning is called *closed-loop* learning.

In summary, in MTL, learning is viewed as a task-oriented inference process that generates desired knowledge from available premises, and stores it for future use. Not all knowledge generated is stored, but only this that is evaluated as potentially useful for some future tasks. It may appear that viewing learning as an inference process does not apply to non-symbolic learning methods, such as those used in connectionist systems or genetic algorithms, because these systems do not seem to perform inference in the conventional sense. A short answer is that they do apply, because both the latter approaches do peform generalization, specialization or other knowledge transformations, except that they do them implicitly rather than explicitly.

3. THE METHODOLOGY

As indicated above, the basic question in implementing MTL is how to develop a mechanism for recognizing what strategy, or strategies, may be most appropriate for a given combination of the input information, the learner's goal and the background knowledge (BK). Given an input (an example, a fact, a rule, etc.), one can distinguish five types of relationship between the input and the learner's BK. First, the input may represent new or partially new information to the learner, neither confirming nor disconfirming the learner's BK. In the case when it is not economical to make a complete test for this property, one assumes that the input is new. Second, the input may be entailed (or may entail) some segment of the learner's BK. Third, the input may contradict some segment (a rule or a ruleset) of the learner's BK. Fourth, the fact may be similar in certain aspects (in particular, in terms of abstract relations, rather than low level attributes) to some segment of the learner's knowledge. Fifth and finally, the fact may be already known to the learner (i.e., strictly match some knowledge segment).

Empirical learning and constructive induction systems are concerned primarily with handling the first and the third cases. "Pure" explanation-based learning is concerned with the second case. The more recent methods of explanation-based learning attempt to address situations in which the learner's knowledge is incomplete (first case), or inconsistent with the BK (third case), or intractable (first case). Learning by analogy and case-based reasoning are concerned with handling the fourth case. The fifth case is handled by symbolic learning methods usually by ignoring such inputs (with some exceptions, e.g., Slimmer and Granger, 1986). In neural networks and genetic algorithms repeated inputs are not ignored; they are handled, however, as any other inputs. The multistrategy task-adaptive learning methodology is intended to ultimately handle all five cases.

Let us explain how the MTL methodology could handle these cases. It is assumed that the input may be a specific fact, a concept example or a rule supplied by an external source, or information resulting from an impasse in processing of an input according to some strategy. The latter case may require activating another learning strategy. For example, in the process of determining if a fact is implied by BK (i.e., in attempting to explain the fact), the learner may find that some parts of it are explainable by BK, and some other parts represent new information. The parts that are explainable are processed by an analytical learning strategy, and the parts that are new would activate a synthetic learning process. It is also assumed that in the absence of a specific task, the learner uses a general default goal, which is to derive "useful" information from the input, make "sense" of it, and assimilate it into the knowledge base. More specific learning tasks, such as to generalize facts to generate a rule, to create a conceptual classification of given facts, to reformulate a part of BK into a more efficient knowledge, to determine new knowledge on the basis of an analogy between the input and past knowledge, etc. are supplied from a supervisory control system.

Presented ideas are concerned only with aspects of building or updating a knowledge base, and not with issues of using the knowledge for problem solving. Given any input information, the learner analizes its relationship to BK and to the learning task in order to determine which of the five cases above ("processing methods") is involved. The rules and segments in BK are indexed in various ways to facilitate this process. The learner performs a "deductive" matching of the information with BK to determine if it satisfies (or is satisifed by) some rule, or at least is consistent with the rules. Such matching is called "deductive" because it may involve several steps of deduction. It is assumed that a limited amount of resources is available for this process, and if they are exceeded, a failure is communicated. In such a case, the information is assumed to be (pragmatically) new to the system.

1. The input represents pragmatically new information

Generally, this case handles situations that require some form of synthetic learning (empirical learning or constructive induction), or learning by instruction. Given an input, the learner searches for a part of BK that is "hierarchically related" to it. For example, it may be a part describing the concept being exemplified by the input, but neither entailing it, nor contradicting it. If this effort succeeds, the relevant part is generalized, so that it accounts for this input and possibly other information stored previously. The resulting generalizations and the input facts are evaluated for "importance," and those that pass an *importance criterion*, are stored (a process that involves storing representative past facts is called *learning with partial memory of the past*). If there is no knowledge "hierarchically related" to the input, the input is stored, and the control is passed to case 4.

