Reasoning about Evidence in Causal Explanations

Phyllis Koton*
MIT Lab for Computer Science
545 Technology Square
Cambridge, MA 02139
elan@OZ.AI.MIT.EDU

Abstract

Causal models can provide richly-detailed knowledge bases for producing explanations about behaviors in many domains, a task often termed interpretation or diagnosis. However, producing a causal explanation from the model can be time-consuming. This paper describes a system that solves a new problem by recalling a previous, similar problem and modifying its solution to fit the current problem. Because it is unlikely that any new problem will exactly match a previous one, the system evaluates differences between the problems using a set of *evidence principles* that allow the system to reason about such concepts as alternate lines of evidence, additional supporting evidence, and inconsistent evidence. If all differences between the new situation and the remembered situation are found to be insignificant, the previous causal explanation is adapted to fit the new case. This technique results in the same solution, but with an average of two orders of magnitude less effort. The evidence principles are domain independent, and the information necessary to apply them to other domain models is described.

1 Explanation transfer

Causal models are frequently proposed for knowledge-based systems because they have a wide range of applicability, they are robust, and they contain detailed information for providing explanations of their reasoning. In practice they are not widely used because they are inefficient compared to associational rules or other types of "compiled" knowledge. This inefficiency could be reduced by using a paradigm such as Case-Based Reasoning [Kolodner, 1985], which uses a memory of previously-solved problems to avoid unnecessarily reproducing complex reasoning. When presented with a new problem, case-based reasoning programs recall a similar problem and adapt its solution to the new case. However, the match between a new problem and a previously solved problem usually is only partial. This presents a difficulty when producing a causal explanation. Some feature of the previously-solved problem that was used as evidence in the causal explanation may be absent from the new problem. Similarly, the new problem may exhibit features that are absent from the previously-solved problem and which must be explained. This requires that the program have a set of principles for reasoning about dependencies between pieces of evidence and the states that they support in the causal explanation, and about the relationships (such as equivalence or incompatibility) between different pieces of evidence. The program can then determine whether a new problem with a somewhat different set of features can still be explained by a previous causal explanation. I have developed and implemented such a set of principles in a new system, CASEY.

*The work reported here has been supported in part by National Institutes of Health grants R01 LM 04493 from the National Library of Medicine and R01 HL 33041 from the National Heart, Lung, and Blood Institute.

2 The causal model and causal explanation

CASEY integrates case-based and causal reasoning techniques with a model-based expert system for managing patients with cardiac disease, the Heart Failure program [Long *et al.*, 1987]. The building blocks of the Heart Failure model are *measures*, *measure values*, and *states*. Measures correspond to observable features, such as heart rate, or laboratory results. Measure values are the input values of the measures, for example, "68" for the patient's heart rate, and are entered by the user. The combination of a measure and a measure value is referred to as a *finding*. States can represent three types of information: specific qualitative assessments of physiological parameters, for example HIGH LEFT ATRIAL PRESSURE; the presence of diseases ("diagnosis" states), for example PERICARDITIS; and therapies given to the patient, for example NITROGLYCERIN. The model recognizes two kinds of relationships. It can indicate that one state causes another with a given probability. It can also indicate that a state is associated with a particular finding with a given probability.

The Heart Failure program produces a *causal explanation*, represented as a graph, consisting of a set of measures, states, and directed links. The causal explanation describes the relationship between findings and the states in the model which cause them. A link between two states, or a state and a measure, indicates that one causes the other. Only abnormal findings are explained, but the program may not be able to explain all the abnormal findings. The causal explanation is derived through a complicated process which involves causal, probabilistic, and heuristic reasoning.

