Plan Recognition Strategies in Student Modeling: Prediction and Description

Bob London and William J. Clancey

Heuristic Programming Project, Computer Science Department
Stanford University, Stanford CA 94305

Abstract

This paper describes the student modeler of the GUIDON2 tutor, which understands plans by a dual search strategy. It first produces multiple predictions of student behavior by a model-driven simulation of the expert. Focused, data-driven searches then explain incongruities. By supplementing each other, these methods lead to an efficient and robust plan understander for a complex domain.

1. Basic problem: Modeling strategic problem solving

Diagnostic problem-solving requires domain knowledge and a plan for applying that knowledge to the problem. A hypothesis-directed diagnostic plan is a rationale for focusing on diagnoses (partial solutions) and for gathering data to solve the problem. The plan is thus a strategy for selecting and ordering the application of domain knowledge.

Teaching diagnosis involves recognizing the intent behind a student's behavior, so that missing knowledge can be distinguished from inappropriate strategies. The teacher *interprets* behavior, *critiques* it, and provides *advice* about other approaches. To do this successfully and efficiently in a complex domain, the teacher benefits from multiple, complementary modeling strategies.

GUIDON2 is a tutoring program that uses the case method approach to teach medical diagnosis [5]. The system divides this task among three components: an "expert," a student modeler, and an instructional manager (see Figure 1-1). Its expert component, NEOMYCIN [4], separately and explicitly represents knowledge about the medical domain and the *domain-independent* strategies of diagnostis. The student modeler, a subprogram called IMAGE, interprets the student's behavior by using NEOMYCIN's knowledge, evaluates the student's skill, and produces alternatives. The instructional module of GUIDON2 will then apply discourse and teaching strategies in deciding whether to interrogate or advise the student.

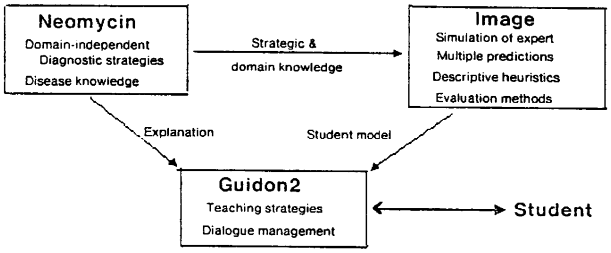

Figure 1-1: Components of the GUIDON2 teaching system

A model of student strategies in medical diagnosis must disambiguate the possible purposes and knowledge underlying the student's actions. The approaches followed by other plan recognizers and student modelers are not sufficient here because:

1. the complex domain makes thorough searches impractical, whether top-down or bottom-up;
2. we are not modeling only facts and rules used in isolation, but also the procedures for applying them;
3. every one of the student's actions must be monitored in case the teaching module decides to interrupt;
4. his behavior must be evaluated and not just explained; and
5. we might not have any explicit goal statements from the student, so we expect to rely only on his queries for problem data as evidence for his thinking.

A top-down, model-driven search works well in an area where the number of plausible solutions is small, and the cost of computing them is manageable. In the SPADE-0 advisor for designing simple programs [9], Miller could use a narrow-branching, context-free "problem-solving grammar" to recognize next steps. Medical diagnosis does not generally fit this requirement. Tracking down a single solution can be very expensive, and many possible answers may exist. However, if the model of expertise offers a way to rank-order strategic decisions, then it can be used by a top-down search to suggest some range of solutions. Problems include: how to apply the model, how far to go in tracing a possible answer, and how many such solutions to generate.

A bottom-up, data-driven search is best in domains where it is easy to recognize the reasons underlying a solution. But in medical diagnosis, an "upward" search often leads to excessive combinatorics. If the student asks how long the patient's headache has lasted, NEOMYCIN links could show that he is testing a hypothesis of viral or bacterial meningitis, or hemorrhage, hematoma, migraine, etc. Or his diagnosis might be more inclusive (meningitis or vascular disorders in general). He might not be testing any specific diagnoses, but routinely following up recent data, or exploring for new hypotheses. Even if his focus can be specified, we would still have to surmise his overall purposes by searching for patterns in his previous actions: we must account for his *planning* -- not only his knowledge -- if we are going to teach procedures of diagnosis.

