Detecting and Correcting Errors of Omission After Explanation-based Learning

Michael J. Pazzani
Department of Information and Computer Science
University of California,
Irvine, CA 92717
pazzani@ics.uci.edu

Abstract

In this paper, we address an issue that arises when the background knowledge used by explanation-based learning is incorrect. In particular, we consider the problems that can be caused by a domain theory that may be overly specific. Under this condition, generalizations formed by explanation-based learning will make errors of omission when they are relied upon to make predictions or explanations. We describe a technique for detecting errors of omission, assigning blame for the error of omission to an inference rule in the domain theory, and revising the domain theory to accommodate new examples.

1 Introduction

An important issue arises when the background knowledge for explanation-based learning is acquired via empirical methods: inaccuracies in the background knowledge create inaccuracies in the "compiled" theory. In this paper, we describe a technique for detecting errors in the compiled theory and revising the background knowledge. We consider the case in which the domain theory may be overly specific. Such domain theories are formed by one-sided inductive learning algorithms [1] [2]. In a system such as OCCAM [3], which integrates empirical and explanation-based learning, two kinds of examples can be distinguished:

- *Foundational.* Examples from which background domain knowledge can be learned. For example, OCCAM is presented with examples of parents helping their children and examples of strangers not assisting children. From these foundational examples, it acquires a schema that indicates that parents have a goal of preserving the health of their children. This schema explains why a parent pays a ransom to a kidnaper.
- *Performance.* These training examples are instances of the performance task. In OCCAM, one performance task is to infer the goals of the participants in kidnaping episodes. For this performance task, examples of actual kidnaping incidents are used as performance examples. Because OCCAM has learned the relevant background knowledge, it can use explanation-based learning on these performance examples.

Note that the classification of an example as a performance or foundational example is relative to a particular performance task. For example, if the performance task were to predict who might be a potential candidate for purchasing ransom insurance, kidnaping incidents would serve as foundational examples. The foundational examples differ from the performance examples in that they can be viewed as a subproblem of the performance task.

In previous work [4], we have described a technique for detecting errors in the background theory when foundational examples are misclassified and we have shown how these changes can be propagated to the compiled theory. In this paper, we demonstrate a procedure which can detect errors in the compiled theory when performance examples are misclassified, and assign blame for the error to an inference rule in the background theory. The technique relies upon recording the dependency between a feature present in a generalization formed by explanation-based learning and the inference rule in the domain theory that is responsible for including the feature in the generalization.

An error of omission in the compiled theory occurs when an performance example is not correctly classified or explained. For example, a kidnaping schema might indicate that a ransom will be demanded of a wealthy parent of the hostage. This schema will not be able to explain a case in which the ransom note is sent to a wealthy grandparent. From an example such as this, it is desirable that a learning system be able to notice that its rule which indicates that parents have a goal of preserving the health of their children is

This research is supported in part by a Faculty Research Grant from the University of California, Irvine.

overly specific and to generalize this rule to include grandparents. Such revisions of the domain theory distinguishes this work from previous work [5] which corrects only the compiled theory. We investigate errors of omission rather than errors of commission for two reasons:

- Errors of omission are the only kind of error which occurs in one-sided algorithms that incrementally generalize their hypothesis as little as possible to account for new positive examples. This class of learning algorithms has been the subject of much recent theoretical analysis (e.g. [1] [2]) and has been proposed as a model which can account for some varieties of human learning [6] [7].

- Blame assignment for errors of omission in compiled theories can be done incrementally as new examples are encountered. Two recent techniques which assign blame to inference rules in the background theory [8] [9] non-incrementally process a collection of examples when errors of commission are detected. When there is an error of commission, at least one rule must be modified by including an additional condition to prevent. However, any rule which played a part in the explanation is a potential candidate for modification. Correlational techniques are used to assign blame to a particular rule and to construct a description of the condition which must be added. In contrast, with errors of omission, a condition must be dropped from at least one rule that was used to chain together an explanation. This condition can be found by recording a dependency between rules in the domain theory and features in the compiled theory.