A graphical representation of the Heart Failure program's causal explanation for a patient, David, is shown in Figure 1. David was diagnosed as having aortic stenosis and unstable angina. The causal explanation illustrates how his symptoms (unstable anginal chest pain, evidence

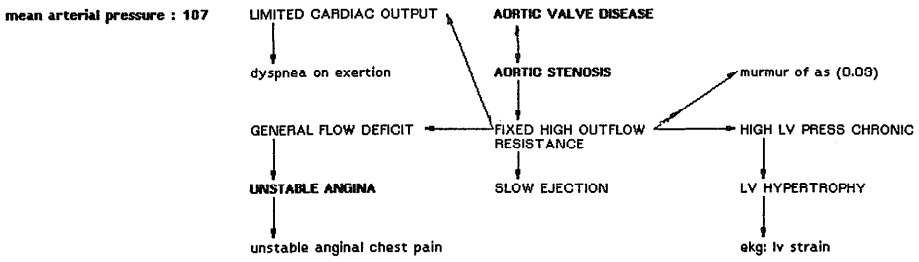

Figure 1: Causal explanation produced by the Heart Failure program for David.

of LV strain on EKG, a heart murmur, and dyspnea on exertion) are caused by these diseases. David also has a high arterial pressure, but it is not explained.

3 Overview of CASEY

CASEY contains a self-organizing memory system [Kolodner, 1983a] for storing previously-seen problems (called *cases*). The memory contains descriptions of patients the program has seen and generalizations derived from similarities between the patients. The patient description is comprised of *features*. These include both input data, such as symptoms, test results, and medical history, and solution data, such as the causal explanation for the patient's symptoms, therapy recommendations and outcome information. A typical patient description presented to CASEY contains about 40 input features. For example, Table 1 shows a fragment of the patient description for Newman, a new patient presented to the system.

CASEY produces the same causal explanation for a new patient as the Heart Failure program, but does so differently, using a five-step process.

Retrieval. CASEY finds a case similar to the new patient in its case memory.

Justification. CASEY evaluates the significance of any differences between the new case and the retrieved case using information in the Heart Failure model. If significant differences are found, the match is invalidated. If all differences between the new case and the retrieved case are judged insignificant or if the solution can be repaired to account for them, the match is said to be *justified*.

Adaptation. If none of the differences invalidate the match, CASEY adapts the retrieved solution to fit the new case. If all matches are ruled out, or if no similar previous case is found, CASEY uses the Heart Failure program to produce a solution for the case *de novo*.

Storage. The new case and its solution are stored in CASEY's memory for use in future problem solving. The user has the option of rejecting CASEY's solution, in which case Heart Failure program is used to produce a causal explanation, which is then stored in memory.

Feature evaluation. Those features that were causally important in the solution of this problem are noted in the memory.

This paper focuses on evaluating and adapting a retrieved solution.

David's case is retrieved as a precedent for the new patient Newman.[1] The cases of David and Newman have many similarities.They also have some differences, which are shown in Table 2. CASEY must decide if these differences are serious enough to rule out the match. CASEY analyzes the significance of differences between patients using information about the cardiovascular system contained in the Heart Failure program's model. In this example, the first four differences are insignificant. The remaining differences are important but the precedent solution can be adapted to explain them. This will be shown in section 4.

4 Principles for reasoning about evidence

Most new cases will not exactly match any previous case in the memory. To allow partial matches, differences between cases must be evaluated. Two cases might have many similar features yet have one critical difference that invalidates the match. Alternatively, many differences may not affect the validity of a match. The difference may have no relation to the solution of the case (the patient's name, for example) or the difference may be explainable. The module in CASEY that performs this evaluation is called the *justifier* because it must justify using a retrieved case as a precedent for the new case. The justifier relies on a set of principles for reasoning about evidence, termed *evidence principles*, that are presented below. There are two basic types of differences that must be evaluated: (1) evidence that supported a state in the previous case's explanation might be missing in the new case, and (2) the new case might contain additional symptoms that must be accounted for. These differences are handled by the first five evidence principles, which attempt to show that the difference in question is insignificant or repairable. The last three evidence principles handle features which have special values that are easy to reason about.

