Proceedings of the Eighth Annual Florida Artificial Intelligence Research Symposium, Melbourne, FL, 1995, pp. 218-222.

AN ARCHITECTURE FOR GOAL-DRIVEN EXPLANATION

Raja Sooriamurthi & David Leake
Computer Science Department
215, Lindley Hall, Indiana University
Bloomington, IN 47405
raja@cs.indiana.edu leake@cs.indiana.edu

Abstract

In complex and changing environments explanation must be a dynamic and goal-driven process. This paper discusses an evolving system implementing a novel model of explanation generation — Goal-Driven Interactive Explanation — that models explanation as a goal-driven, multi-strategy, situated process inter-weaving reasoning with action. We describe a preliminary implementation of this model in GOBIE, a system that generates explanations for its internal use to support plan generation and execution.

1 INTRODUCTION

In standard AI models of explanation, explanation is purely a reasoning process done in isolation from the situation being explained. Likewise in models of explanation that consider explanation goals, goals exert a fixed top-down influence. Our model, termed *goal-driven interactive explanation* (GDIE) [Leake, 1994] allows incremental information obtained by interaction with the environment to influence the explainer's goals and dynamically re-focus the explanation process. In this paper we describe an implemented system in which the explanation process is shaped by ongoing and strategic decisions as the explainer interacts with its environment.

GDIE models explanation in the context of specific motivation and specific circumstances affecting the cost of pursuing the explanation and how it may be used. As an example, a driver trying to diagnose a break-down in an automobile while on the way to an airport would first have to gather additional information to determine whether a repair could be effected in time. Situational factors such as the availability of tools or alternative modes of transport (e.g., taking a taxi) influence the decision as to whether to explain or not. Our model uses the task goals of the system (e.g., get to the airport on time) to reason about whether to explain, which strategies to use to pursue an explanation, what information is needed to establish an explanation and how to obtain it in the current situation. As the explanation process progresses incremental information obtained by interaction with the surrounding situation feeds back to influence the explanation goals and may strategically re-focus the explanation process (possibly even leading to its abandonment).

We describe the architecture of a system, GOBIE[1] developed for GDIE. The system consists of a planner integrated with a case-based explainer functioning in a simple simulated world. GOBIE generates explanations as needed to aid in the formulation and execution of its plans. The next section outlines the capabilities that GDIE requires. Subsequent sections describe our implemented system with an illustrative example.

2 REQUIREMENTS FOR GOAL-DRIVEN EXPLANATION

To support goal-driven explanation a system must have the following capabilities.

- *Ability to recognize information needs.* An explanation bridges the gap between unexpected observed events and the explainer's internal world model. For instance, to a person who *knows* that her car battery is dead the fact that the car doesn't start will not be unexpected. But to a person whose knowledge about the status of the car is incomplete failure to start would be unexpected requiring the acquisition of information to explain the malfunction. The central requirement of our model is that the information seeking actions during explanation are guided by explanation goals [Leake, 1992] (representing *what* the system is trying to explain) and task goals (representing *why* the system is trying to explain) [Ram and Leake, 1994]. These goals should be accessible to both the process requesting explanation (in our system the planner) and the explainer.
- *Interleaved explanation, planning and action.* The formation of an explanation goal initiates explanation. To complete the explanation the system needs to gather additional corroborating information by first reasoning about what information

[1] GOal Based Interactive Explanation

is needed, planning to acquire that information [Hunter, 1990, Pryor and Collins, 1991] and then executing the plan to gather the information. This action may either be in the external world (e.g., running a diagnostic and observing the result) or an internal reasoning action (e.g., reminding, inference, transformation). In planning to acquire information we need different *knowledge acquisition strategies*.

- *Multi-strategy processing.* By situating the explanation process in the environment in which it will be used we gain flexibility by being able to formulate explanations by a variety of approaches as influenced by the extant situation (e.g., being reminded of a prior similar situation, gathering information from a knowledge source (an expert, manual etc), and as a final resort in an unconducive situation, chaining an explanation together from scratch).

