Selecting the Best Explanation for Explanation-Based Learning

Michael Pazzani
UCLA Artificial Intelligence Laboratory
3531 Boelter Hall
Los Angeles, CA 90024
pazzani@cs.ucla.edu

Introduction

In this paper, I consider the problem of selecting the best explanation for explanation-based learning [3, 11]. There are several different ways of defining a "best" explanation. For example, in PRODIGY [10], the best explanation is one selected from an implicit set of mutually consistent explanations that will result in a generalization providing the largest performance increase (e.g., by reducing matching costs). In this research, the best explanation refers to the explanation selected from a set of inconsistent explanations which is most plausible given the observed data and existing knowledge. I present several heuristics to rate the plausibility of explanatory hypotheses.

To some, the concept of the multiple inconsistent explanations may seems strange. Many people actively researching explanation-based learning have considered an explanation to be a deductive proof (e.g., [5, 8]). In this case, inconsistent explanations can result from an inconsistent domain theory (or a theory with defaults) [11]. However, explanation cannot properly be viewed as a deductive process [2, 6, 9]. Instead, explanation should be viewed as an abductive process which generates hypotheses to account for facts. Explanation cannot be deduction, because in the typical case, information is missing which would allow a deductive proof to be completed [14]. When a domain theory is incomplete, missing information must be assumed, and multiple inconsistent explanations arise because it is possible to make multiple inconsistent assumptions.

In this paper, I consider one form of abduction: *plan recognition.* Plan recognition is the process of constructing an explanation which relates actions and intermediate goals which result in the achievement of a goal. There have been numerous approaches to plan recognition. However, most existing approaches either return a set of possible explanations and do not address the issue of selecting the best explanation from the set (e.g., [7]), cannot deal with missing information (e.g., [17]) or commit to one explanation without considering alteratives (e.g., [19]). A noteworthy exception is [18] which provides a general framework of reasoning about alternative hypotheses. In this paper, I suggest specific heuristics for favoring one hypothesis over another.

There are three sources of ambiguity in plan recognition:

1. Refinement- A single action may be a specific means of accomplishing more than one goal. For example, a parent licking a child's ice cream cone could be viewed as the parent eating the ice cream or the parent fixing the cone so it doesn't drip.
2. Decomposition- A plan can be decomposed into several different steps. An initial subset of the steps for accomplishing one goal might be identical to those of another goal. For example, a mailman walking into Burger King is the initial step of at least two plans: delivering mail and purchasing food.
3. Composite Plans- If more than one plan is executed simultaneously, then the problems caused by plan decomposition are more severe. In particular, there may be multiple sets of plans which can account for a series of actions. [16]

In this paper, I deal with only ambiguities resulting from plan refinements.

Selecting the best explanation

Consider the following problem: There is a new a doll on the market (called a Hollywood Girl). When this doll is dipped in hot water, its hair changes from yellow to blue (see Figure 1). There are at least three plans which can be achieved by dipping something into hot water:

- Putting something in hot water is a plan for heating something (see Figure 2).
- Putting something in a hot liquid is a plan for cooling the liquid (see Figure 3).
- Putting something in a liquid is a plan for causing a chemical reaction with the liquid (see Figure 4).

```
put(doll1,water1)
hollywood_girl(doll1)
water(water1)
temp(water, 90)
temp(doll1, 70)
height(doll1, 6)
hair(doll1, hair1)
color(hair1,yellow)
state-change(color(hair1,yellow),color(hair1,blue))
```

Figure 1: A description of a situation in which the color of a doll's hair changes when it is dipped in water.

```
goal:           state-change(temp(X,TX),temp(X,NTX))
plan:           heat(X)
realization:    put(X,Y)
effects:        state-change(temp(X,TX),temp(X,NTX))
                state-change(temp(Y,TY),temp(Y,NTY))
constraints:    water(Y)
                temp(X,TX)
                temp(Y,TY)
                TX < TY
                NTX > TX
                NTY < TY
```

Figure 2: A plan for increasing the temperature of an object. Variables are capitalized.

```
goal:           state-change(temp(Y,TY),temp(Y,NTY))
plan:           cool(X)
realization:    put(X,Y)
effects:        state-change(temp(Y,TY),temp(Y,NTY))
                state-change(temp(X,TX),temp(X,NTX))
constraints:    liquid(Y)
                temp(X,TX)
                temp(Y,TY)
                TX < TY
                NTY < TY
                NTX > TX
```

Figure 3: A plan for decreasing the temperature of a liquid.