1. *Rule out.* A state is ruled out from the transferred solution if there is some feature in the new case which is incompatible with that state. This is detected when the feature has zero probability for some state in the

[1] CASEY will often retrieve more than one case that matches the new case. CASEY's method of choosing among retrieved cases is described in [Koton, 1988].

```
(defpatient "Newman"
HISTORY
(age . 65)
(sex male)
(dyspnea on-exertion)
(orthopnea absent)
(chest-pain anginal)
(anginal within-hours unstable)
(syncope/near-syncope on-exertion)
(cough absent)
LABORATORY-FINDINGS
(ekg lvh normal-sinus)
(cxr calcification)
(calcification mitral aortic-valve))

VITAL-SIGNS
(blood-pressure 138 80)
(heart-rate . 90)
(arrhythmia-monitoring normal)
(resp . 20)
(temp . 98.4)
PHYSICAL-EXAM
(abdomen normal-exam)
(pulse slow-rise)
(extremities normal-exam)
```

Table 1: Patient description for Newman

Feature name	Value for David	Value for Newman
age	72	65
pulse-rate	96	90
temperature	98.7	98.4
orthostatic-change	absent	unknown
angina	unstable	within-hours & unstable
mean-arterial-pressure	107	99.3
syncope	none	on exertion
auscultation	murmur of AS	unknown
pulse	normal	slow-rise
ekg	normal sinus & lv strain	normal sinus & lvh
calcification	none	mitral & aortic

Table 2: Differences between patients David and Newman.

retrieved solution. For example, a heart rate of 40 beats per minute is incompatible with the state HIGH HEART RATE.

2. *Other evidence.* This is used when a feature present in the retrieved case is missing in the new case. It tries to find a piece of evidence in the new case which supports the state that the missing feature supported.
3. *Unrelated oldcase feature.* This is used when a feature is present only in the retrieved case. If the feature was not used in the causal explanation, its absence has no effect on any states in the explanation, so it can be ignored.
4. *Supports existing state.* This principle is used when a feature is present in the new case but not in the retrieved case. It tries to attribute this feature to a state in the retrieved causal explanation.
5. *Unrelated newcase feature.* This is also used when a feature is present only in the new case. If the feature is abnormal, but is not evidence for any existing state and does not strongly suggest a new state, then add it to the explanation as an unexplained feature.
6. *Normal.* Normal values are not explained by the Heart Failure program, so a normal value in the new case does not need to be explained. (Note that if a model did reason using normal values, this rule would be eliminated).
7. *No information.* If there is no information given about a feature in one of the cases and it is known to be absent in the other case, then assume that it is also absent in the former case.
8. *Same qualitative region.* CASEY evaluates differences between features with numerical values by translating them into qualitative value regions. For example, a blood pressure of 180/100 becomes "high blood pressure." Features whose values fall into the same qualitative region are judged not to be significantly different. The regions are determined using range information for the corresponding measure in the Heart Failure model.

CASEY can reject a match on either of two grounds: a significant difference could not be explained, or all the diagnosis states in the retrieved solution were ruled out. If all differences between the new case and the retrieved case are insignificant or repairable, then the transfer of solutions from the precedent to the current case proceeds.

Some of the inferences about the differences between patients David and Newman that CASEY makes are:

- Both patients' heart-rates are in the same qualitative region (moderately high heart rate) so the difference is considered insignificant.
- David's mean arterial pressure is high, but Newman's is not. However, this feature was not accounted for by the causal explanation, so it is judged insignificant by the rule *unrelated oldcase feature.* Newman's mean arterial pressure is normal, so it does not have to be explained.