- *Utility-based processing.* The cost of exploring a candidate explanation is determined by the knowledge acquisition actions that need to be performed to determine the validity of the candidate. Contrary to other models of explanation which focus on a single explanation, in GDIE multiple candidate explanations are explored concurrently. Explanation effort is continuously re-focussed on that candidate whose pursuit is estimated to have the better utility in the current situation. Also, the depth to which an explanation is pursued is determined by the task goals. The guiding principle is to *explain until the explanation satisfies the system's need for information*.

- *Dynamic adjustment of goals.* Purely top-down goal-based guidance is not sufficient. The explanation process should be able to feedback to affect the goals of the system. For instance, in trying to diagnose a faulty appliance with the intention of repairing it, it may be determined that the problem is beyond repair. The goal of the system may then change to that of determining if the appliance could be sold for spare parts.

- *Dynamic pursuit of goals.* In a goal-based system all the goals will not be of the same priority. Lower priority goals should be suspended allowing the system to focus on achieving higher priority ones. At the same time an interactive explanation system situated in and capable of action in the world should be able to opportunistically notice situations for satisfying the suspended goals [Hammond and Seifert, 1994].

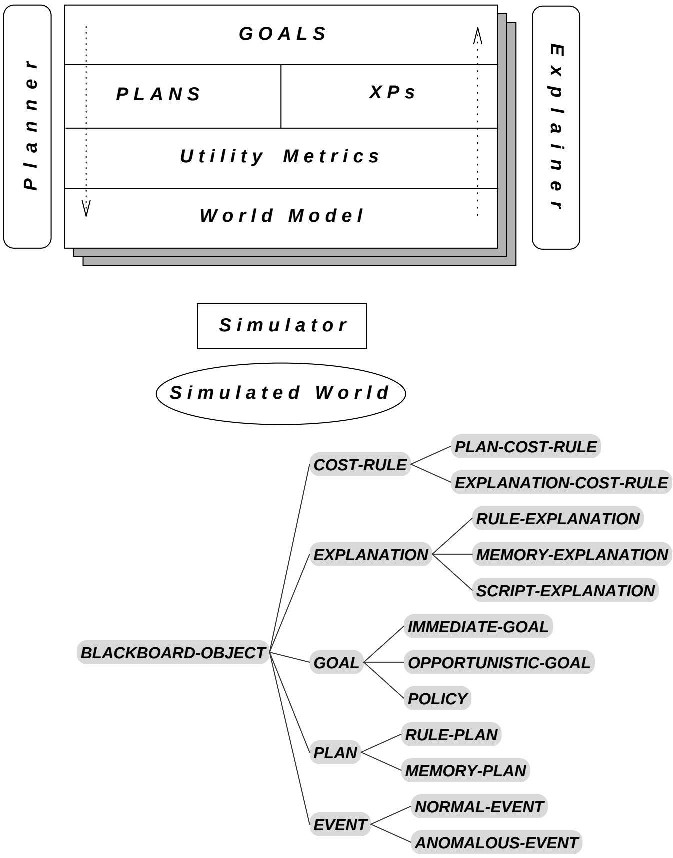

Figure 1: The blackboard framework of the system

3 SYSTEM ARCHITECTURE

GOBIE is an evolving implementation of GDIE. To provide the previously described functionality the system is organized around a blackboard architecture [Engelmore and Morgan, 1988]. GOBIE consists of a STRIPS style planner [Fikes and Nilsson, 1971] functioning in a simple simulated world and a case-based explainer [Schank and Leake, 1989, Schank et al., 1994]. Case-based explanation is the application of the case-based reasoning paradigm [Kolodner, 1994] to the problem of creating explanations: new explanations are formed by adapting explanations that had been applied to similar prior problems.

The schematic representation of the blackboard and the class hierarchy of the blackboard objects is given in Figure 1. Information posted on the blackboard is available to the planner and the explainer. World information is represented in two ways in the system. The *simulated world* represents the condition of the "real world". This information is available only to the simulator. The generated plans and explanation actions are executed by the simulator against the backdrop of its simulated world.

The success or failure of the various simulated actions is probabilistically determined and the outcomes of these simulated actions are registered as the *world model* on the blackboard. These world events form the situation in which planning and explanation is done by the system. For example, the simulated world could have a feature wherein the car ignition may start with a probability of 0.6. This information is not directly accessible to the planning system. The planner and the explainer thus have *incomplete* information about the world.