Other student modelers and plan recognizers have offered useful tools and have also shown why particular features prevent their direct application for teaching strategies of medical diagnosis. If a student modeler is infrequently invoked (say when the user explicitly asks for help) then a thorough multi-technique, multi-pass search is practical. Genesereth's MACSYMA advisor [6] takes this approach; it also has the user's explicit statement of his goal as a guide in the search. The BELIEVER program [10] predicts the subject's current plan and updates the plan's details after observations, but does not judge the appropriateness of his behavior. Its single predicted plan, plus data-driven completion of details and repairs, is appropriate in its domain of common-sense actions where there are few probable interpretations of any given action; since the predicted plan is unlikely to be far off, the need for repairs is relatively minor. Another ICAI program, BUGGY [1], succeeds in a forward, data-driven search of a "procedural net" because the domain (children's subtraction) was completely described by about 200 rules. The student's skills could thus be mapped or "overlayed" onto the procedural net. Goldstein and Carr's WUMPUS coach [2, 7], and the student modeler for the first version of GUIDON [3], also "overlay" estimates of the student's performance onto semi-independent rules of problem knowledge, in a primarily data-driven way.

The IMAGE student modeler uses two separate but complementary apporaches to infer and evaluate the student's plan, under our requirements as listed above. It first forms a model-driven *range of predictions*, then data-driven *descriptions* about the student's behavior. (IMAGE also takes two further steps: describing the student's level of domain knowledge, and evaluating the success of the student model itself; these are discussed elsewhere [8].)

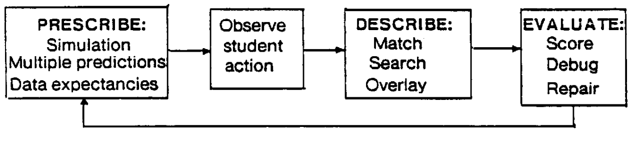

Figure 1-2: Phases of the modeler

2. Understanding strategies: An example

In a test case, the tutor presented a patient who complained of headaches, vomiting, and bouts of irrationality and apathy. The student then asked several questions about the patient's headache, and whether he had been irritable or depressed (both negative). At this point IMAGE, using NEOMYCIN's knowledge, infers that the student's **active hypotheses** are brain pressure and tension headache; and that his current task (purpose) is trying to set up a group of hypotheses to consider. (Refer to Figure 2-2 for IMAGE's resolution of these actions into a global plan.)

Since IMAGE believes the student is not yet focusing on any one diagnosis, it predicts that he will continue to ask questions which follow up previous data, rather than focus on hypotheses. (We will see below how the predictions are generated.) At this point, the data **expectancies** are:

Stiff neck; Fever; Precipitating factors of headaches; Abruptness of headache onset; Apathetic; Confused; Amnesiac; Dysphasic; Aphasic

These represent the choices NEOMYCIN would favor, in the partial order given. However, the student's next query is: "Does the patient show focal neurological signs?" This does not match the expectancies: the model-driven phase has not explained the student's thinking here. So IMAGE begins a *rule-based, bottom-up search strategy* to understand his behavior. One focused search is guided by the rule shown in Figure 2-1. This search succeeds in finding a new diagnosis (brain mass-lesion) which is a refinement of an active hypothesis (brain pressure) and is also related to the query (focal signs). Thus IMAGE assumes the student is "refining" (specifying) his diagnosis.

```
Rule-20: Refined hypothesis
IF some untested hypotheses that are closely relevant
    to student's data query are related as causal or
    "taxonomic" descendents of any members of his
    set of active hypotheses,
THEN assume he is "Refining" an active hypothesis;
    if it can be pinpointed to one hypothesis, then
    consider "refining" that node in the student-model
```

Figure 2-1: Example of a rule for Descriptive phase

Explaining the student's action leads to updating the student model in several ways. Since he seems to be testing a diagnosis, IMAGE infers that the student is no longer gathering initial data ("Identify-problem" node in Figue 2-2); now he is trying to focus on a few diagnoses ("Establish-hypothesis-space" node). (Further modeling here includes "overlaying" domain knowledge, and student evaluation [8].)

IMAGE now **predicts** the student's behavior by a *model-driven generation of multiple expectations*, since the student has entered a new stage in his problem solving. By **simulating** NEOMYCIN, IMAGE finds that the expert model's preferred plan would be to pursue the current focus (brain pressure). Its secondary choice is to pursue the other active hypothesis. These preferred plans lead to the following data **expectancies** (in order of evidential strength):

For hypothesis **brain pressure:** Papilledema; Enlarged head; Diplopia; Seizures

For hypothesis **tension headache:** Headaches sensitive to emotional disturbances; Headache pressure; Headache throbbing?; Fever (disconfirmatory)

As it turns out, the student asks whether the patient has a fever. IMAGE confirms its prediction: it believes that one part of the student's plan is to test (by mildly disconfirming evidence) the diagnosis of tension headache. The top-down prediction produces an immediate, likely explanation. (This step is incorporated into the global plan as the "Test-hypothesis" node in Figure 2-2. It would have been very difficult to pin down the student's thinking with a bottom-up search starting with "fever" because of the multiple purposes such a datum could serve.)