2. The input is implied by, or implies a part of BK

This case represents a situation when it is determined that there is a part of BK that accounts for the input, or is a special case of it. The learner creates a derivational explanatory structure that links the input with the involved BK part. Depending on the learning task, this structure can be used to create a new ("operational") knowledge that is more adequate for future handling of such cases. If the new knowledge passes an "importance criterion," it is stored for future use. This mechanism is related to the ideas on the utility of explanation based-learning (Minton, 1988). If the input represents a "useful" result of a problem solving activity, e.g.," for given state x, it was found that a useful action is y", then storing such a fact as a rule is similar to chunking in SOAR (Laird, Rosenbloom, and Newell, 1986). If the input information (e.g., a rule supplied by a teacher) implies some part of BK, then an "importance criterion" is applied to it. If the input passes this criterion, it is stored, and an appropriate link is made to the part of BK that is implied by it. In general, this case handles situations requiring some form of analytic learning.

3. The input contradicts some part of the learner's BK

The system identifies the part of BK that is contradicted by the input information, and then attempts to specialize this part. If the specialization involves too much restructuring, and/or the confidence in the input is low, no change to this part of BK is made, but the input is stored. When some part of BK has been restructured to accommodate the input, the input is also stored, but only if it passes an "importance criterion." If contradicted knowledge is a specific fact, this is noted, and any knowledge that was generated on the basis of the contradicted fact may have to be revised. In general, this case handles situations requiring a correction of BK through some form of synthetic learning and, generally, managing inconsistency.

4. The input evokes an analogy to a part of BK

This case represents a situation when the input does not match any background fact or rule exactly, nor is "hierarchically" related to any part of BK, but there is a similarity between the fact and some part of BK at a higher abstraction level. That is, unlike in case 1, in which the system tries to directly match the fact with a knowledge segment, in this case, matching is done at a higher level of abstraction, using generalized attributes or relations. If the fact is "sufficiently important" it is stored with an indication of an similarity (analogy) to a background knowledge segment, and with the specification of the aspects (abstract attributes or relations) defining the analogy. For example, an input describing a lamp may evoke an analogy to the part of BK describing the sun, because both lamp and sun match in terms of an abstract attribute "produces light."

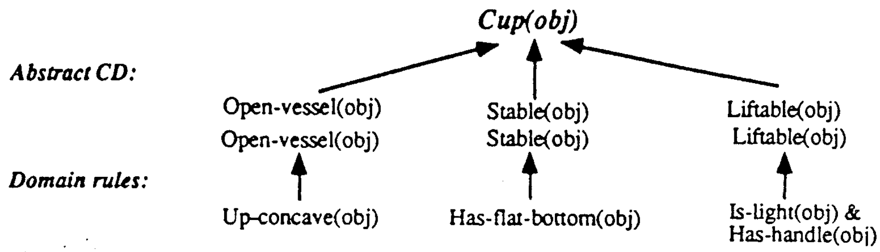
5. The input is already known to the learner

This case occurs when the input matches exactly some part of BK (a stored fact, a rule or a segment). In such a situation, a measure of confidence associated with this part is updated.

Summarizing, in multistrategy task-adaptive learning, any act of receiving information activates some learning strategy. The learner employs deductive inference when an input fact is consistent with, implies, or is implied by the background knowledge; analogical inference when it is similar to some part of past knowledge; and inductive inference when there is a need to hypothesize a new and/or more general knowledge. It also learns when input facts confirm its knowledge, by reinforcing current beliefs.

4. SIMPLE EXAMPLE

To illustrate some of the ideas described above, let us use a well-known example of learning the concept of "cup" (Mitchell, Keller and Kedar-Cabelli, 1986). The example is not very realistic, but illustrates well the ideas (Figure 1).



Example (Specific OD):

Up-concave(OBJ) & Has-flat-bottom(OBJ) & Is-light(OBJ) & Has-handle(OBJ) & Color(OBJ) = red & Owner(OBJ) = RSM & Made-of(OBJ) = glass &...<---> Cup(OBJ)

Abstract OD:

Open-vessel(OBJ) & Stable(OBJ) & Liftable(OBJ) & Has-handle(OBJ) & ... <---> Cup(OBJ)

Operational CD:

Up-concave(obj) & Has-flat-bottom(obj) & Is-light(obj) & Has-handle(obj) <---> Cup(obj)