2 Errors of omission in OCCAM

OCCAM maintains a hierarchy of schemata that can be used to make predictions or give explanations. A schema which applies to a new experience is found by a memory search that starts at a very general node in memory (e.g., coercion). By following indices (e.g., threat = refuse to sell) from the general schemata, more specific schemata that provide additional inferences and expectations (e.g. economic-sanctions) are retrieved. In this extension to OCCAM, there are two types of indices:

- *Predictive.* A predictive index is traversed to find a schema that describes the result of a given event.
- *Explanatory.* An explanatory indexed is traversed to find a schema that describes the cause (i.e., the explanation) of a given event.

These indices are similar in spirit to the predictive and predictable features used in UNIMEM [10]. However, there is one important difference. In UNIMEM, this information is determined empirically: the predictive features are those characteristic features of a generalization that are unique (or nearly unique) to that generalization. The predictable features appear in many generalizations. The idea is that the predictive features are likely to be the cause of the predictable features [11]. In OCCAM, this information is derived analytically. The predictive features are those that appear in the antecedent of an inference rule which is used to create a generalization with explanation-based learning. The explanatory features appear in the consequent of an inference rule. Figure 1 illustrates one inference rule which is used in part of an explanation to explain why the US grain embargo with the Soviet Union failed.

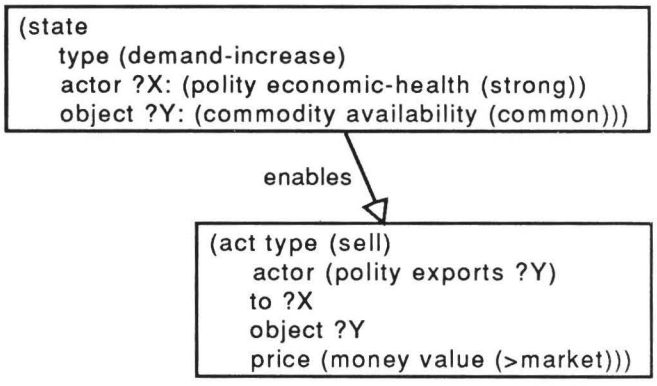

Figure 1. An inference rule that indicates that an increased demand for a commonly available commodity by a country with a strong economy enables a country which exports the commodity to sell the commodity at a greater than market rate.

When a generalization of the US grain embargo is constructed by explanation-based learning, it is indexed by predictive features (e.g., `economic-health of the target country = strong, availability of the commodity = common`). The schemata formed from generalizing the US grain embargo can be retrieved via predictive indices to make predictions about future or hypothetical cases. For example, OCCAM answers the question "What would happen if the US refused to sell computers to South Korea, if South Korea continues to export automobiles to Canada?", by retrieving the generalization via predictive indices. The generalization will also be indexed by explanatory features (e.g., `price of the commodity = >market`). Explanatory indices could be followed to find an answer to a question such as "What could cause the price of oil to rise?". Similar distinctions have been made by Pearl [12] who distinguishes causal and evidential support and by Tversky and Kahneman [13] who find that people treat causal and diagnostic evidence differently.

2.1 Detecting Errors of Omission

In OCCAM, when a new event is encountered, predictive indices are followed to determine if there is a schema which can predict the observed result of the new event. If the observed result is predicted, then there is nothing that can be learned from the new event. If there is no schema which predicts the result, OCCAM attempts to create one via explanation-based learning.[1] If this fails, because the domain theory cannot explain the result of the new event, then OCCAM checks for an error of omission. Schemata are retrieved by following explanatory indices. If a single schema is retrieved which could explain the event, then this schema may be making an error of omission. That is, the conditions which predict the outcome of the schema may be overly-specific. Overly-specific predictive indices would prevent the schema from being retrieved by predictive indices, but not by explanatory indices. The general idea is that the current domain theory does not predict the outcome of the new event. However, the current domain theory predicts some conditions (which differ from the current event) under which that outcome could occur. It may be the case that a slight modification of these conditions could account for the new event. The search for a modification focuses on the differences between the new event and the schema retrieved by explanatory indices.

