# Two Paradigms of Complementary CBR Integrations[1]

**L. Karl Branting**
Department of Computer Science
University of Wyoming
Laramie, Wyoming 82071-3682
karl@uwyo.edu

## Introduction

A wide variety of multiple-paradigm reasoning systems have been developed in recent years. Case-based reasoning (CBR) is very frequently one of the reasoning paradigms in such integrations. Three justifications for the integration of multiple reasoning paradigms can be distinguished.

First, the domain may be characterized by multiple knowledge representations. Reasons for multiple knowledge representations include the following:

- Institutional. Law typifies a domain with multiple distinct knowledge sources—statutes and precedents—that are not reducible to one another.
- Cognitive. In some domains humans typically reason with multiple representations, e.g., pictures, text, and equations. Modeling human problem-solving behavior in these domain requires using these multiple representations as well.
- Computational. A single domain may give rise to multiple representations specialized for various distinct subtasks.
- Domain theory incompleteness. A domain theory may consist of several distinct knowledge sources. If the domain theory is incomplete, it may be unclear how to reduce these multiple knowledge sources to a single representation.

A second justification is to improve computational efficiency. For example, search can often be reduced by caching macros or cases so that previous search episodes can be reused. Conversely, an induction algorithm may be applied to induce general rules from cases. However, it is sometimes desirable to retain the cases as well as the induced rules.

[1]This research was supported in part by NSF Career Grant IRI-9502152.

The third justification is to improve problem-solving completeness or accuracy. In many domains, no single knowledge source or problem-solving method is individual sufficient for accurate and efficient problem solving. Moreover, explanations in some domains require reference to multiple methods or knowledge sources. In these cases, the problem-solving methods are *complementary* in the since that each component compensates work a weakness of the other.

This paper describes two general paradigms of complementary integrations involving CBR. In the first, *approximate-model-based adaptation*, cases and models represent opposite extremes on a continuum of possible tradeoffs between certainty and generality. In the second, integration of rules and cases in weak-theory domains, both rules and cases reduce the uncertainty in the other's applicability conditions. Several systems representative of each paradigm will be described.

## Approximate-model-based adaptation

Model-based adaptation (MBA) consists of using case-based reasoning to find an approximate solution and model-based reasoning to adapt this approximate solution into a more precise solution. The first systems using model-based adaptation assumed the existence of a perfect domain model, e.g., (Koton, 1988; Goel, 1991; Bhatta and Goel, 1996). By contrast, a more recent form of MBA is premised on the **absence** of a complete and correct domain model. In this approach, cases compensate for incompleteness in the model by providing a set of reference points with known solutions. Conversely, models compensate for insufficient case coverage by permitting the solutions associated with cases to be adapted to sufficiently similar situations. This form of MBA is termed *approximate-model-based adaptation.*

Approximate-model-based adaptation is suited for domains in which cases have high certainty but low generality, and models have low certainty but high generality. As argued in (Branting, 1998), many forms of

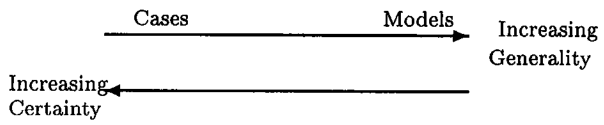


Figure 1: In domains suited for approximate-model-based adaptation cases have high certainty but low generality, and models have low certainty but high generality.

control planning for physical systems require use of cases and models with these characteristics. For example, models of many ecological, biological, chemical systems encountered in engineering, manufacturing, and agriculture, are incomplete, either because a complete state description for such systems cannot be determined or because the number and type of interactions between system elements are poorly understood. Moreover, while historical data often exist for such systems, they are often insufficient for accurate prediction and control using empirical methods. In such systems, both models and empirical data exist, but neither is *per se* sufficient for accurate control decisions. Accurate prediction and control of the behavior of such systems requires exploitation of multiple, individually incomplete, knowledge sources.

