~~Original~~ 89-22

D. R. S. Michalski

PROCEEDINGS OF THE SIXTH INTERNATIONAL WORKSHOP ON MACHINE LEARNING

Cornell University
Ithaca, New York

June 26–27, 1989

EDITOR/WORKSHOP CHAIR

Alberto Maria Segre, *Cornell University*

ORGANIZING COMMITTEE

Jaime Carbonell, *Carnegie Mellon University*
Gerald DeJong, *University of Illinois at Urbana-Champaign*
Tom Dietterich, *Oregon State University*
John Laird, *University of Michigan*
Pat Langley, *University of California at Irvine*
Ryszard Michalski, *George Mason University*
Tom Mitchell, *Carnegie Mellon University*
Jack Mostow, *Rutgers University*
J. Ross Quinlan, *University of Sydney*
Paul Rosenbloom, *University of Southern California*

PROGRAM COMMITTEE

Steve Chien, *University of Illinois at Urbana-Champaign*
Gerald DeJong, *University of Illinois at Urbana-Champaign*
Doug Fisher, *Vanderbilt University*
Allen Ginsberg, *AT&T Bell Laboratories*
John Grefenstette, *Naval Reserach Laboratory*
Michael Pazzani, *University of California at Irvine*
Claude Sammut, *University of New South Wales*
Jeff Schlimmer, *Carnegie Mellon University*
Devika Subramanian, *Cornell University*
Rich Sutton, *GTE Laboratories*
Paul Utgoff, *University of Massachusetts at Amherst*

SPONSORS

Office of Naval Research, Artificial Intelligence and Robotics Program
Office of Naval Research, Cognitive Science Program
National Science Foundation, Knowledge Models and Cognitive Systems Program
Cornell University Department of Computer Science

MORGAN KAUFMANN PUBLISHERS, INC.
SAN MATEO, CALIFORNIA

A DESCRIPTION OF PREFERENCE CRITERION IN CONSTRUCTIVE LEARNING: A Discussion of Basic Issues

Jianping Zhang & Ryszard S. Michalski
Center For Artificial Intelligence
George Mason University
4400 University Drive
Fairfax VA, 22030

1. INTRODUCTION

The criterion for preferring one concept description over the other plays an important role in both empirical and analytical learning systems. Without the preference criterion, these learning systems would lack the means for deciding which of the many alternative candidate descriptions should be chosen. For example, an empirical learning system could simply take the disjunction of all positive examples as the learned description of the concept while an analytical learning system could take the goal concept description as its learned concept description.

In inductive learning, given examples, background knowledge, and optionally, an initial concept description, the system hypothesizes a general concept description. Usually a large number of general descriptions can be generated for any set of examples and/or initial concept descriptions. For an analytical learning system, given a complete domain theory, a goal concept description, and a positive instance, a set of descriptions which explains the given instance are deductively generated from the goal concept description. To choose among the candidate descriptions in both systems one needs a criterion for preferring one description over the other.

The problem of evaluating description is not new to empirical learning systems, and a number of measures of description quality have been developed in the past. Some of them concentrate solely on the aspects of completeness and consistency. Other measures include also additional criteria, such as the simplicity and the cost of evaluating the learned descriptions. Broader aspects of the problem of what should be the preference criterion for judging competing inductive hypotheses are discussed in (Michalski, 1983; Utgoff, 1986; Bergadano et al., 1988).

In early analytical learning systems (e.g., Mitchell et al., 1986), operationality was the only criterion used for selecting the target concept description. Recently some researches (e.g., Segre, 1987) called for considering a trade-off between operationality and generality.

In order to design a learning system that integrates empirical and analytical methods, an unified preference criterion that combines the criteria used in both methods has to be defined first. This paper proposes a preference criterion that can be used in a constructive learning (CL) system (Michalski et al., 1988) which will be discussed in next section. The proposed criterion combines four basic criteria: accuracy, operationality, generality, and simplicity.

2. CONSTRUCTIVE LEARNING

Currently, there are two main approaches to concept learning: syntactic learning and analytical learning (Michalski & Kodratoff, 1989). The simplest form of syntactic learning is empirical learning which performs inductive inference from a training set of concept instances, without use of extensive amount of background knowledge. In this from of learning, the systems usually generates descriptions that use descriptors (attributes, predicates and terms) selected among those present in the descriptions of learning examples (selective induction). In more advanced syntactic methods, the systems perform constructive induction which is able to generate new descriptors that are not present in the input data. Constructive induction is usually guided by domain theory or domain independent heuristics.