	Given:		To be learned:
Constructive Deduction (Abstraction)	Example Domain rules	$\triangleright$	Abstract OD
Explanation-based Learning	Abstract CD Domain rules Example	$\triangleright$	Operational CD
Empirical Induction	Examples Partial BK'	$\mid\!<$	Operational CD
Constructive Induction (Generalization)	Domain rules Example(s)	$\mid\!<$	Abstract CD
Constructive Induction (Abduction)	Example(s) Abstract CD	$\mid\!<$	Domain rules
Multistrategy Constructive Learning	Any of the above and other combinations, depending on what is the input, what the learner knows already and what is to be learned		

OD and CD denote object and concept descritpion, respectively. OBJ stands for a specific object; obj denotes a variable. BK' denotes some partial background knowledge, e.g., a specification of the value sets of the attributes and the type of the attributes. Operators $\triangleright$ and $\mid\!<$ denote deduction and induction, respectively.

Figure 1. Learning various aspects of the concept of "cup" using different strategies.

The top part of Figure 1 presents an abstract concept definition (abstract CD) for the concept "cup," the domain rules, a description of one example of a cup (characterized as the "Specific object description" or OD), an abstract object description (abstract OD), and an operational concept description (operational CD). The bottom part of the figure summarizes information that is assumed to be given and to be learned by different learning strategies: constructive deduction (abstraction), explanation-based learning, empirical induction, constructive induction (both, constructive generalization and abduction), and the proposed multistrategy task-adaptive learning. The example does not illustrate the mechanism of updating the strength of the rules, nor learning by analogy. Figure 2 explains the case of constructive learning based on abduction. It is assumed that the learner knows an abstract concept description (Abstract CD), but has an incomplete domain rules (the rule defining the stability is missing).

Learning Strategy: **EX & AC --> DR**

Constructive Induction (Case of abduction)	Example(s) Abstract CD	$\vdash$	Domain rules

---

Given:

1. INPUT ***New example***

Up-concave(CUP1) & Is-light(CUP1) & Has-handle(CUP1) & Owner(CUP1)= Learner & Color(CUP1)=red & Made-of(CUP1)=glass & Has-flat-bottom(CUP1) <----> Cup(CUP1)

2. BACKGROUND KNOWLEDGE

Abstract CD

Cup(obj)

Open-vessel(obj) Stable(obj) Liftable(obj)

Incomplete domain rules

Open-vessel(obj) Stable(obj) Liftable(obj)

Up-concave(obj) ? Is-light(obj) & Has-handle(obj)

Additional knowledge

Stable(obj) <--/--> Owner(obj) & Color(obj) & Material(obj)
Material(obj)=glass <---- Made_of(obj)=glass

3. TASK

To extend the background knowledge so that it entails the example.

To be learned:

A new domain rule: Stable(obj) <---- Has-flat-bottom(obj)

STEPS:

1. Determine the relationship between the input and BK
2. If BK is insufficient to entail the example, use the information in the example and BK to hypothesize an additional rule(s) that together with BK would entail the example.

Figure 2. An illustration of abduction.

The BK includes rules stating that the stability of an object does not depend on who is the object's owner, on its color nor on its material (the latter is a simplication). The input is an example of a cup. In order to consolidate the example of a cup with the current definition of the cup, the learner creates, by abductive reasoning, a hypothesis that if an object has flat bottom then it is stable.

A more realistic example of some asapects of MTL is described in ((Ko, 1989; Ko and Michalski, 1989). The example, based on the implemented program NOMAD, shows how a system can learn a general schema for creating a plan for putting together simple assemblies, for example, a bell. The schema is developed by an incremental improvement and testing of intermediate schemas.

4. CONCLUSION

The proposed MTL methodology stems from the inference-based theory of learning that considers learning as an inference process, whose useful results are stored for future use. Such a process involves input information, the learner's background knowledge, and the task of learning. It may employ any kind of inference - deductive, analogical or inductive.

We have outlined a theoretical framework for unifying basic learning strategies, and discussed several theoretical aspects of implementing an MTL system. This work is motivated by our belief that machine learning systems should, ultimately, be capable, like people, of employing any leanring strategy depending on the task at hand. Among underlying assumptions are that a learning system should be capable of acquiring knowledge from any input, and be able to use the knowledge gained in one learning task in any new learning task, i.e., be capable of a "closed-loop" learning.