The planner and explainer find out about the world by executing simulated actions and noticing the results. It is only when the planner initiates the action of starting the car does the feedback from the simulator inform the system about the current world condition. The plans generated by the planner, the candidate explanations (XPs) [Schank, 1986] under consideration by the explainer, the utility-metrics used to determine which candidate to explore further are all posted to the blackboard. In the current implementation the explainer is a case-based reasoner: given the description of an unexpected event it retrieves similar prior cases which form the starting points for candidate hypothesis. These candidates are then pursued by the explainer which plans and acts to gather corroborating information. The interweaving of explanation, planning and action is obtained through the central control of the blackboard which is driven by the posted goals. This is described in further detail in the next section.

3.1 Goal-Directed Processing

The overall processing of the system is governed by the explanation and system goals.

Goal types: Depending on their priority, goals are categorized into three types. *Immediate goals* are those the system takes explicit effort to achieve. These goals drive the primary processing in the system. *Opportunistic goals* are indexed by the features of the situation in which they may be achieved [Hammond and Seifert, 1994]. When the system notices the occurrence of these situational features in the simulated world it triggers the achievement of these goals. *Policies* influence the broad nature of the processing. For example, a policy of "conserve cost" will lead the system to favor the cheapest (but not necessarily the best) amongst alternatives.

Goal and situational influence on processing: As proposed by the GDIE model, goals exert a *top-down influence* on the explanation process indicative of the desired outcome (e.g., need to catch a flight) while the world events exert a *bottom-up influence* indicative of the "realities of the situation" (in the simulated world) and what can be achieved (e.g., car doesn't start, appropriate tools are not available). The presence of an immediate goal on the blackboard triggers the planner to generate a plan to attain the goal. During the simulation of a generated plan the occurrence of a failed plan step initiates the explanation process. By examining the overarching goals of the planner and the current situation (as represented in its world model) the explainer first has to determine whether to try to explain the failure or whether it would be better to abandon the current plan and choose an alternative plan (if available). Starting off with incomplete information the explainer has to strategically elaborate its world model to formulate an effective explanation. The case-based explainer uses its prior experience and available knowledge about the current situation (e.g., there is not much time left to catch the flight, the flight has been delayed) to determine whether to explain or not. Once an explanation is initiated the system uses its case memory to recall previous similar episodes.

Utility-based reasoning: A characteristic feature of our model is that these alternatives are examined in a utility-based manner depending on the explanation goals: explaining to (a) determine the cause of the failure (b) effect a repair (c) prevent similar future failures. The explainer then uses various knowledge acquisition strategies to gather corroborating or refuting information to establish a viable hypothesis. Two such strategies currently implemented are: (i) *Diagnostic search.* The idea here is to run diagnostic tests to gather confirming or refuting information for a hypothesis. For example to determine if a car battery is dead or not a simple test would be to try the car radio. (ii) *Analogical search.* The strategy here is to substitute a required but inaccessible piece of information with an accessible equivalent (e.g., substituting a 5/8-wrench needed for a repair with a monkey wrench). As with the decision of whether to explain or not, the system goals and encompassing situation determine whether to continue examining a candidate hypothesis or not. The goals and situation also determine to what extent a candidate is examined. The extent to which a problem is explained may give rise to explanations of different levels of granularity termed hierarchical explanations.

4 AN IMPLEMENTED SCENARIO

A scenario implemented in our system concerns the goal of reaching an airport under time restrictions. The system is initialized with the following goals: (a) a policy of reducing cost and (b) an immediate goal of catching a flight. The unsatisfied immediate task goal initiates planning to achieve it. Having no initial plans in memory the planner builds two plans from scratch to achieve the goal of catching a flight: driving or taking a taxi (Figure-2).

The process of planning involve two costs: that of plan generation and that of plan execution. While in GOBIE the cost of generating both of these plans is the same their expected execution costs differ. It chooses the cheaper (and of higher utility) alternative of driving, and the plan steps are simulated. A problem is encountered when the system attempts to turn on the ignition in starting the car: the vehicle fails to start. This failure interferes with the attainment of the planners goal of catching the flight and results in an explanation goal of trying to determine the reason for the vehicle failing to start.