2.2 Blame assignment

After an error of omission has been detected and a schema which could be modified is retrieved, an attempt is made to account for the differences between the new event and the existing schema. The following process isolates one rule in the domain theory which can be modified to account for the new event:

1. Rederive the justification for the current schema by reexplaining how the generalized event of the schema results in the outcome of the schema. This explanation process differs from the standard explanation-based learning algorithms [14] [15] in one crucial aspect: a dependency is recorded between each feature which is included in the generalized event and the inference rule(s) in the domain which are responsible for including the feature in the generalized event. There is a trade-off between storing this dependency with every schema that is created and rederiving this dependency when a possible error of omissions has been detected. We rederive this information, because the psychological evidence does not indicate that people are aware of the assumptions behind their beliefs [16].

2. The features which differ between the retrieved schema and the new event are found.

3. The inference rules which are responsible for the features which differ are collected.

4. If exactly one inference rule is responsible for all the differences (e.g., if there was only one difference), then it assumed that this inference rule is making an error of omission. Recall that OCCAM learns by observation. If it had a teacher that could indicate the correct explanation, then it would be possible to correct all the inference rules responsible for every difference. Without a teacher, it is necessary to find an explanation that is "close" to explaining the new event. Here, "close" means that there is only one schema that is retrieved via explanatory indices and there is only one inference rule responsible for all the differences between the retrieved schema and the new event. In effect, when an error of omission in a compiled theory occurs, this step determines that an inference rule in background theory is committing an error of omission.

5. A new training example for the inference rule in the background theory is derived from the new event. The general idea here is that the new event is a performance example. A foundational example is constructed from the performance example so that the background theory can be revised. This is accomplished by exploiting variable constraints between the inference rule used in the explanation and the generalization produced by explanation-based learning.

2.2.1 Blame assignment: an example

In this section, we provide an example of the way that blame assignment works. The antecedent of a rule which indicates that refusing to sell a commodity results in an increased demand for the commodity was modified so that the countries which sell and buy the product must have free economies (see Figure 2).

OCCAM is presented with the following example and a new schema is created with explanation-based learning:

Australia and France, 1983

In 1983, Australia refused to sell uranium to France, unless France ceased nuclear testing in the South Pacific. France paid a higher price to buy uranium from South Africa and continued nuclear testing.

1. Explanation-based learning in OCCAM is similar to other EBL algorithms (e.g., [14] [15]). Rules in OCCAM are stored as schemata in memory. Predictive indices can be followed to retrieve a rule whose antecedent matches an explanatory goal. A complex event is explained by chaining together several schemata. Explanation-based learning creates a new schema so that in the future, an explanation can be found by memory search rather than chaining.

```
(ACT TYPE (SELL)
     ACTOR (POLITY EXPORTS ?Y
                   ECONOMY (FREE))
     TO ?X:(POLITY IMPORTS ?Y
                   ECONOMY (FREE))
     OBJECT ?Y:(COMMODITY)
     MODE (NEG))
```

result

```
(STATE TYPE (DEMAND-INCREASE)
       ACTOR ?X
       OBJECT ?Y)
```

Figure 2. An overly specific inference rule that indicates that a country with a free economy refusing to sell an exported commodity to a country with a free economy that imports the commodity will result in an increased demand for the commodity.

The outcome of this event (France possessing the uranium) is explained by chaining together three inference rules. Refusing to sell the uranium results in an increased demand for the uranium (see Figure 2), an increased demand for the uranium enables a country to sell the uranium at a higher price (see Figure 1), and selling the uranium results in France possessing the uranium. The generalization indicates that three features of the target (i.e., France) are essential:

- The target imports the commodity. (The rule in Figure 2 is responsible for this constraint.)
- The target has a strong economy. (Figure 1)
- The target has a free economy. (Figure 2)

By keeping track of the source of each feature included in a generalization, OCCAM can assign blame when an error of omission occurs. Consider what happens when OCCAM is presented with the example of the US grain embargo with the Soviet Union:

US and USSR, 1980

In 1980, the US refused to sell grain to the Soviet Union if the Soviet Union did not withdraw its troops from Afghanistan. The Soviet Union paid a higher price to buy grain from Argentina and did not withdraw from Afghanistan.

