Machine Learning, 10, 179–183 (1993)

Book Review

Creating a Memory of Causal Relationships, by Michael Pazzani. Hillsdale, NJ: Lawrence Erlbaum, 1990.

WILLIAM W. COHEN WCOHEN@RESEARCH.ATT.COM
AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974

Editor: Alberto Segre

1. Overview

Creating a Memory of Causal Relationships describes the OCCAM program. OCCAM is an incremental learning system that inputs examples of sequences of causally connected events and constructs a causal model of the processes that govern the events. A variety of learning methods are used in OCCAM, ranging from strong, knowledge-intensive methods like explanation-based learning to weak, knowledge-free learning methods like similarity-based learning. OCCAM has two major goals: to model how humans do learn, and to propose an effective method by which computers can learn causal relationships.

This review attempts to evaluate the research described in *Creating a Memory of Causal Relationships*. This evaluation is primarily from the prospective of AI machine learning, although I have also made an attempt to evaluate OCCAM's contribution as a model of human learning.

2. About the book

2.1. Knowledge representation in OCCAM

The first two chapters of Pazzani's book are a detailed description of the learning problem that OCCAM is intended to solve. These chapters also describe the scheme that OCCAM uses to represent knowledge. In OCCAM, sequences of events (in particular, OCCAM's training examples) are represented as networks of Conceptual Dependency (CD) schemata; OCCAM's memory, which generalizes these sequences of events, is represented as a hierarchy of networks of abstract schemata. Each schema in a network represents an *event*, a *state*, or a *goal*; schemata are linked by binary relations such as *enable* or *result*. The hierarchy of schemata networks is organized with the most general networks at the top; it is used by traversing it, in a top-down manner, and retrieving the most specific schemata network in the hierarchy that is relevant to an example. Instantiating the schema against the example yields certain predictions, which may be either true or false. Notice that organizing learned schemata into a general-to-specific hierarchy, and basing predictions on the most specific

schemata in the hierarchy, means that knowledge has a non-monotonic character: a schema is assumed to hold unless some more specialized schema encoding a class of exceptions also holds.

In addition to prediction, the OCCAM's memory may also be used for *explanation* of an example. Explanations can be constructed either by instantiating a schema or by chaining together rules derived from schemata; however, for reasons of efficiency, the second type of explanation is performed only under certain circumstances. In particular, while rules can be chained together to explain an observed event, they are not chained together to form predictions.

2.2. Learning methods

The next three chapters describe the three learning methods used in OCCAM. Similarity-based learning (SBL) is the weakest method, and requires no background knowledge (beyond that implicit in the choice of representation). The SBL method used in OCCAM consists of clustering similar examples, and then finding the maximally specific conjunctive generalization of each of the clusters. Theory driven learning (TDL) uses *generalization rules*, which are a sort of rule schema, to generalize a cluster of examples. Generalization rules are used to encode fairly general types of causal knowledge, such as the knowledge that a cause must precede its effect. TDL thus incorporates a stronger bias than SBL, as it will only output a hypothesis that is consistent with this pre-existing causal knowledge. Finally, explanation-based learning (EBL) is also used in OCCAM. Pre-programmed (or previously-learned) rules can be chained together to explain an observed sequence of events; the resulting explanation is then generalized and cached in OCCAM's memory.

2.3. Integration

The final section of the book describes how these three learning systems are integrated in OCCAM, and also gives some extended examples of how they interact on various learning problems. Whenever OCCAM is given a new example, it first searches its memory for the most specific relevant schema. If the prediction made by that schema is correct, then no change is made to OCCAM's memory.

If OCCAM is given a new example that is not consistent with its existing memory, it first attempts to use EBL. If the differences between the prediction made by the retrieved schema and the example can be explained analytically, then EBL is invoked, and the generalization produced by EBL is added to the memory as a specialization of the existing schema. The effect of this is that the original schema is retained, but qualified by a description of a class of possible exceptions.