- Orthostatic change is absent in David, but not specified for Newman, so the rule *no information* concludes that it is also absent in Newman.
- Newman's finding of angina within hours is additional evidence for the state UNSTABLE ANGINA. His finding of syncope on exertion is additional evidence for the state LIMITED CARDIAC OUTPUT.
- Murmur of AS, which is absent in Newman, is evidence for the state FIXED HIGH OUTFLOW RESISTANCE in the precedent solution, but this state has other evidence supporting it in the new case.
- LV strain on David's EKG is evidence for the state LV HYPERTROPHY. Newman's EKG shows LVH, which is evidence for the same state.
- Newman's finding of aortic calcification is evidence for only one state AORTIC VALVE DISEASE, so this state is added to the causal explanation, and similar reasoning applies to the symptom of mitral valve calcification.

All the differences between David and Newman are insignificant or repairable, so the match is justified.

5 Adapting the solution

CASEY uses *repair strategies* to adapt a previous solution to a new case. Associated with each type of repairable difference detected by the evidence principles is an explanation repair strategy which modifies the precedent causal explanation to fit the new case. Repair strategies modify the transferred causal explanation by adding or removing nodes and links. CASEY makes seven types of repairs:

1. *Remove state.* This strategy can be invoked in two circumstances: either the state is known to be false, or all of the evidence that previously supported the state has been removed (the removed evidence could be either features missing in the new patient, or states ruled out during justification). In the first case, this strategy is invoked by the *rule out* evidence principle. In the second case, when all the evidence for a state is missing in the new case, or if the only cause of a state has been removed from the transferred causal explanation, CASEY removes that state from the explanation. CASEY also determines whether states caused by this state must now be removed.
2. *Remove evidence.* This repair strategy is invoked by the principles *other evidence* and *unrelated oldcase feature.* When a piece of evidence that was used in the retrieved case is absent in the new case, this removes the feature and any links to it.
3. *Add evidence.* This repair strategy is invoked by the principles *other evidence* and *supports existing state.* It adds a piece of evidence to the causal explanation, and links it to those states for which it is evidence.
4. *Substitute evidence* is invoked by the *same qualitative value* principle. When two numerical values have the same qualitative value, this repair strategy replaces the old value with the new value as evidence for some state.
5. *Add state.* The only time CASEY adds a state to the causal explanation is when the feature it is attempting to explain has only one cause. This repair strategy is invoked by the principle *supports existing state*, because the fact that a feature has only one cause is discovered while CASEY is searching for existing states that cause this feature. When the evidence has only one possible cause, that state is added to the causal explanation. CASEY then tries to link it to existing states and features in the causal explanation (using *add link*).
6. *Add link* is invoked by the *add state* repair strategy, and is used to add a causal link between two states.
7. *Add measure* is invoked by *unrelated newcase feature.* This adds an abnormal feature which CASEY cannot link to the causal explanation.

Some of the repair strategies invoked by the justifier in order to adapt the explanation transferred from David to fit the data for Newman are:

```
(substitute-evidence hr:90  hr:96)
(remove-evidence mean-arterial-pressure:107)
(add-evidence within-hours unstable-angina)
(add-evidence syncope-on-exertion
        limited-cardiac-output)
(remove-evidence murmur-of-as)
(remove-evidence lv-strain)
(add-evidence lvh lv-hypertrophy)
(add-state aortic-valve-disease)
(add-evidence aortic-calcification
        aortic-valve-disease)
(add-state mitral-valve-disease)
(add-evidence mitral-calcification
        mitral-valve-disease)
```

CASEY's causal explanation for Newman is identical to the solution produced by the Heart Failure program. However, CASEY examined 674 states in the model to obtain this solution, while the Heart Failure program examined approximately 76,000 states.

6 Results

CASEY's performance was evaluated on two counts: *efficiency*, and *quality* of the solution. The program was tested on a set of 45 patients with symptoms of heart failure covering about 15 different diseases.