There is no schema in memory which could predict the observed outcome of this incident and an explanation cannot be created for the outcome by chaining together political or economic rules from the domain theory (since the Soviet Union does not have a free economy). However, the schema created from the sanction incident involving Australia and France can be retrieved by explanatory indices. The following trace shows OCCAM detecting and assigning blame for this error of omission:

```
Cannot explain SANCTION-80-1.
Trying retrieval by explanatory indices.
Found schema:
(COERCE
   ACTOR (POLITY EXPORTS =OBJECT
                 ECONOMY (FREE))
   TARGET (POLITY ECONOMIC-HEALTH (STRONG)
                  ECONOMY (FREE)
                  IMPORTS =OBJECT)
        ...
        RESPONSE (ACT
                  TYPE (SELL)
                  ACTOR (POLITY
                          BUSINESS-REL =TARGET
                         EXPORTS =OBJECT)
                  OBJECT =OBJECT
                  PRICE (MONEY VALUE (>MARKET))
                  TO =TARGET)
        OUTCOME (GOAL-OUTCOME TYPE (FAILURE)
                              ACTOR =ACTOR))
Difference between SANCTION-80-1 and schema:
(COERCE TARGET (POLITY ECONOMY (FREE)))
Responsible rule is:
 <<CD-65 ACT> RESULT <CD-66 STATE>>
New example:
(ACT TYPE (SELL)
     ACTOR (POLITY TYPE (COUNTRY)
                   NAME (US)
                   ECONOMY (FREE) ...)
     TO (POLITY TYPE (COUNTRY)
                NAME (USSR)
                ECONOMY (CONTROLLED)...)
     OBJECT (COMMODITY AVAILABILITY (COMMON)
                       TYPE (GRAIN))
     MODE (NEG))
RESULT
(STATE TYPE (DEMAND-INCREASE)
       ACTOR (POLITY NAME (USSR) ...)
       OBJECT (COMMODITY TYPE (GRAIN) ...)
```

In this example, OCCAM retrieves a schema by following explanatory indices, determines that the only difference is the type of economy of the target and finds the rule which indicates that the target must have a free economy. Next, it creates a foundational example from the performance example so that the rule may be revised. In effect, what has occurred is that a performance example of the Soviet Union thwarting a sanction attempt by bidding the price of the grain up is converted to a foundational example of the US refusing to sell grain to the Soviets resulting in an increased demand for the grain by the Soviets. An inference rule in the domain theory can be revised to account for the derived foundational example.

2.3 Correcting the domain theory.

Once a foundational training example has been constructed, the inference rule in the domain theory which is responsible for error of omission can be revised. The central problem addressed in this paper is the assignment of blame for an error of omission. A number of techniques could be used to revise the domain theory. If the learning system is acquiring deterministic causal rules and there is no noise in the data, then an inference rule can be revised as soon as one exception is seen. Although, we acknowledge that this situation rarely occurs in human learning or in real-world domains, we make this assumptions in the experimental study described in Section 3. This is, in part, necessitated by the small number of economic sanctions incidents that have occurred. An error of omission is caused by an inference rule being too specific. This is corrected by generalizing the antecedent inference rule as little as possible to accommodate the training example. Those features that are present in the inference rule but are not present in the new example are dropped. The following trace continues the example from Section 2.2.1:

```
Generalizing antecedent to accommodate example
Old:
(ACT TYPE (SELL)
     ACTOR (POLITY EXPORTS ?Y
                   ECONOMY (FREE))
     TO ?X:(POLITY IMPORTS ?Y
                   ECONOMY (FREE))
     OBJECT ?Y:(COMMODITY)
     MODE (NEG))
New:
(ACT  TYPE (SELL)
      ACTOR (POLITY EXPORTS ?Y
                    ECONOMY (FREE))
      TO ?X:(POLITY IMPORTS ?Y)
      OBJECT ?Y:(COMMODITY)
      MODE (NEG))
```

