Detecting and Coping with Failure

William J. Clancey
Institute for Research on Learning
2550 Hanover St.
Palo Alto, CA 94304

Proceedings of the Symposium on Explanation-Based Learning,
AAAI Spring Symposium Series,
March 1988.

Extended Abstract

A learning method is described by which a program can detect problem-solving failures and track them back to gaps in domain knowledge, which are then reformulated as questions to ask a teacher or to be empirically tested.

Apprenticeship learning research has considered how a knowledge acquisition program can help a human teacher debug problem-solving failures (Mitchell, et al., 1985, Smith, et al., 1985). This research demonstrates that learning is more efficient when a program can help determine what it needs to know. The method presented here is similar, but derives from an explicit representation of *the form of an adequate solution.* In our application, this representation consists of the constraints a good diagnostic causal model should satisfy, tied to a declarative representation of a diagnostic procedure for improving a partial model. Inadequacies are tracked back to failed diagnostic tasks (such as failure to discriminate among hypotheses). From this, gaps in the knowledge base are posed as possible explanations of why problem-solving steps failed to apply. Justification for alternative explanations is then sought from an outside source, such as a human teacher.

The learning procedure has the following steps:

- Detect possible failures (unsatisfied constraints) in the inferred causal model:
  - unable to test or refine a hypothesis;
  - unable to explain an abnormal finding;
  - finding explained by two or more hypotheses;
  - two or more hypotheses explain exactly the same findings and evidence doesn't discriminate between them, or they explain findings uniquely;
  - the situation-specific model is not specific enough to select among or construct action plans (such as alternative therapies).
- Reason backwards to say what task, if it had succeeded, would have prevented this failure, and what facts (the hypothesized gaps in the domain knowledge), if true or proved false, would allow the metarule to succeed.
- Prune alternative explanations using knowledge of what beliefs typically could be wrong or might be true, but which were not explicitly learned before. Domain-general knowledge about disorder processes can be used to focus the search for plausible facts.
- Ask the teacher questions to gain missing knowledge or validate hypothesized facts.

In practice, it may also be necessary to relax the constraints imposed on the diagnostic model, given the pragmatic requirements of how the model will be used (e.g., the action plans it must discriminate between) and the inability to confirm hypothesized facts (e.g., lack of scientific understanding of causal mechanisms). Resolving these uncertainties and filling in

details is now being investigated in a program called GUIDON-DEBUG, based on the architecture of NEOMYCIN.

Keller's explanation-based learning approach (Keller, 1986) resembles the learning method presented here. His program uses contextual knowledge about how concepts are used, in order to formulate which concepts need to be learned. Keller's program incorporates knowledge about the "performance procedure and objective," which corresponds to the diagnostic procedure and constraints on the form of a diagnosis in GUIDON-DEBUG. The objective provides a criterion for determining the usefulness of a concept, called the *operationality criterion*. For most explanation-based learning, which focuses on deriving a relation that is already implicit in the knowledge base (Dietterich, et al., 1986, Dietterich, 1986), the operationality criterion concerns efficiency of the inference procedure. That is, the goal of learning is to make the program able to solve a search problem that was previously too time-consuming.

The learning method described here does not involve simply chaining together previously known facts and procedures, but conjecturing new facts or conjecturing the need for a certain type of knowledge. The operationality criterion is the description of a diagnostic solution, particularly its form as a causal model and how it will be used to select action plans (repairs). Learning is driven by failure to satisfy these constraints. GUIDON-DEBUG cannot automatically refine or generalize a previously known concept, as in other EBL programs. Rather, the explanation of problem-solving failure is analyzed to determine the concepts or relations that could prevent the failure, which must then be confirmed or supplied in more detail.

This analysis is a form of *goal regression*, a technique found in many EBL programs: We reason from the failed goal (unsatisfied diagnostic model constraint) to the task that could have satisfied the goal if it succeeded, back through failed metarules and failed metarule preconditions to other failed subtasks, eventually reaching ground facts about the world that are not in the knowledge base (e.g., the subtypes of a disease or what might cause it).

This learning method thus extends previous EBL in several ways:

- Goal regression involves reasoning through the problem-solving procedure itself (tasks and metarules), rather than a separate description of the procedure (as in Keller's program).
- The operationality criterion is described in terms of the form of a solution and how it will be used, rather than in terms of computational efficiency.
- Learning is based in explaining problem-solving failures, as detected by the program itself, not in explaining why a supplied example is correct. That is, the method involves determining what needs to be learned in order to properly solve problems, not just to increase problem-solving speed. Thus, this method bridges a gap between EBL and cognitive models of failure-driven learning (Schank, 1981,

Kolodner and Simpson, 1984).

- The problem-solving procedure involves a schema-model of the world (the diagnostic relations of the knowledge base), which constitutes an incomplete theory, in contrast with the inherently axiomatic theory of domains like calculus (see (Clancey, 1986) for further discussion).

Further comparisons can be drawn with MORE (Kahn, et al., 1985), MOLE (Eshelman, et al., 1986), and the LEARNING APPRENTICE SYSTEM (Smith, et al., 1985). GUIDON-DEBUG is distinguished from these programs by detecting its own failure to solve a problem and reasoning through its inference procedure to explain how additional domain knowledge might have prevented the failure. The inference procedure is either implicit in learning heuristics used by the other programs or redundantly encoded in both the performance and learning programs.

References

Clancey, W.J. Qualitative student models. In Traub, J.F. (editor), *Annual Review of Computer Science*, pages 381-450. Annual Reviews, Inc., Palo Alto, 1986.

Dietterich, T. G. Learning at the knowledge level. *Machine Learning*, 1986, *1(3)*, 287-316.

Dietterich, T. G., Flann, N. S., and Wilkins, D. C. *A summary of machine learning papers from IJCAI-85*. Technical Report 86-30-2, Oregon State University, 1986.

Eshelman, L., Ehret, D., McDermott, J., and Tan, M. *MOLE: A tenacious knowledge acquisition tool*, in *Proceedings of Knowledge Acquisition for Knowledge-Based Systems Workshop*, pages 13-1--13-12, 1986.

Kahn, G., Nowlan, S., and McDermott, J. *MORE: An intelligent knowledge acquisition tool*, in *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 581-584, Los Angeles, August, 1985.

Keller, R. M. *Deciding what to learn*. Technical Report ML-TR-6, Rutgers University, 1986.

Kolodner, J.L. and Simpson, R.L. *Experience and problem solving: a framework*, in *Proceedings of the Sixth Annual Conference of the Cogtnitive Science Society*, pages 239-243, Boulder, June, 1984.

Mitchell, T.M., Mahadevan, S., Steinberg, L.I. *LEAP: A learning apprentice for VLSI design*, in *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 573-580, Los Angeles, August, 1985.

Schank, R. C. Failure-driven memory. *Cognition and Brain Theory*, 1981, *4(1)*, 41-60.

Smith, R. G., Winston, H. A., Mitchell, T. M., and Buchanan, B. G. *Representation and use of explicit justifications for knowledge base refinement*, in *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 673-680, 1985.