The occurrence of an explanation goal initiates explanation. Using the information that the vehicle failed to start the explainer retrieves three candidate explana-

```
1. Plan-2  [Rule-plan]
   Goal:   [Goal-4: Immediate (catch rob flight-701)]
   1. Rob Gets Into Car At Home
   2. Starts The Car
   3. Drives From Home To Air-Port
   4. Rob Boards The Plane At Air-Port
       Expected cost:    8
       Usefulness:      10
       UTILITY:        125

2. Plan-1  [Rule-plan]
   Goal:   [Goal-4: Immediate (catch rob flight-701)]
   1. Rob Gets A Phone Directory
   2. Looks Up The Taxi Phone Number
   3. Picks Up The Phone
   4. Calls A Taxi
   5. Takes A Taxi From Home To Air-Port
   6. Rob Boards The Plane At Air-Port
       Expected cost:   22
       Usefulness:      10
       UTILITY:         45
```

Figure 2: Sample plans generated by the system. Utility = 100*Usefulness/Expected-cost

tions: fouled spark-plugs, an engine priming problem and a clogged fuel pump. A typical explanation retrieved from the systems case memory consists of the following components:

(i) an estimated cost to examine the candidate. The cost of examining a candidate consists of a static pre-assigned cost and a situated cost which depends on the context of the explanation (e.g., the availability of a needed tool etc)

(ii) a collection of *situational modifiers* (e.g., the car hood is open, a 5/8-wrench is not available) that influence the cost and applicability of the explanation in the current situation

(iii) An estimate of the utility of the explanation

(iv) a collection of diagnostic actions needed to either establish or refute the explanation.

Using diagnostic queries associated with the retrieved candidates and with the policy of minimizing cost the alternatives are examined in parallel in a utility-driven manner using either action in the (simulated) world or internal reasoning. Each action required by the explainer is simulated. As mentioned earlier the outcomes of these actions in the simulator is determined probabilistically (e.g., with a probability of 0.8 the car might not have a tool-box).

Following are brief descriptions outlining the different paths along which an explanation might proceed:

- ▷ In one situation the necessary tools to examine a candidate explanation are not available in the simulated world and hence explanation effort is abandoned. The still extant goal of catching a flight then results in the plan of driving being abandoned and the alternative of taking a taxi being executed.

- ▷ In another situation, a wrench needed to examine the spark-plugs isn't available. The system performs an *analogical search* in the semantic memory representing its world model for a similar object. By action in its world the alternative wrench is obtained and the explanation effort completed. Successful explanation leads to the suspended plan being resumed.

- ▷ In a third situation, where the reason the vehicle does not start is a dead battery the explainer could produce a cheap to generate explanation of the form "something is wrong with the engine" to the more useful "the battery is dead" to the still more effective "the battery fluid is low" (which then suggests a concrete repair action). Which explanation is generated depends on the explanation goals and the situation. For example, the explanation goal of "explaine to effect a repair" produces the explanation of a dead battery while a situational feature such as "there is not much time left to catch the flight" curtails the explanation effort.

5 RELATED WORK

Previous work [Kedar-Cabelli, 1987] has considered how to generate target concepts for explanations from the need to select artifacts to use in plans. In GOBIE the explanation process is monitored in order to possibly refine the target concept as explanation proceeds, based on information gathered during explanation. The use of explicit goals to direct the processing of our system is in the spirit of [Krulwich et al., 1990] and the more recent work on goal directed diagnosis (GDD) [Rymon, 1993]. Our approach differs from theirs in that we use a multi-strategic framework and interaction with the environment is an integral part of our model. Also, in other approaches a single hypothesis is examined until it is either accepted or ruled out. Our approach involves the parallel utility-based pursuit of the alternative possibilities. The process of interleaving reasoning and action during interaction with an expert in novice learning has been studied in [Redmond, 1992]. The use of a memory of previous diagnosis stored as schemas to aid diagnostic reasoning and making context sensitive choices of what to diagnose first has been examined in [Turner, 1994] and [Stern and Luger, 1992].