If the differences between the example and the retrieved schema cannot be explained, then one of three things can happen. If there are few other exceptions to the schema, the example is simply stored as an exception. If there are many exceptions of a schema that has not often been used successfully, then the retrieved schema is simply deleted from memory, and relearned using the new and larger set of examples. Finally, if there are many exceptions and the schema has been used successfully many times before, then an attempt

is made to generalize the exceptions using the other learning methods. First, TDL is invoked. If TDL fails—that is, if no hypothesis can be formed that is consistent with the causal knowledge encoded by the generalization rules used by TDL—then SBL is invoked.

The OCCAM architecture allows for many interesting interactions between the three learning techniques, some of which are explored in the last section of Pazzani's book. For example, TDL can be used to learn a domain theory for EBL. If learning is incomplete, then when EBL is applied, the domain theory may still be buggy; in this case the generalization produced by EBL will be incorrect. However, if later examples show the schema to be inaccurate, OCCAM will eventually remove the incorrect schema under the procedure described above, and re-learn it using a newer and more correct version of the domain theory.

2.4. Presentation

Throughout, Pazzani's book is clearly written. There is a large amount of motivational material, and each chapter includes several detailed examples. This is by far the best source for learning about OCCAM; although much of the technical material appears elsewhere, the results are scattered across several short papers, and are not presented in as much depth as in the book.

One can also obtain, for a small additional charge, Common LISP source code for OCCAM-LITE, a micro-version of OCCAM which implements most of the main ideas; I also tested this code (on a Macintosh II under Allegra Common LISP) and found it to work as advertised. Of course, having a working and easily-understandable implementation of OCCAM available would be very valuable for a researcher wishing to build on, or more deeply understand, this work.

3. Methodology

Pazzani's primary goal in developing OCCAM is to develop a model of human learning; in particular, a model of how humans learn causal relationships. This is an important question because (by Pazzani's hypothesis) humans use some prior knowledge in learning causal relationships: understanding how humans use such prior knowledge may thus help to illuminate the question of how prior knowledge can affect learning, and thus illuminate the process of "learning to learn."

OCCAM's psychological plausibility is bolstered by several experiments. In one experiment, humans are shown to learn faster (i.e., using fewer examples) in domains in which prior knowledge suggests an answer than in domains in which no prior knowledge exists; in a parallel experiment, OCCAM's EBL method is shown to learn faster than SBL. In a second experiment, humans are shown to learn faster in domains in which the expected answer is consistent with general knowledge of causality; in a parallel experiment, OCCAM's TDL is shown to learn faster than SBL. Pazzani also notes that OCCAM's SBL strategy is similar to that used by most people, as determined by previous psychological experiments.

In a final set of experiments, Pazzani shows that the OCCAM strategy of preferring EBL to SBL is computationally advantageous (that is, better than using either EBL or SBL alone)

in most settings. This is not a psychological argument per se, but it could be argued that evolution favors more computationally effective learners, and hence that a simple but computationally effective algorithm is, at the very least, more plausible than an ineffective one. These results also support Pazzani's claim that OCCAM (or at least, OCCAM's method for integrating of SBL and EBL) is an effective learning algorithm.

It should be noted, however, that these experiments show only that OCCAM has the right rough qualitative behavior; no evidence is given to show (for example) that OCCAM's learning rates are quantitatively similar to human learning rates. Also, OCCAM makes several additional detailed predictions about learning rates that are not experimentally tested. For example, OCCAM's TDL algorithm makes use of several types of generalization rules: first "exceptionless" rules, then "dispositional" rules, and finally "historical" rules. Thus OCCAM predicts that the sorts of causal relationships expressible by exceptionless generalization rules are more easily learned than causal relationships expressible by dispositional generalization rules, and so on: this prediction, however, is not tested. Pazzani also discusses cases (such as "illusory correlations" and the differing impacts of earlier versus later training examples) in which human learning is clearly suboptimal, and suggests that OCCAM would display the same behavior in these cases. However, this is not demonstrated experimentally. Finally, there is no rigorous investigation of either the computational effect of learning "dispositions," or of their psychological validity.[1]