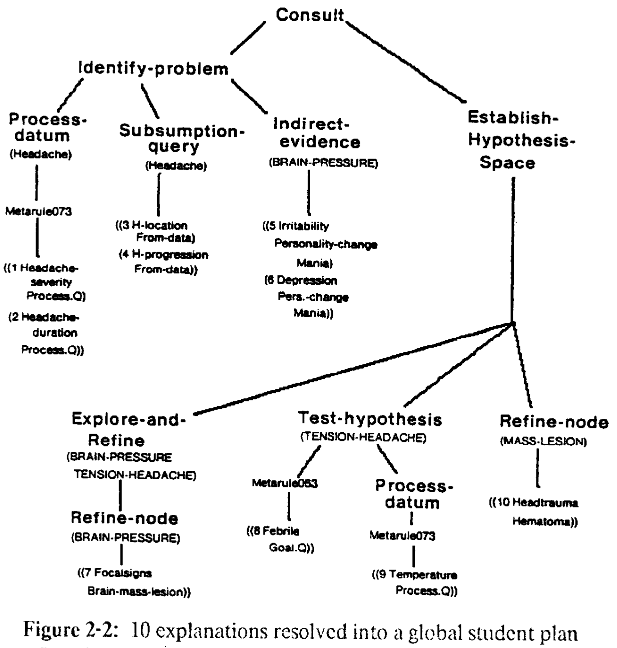
Consult
Identify-problem
Process-datum
(Headache)
Metarule073
((1 Headache-severity Process.Q)
(2 Headache-duration Process.Q))
Subsumption-query
(Headache)
((3 H-location From-data)
(4 H-progression From-data))
Indirect-evidence
(BRAIN-PRESSURE)
((5 Irritability Personality-change Mania)
(6 Depression Pers.-change Mania))
Establish-Hypothesis-Space
Explore-and-Refine
(BRAIN-PRESSURE TENSION-HEADACHE)
Refine-node
(BRAIN-PRESSURE)
((7 Focalsigns Brain-mass-lesion))
Test-hypothesis
(TENSION-HEADACHE)
Metarule063
((8 Febrile Goal.Q))
Process-datum
Metarule073
((9 Temperature Process.Q))
Refine-node
(MASS-LESION)
((10 Headtrauma Hematoma))

Figure 2-2: 10 explanations resolved into a global student plan (Boldface nodes are tasks invoked by NEOMYCIN meta-rules; small capitals denote active hypotheses; numbers are followed by student's data query; remaining terms refer to kinds of strategies and contexts)

3. Prediction and description: Discussion

IMAGE's predictions are termed "**prescriptive**" because they represent the range of plans that the student *should* be doing, because they are what the NEOMYCIN expert *would* do. They pin down the most likely possibilities of student behavior immediately. If the student's actions violate the predictions, then slower data-driven processing is required to explain the data. But when the observations match the predictions, student behavior is quickly explained.

IMAGE generates its predictions by **simulating** the expert at key points: applying the domain-independent tasks and strategic meta-rules to the inferred context of student thinking. However, it makes several adjustments to increase (1) likelihood of successful *recognition* and *evaluation* of student behavior, (2) depth of *detail*, and (3) computational *efficiency*. Also, we want (4) a *robust* model: it should perform reasonably well even if the student acts in unusual ways, and it should be able to recover from its own errors. (See [8] for discussion of latter two issues.)

To gain the most benefit from its predictions in terms of *understanding* and *evaluating* student behavior, IMAGE generates **multiple expectations.** It does not stop at the first action that NEOMYCIN would take; it finds both the set of near-equivalent favorite actions and a range of secondary alternatives. This increases the chance of matching the student's observed action, and provides a spectrum of behaviors against which his behavior can be judged.

Of what value to the teaching module is a confirmed prediction? IMAGE has not simply generated data **expectancies** in each prediction, but has kept a trace of the strategies and domain rules used in the simulation of NEOMYCIN along the way. For example, the prediction leading to the expected query of "fever" also records the following:

Task:	Test the focused hypothesis
Strategy:	Strategy-rule063: If "trigger" and "enabling" data has already been tried, then consider any other evidence available.
Focus:	Tension headache
Domain rule:	Rule156: if patient has a fever, then his headache is not a tension-headache [.2 belief]

This information about the student's behavior is passed to the instructional executive, *along with analogous information about the other predictions*, many of which could be judged *more appropriate* at the moment. The generation of multiple predictions, with their traces of strategical decisions, thus add a *normative* element to the understanding of student plans. Since we can often identify one strategical choice as preferable to another, we can group behavioral predictions by their desirability. This provides a ready basis for advising the student.