6 CONCLUSIONS

In GOBIE explanation is modeled as a goal-driven activity with the aim of serving the over-arching process that initiates the explanation effort. When the explanation process is initiated in real world contexts the explainer often lacks sufficient information to form an explanation. Our model enhances the process with the ability to reason about its information needs, to actively seek needed information and to be sensitive to changing circumstances. In the pursuit of such information the system uses the utility of the possible actions to determine which candidate to explore further and how deep to examine a candidate. By situating this process in the environment in which the explanation will be used explanation formation becomes a multi-strategy process inter-weaving reasoning with action. Our current implementation is a step towards a richer model of explanation as a situated, goal-driven, utility-based, hierarchical process.

References

[Engelmore and Morgan, 1988] Engelmore, R. and Morgan, T. (1988). *Blackboard systems*. Addison-Wesley, Reading, Massachuset.

[Fikes and Nilsson, 1971] Fikes, R. E. and Nilsson, N. J. (1971). Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:189–208.

[Hammond and Seifert, 1994] Hammond, K. and Seifert, C. (1994). Opportunistic memory. In Schank, R. C. and Langer, E., editors, *Beliefs, reasoning, and decision making : psycho-logic in honor of Bob Abelson*, chapter 4, pages 111–142. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

[Hunter, 1990] Hunter, L. (1990). Planning to learn. In *12th Annual Conference of the Cognitive Science Society, Cambridge, Massachusetts*, pages 261–268. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

[Kedar-Cabelli, 1987] Kedar-Cabelli, S. (1987). Formulating concepts according to purpose. In *Proceedings of the Sixth Annual National Conference on Artificial Intelligence*, pages 477–481, Seattle, WA. AAAI.

[Kolodner, 1994] Kolodner, J. (1994). *Case-Based reasoning*. Morgan Kaufmann.

[Krulwich et al., 1990] Krulwich, B., Birnbaum, L., and Collins, G. (1990). Goal-directed diagnosis of expectation failures. In O'Rorke, P., editor, *Working Notes of the 1990 Spring Symposium on Automated Abduction.*, pages 116–119. AAAI.

[Leake, 1992] Leake, D. B. (1992). *Evaluating Explanations: A content theory*. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

[Leake, 1994] Leake, D. B. (1994). Issues in goal-driven explanation. In desJardins, M. and Ram, A., editors, *Working notes of the AAAI Spring Symposium in goal-driven learning*, pages 72–79.

[Pryor and Collins, 1991] Pryor, L. and Collins, G. (1991). Information gathering as a planning task. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, pages 862–866, Chicago, IL. Cognitive Science Society.

[Ram and Leake, 1994] Ram, A. and Leake, D. B. (1994). *Learning, goals and learning goals*. MIT press. (in press).

[Redmond, 1992] Redmond, M. A. (1992). *Learning by observing and understanding expert problem solving*. PhD thesis, Georgia Institute of Technology. (GIT-CC-92/43).

[Rymon, 1993] Rymon, R. (1993). *Diagnostic Reasoning and Planning in Exploratory-Corrective Domains*. PhD thesis, University of Pennsylvania.

[Schank, 1986] Schank, R. C. (1986). *Explanation Patterns: Understanding Mechanically and Creatively*. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

[Schank et al., 1994] Schank, R. C., Kass, A., and Riesbeck, C. K. (1994). *Inside Case-Based Explanation*. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

[Schank and Leake, 1989] Schank, R. C. and Leake, D. B. (1989). Creativity and learning in a case-based explainer. *Artificial Intelligence*, 40((1-3)):353–385. Also in Carbonell, J., editor, *Machine Learning: Paradigms and Methods*, MIT Press, Cambridge, MA, 1990.

[Stern and Luger, 1992] Stern, C. and Luger, G. (1992). A model for abductive problem-solving based on explanations templates and lazy evaluation. *International Journal of Expert Systems*, 5(3):249–265.

[Turner, 1994] Turner, R. M. (1994). *Adaptive reasoning for real-world problems: A schema-based approach*. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.