The feature that was removed by the new example indicated that the target of the coercion must have a free market economy. Once the domain theory has been revised, the schema formed via EBL with the old theory is deleted, and a new, more accurate schema is created. In the current example, this schema still contains an erroneous feature that indicates that the actor of the coercion (i.e., the country that refuses to sell the product) have a free market economy, since this was part of the initial domain theory and no counter example has been seen.

3 Experimental results on OCCAM

To understand how much of an improvement can be gained by the algorithms described in this paper, we intentionally introduced errors into OCCAM's background knowledge. Every inference rule which indicates the result of threat (e.g., refusing to sell a product, or cutting of economic aid) was modified to also include the features which are present in the other inference rules in this class. In effect, this simulates an approach in which the potentially relevant features are known *a priori,* but the contexts in which the features are relevant is not known (cf. [17]). Figure 3 illustrates the accuracy of OCCAM, OCCAM with the erroneous knowledge base, and OCCAM using the algorithms for detecting and correcting errors omission on the erroneous knowledge base. We measured the accuracy of predicting the outcome on unseen cases after 10 and 15 training examples. The values are averaged over 10 runs. In each run, a collection of cases was selected randomly from a database of 15 cases. The remaining cases and five hypothetical cases analyzed by a political analyst at the Rand Corporation were used as test examples. While this may seem like a small number of training examples, there have not been many major economic sanctions incidents. It is exactly in this type of situation that knowledge-intensive strategies are best suited.

The figure clearly shows that the algorithm described in this paper increases the accuracy of predicting the outcome of new cases. With 10 training examples, correction improves the performance significantly ($p<.005$, $t(18)=3.16$). The difference after 15 examples is significant at the .001 level. The correction strategy does not achieve the same performance as OCCAM with a correct domain theory because in some cases, there are no precedents which contain the necessary features to force the revisions of the incorrect domain theory.

4 Conclusion

We have described a technique for detecting errors of omission in compiled theories, assigning blame for the error of omission on a rule (or rules) in the domain theory and revising the domain theory to account for the new example. Although we have explored this technique in the context of explanation-based learning, it appears to be applicable to case-based reasoning systems which make use of domain knowledge [18] as well as rule-based expert systems [19]. One advantage of this technique that we have not yet demonstrated is that a more accurate domain theory can increase the accuracy of explanation-based learning of concepts other than the one in which the error was detected. This will occur when two compiled concepts share some foundational background knowledge.

Acknowledgements

I'd like to thank Wendy Sarrett and Kamal Ali for comments on an earlier draft of this paper.

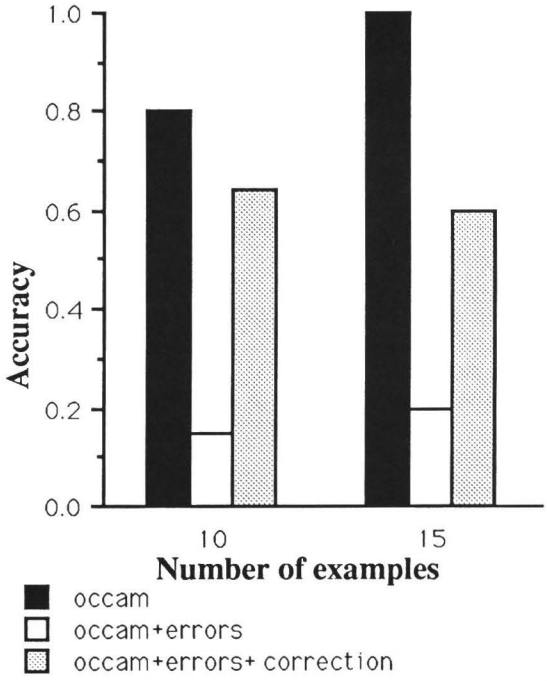

Figure 3. Accuracy of the learning algorithms at predicting the outcome of unseen sanction incidents.

