Autonomous Discovery In Empirical Domains

Gary Livingston

University of Pittsburgh
Room 202 Mineral Industries Building
Pittsburgh, PA 15260
Email: gary@cs.pitt.edu; WWW: www.cs.pitt.edu/~gary

Bruce G. Buchanan

University of Pittsburgh

We have tested a hypothesis that the agenda-based architecture used in AM (Lenat 1982) can be adapted to perform autonomous discovery in empirical domains. Our preliminary evaluation of our adaptation, HAMB (Heuristic, Autonomous Model Builder), suggests that the architecture is practical and sufficient for empirical discovery. HAMB was able to make many discoveries and rediscoveries from the domain of macromolecule crystal-growing experiments (Gililand 1987).

Adapting AM to Empirical Discovery

We made three types of changes to AM's basic structure:

- **New concept-space.** Instead of AM's concept-space of set- and number-theoretic concepts, we substituted concepts encountered while performing rule-induction: attributes, rules, conjuncts, example-sets, rule-sets, and example-classes.
- **New operation-types.** Because we changed the fundamental discovery-type and concept-space, we provided new operation-types.
- **New heuristics.** Because we changed the operation-types, we also changed the heuristics performing them.

HAMB's basic operation is identical to AM's: (1) select and remove the most "plausible" task from an agenda of tasks, (2) perform the task using heuristics provided for performing tasks, and (3) repeat steps (1) and (2) until a set of stopping criteria are satisfied. Tasks are put onto the agenda during initialization and while executing heuristics when performing tasks.

Calculating Plausibility: Reasons and Ratings

When putting a task onto the agenda, a heuristic provides *reasons*, text justifications, for performing the task, and corresponding *ratings*, numeric values indicating the relative strengths of the reasons. No task is allowed on the agenda without a reason for performing it. If a task is suggested that is already on the agenda, any new reasons and corresponding ratings are associated with the existing task, increasing its plausibility (see below).

A task's *plausibility* is calculated from the *ratings* of its reasons and the *worths* (interestingnesses) of the items in the task. *We feel that this use of a task's reasons and ratings as well as the worths of its items to calculate its plausibility makes the architecture particularly well suited for empirical discovery and discovery in general.*

The architecture's use of reasons and ratings has the following properties (ceteris paribus):

- No task is performed without a reason for performing it.
- A task with more or stronger reasons is more plausible.
- A task with more interesting items is more plausible.
- As additional reasons are identified for performing a task, the task's plausibility increases.

The reasons and ratings also provide two additional benefits. Because they are intuitive and heuristic in nature, they aid the creation of heuristics and task-types and increase the expressiveness of the task-types. And, reasons and ratings provide documentation for HAMB's choices of tasks, facilitating comprehension of its actions.

Conclusion

We have adapted Lenat's AM to autonomous empirical discovery. Our tests indicate that our adaptation, HAMB, is practical and sufficient for empirical discovery, making several significant discoveries and rediscoveries in the domain of macromolecule crystal-growing.

Bibliography

Gililand, G. C. 1987. A Biological Macromolecule Crystallization Database: A Basis for a Crystallization Strategy. In *Proceedings of the Second International Conference on Protein Crystal Growth.* Bischenberg, Strasbourg, France, North Holland.

Lenat, D. 1982. "AM: Discovery in Mathematics as Heuristic Search." In Lenat, D., and Davis, R., Eds., *Knowledge-Based Systems in Artificial Intelligence.* New York, New York, McGraw-Hill: 3-225.