A final criticism of the experimentation is that all of the experiments with OCCAM are performed on hand-constructed learning problems and/or using hand-constructed data. For example, the TDL component was evaluated by learning what class of people can open a refrigerator door, and five of the examples in the economic sanctions problem used to test the EBL component were hypothetical cases constructed by hand. While testing on natural problems is not always essential—for example, I found the study of OCCAM's integration of SBL and EBL to be quite informative—I would argue that testing a learning algorithm only on a small number of hand-crafted problems is never adequate. Hand-crafted problems might not adequately test the weaker components of the learning system; in OCCAM's case, for example, the TDL and SBL algorithms are dependent on a clustering procedure that seems rather ad hoc. More rigorous experimental validation of these learning algorithms would be desirable; such validation would certainly strengthen OCCAM's claim to be an effective machine learning algorithm.

The above remarks should perhaps be tempered with the observation that since OCCAM is the first attempt to model this aspect of learning, it may be unreasonable to expect more than a qualitative match with psychological data. Also, as a machine learning method, OCCAM attacks a recently-identified problem in a novel way; again, one cannot expect the sort of strong experimental results associated with more mature research efforts.

4. Research issues raised

Like most research projects, OCCAM raises, as well as answers, many questions. One topic for further research discussed by Pazzani is the problem of learning the generalization rules required by TDL. These generalization rules, which encode general knowledge of causality, are currently hand-coded into OCCAM; however, psychological evidence suggests

that this knowledge is not innate. This flaw in OCCAM (as a model of human learning) could be addressed by incorporating the ability to learn TDL generalization rules.

A second question raised by OCCAM pertains to the role and meaning of the domain theories used by EBL. It is perfectly possible for OCCAM to be in a semantically inconsistent state, in the sense that some schemata contradict the logical consequences of other schemata. (In fact, when OCCAM discovers that a schema makes an incorrect prediction p, the first thing it tries to do is *prove*, by chaining together other rules, that the prediction is incorrect, i.e., that $\neg p$ holds). An interesting question is: why is it that this logical inconsistency is, in practice, *not* problematic?

The answer to this question seems to be related to the fact that OCCAM's reasoning mechanism is incomplete, in the sense that it is not possible to derive all of the logical consequences of the schemata stored in its memory. The only time OCCAM will predict some result p is if the most specific schema that is explicitly stored in memory predicts p; this in turn can only occur if, in previous examples, p was actually a result. Thus, as OCCAM learns using EBL, and as the schema memory develops into a progressively larger subset of the deductive closure of the existing theory, inconsistencies in its knowledge are gradually resolved in favor of those cases that actually occur. This is a fairly subtle aspect of OCCAM's behavior, and one that is not discussed at length in Pazzani's book.

OCCAM includes several other results that I found interesting, although Pazzani did not discuss them at length. For example, OCCAM's SBL and TDL methods, since they operate on CD schemata, allow one to learn a constrained type of relational concept. This makes the integration of TDL and EBL far more powerful, since the rules used by EBL are not restricted to be propositional. Finally, OCCAM's integration of TDL and EBL allows EBL to be applied with an initially buggy domain theory.

5. Summary

Creating a Memory of Causal Relationships is an investigation of the specific problem of learning causal relationships. More generally, the book is an investigation of the problem of learning using prior background knowledge, and as such would be of interest to anyone interested in this problem.

Primarily, the book is a clear and detailed description of OCCAM, a program for learning causal relationships. OCCAM incorporates many interesting and novel ideas. The ideas that are emphasized by Pazzani, and that are explored in depth in the book, are a novel integration of explanation-based and similarity-based learning methods; a third learning technique, theory-based learning or TDL, which is in some ways an intermediary between EBL and SBL; and a characterization of (at least some of) the background knowledge that is needed to learn causal relationships.

Notes

1. "Dispositions" are feature-preferences that are learned as means of tuning the TDL component of OCCAM.