Integrating Cases and Models Through Approximate-Model-Based Adaptation[1]

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

Introduction

This paper describes an approach to the integration of case and models called *model-based adaptation* (MBA). Under this approach, case-based reasoning is used to find an approximate solution, and model-based reasoning is then used to adapt this approximate solution into a more precise solution. This paper distinguishes two distinct forms of MBA—*perfect-model-based adaptation* and *approximate-model-based adaptation*—sets forth the applicability conditions and potential benefits of each, and describes three systems that use approximate-model-based adaptation.

Forms of Case-Based Reasoning

"Case-based reasoning" (CBR) refers to two related but distinct problem-solving paradigms: (1) reasoning with exemplars; and (2) reuse of sequences of problem-solving operators or configurations. The first CBR paradigm is typically used in domains having an incomplete domain theory, such as medicine (Porter et al., 1990) and law (Ashley, 1990; Branting and Porter, 1991). The underlying task is typically some form of classification, *e.g.,* selecting the appropriate diagnostic category or determining the applicability of a legal predicate. Although there may be extensive inference in the process of matching, *e.g.,* as in Protos (Porter et al., 1990), there is typically little adaptation, because classification does not lend itself to adaptation.

The second CBR paradigm has typically been used to reduce search, *e.g.,* (Veloso, 1992; Bergmann and Wilke, 1996; Koton, 1988). The underlying task is typically a configuration task, such as STRIPS planning (Veloso, 1992), fabrication planning (Bergmann and Wilke, 1995), or constructing a causal model that explains a set of symptoms (Koton, 1988). Typically, this second form of CBR assumes a perfect domain theory and attempts to reduce the computational expense of using this theory *ab initio*.

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

Model-based adaptation

Model-based adaptation 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. The purpose of CBR in these systems is to reduce the high computational costs associated with a purely model-based approach. For example, Casey (Koton, 1988) assigned a causal explanation and diagnosis to patients based on the explainable similarity of their symptoms to those of previous cardiac cases. Casey used causal rules derived from the Heart Failure program, a model-based expert system for cardiac disorder diagnosis, to reason about whether the symptoms of a new case were consistent with the causal model that accounted for the symptoms of a precedent. The underlying Heart Failure model generated the same solutions as Casey. Casey's benefit consisted entirely of reducing computation time.

Similarly, KRITIK (Goel, 1991) uses model-based case adaptation for design. In KRITIK, each case is associated with a qualitative model that expresses how a solution satisfies a set of constraints. If the function of a retrieved design case differs from the desired function, KRITIK retrieves an abstract plan for generating appropriate structural modifications. The abstract plan is then instantiated with the specific facts of the case to be adapted and applied to generate candidate structural modifications.[1]

The primary purpose for CBR in this form of MBA is to reduce search by substituting case retrieval and search in the portion of the problem space surrounding the case for a more expensive *a initio* search (although a secondary purpose may be that this approach has greater cognitive verisimilitude and is therefore more comprehensible to users). Since this approach is predicated upon a complete and accurate model of the domain, it is termed *perfect-model-based adaptation.*

[1]A more recent application of structure-behavior-function models for adaptation is described in (Bhatta and Goel, 1996). Another approach to using functional knowledge for adaptation is described in (Sycara et al., 1992).

Perfect-model-based adaptation falls squarely within the second CBR paradigm, since the primary purpose of CBR is search reduction rather than compensating for an incomplete domain theory.

By contrast, the second 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 adaption represents a hybrid of the two main CBR paradigms in that it uses cases to compensate for an incomplete domain theory, as in paradigm 1, but the solutions associated with cases are often operator configurations, as in paradigm 2. Approximate-model-based adaption appears to be particularly well suited for tasks that require prediction of the behavior of complex physical systems.

Control Planning for Physical Systems

Many management and synthesis tasks require selecting a configuration of operators that makes the behavior of a physical system conform to a given set of specifications. This operator configuration may consist of a sequence of control actions or a set of parameter values. For example, decision-support in agriculture and natural resources management requires finding a set of management actions that optimizes some criteria, *e.g.,* maximizing crop production or minimizing expected forage consumption by pests. Similarly, bioprocess recipe planning requires finding a sequence of parameter values for a recipe that satisfies a given set of process specifications.

In systems for which a precise model exists and accurate values of state variables can be determined, simulation can be used to guide search for an appropriate operator configuration. Indeed, if the theory underlying the behavior of the physical system is sufficiently complete, no search is necessary, *e.g.,* no search is required to determine the velocity of a spacecraft to enable it to arrive at Saturn on a given date at a given location. Even if the theory of the behavior of a physical system is incomplete, empirical methods such supervised classifier induction can be used to guide the operator selection.

However, models of many systems, including 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 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 for grasshopper physiology and life cycles and for 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.

CARMA elicits the relevant features of a new case from the user (*e.g.,* observed color and numbers of grasshoppers), infers a set of abstract case features (*e.g.,* grasshopper species and density), retrieves the best matching prototypical case from a case library derived from the judgments of expert pest managers, adapts the forage loss prediction associated with the prototypical case using a simulation model of grasshopper growth, attrition, and forage consumption, determines the acceptable treatment options, estimates the value of the forage saved and the likelihood of reinfestation under each treatment option, and presents a cost/benefit analysis for each option.

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 is 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 tested each in isolation and compared the results to the performance of the full CARMA prediction system.