## Examples of Systems Using Approximate-Model-Based Adaptation

### CARMA

CARMA is an advisory system that helps ranchers and pest managers determine the most economical response to grasshopper infestations. Grassland pest management is typical of the physical-system control planning tasks described in the previous paragraph. The task is to select a set of control actions, such as applying chemical or biological control agents, to minimize expected forage consumption by pests. Numerical models of grasshopper physiology and life cycles and of grassland ecology have been developed by entomologists, but these models are incomplete and require more data than is typically available for management decisions (Lockwood and Lockwood, 1991). Some specific cases are available, but the absence of systematic record-keeping on rangeland pest infestations means that the number of cases is relatively small. Thus, neither the domain model nor empirical data are individually sufficient for accurate prediction.

The design of CARMA's forage consumption component was based on the hypothesis that an integration of model-based and case-based reasoning can lead to more accurate forage consumption predictions than the use of either technique individually. This hypothesis was based on the observation that neither the causal model nor the empirical data available for rangelands are individually sufficient for accurate prediction. To test this hypothesis, CARMA's empirical and model-based knowledge components were each tested in isolation and the results compared to the performance of the full CARMA prediction system.

In an empirical evaluation described in (Branting et al., 1997), the ability of CARMA's forage consumption module to duplicate the predictions of entomologists experienced in grasshopper management and ecology was compared to the following:

- CBR only: CARMA's forage consumption module with all model-based adaptation disabled.
- Other purely empirical methods
  - Decision-tree induction using ID3.
  - Linear regression.
- Model only: a numerical simulation based on CARMA's model of rangeland ecology.

The accuracy of each approach was tested using leave-one-out testing for the sets of predictions by each expert and for a data set consisting of the median of the predictions of individual experts on each case. The full CARMA prediction system was tested using both global adaptation weights (CARMA-global) and case-specific adaptation weights (CARMA-specific). The root-mean-squared error for each of the methods, set forth in Figure 2, provides initial confirmation of the hypothesis that integrating model-based and case-based reasoning through model-based adaptation leads to more accurate forage consumption predictions than the use of either technique individually.

### FormTool

FormTool is a system for selecting colorants for plastic coloring (Cheetham and Graf, 1997). A coloring theory, Kubelka-Munk color theory, can be used for this task, but it requires exhaustive search, fails to take into consideration all the important attributes of a color match, and typically yields a nonoptimal solution. While a large library number of formulae exists, it is very unlikely that a new color will precisely match an existing formula.

FormTool retrieves the color formula that most closely matches the a given sample using a fuzzy similarity metric. Kubelka-Munk color theory is then used to predict the effect of incremental changes in the retrieved formula. Adaptation consists of hill-climbing to reduce to difference in between the sample's color and the color of the adapted formula as predicted by Kubelka-Munk color theory.

| | CARMA | | Empirical Only | | | Model-Based Only |
|---|---|---|---|---|---|---|
| | Specific weights | Global weights | CBR only | ID3 | Linear regression | Numerical simulation |
| Expert sets | 13.3 | 14.2 | 21.1 | 34.9 | 25.6 | 29.6 |
| Median set | 9.7 | 10.0 | 22.8 | 35.2 | 11.9 | 28.8 |

Figure 2: Root-mean-squared errors (in %) for leave-one-out-test results.

As with rangeland pest management, colorant selection is a task for which there is a model which is powerful but not accurate enough to use independently and a large but insufficient collection of cases. The integration of these two knowledge sources, however, yields an effective system.

### Sophist

Sophist is a system for bioprocess recipe planner (Aarts and Rousu, 1996; Rousu and Aarts, 1996; Aarts and Rousu, 1997). Although many bioprocesses, such as beer brewing, have fairly well-understood models, these models are typically not precise enough to permit bioprocess planning through model-based reasoning alone. Numerous effective recipes are known, but new conditions require development of new recipes. Bioprocess recipe planning is therefore a physical-system control planning task. Although there are both extensive cases and a powerful model, these knowledge sources are individually insufficient and must therefore be integrated.

Sophist uses a domain model expressed in Qualitative Process Theory for case adaptation. Cases are indexed by a discrimination net. Cases are adapted by using a qualitative model to identify a set of *adaptation goals*, changes that would reduce the difference between the results of the retrieved case and the current specifications. The adaptation actions associated with each adaptation goal are then performed in order of expected benefits.

The three approximate-model-based adaptation systems are summarized in Figure 3.