The fact that empirical approach does not need much domain knowledge is both an advantage and a disadvantage. On one side, since it primarily relies on the examples given to the system, and examples are often easy to obtain, this approach is very attractive for many applications. On the other side, because it utilizes little domain knowledge, it can hardly be applied to learning in complex, knowledge intensive domains.

Analytical learning approach, in particular, its most popular form -- explanation-based learning (EBL), produces an operational concept description starting from a domain theory, a goal concept description and, a single concept instance. In order to obtain a concept description, complete and consistent domain knowledge is required. When incomplete, inconsistent or uncertain domain knowledge is present, pure analytical learning systems fail to generate any concept description. The requirement of complete and consistent domain knowledge is often too hard to satisfy in the real world.

As many researchers (e.g., Pazzani, 1988) pointed out, both pure empirical and analytical learning methods fail as general theories of learning. They should not compete against each other, instead they should complement to each other in an integrated learning paradigm. Some systems that integrate these two learning strategies have been described in the literature (e.g., Pazzani, 1988). The results generated from these systems have confirmed that a multistrategy methodology is worthwhile to pursuit.

In the following, we will briefly describe the fundamental ideas underlying *constructive learning* and then concentrate on criterion for evaluating the quality of a concept description.

Constructive learning aims at developing concept learning systems that unify and generalize syntactic and analytical learning. In constructive learning, the system, given a learning task, explores first the relationship between its background knowledge, the goal of learning and the task to be performed. Whenever prior knowledge is incomplete, incorrect or not useful for the task, the system executes a form of inductive learning to derive required knowledge. This form of learning may be selective induction, if the prior knowledge is very limited or it can be constructive induction, that utilizes background knowledge to construct new descriptors and concepts in the process of learning (Michalski, 1983). If the system has sufficient knowledge for the task, then it acts as an analytical learning system. The knowledge the system acquires is always assimilated in the knowledge base so that it can be used in subsequent learning. Such a feature is called closed-loop learning. Thus, a constructive learning system integrates selective and constructive induction, analytical learning, and is also capable of closed-loop learning. For more detail, see (Michalski et al., 1988).

3. INDIVIDUAL CRITERIA AND THEIR RELATIONSHIPS

This section first discusses four single criteria: accuracy, operationality, generality, and simplicity, that enter into the unified preference criterion proposed in this paper.

Accuracy represents the description's ability to produce correct classifications. It is measured by the correctness of a concept description with respect to the information and the knowledge available to the learning system. The basic and easy-to-measure criterion that relates to **accuracy** of a conjunction is the ratio of the number of the positive training instances covered by the conjunction and the number of total instance covered by the conjunction (negative and positive). Most of learning systems generate completely accurate descriptions. It is not necessary to generate such completely accurate descriptions for the following reasons:

1 the concept itself may be imprecise and flexible,
2 some examples may have been erroneous,
3 an approximate description is acceptable if an accurate description is hard to find (or expensive to use).

For these reasons, a less accurate description may be better than an accurate one. In fact, in order to achieve completeness and consistency (total accuracy) in the presence of noise, one may have to generate overly complex and detailed descriptions. Such descriptions, however, may not perform well in future cases and examples. This is the well known phenomenon of overfitting

Operationality The goal of explanation-based learning is to obtain an operational concept description from a non-operational goal concept description by analyzing a particular instance of that concept. Generally speaking, an descriptor is operational, if it can be used efficiently by the performance element. We will use a degree of operationality instead of binary operationality used by most of EBL systems.

Generality There are two aspects of **generality** of a concept description in an inductive learning system. The first is related to the coverage of positive instances of concept. The more positive instances a concept description covers, the more general and more desirable the description. The second is related to the coverage of the unseen instances. This aspect of generality affects the predictive power of the description. The more general a description, the more predictive it is and there is more chance to commit errors when using it. Relative generality is easily defined. If the system has a domain theory with logical axioms, then a concept q can be shown to be at least as general as the concept p, if it can be shown that p --> q.

Simplicity is related to the comprehensibility. An important requirement from an AI system is that its knowledge should be explicit and easily understandable by human experts. This is crucial for systems that need to communicate with experts. A black box classifier will not be accepted by experts as a help in their work, even if it performs very well. Knowledge acquired automatically should be easy to understand, contain descriptors used by experts, and not be syntactically too complex.

Generally speaking, these four criteria are separate and distinct, one constitutes different dimensions in our unified criterion. In many situations, in order to satisfy one criterion, one has to sacrifice another. We will now discuss some of the relationships among these criteria.