Recall the example in Section 2: the student apparently considered brain pressure in his first query, then moved to a new hypothesis. IMAGE's predictions show that the expert would have continued to pursue the more likely hypothesis of brain pressure; and if it did pursue tension headache, stronger evidence could have been chosen (such as the role of emotional factors). The teaching module could use one of these alternatives as a basis for advising or testing.

The rule-based, bottom-up searches are a valuable complement to the top-down predictive phase. But they have a disadvantage: since the bottom-up searches are only practical under tightly focusing heuristics, we cannot get alternative answers. So they give no ready basis for comparison of possibilities, as do the top-down predictions. The only way to evaluate the appropriateness of our bottom-up explanation is by incorporating explicit "buggy" links [1, 11]. Buggy rules have not yet been added to the NEOMYCIN expert model.

4. Conclusion and current status

Preliminary tests of the IMAGE student modeler have indicated that the complementary search strategies of model-driven **predictions** and data-driven **descriptions** yield highly plausible analyses of students' strategical behavior. The *efficiency, detail,* and *robustness* of the modeling have also satisfied initial demands. Even with occasional unusual queries, IMAGE almost never yields implausible explanations: this is because the (1) bottom-up searches are highly focused, and (2) if an explanation is not confidently believed, only partial results are saved in order to help disambiguate the next observation. We are now arranging to run controlled experiments with medical students and experts, in which we will test built-in methods of localizing inconsistencies to either the student, the student model, or the expert model [8].

A few plan understanding programs include a **predictive** phase (such as BELIEVER, for common-sense plans [10]). Very few plan understanders generate **multiple predictions**; for many applications this would be inefficient. Multiple predictions are useful in domains where (1) either the number and cost of likely solution paths (from high-level strategies down to result) are not very large, or else the paths can be ranked by appropriateness (so that generation of predictions can be selective), and (2) recognizing solution paths by observing final data is often combinatorially impractical. Medical diagnosis fits this description. Bottom-up searches are not ruled out; in fact, they complement the model-driven predictions by often explaining observations that violate expectations.

We have shown how the **multiple prediction** strategy can aid plan recognition for teaching medical diagnosis in several ways: depth of *detail* in plan recognition; student *evaluation* (using ranked groups of "prescribed" behaviors as a standard); and complete alternatives ready to serve as *advice* or as a basis for *testing* the student. With the student's behavior explained as strengths and weaknesses in problem-solving *strategy* (ordered tasks and methods, hypothesis management and focus) as well as in domain-specific *knowledge* (hypotheses, rules, and relations), the tutor is then in a position to pinpoint its instruction to the areas in most need of attention.

5. Acknowledgments

This research has been supported in part by ARPA and ONR contract N00014-79C-0302. Computational resources have been provided by the SUMEX-AIM facility (NIH grant RR 0078).

We wish to thank our GUIDON collaborators Reed Letsinger, Bruce Buchanan, and Dr. Timothy Beckett, MD. Special thanks to Jim Bennett, Barbara Hayes-Roth, and Diane Kanerva for their helpful comments on drafts of this paper.

References

[1] Brown, J.S. and Burton, R.R., Diagnostic models for procedural bugs in basic mathematical skills, *Cognitive Science 2* (1978), 155-192.

[2] Carr, B. and Goldstein, I., *Overlays: A theory of modelling for computer aided instruction,* AI memo 406, (MIT AI Lab, 1977).

[3] Clancey, W.J., *Transfer of rule-based expertise through a tutorial dialogue,* Ph.D. Thesis, Stanford University (August 1979).

[4] Clancey, W.J. and Letsinger,R., *NEOMYCIN: Reconfiguring a rule-based expert system for application to teaching*, pp. 829-836, Proceedings of the Seventh IJCAI, (1981).

[5] Clancey, W.J., *Methodology for building an intelligent tutoring system,* Rep. STAN-CS-81-894, (Stanford Univ., 1981). To appear in *Methods and Tactics in Cognitive Science*, Kintsch, Polson and Miller, editors.

[6] Genesereth, M.R., *Automated consultation for complex computer systems,* Ph.D. Thesis, Harvard University (November 1978).

[7] Goldstein, I., *Developing a computational representation for problem solving skills,* AI memo 495, (MIT AI Lab, 1978).

[8] London, B., *Plan recognition, overlays, and evaluation methods in student modeling,* Technical report, Stanford Heuristic Programming Project, in preparation.

[9] Miller, M.L., A structured planning and debugging environment for elementary programming, *International Journal of Man-Machine Studies 11* (1979), 79-95.

[10] Schmidt, C.F., *Plan recognition and revision: Understanding the observed actions of another actor,* AI memo CBM-TR-115, (Rutgers Univ., September 1980).

[11] Sleeman, D.H. and Smith, M.J., Modelling student's problem solving, *Artificial Intelligence Journal 16* (1981), 171-187.