## Integration of rules and cases domains with non-operational concepts

A second paradigm for complementary integration is the integration of rules and cases in domains characterized by non-operational concepts. This non-operationality typically arises from uncertainty about whether abstract terms are satisfied by particular collections of facts (Porter et al., 1990). In legal reasoning, this phenomenon is termed *open-texture* (Hart, 1958). For example, determining whether a rule for negligence liability applies in a given case requires determining whether there was a "failure to exercise reasonable care," but there may be considerable uncertainty concerning whether a particular set of facts satisfies the abstract concept "reasonable care." A wide range of human concepts are characterized by "polymorphy" (?) or "graded structure" (?) that necessitates use or prototypes or exemplars for categorization (?).

In domains with non-operational concepts there is uncertainty concerning the conditions for applicability of both rules and cases to the top-level goal. Uncertainty in the applicability conditions of rules can arise either from the absence of inference rules connecting abstract concepts to specific facts (i.e., "the rules run out" (Gardner, 1984)) or because rules are over-general (Golding and Rosenbloom, 1996). Uncertainty in the applicability conditions of cases arises from the absence of a well-defined relevance criterion capable of determining the importance of case differences (Ashley and Rissland, 1988).

A number of different approaches to complementary integration of rules and cases have been implemented. In one approach, typified by Protos (Bareiss et al., 1990),general knowledge in the form of rules is used in the assessment of case similarity. For example, background knowledge of causal connections between physiological conditions and symptoms could be used to to reason about whether the symptoms of a new case are consistent with the causal model that accounted for the symptoms of a precedent. This approach was termed *case elaboration* in (Branting and Porter, 1991).

A second approach uses rules to combine multiple case-based reasoning steps or reformulate a top-level goal in such a way as to improve case matching. This approach was termed *term reformulation* in (Branting and Porter, 1991). For example, CABARET (Skalak and Rissland, 1992) used a blackboard architecture together with a sophisticated set of heuristics for choosing among the rules and cases applicable each current goal. A related approach was used in ANAPRON, a system for proper-name pronunciation, in which cases represented exceptions to general pronunciation rules (Golding and Rosenbloom, 1996).

Several hybrid architectures, including GREBE (Branting and Porter, 1991) and EXPANDER (Walker, 1992), have been devised that permit both

| System | CARMA | FormTool | Sophist |
|---|---|---|---|
| Task | Forage loss prediction | Colorant recipe planning | Bioprocess recipe planning |
| Model | Grasshopper simulation + adaptation weights | Kubelka-Munk color theory | Qualitative Process Theory |
| Cases | Grassland infestations | Colorant recipes | Bioprocess recipes |

Figure 3: Three systems for approximate-model-based adaptation.

CBR as a subgoal of rule-based reasoning and rule-based reasoning to assist case matching at any stage of the problem-solving process.

All these systems are characterized by a domain theory in which neither rules nor cases are individually sufficient. This insufficiency arises because of the absence of either a unique, correct rule for the top-level goal or of a unique case that clearly matches the current facts. Under these circumstances, the most persuasive and likely result requires reasoning with both rules and cases.

## Summary

This paper has described two general paradigms of complementary integrations involving CBR: approximate-model-based adaptation, and integration of rules and cases in domains with non-operational concepts. Approximate-model-based adaptation is appropriate when cases have high certainty but low generality, and models have low certainty but high generality. Control planning tasks for physical systems, such as ecological, biological, chemical systems, often give rise to these conditions.

Techniques for integrating of rules and cases, such as case elaboration and term reformulation, are appropriate in weak-theory domains in which both rules and cases can provide only uncertain inferences. A number of systems using these techniques have been implemented, including systems in law, medicine, and pronunciation.

## References

Aarts, R. and Rousu, J. (1996). Toward cbr for bioprocess planning. In *Proceedings of the Third European Workshop on Case-Based Reasoning (EWCR-96)*, pages 16–27, Lausanne, Switzerland.

Aarts, R. and Rousu, J. (1997). Qualitative knowledge to support reasoning about cases. In *Proceedings of the Second International Conference on Case-Based Reasoning*, pages 489–498, Providence, Rhode Island. Springer.

Ashley, K. and Rissland, E. (1988). Waiting on weighting: A symbolic least commitment approach. In *Proceedings of AAAI-88*, Minneapolis. American Association for Artificial Intelligence.