There is not enough information to deduce what caused the color of the hair to change. To create an explanation, one must assume that the hair contains a substance whose color changes when heated, or a substance whose color changes when water is cooled, or a substance whose color changes when it reacts with water. Once such an assumption is made, explanation-based learning can be performed using any of the standard techniques (e.g., [5, 12, 8]). For example, if the assumption that the hair contains a substance which changes color when

```
goal:          state-change(composition(X,CX), composition(X,Z))
plan:          react(X,Y,Z)
realization:   put(X,Y)
effect:        state-change(composition(X,CX), composition(X,Z))
               state-change(composition(Y,CY), composition(Y,Z))
constraints:   CX <> Z
               CY <> Z
               liquid(Y)
```

Figure 4: A plan for causing a chemical reaction with a liquid.

heated is made, the plan for changing the color of the doll's hair in Figure 5 can be constructed. The different assumption will result in different plans which would predict different results is some situations. For example, the plan in Figure 5 would predict that the hair would change color if heated in a different manner, such as placing the doll in a heated oven. On the other hand, if a chemical reaction with water caused the hair to change color, placing the doll in the oven would not change the hair color.

```
goal:          state-change(color(X,yellow),color(X,blue))
plan:          g0001(Z)
realization:   heat(Z)
constraints:   hair(Z,X)
               hollywood_girl(Z)
               color(X,yellow)
```

Figure 5: A plan for changing the color of a doll's hair.

Four strategies have been identified to select between alternative explanations:

- Avoid hypotheses which account for the observed change, but predict other effects which were not noticed. In the doll example, this would result in the rejection of the hypothesis that a chemical reaction caused the doll's hair to change color. If this happened, then in addition to the doll's hair changing color, the water would also change color. This effect did not occur. While it is possible that some other process inhibited this effect, this should not be considered if an alternative simpler explanation is feasible.

- Avoid hypotheses which violate a general theory of causality. People have general knowledge of the kinds of situations which appear to be causal [13]. For example, one constraint on causal relationships indicates that in order for action to result in a state change for an object, the action must operate on the object. In the doll example, a general theory of causality would rule out the explanation that cooling water results in the doll's hair changing color.

- Favor hypotheses based upon assumptions for which an exemplar is known. While logically one is justified in using evidence to rule out assumptions, people tend to give more credence to a chain of reasoning which results in a conclusion similar to a known example [1]. In the doll example, knowledge of the existence of a thermometer placed on the forehead which changes color to indicate body temperature provides corroborative evidence for the hypothesis that doll's hair contains a substance which changes color when heated [4].

- Favor hypotheses which account for a larger number of observed changes.[1] For example, if in the doll example, the water and the hair both had changed color, then the hypothesis that heat makes the hair change color does not account for the water changing color. However, the hypothesis that a chemical reaction resulted in the hair color changing also accounts for the water changing color. This strategy is particularly important for assemblying composite hypotheses.

- Collect more data (by designing an experiment) to rule out alternative hypothesis. In principle,

[1]Thomas Fawcett suggested that I include this strategy.

running experiments could suffice as the only strategy for eliminating hypotheses. However, in practice, it may be the most expensive strategy.[2] For example, it would be wasteful to run an experiment to determine if the doll's hair will change color if water is cooled by placing ice cubes in the water. It is unlikely that this hypothesis will be correct because it is not consistent with general knowledge of causal relationships. Two types of experiments have been identified:

1. Change initial conditions so that one explanation does not apply. In the doll example, the doll could be placed in water which is colder than the doll. The approach to experimental design in [15] appears to be of this type.
2. Change the realization of the plan so that one goal is achieved but another is not. For example, there are many actions which result in heating an object. In addition to putting the object in hot water, one could put the object in an oven or under a hair drier. The idea here is instead of repeating the same action in a different condition (as in the first strategy), a different action which does not achieve at least one alternative goal is attempted.