Because of the absence of empirical data on forage consumption, the judgments of expert entomologists were used as the standard of evaluation for forage prediction. Questionnaires were sent to 20 entomologists (including pest managers) recognized for their work in the area of grasshopper management and ecology. Each expert received 10 cases randomly selected from a complete set of 20 hypothetical cases set in northern

Wyoming. The descriptions of the 20 cases contained at least as much information as is typically available to an entomologist from a rancher seeking advice. The questionnaire asked the expert to predict quantitative forage loss and the most appropriate course of action.

Eight sets of responses were received from Wyoming experts, who had a mean of 18.0 years experience. Although there was considerable variation in the consumption predictions of the Wyoming experts (the mean standard deviation of forage loss predictions was 12.4%) the variation was less than for all respondents (whose forage loss predictions had a mean standard deviation of 18.8%).

Experimental Design Each predictive method was tested using a series of leave-one-out tests in which a set of cases (S) from a single expert was split into one test case (C) and one training set (S − C). The methods were trained on the forage-loss predictions of the training set and tested on the test case. This method was repeated for each case within the set (S).

CARMA's empirical component was evaluated by performing leave-one-out-tests for CARMA's forage consumption module with all model-based adaptation disabled. CARMA's forage consumption module with model-based adaptation disabled is termed *factored nearest-neighbor prediction* (factored-NN), because under this approach prediction is based simply on the sum of nearest neighbor predictions for each subcase. Two other empirical methods were evaluated as well: decision-tree induction using ID3[2] (Quinlan, 1993) and linear approximation using QR factorization (Hager, 1988) to find a least-squares fit to the feature values and associated predictions of the training cases.

The predictive ability of CARMA's model-based component in isolation was evaluated by developing a numerical simulation based on CARMA's model of rangeland ecology. This simulation required explicit representation of two forms of knowledge implicit in CARMA's cases: the forage per acre based on the range value of the location, and the forage typically eaten per day per grasshopper for each distinct grasshopper overwintering type and developmental phase.

Results The accuracy of each approach was tested using leave-one-out testing for each of the eight Wyoming Expert Sets and for a data set consisting of the median of the predictions of the Wyoming 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 are set forth in Table 1. These provide initial confirmation for 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. The smallest root-mean-squared error rate was obtained by CARMA-specific. On the Wyoming Expert Sets, the root-mean-squared error rate was 13.3% for CARMA-specific and 14.2% for CARMA-global. The root-mean-squared error rate was higher both for the empirical approaches—21.1% for factored-NN, 34.9% for ID3, and 25.6% for linear approximation—and for the purely model-based approach—29.6%. CARMA-specific and CARMA-global were also more accurate than the alternative methods on the Wyoming median set, although linear approximation was only slightly less accurate.

The 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 is tentative because the low level of agreement among experts and the absence of any external standard give rise to uncertainty about what constitutes a correct prediction. A detailed description of the design and evaluation of CARMA is set forth in (Branting et al., 1997).

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.

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

[2]ID3 classified cases into 10 qualitative consumption categories representing the midpoints (5, 10, 15, ... , 95) of 10 equally sized qualitative ranges. ID3's error was measured by the difference between the midpoint of each predicted qualitative category and the expected quantitative consumption value.

	CARMA		Empirical Only			Model-Based Only
	Specific weights	Global weights	Factored-NN	ID3	Linear appr.	Numerical simulation
Wyoming expert sets	13.3	14.2	21.1	34.9	25.6	29.6
Wyoming median set	9.7	10.0	22.8	35.2	11.9	28.8

Figure 1: Root-mean-squared errors (in %) for leave-one-out-test results.

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.

Summary

Model-based adaptation is a technique for integration of case-based and model-based reasoning. In domains having a perfect model, MBA can reduce search by restricting the search space to the region surrounding the case closest to the new problem. In domains with an imperfect model, MBA permits two individually insufficient knowledge sources—cases and models—to be integrated in a manner that leads to better accuracy than can be obtained with either in isolation.

This paper has described three systems that use approximate-model-based adaptation: CARMA, a system for rangeland pest management; FormTool, a system to select colorants for plastic; and Sophist, a bioprocess recipe planner. An empirical evaluation of CARMA provided initial confirmation of the hypothesis that approximate-model-based adaptation can lead to better performance than models or cases in isolation. The three approximate-model-based adaptation systems are summarized in Table 2.

It is my belief that there are large number of complex physical systems characterized both by incomplete models and limited empirical data. Prediction and control planning in such domains requires exploitation of multiple, individually incomplete, knowledge sources. I believe that approximate-model-based adaptation may be applicable to a wide range of such prediction and planning tasks in such systems.

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. (1990). *Modelling Legal Argument: Reasoning with Cases and Hypotheticals.* MIT Press, Cambridge, Massachusetts.

Bergmann, R. and Wilke, W. (1995). Building and refining abstract planning cases by change of representation language. *Journal of Artificial Intelligence Research*, 3(53–118).

Bergmann, R. and Wilke, W. (1996). On the role of abstraction in case-based reasoning. In *Proceedings of the Third European Workshop on Case-Based Reasoning (EWCR-96)*, pages 28–43, Lausanne, Switzerland.

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., 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.

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

Hager, W. (1988). *Applied Numerical Linear Algebra.* Prentice Hall, Englewood Cliffs, New Jersey.

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 dy-

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 2: Three systems for approximate-model-based adaptation.

namics: 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).

Quinlan, R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, California.

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.

Sycara, K., Navinchadra, D., Guttal, R., Koning, J., and Narashimhan, S. (1992). Cadet: A case-based synthesis tol for engineergin design. *International Journal of Expert Systems*, 4(2):157–188.

Veloso, M. (1992). *Learning by Analogical Reasoning in General Problem Solving*. PhD thesis, Carnegie Mellon University.