The quality of CASEY's solution was evaluated by comparing its explanation to the Heart Failure program's explanation for the same patient. A solution was considered *successful* if it was identical to the Heart Failure program's solution. A solution was considered *satisfactory* if it was identical to the Heart Failure program's solution except for the features which CASEY could not explain. In these latter cases, CASEY had already performed most of the task of deriving the causal explanation, and the Heart Failure program could be used to incrementally account for the remaining features. CASEY produced a solution that was either successful or satisfactory for 86% of the test cases for which there was a similar case in its memory. CASEY produced a solution identical to the Heart Failure program's solution in 14 out of the 45 test cases. It produced a satisfactory explanation for an additional 18 test cases. It

gave up on six of the test cases, and produced an incorrect causal explanation for seven test cases. An examination of the test cases for which CASEY failed to reproduce even part of the Heart failure program's solution revealed that each one of these cases had a causal explanation that was completely different from any other patient in the memory. Even on these cases, CASEY could often produce part of the causal explanation, but could not account for the combination of features seen in the patient.

CASEY's efficiency was evaluated by comparing the number of states (of the Heart Failure program) it examined to the number states examined when the Heart Failure program solved the same problem. CASEY always examined fewer states than the Heart Failure program by *at least an order of magnitude, and often by two or three orders of magnitude.* Cases that required relatively more effort by CASEY to solve did not necessarily correspond to cases that the Heart Failure program required a lot of effort to solve. Problems that can be solved quickly by the Heart Failure program have features which are specific to only one (or a small number) of states. Problems that require a lot of effort for the Heart Failure program are those with many symptoms that are evidence for a large number of states, which generate a large number of possible explanations that must be evaluated. By contrast, a simple case for CASEY is one in which there are few differences between the precedent and the new case. A difficult case for CASEY is one in which many differences between the precedent and the new case must be analyzed. A consequence of this difference is that as the number of cases solved by CASEY increases, it requires less effort to solve subsequent cases because it is more likely to find a close match. The Heart Failure program, conversely, cannot increase its efficiency except by re-implementation.

7 Discussion

7.1 Related work

Retrieving, adapting, and storing cases are standard procedures of a case-based reasoner. CASEY differs from previous case-based reasoning systems because it incorporates reasoning from its causal model in each of these steps.

Most case-based reasoning systems use a fixed and often *a priori* ranking that indicates which features of a new case are important for matching against cases in the memory (e.g., [Bain, 1986], [Hammond, 1986], [Simpson, 1985]). It is not always possible to determine in advance which features are going to be important, and furthermore, the important features may vary from case to case. CASEY therefore matches a new case against cases in its memory using every feature in the patient description. Using knowledge of which features were important in determining the causal explanation of previous cases, CASEY then determines the important features of the new case, and gives these features greater weight for matching.

During justification, model-based reasoning is used to judge the significance of differences between the new and previous cases. Because the match between a new problem and a previously solved problem usually is only partial, there may be differences between the two cases that preclude using even a modified version of a retrieved solution for a new problem. The justification step proves that a retrieved solution can be supported by the features of the new problem.

Feature evaluation uses the causal explanation of the new case to determine its important features. These are then recorded as part of the case's representation in memory. Determining which features of the new problem were important to the solution helps the program make better matches in the future, because it allows the program to distinguish between extraneous and important features.

7.2 Generalizing the results

The evidence principles do not depend in any way on the specific domain information in the model. The evidence principles do depend on the *form* of the model, namely a causal inference network. In order to use the evidence principles, a model must provide the following information:
$\mathcal{S}$, a finite set of *states.*
$\mathcal{F}$, a finite set of *features* which can be evidence for the states in $\mathcal{S}$. $f \in \mathcal{F}$ is what up till now has been referred to as a feature-value pair.
$C \subseteq (\mathcal{S} \times \mathcal{F}) \cup (\mathcal{S} \times \mathcal{S})$. The relation C in the Heart Failure model is used to imply causality. In fact, it is not even necessary that the relation be causal. For CASEY's evidence principles it is sufficient that $(s, f) \in C$ is associational and s temporally precede f (similarly for $(s_1, s_2) \in C$).
The problem presented to CASEY is then:
$\mathcal{F}^+ \subseteq \mathcal{F}$, some subset of the features which has been observed.