Bareiss, E. R., Porter, B. W., and Wier, C. C. (1990). Protos: An exemplar-based learning apprentice. In Kodratoff, Y. and Michalski, R., editors, *Machine Learning*, volume 3, pages 112–139. Morgan Kaufmann Publishers, Inc., San Mateo, California.

Bhatta, S. and Goel, A. (1996). From design experiences to generic mechanisms: Model-based learning in analogical design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 10:131–136.

Branting, K. (1998). Integrating cases and models through approximate-model-based adaptation. In *Proceedings of the AAAI 1998 Spring Symposium on Multimodal Reasoning, Palo Alto, California, March 23–25.*

Branting, K., Hastings, J., and Lockwood, J. (1997). Integrating cases and models for prediction in biological systems. *AI Applications*, 11(1):29–48.

Branting, K. and Porter, B. W. (1991). Rules and precedents as complementary warrants. In *Proceedings of Ninth National Conference on Artificial Intelligence*, Anaheim. AAAI Press/MIT Press, Menlo Park, California.

Cheetham, W. and Graf, J. (1997). Case-based reasoning in color matching. In *Proceedings of the Second International Conference on Case-Based Reasoning*, pages 1–12, Providence, Rhode Island. Springer.

Gardner, A. (1984). *An Artificial Intelligence Approach to Legal Reasoning.* PhD thesis, Stanford University.

Goel, A. (1991). A model-based approach to case adaptation. In *Thirteenth Annual Conference of the Cognitive Science Society*, pages 143–148.

Golding, A. and Rosenbloom, P. (1996). Improving accuracy by combining rule-based and case-based reasoning. *Artificial Intelligence*, 87(1–2):215–254.

Hart, H. L. A. (1958). Positivism and the separation of law and morals. *Harvard Law Review*, 71:593–629.

Koton, P. (1988). *Using Experience in Learning and Problem Solving.* PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts. Department of Electrical Engineering and Computer Science.

Lockwood, J. and Lockwood, D. (1991). Rangeland grasshopper (Orthoptera: Acrididae) population dynamics: Insights from catastrophe theory. *Environmental Entomology*, 20:970–980.

Porter, B. W., Bareiss, E. R., and Holte, R. C. (1990). Concept learning and heuristic classification in weak-theory domains. *Artificial Intelligence*, 45(1–2).

Rousu, J. and Aarts, R. (1996). Adaptation cost as a criterion for solution evaluation. In *Proceedings of the Third European Workshop on Case-Based Reasoning (EWCR-96)*, pages 354–361, Lausanne, Switzerland.

Skalak, D. and Rissland, E. (1992). Arguments and cases: An inevitable intertwining. *Law and Artificial Intelligence*, 1(1).

Walker, R. (1992). *An Expert System Architecture for Heterogeneous Domains.* PhD thesis, Vrije University.

## Appendix

1. Integration name/category
   (a) CARMA, FormTool, and Sophist
   (b) GREBE, CABARET, Protos (and others)
2. Performance Task:
   (a) Control planning tasks for physical system
   (b) Argument generation
3. Integration Objective:
   (a) Cases compensate for low certainty of model; model compensates for low generality of cases.
   (b) Cases operationalize rule antecedents; rules assist case matching
4. Reasoning Components
   (a) Cases + model used for adaptation
   (b) Cases + rules
5. Control Architecture
   (a) CBR as master
   (b) Complete reciprocity between rules and cases (in GREBE, CABARET, and EXPANDER)
6. CBR Cycle Step(s) Supported:
   (a) Adaptation
   (b) Selection of the goal to which CBR is applied, determining degree of match.
7. Representations
   (a) CARMA: Grassland/Grasshopper population dynamics model, adaptation weights (determined through wrapper procedure), match weights (determined through information gain), feature inference rules
   (b) GREBE: Legal and common-sense rules, taxonomic hierarchies, semantic network representation of cases
8. Integration Status
   (a) Empirically evaluated (CARMA), fielded (CARMA, FormTool, Sophist)
   (b) Empirically evaluated (GREBE)
9. Priority future work
   (a) Determine range of applicability
   (b) Develop knowledge-acquisition tools for domains with complex concept structure