Segre (1987) discusses the trade-off between operationality and generality in EBL systems. This trade-off plays an important role in our constructive learning system. The goal concept description used in EBL can be generated by an inductive learning system, if operationality of descriptors is not considered. This is because the goal concept description is usually most general, accurate and simplest. Yet, it may be too general to be used by a performance element. On the other hand, an EBL system may generate an operational concept description which may be too specific to be used for the future.

In some applications, there may not exist both very general and operational description. There are three reasons that a too specific concept description may not be efficient, even it is operational. First, it may take a lot of memory to save a complete concept description. Second, it may take time to search a conjunction from the whole concept description which is matched with the current instance to be classified. Finally, a too specific description may have not enough predictive power, and be not applied to some cases in the future. Consider a concept description that is the disjunction of all instances, each conjunction, an instance, is operational. It is obvious that it is not what we want to learn. In the case discussed above, a less operational but more general and simpler description may be used much more efficiently by a performance element.

There is one more reason to prefer a less operational, but more general and simpler concept description. It is not necessary that a concept learning system only learns concept descriptions for the performance element. In a domain which is not completely known, a human user may want to know the principles behind some concepts in the domain and understand the concept completely. In this case, a less operational description preferred. The trade-off between operationality and generality is dependent on the goal of learning a concept.

The relationship between operationality and accuracy is less obvious. In fact, they are two independent criteria. But in some domains, like medical domains, one can gain operationality of a description at the expense of accuracy, or one can gain accuracy by sacrificing operationality. This is because some approximations may have to be made to obtain an operational description. For example, to have an accurate diagnosis in some situations, expensive tests have to be made.

In the context of learning an imprecise and flexible concept, it is generally true that the more general a description is, the less accurate it is. This is because that a more general description may cover more exceptions or rare cases of an imprecise and flexible concept.

A more general description usually is a simpler description. The relationship between simplicity and the other two criteria (accuracy and operationality) is similar to the relationship between generality and these two criteria.

We have discussed all four components in the quality measure and their relationships. These individual criteria need to be combined into a single evaluation procedure that can be used to compare different concept descriptions. Several mechanisms, such as linear weighted function, lexicografic evaluation function (Michalski, 1972), or the combination of these two, can serve for this purpose. The mechanism used to combine these criteria should be flexible enough to allow the user to specify his own preference. And also the system should have the capability to adjust the the description quality criterion to adapt to the learning task at hand.

Acknowledgements

This research was done in the Center for Artificial Intelligence of George Mason University. The activities of the Center are supported in part by Defence Advanced Projects Agency under grant, administered by the Office of Naval Research, No. N00014-87-K-0874, in part by the Office of Naval Research under grant No. N00014-88-K0397, and in part by the Office of Naval Research under grant No. N00014-88-K-0226. The authors thank Tak Wai Chen and Pawal Stefanski for useful criticism.

Reference

Bergadano, F., Matwin, S., Michalski. R. S. and Zhang, J., "Measuring Quality of Concept Descriptions", Proceedings of the third European Working Session on Learning, Pitman, 1988.

Michalski, R. S., "A Variable-Valued Logic System as Applied to Picture Description and Recognition", in "Graphic Languages", Nake, F. and Rosenfield, A. (Eds.), North Holland, 1972.

Michalski, R. S., "A Theory and Methodology of Inductive Learning", Chapter in the book "Machine Learning, an Artificial Intelligence Approach", Michalski, Carbonell, Mitchell (Eds.), Tioga Pub. Co., Palo Alto, CA, 1983.

Michalski, R. S. and Watanabe, L. M., "Constructive Learning: Fundamental Ideas and Examples", Tech Report, Center for Artificial Intelligence, George Mason University, 1988.

Michalski, R. S. and Kodratoff, I., "Introduction" Chapter to Machine Learning: An Artificial Intelligence Approach Vol. III, Morgan Kaufmann 1989 (to appear).

Mitchell T. M., Keller R. M. and Kedar-cabelli S. J., " Explanation-Based Generalization: A unifying view", Machine Learning, 1, 1986.

Pazzani, M. J., "Integrated Learning with Incorrect and Incomplete Theories",Proceedings of the Fifth International Conference on Machine Learning, Ann Arbor, MI: Morgan Kaufmann, 1988

Segre, A. M., On the Operationality/Generality Trade-off in Explanation-Based Learning",Proceedings of Tenth International Joint Conference on Artificial Intelligence, Milano, Italy, 1987.

Utgoff, P. E., "Machine Learning of Inductive Bias", Kluwer Academic Pub., 1986.