The MTL methodology is intended to ultimately integrate capabilities for empirical learning, constructive induction, abduction, analytic learning, reinforcement learning and learning by analogy.The presented ideas are at an early state of development and many issues have not been resolved. Among the most important unresolved issues are: the development of a method for automatically determining the most suitable learning strategy in any given sistuation, the design of a flexible control of the execution of different learning strategies, handling input information whose different components need to be processed separately, but in a globally coordinated way, and the development of an appropriate knowledge representation for supporting the integration of different strategies (for early ideas on the latter subject see *parameterized association rules* in Michalski, 90).

In closing, our goals in developing the MTL methodology are to explore research issues involved in the integration of diverse learning strategies, and to understand how various strategies can best be utilized and how they can support each other in different learning situations. This understanding is important for building powerful and efficient multistrategy learning systems. Such systems are needed for many practical problems in which the process learning needs to involve an intricate interaction between new information, background knowledge and the learner's task. Among examples of such problems are robot navigation, automated assembly, diagnostic decision making, economical prediction, resource management and sensory signal interpretation.

ACKNOWLEDGMENTS

The author expresses his gratitude to Hugo De Garis, Ken De Jong, Kejitan Dontas, Heedong Ko, David Littman, Elizabeth Marchut, Janusz Wnek and Jianping Zhang for comments and suggestions on the earlier draft of this paper. This research was supported in part by the Office of Naval Research under grants No. N00014-88-K-0397 and No. N00014-88-K-0226, and in part by the Defense Advanced Research Projects Agency under the grant administered by the Office of Naval Research No. N00014-K-85-0878.

REFERENCES

Collins, A., and Michalski, R.S., "The Logic of Plausible Reasoning: A Core Theory," *Cognitive Science*, Vol. 13, No. 1, pp.1-49, 1989.

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

DeJong, G. and Mooney, R., "Explanation-Based Learning: An Alternative View," *Machine Learning Journal*, vol 2, 1986.

DeJong, G. F. (Ed.), *Proceedings of the 1988 Symposium on Explanation-Based Learning*, Stanford University, March, 1988.

DeJong, K.A., and Spears, W., "Using Genetic Algorithms to Solve NP-Complete Problems," *Proceedings of the 3rd International Conference on Genetic Algorithms and their Applications*, Fairfax, VA, 1989.

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.

Holland, J.H., "Escaping Brittleness: The Possibilities of General-Purpose Learning Algorithms Applied to Parallel Rule-based Systems," 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.), pp. 499-570, 1986.

Ko, H., "Assembly Planning Through Constructive Learning Approach," to appear in *Machine Learning and Inference Reports*, Center for AI, George Mason University, 1990.

Ko, H. and Michalski, R.S., "Types of Explanation and Their Role in Constructive Learning," to appear in *Machine Learning and Inference Reports*, Center for AI, George Mason University, 1990.

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

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.

R.S. Michalski, D. Boehm-Davis, and K. Dontas, "Plausible Reasoning: An Outline of Theory and Experiments," *Proceedings of the Fourth International Symposium on Methodologies for Intelligent Systems*, October 12-14, 1989, Charlotte, NC, North Holland, 1989.

R.S. Michalski, "Multistrategy Constructive Learning: Toward a Unified Learning Paradigm," an invited paper the *ONR Workshop on Knowledge Acquisition*, Crystal City, VA, November 6, 1989.

R.S. Michalski, "Toward a Unified Theory of Learning: Multistrategy Task-adaptive Learning, *Reports of Machine Learning and Inference Laboratory*, ML-90-1, George Mason University, 1990 (an extended and significantly modified version of the paper above).

Michalski, R.S.and Ko, H., "On the Nature of Explanation or Why Did the Wine Bottle Shatter," *Proceedings of the AAAI Workshop on Explanation-based Learning*, Stanford University, March 1988.

Michalski, R. S. and Watanabe, L., "Constructive Closed-loop Learning: Fundamental Ideas and Examples," *Reports of Machine Learning and Inference Laboratory*, 1988.

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

O'Rorke, P.V., Morris, S., and Schulenburg, D., "Theory Formation by Abduction: Initial Results of a Case Study Based on the Chemical Revolution," *Proceedings of the 6th International Workshop on Machine Learning*, pp. 266-271, Ithaca, NY, 1989.

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

Porter B. W. and Mooney, R.J., *Proceedings of the Seventh International Conference on Machine Learning*, University of Texas, Austiin, June 21-23, 1990.

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