Future Work

In the tradition of explanation-based learning, existing knowledge could be used to reason explicitly about the generality of assumptions. For example, it is possible to assume that a specific doll contains a substance whose color changes, or all Hollywood Girl dolls, or all small dolls in months with an "r" in their name, etc. Knowledge such as "mass produced objects tend to behave similarly" can be used to select the appropriate level of generality for assumptions to complete an explanation.

Currently, I am assuming that all important changes have been noticed before hypothesis generation. In an experimental system, before ruling out a hypothesis because an expected effect was not detected, it would be important to check to see if the expected effect actually occurred.

Conclusion

I have argued that the process of explanation should be viewed as an abductive process. In the typical case, assumptions are necessary because information required to make a deductive proof is missing. The issue of multiple inconsistent explanations arises and I have proposed several strategies for selecting the best explanation.

Acknowledgements

This work was supported in part by a RAND-UCLA Artificial Intelligence Fellowship. Comments by Tom Fawcett on an earlier draft of this paper were instrumental in the the further development of these ideas.

Bibliography

[1] Baker, M., Burstein, M. & Collins, A. Implementing a model of human plausible reasoning. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence.* Morgan-Kaufmann, Milan, Italy, 1987.

[2] Charniak, E. & McDermott, D. *Introduction to Artificial Intelligence.* Addison-Wesley, Reading, Mass, 1985.

[2]Furthermore, I have noticed that young children get upset if you put their doll in the oven to test the hypothesis that heat causes the hair to change color.

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

[4] Goodman, N. *Fact, Fiction and Forecast, fourth edition.* Harvard University Press, Cambridge, Mass, 1983.

[5] Hirsh, H. Explanation-based learning in a logic programming environment. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence.* Morgan-Kaufmann, Milan, Italy, 1987.

[6] Josephson, J., Chandraskaran, B., Smith, J., & Tanner, M. A mechanism for forming composite explanatory hypotheses. *IEEE Transactions on Systems, Man, and Cybernetics* 17(3):445-454, 1987.

[7] Kautz, H. & Allen, J. Generalized plan recognition. In *Proceedings of the National Conference on Artificial Intelligence*, pages 32-37. American Association for Artificial Intelligence, Morgan-Kaufmann, 1986.

[8] Kedar-Cabelli, S. Explanation-based generalization as resolution theorem proving. In *Proceedings of the Fourth International Machine Learning Workshop.* Irvine, CA, 1987.

[9] McDermott, D. *A Critique of Pure Reason.* Computer Science Research Report 480, Yale University, 1986.

[10] Minton, S., Carbonell, J., Etzioni, O., Knoblock C., & Kuokka D. Acquiring effective search control rules: Explanation-based learning in PRODIGY. In *Proceedings of the Fourth International Machine Learning Workshop.* Irvine, CA, 1987.

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

[12] Mooney, R. & Bennett, S. A domain independent explanation-based generalizer. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence.* Morgan-Kaufmann, Los Angeles, CA, 1985.

[13] Pazzani, M. Inducing causal and social theories: A prerequisite for explanation-based learning. In *Proceedings of the Fourth International Machine Learning Workshop.* Irvine, CA, 1987.

[14] Rajamoney, S. & DeJong, G. The classification, detection and handling of imperfect theory problems. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence.* Morgan-Kaufmann, Milan, Italy, 1987.

[15] Rajamoney, S., DeJong, G. & Faltings, B. Automated design of experiments for refining theories. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence.* Morgan-Kaufmann, Los Angeles, CA, 1985.

[16] Reggia, J., Nau, D. & Wang, P. Diagnostic expert systems based on a set covering model. *International Journal of Man-Machine Studies* 19:473-460, 1983.

[17] Sidner, C., & Israel, D. Recognizing intended meaning and speakers' plans. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, pages 203-208. Morgan-Kaufmann, Vancouver, 1981.

[18] Sullivan, M. & Cohen, P. An endorsement-based plan recognition program. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 475-479. Morgan-Kaufmann, Los Angeles, CA, 1985.

[19] Wilensky, R. *Planning and understanding.* Addison-Wesley, Reading, Mass, 1983.