The ability of the technique to produce a meaningful solution depends on selecting a good precedent case. Much research has been done in this area (for example [Kolodner, 1983b], [Simpson, 1985], [Ashley and Rissland, 1987], [Kass *et al.*, 1986]). CASEY uses a novel matching algorithm specifically designed for reasoning about causal explanations. This algorithm gives extra importance to features that played a role in the causal explanation of previous similar cases [Koton, 1988].

7.3 Limitations of the method

CASEY's current implementation has limitations. Some problems presented to the system have a large number of "reasonable" explanations. CASEY does not use all the quantitative information available in the Heart Failure model that would allow it to distinguish between statistically more- and less-likely solutions. For certain applications (e.g. geological interpretation [Simmons and Davis, 1987]), any explanation for the input features is acceptable. In the Heart Failure domain, the users require the most likely explanation. CASEY's justifier will soon be extended to recognize when the solution it is creating is not the most likely one, in which case it can reject the match.

8 Conclusions

CASEY *integrates* associational reasoning, model-based reasoning, and learning techniques in a program which is efficient, can learn from its experiences, and solves commonly-seen problems quickly, while maintaining the ability to reason using a detailed knowledge of the domain when necessary. Furthermore, the methods used by the system are domain-independent and should be generally applicable in other domains with models of a similar form.

Acknowledgments

Robert Jayes, MD kindly provided the example cases. William Long's Heart Failure program provided an excellent testbed for this work. Thanks to Peter Szolovits, Ramesh Patil, and William Long for their supervision of this research, and to Janet Kolodner for her helpful comments on this paper.

References

[Ashley and Rissland, 1987] Kevin D. Ashley and Edwina L. Rissland. Compare and contrast, a test of expertise. In *Proceedings of the National Conference on Artificial Intelligence*. American Association for Artificial Intelligence, 1987.

[Bain, 1986] William M. Bain. A case-based reasoning system for subjective assessment. In *Proceedings of the National Conference on Artificial Intelligence*, pages 523–527. American Association for Artificial Intelligence, 1986.

[Hammond, 1986] Kristian Hammond. *Case-based Planning: An Integrated Theory of Planning, Learning and Memory.* PhD thesis, Yale University, 1986.

[Kass *et al.*, 1986] A. M. Kass, D. B. Leake, and C. C. Owens. Swale: A program that explains. In *Explanation Patterns: Understanding Mechanically and Creatively.* Lawrence Erlbaum Associates, Hillside, NJ, 1986.

[Kolodner, 1983a] Janet L. Kolodner. Maintaining organization in a dymanic long-term memory. *Cognitive Science*, 7:243–280, 1983.

[Kolodner, 1983b] Janet L. Kolodner. Reconstructive memory: A computer model. *Cognitive Science*, 7:281–328, 1983.

[Kolodner, 1985] Janet L. Kolodner. Experiential processes in natural problem solving. Technical Report GIT-ICS-85/22, School of Information and Computer Science, Georgia Institute of Technology, 1985.

[Koton, 1988] Phyllis A. Koton. *Using Experience in Learning and Problem Solving.* PhD thesis, Massachussetts Institute of Technology, 1988.

[Long *et al.*, 1987] William J. Long, Shapur Naimi, M. G. Criscitiello, and Robert Jayes. The development and use of a causal model for reasoning about heart failure. In *Symposium on Computer Applications in Medical Care*, pages 30–36. IEEE, November 1987.

[Simmons and Davis, 1987] Reid Simmons and Randall Davis. Generate, test, and debug: Combining associational rules and causal models. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pages 1071–1078, 1987.

[Simpson, 1985] Robert L. Simpson. A computer model of case-based reasoning in problem solving: An investigation in the domain of dispute mediation. Technical Report GIT-ICS-85/18, Georgia Institute of Technology, 1985.