References

[1] Valiant, L. (1984). A theory of the learnable. *Communications of the Association of Computing Machinery*, 27(11), 1134-1142.

[2] Haussler, D. (1987). *Bias, Version Spaces and Valiant's Learning Framework.* Proceedings of the Fourth International Workshop on Machine Learning, Irvine CA.

[3] Pazzani, M. J. (1989). *Learning causal relationships by integrating empirical and explanation-based learning methods.* Hillsdale, NJ.: Lawrence Erlbaum Associates. (in press)

[4] Pazzani, M. (1988). *Integrated Learning with Incorrect and Incomplete Theories.* Proceedings of the Fifth International Machine Learning Conference. Ann Arbor, MI.

[5] Laird, J. (1988). *Recovering from incorrect knowledge in SOAR.* Proceedings of the Seventh National Conference on Artificial Intelligence. St. Paul, Minn.

[6] Berwick, R. (1986). *Learning from positive only examples: The subset principle and three case studies.* In Michalski, R., Carbonell, J., & Mitchell, T. (Ed.), Machine Learning, Vol 2. Los Altos, Ca.: Morgan Kaufmann Publishers.

[7] Bruner, J.S., Goodnow, J.J., & Austin, G.A. (1956). *A Study of Thinking.* New York: Wiley

[8] Ginsberg, A. (1988). *Theory revision via Prior Operantionalization.* Proceedings of the Seventh National Conference on Artificial Intelligence. St. Paul, Minn, 590-595

[9] Bergadano, F & Giordana, A. (1988). *A knowledge intensive approach to concept induction.* Proceedings of the Fifth International Machine Learning Conference. Ann Arbor, MI. 305-317.

[10] Lebowitz, M. (1986). *Concept learning in an rich input domain: Generalization-based memory.* In Michalski, R., Carbonell, J., & Mitchell, T. (Ed.), Machine Learning, Vol 2. Los Altos, Ca.: Morgan Kaufmann Publishers.

[11] Lebowitz, M. (1986). *Not the path to perdition: The utility of similarity-based learning.* Proceedings of the National Conference on Artificial Intelligence. Philadelphia, PA

[12] Pearl, J. (1987) *Embracing causality in formal reasoning.* Proceedings of the Sixth National Conference on Artificial Intelligence, Seattle.

[13] Tversky, A. & Kahneman, D. (1980) *Causal schemas in judgments under uncertainty.* In M. Fishbein (Ed.), Progress in social Psychology. Hillsdale NJ: Erlbaum.

[14] DeJong, G. and Mooney, R. (1986). *Explanation-based learning: An alternate view.* Machine Learning, Vol. 1(2).

[15] Mitchell, T., Kedar-Cabelli, S. & Keller, R. (1986). *Explanation-based learning: A unifying view.* Machine Learning, Vol. 1(1).

[16] Nisbett, R. & Ross, L. (1978). *Human Inference: Strategies and Shortcomings of Social Judgments.* Engelwood Cliffs, NJ: Prentiss-Hall, Inc.

[17] Kolodner, J. (1984). *Retrieval and organizational strategies in conceptual memory: A computer model.* Hillsdale, NJ.: Lawrence Erlbaum Associates.

[18] Barletta, R., & Mann, W. (1988). *Explanation-based Indexing of Cases.* Proceedings of the Case-Based Reasoning Workshop, Morgan-Kauffman.

[19] Ginsberg, A., Weiss, S. & Politakis (1985) *SEEK2: A generalized approach to automatic knowledge base refinement.* Proceedings of the Ninth International Joint Conference on Artificial Intelligence, Los Angeles.