Learning to Integrate Multiple Knowledge Sources for Case-Based Reasoning*

David B. Leake, Andrew Kinley, and David Wilson
Computer Science Department
Lindley Hall 215, Indiana University
Bloomington, IN 47405, U.S.A.
{leake, akinley, davwils}@cs.indiana.edu

Abstract

The case* based reasoning process depends on multiple overlapping knowledge sources, each of which provides an opportunity for learning. Exploiting these opportunities requires not only determining the learning mechanisms to use for each individual knowledge source, but also how the different learning mechanisms interact and their combined utility. This paper presents a case study examining the relative contributions and costs involved in learning processes for three different knowledge sources—cases, case adaptation knowledge, and similarity information—in a case-based planner. It demonstrates the importance of interactions between different learning processes and identifies a promising method for integrating multiple learning methods to improve case-based reasoning.

1 Introduction

The case-based reasoning (CBR) process solves new problems by retrieving records of problem solving for similar prior problems and adapting their solutions to fit new needs. Learning by acquiring new cases is an integral part of this process: each problem-solving episode itself provides a new case to save for future reuse. However, learning-new cases is only one of many ways to learn within the CBR framework. CBR systems rely on at least four types of knowledge: the case base, indexing scheme, similarity criteria, and case adaptation knowledge. Each of these types of knowledge provides an opportunity for learning. Consequently, a multistrategy learning approach [Michalski and Tecuci, 1994] that improves multiple types of knowledge is promising for improving case-based reasoning. Because the information content of the different types of knowledge in a CBR system may overlap [Richter, 1995], learning that augments one type of knowledge can even help overcome deficiencies in the others. For example, learning new cases might reduce the need for case adaptation knowledge, by enabling the system to start from more relevant cases; conversely, learning new case adaptation knowledge might enable a system to solve a wider range of problems with its existing cases.

Developing the requisite learning methods for each knowledge type requires addressing questions about the learning mechanisms to use, how to integrate them, and the overall utility of adding them to the CBR process. A simple approach is to develop learning strategies for each knowledge type individually and then add them all to the CBR system. Learning methods exist, for example, for refining indexing criteria (see [Kolodner, 1993] for an overview); learning methods have also been applied to case adaptation knowledge [Hanney, 1997; Sycara, 1988]; and some CBR systems already combine multiple forms of learning [Hammond, 1989].

However, simply combining methods may not achieve the desired overall benefits, even if each method is effective individually. For example, Leake, Kinley, and Wilson [1996] describe tests in which case learning, and learning about case adaptation, each independently made solution generation much more effective, but when case learning was added to adaptation learning, the addition yielded minimal improvement over adaptation learning alone. One possible explanation would be that in these tests, adaptation learning alone was almost sufficient for optimal performance, leaving little room for improvement. However, tests described in this paper show that adding learning for another knowledge source can actually degrade performance: when the system learned both new cases and new adaptations, it was unable to retrieve the cases it needed in order to take full advantage of the learned adaptations. This interaction raises questions about how a CBR system can best exploit learning for each of its multiple knowledge sources.

"This work was supported in part by the National Science Foundation under Grant No. IRJ-9409348.

This paper presents a case study examining the relationship of case learning, learning to refine case adaptation, and learning to refine similarity judgments in a case-based planning system. It considers two sets of issues: the requirements for each of the individual learning methods to be effective, and the requirements for realizing their full potential for improving overall system performance. It demonstrates the tight coupling of knowledge sources for CBR and shows that linking similarity assessment to learned adaptation knowledge can yield important benefits for exploiting both case and adaptation learning.

2 Motivations and Issues

This study grew out of research on learning to refine case adaptation. Case adaptation remains the least understood part of case-based reasoning, and experts agree that the state of the art in case adaptation is inadequate for automatic case adaptation to be included in fielded applications of CBR [Barletta, 1994; Mark *et* al., 1996]. One possible way to alleviate this problem is to develop new methods for automatic learning of case adaptation knowledge. The DIAL system, our testbed case-based planner, uses a hybrid approach to learning adaptations [Leake *et* al., 1996], building initial adaptations by reasoning from scratch and then reusing adaptations by case-based reasoning. Learning of adaptation cases takes place in tandem with learning of plan cases to be reused by the normal case-based planning process [Hammond, 1989].

Unlike previous case-based approaches to case adaptation (e.g., [Sycara, 1988]), DIAL'S method reuses adaptations by derivational analogy [Carbonell, 1986; Veloso, 1994], replaying the *derivations* of previous adaptations to generate analogous adaptations, rather than transforming the solutions to prior adaptation problems. When the rationale for a problem-solving process is available, derivational approaches can increase problem-solving efficiency for the broad class of problems with similar derivations [Veloso, 1994]. In addition to recording and replaying the traces of adaptations done from scratch, DIAL also stores traces of user-performed adaptations for problems it cannot adapt, increasing the range of adaptation problems it can solve.

Both case learning and adaptation learning would be expected to reduce the effort expended on case adaptation. Case learning should increase the range of plans available as the starting point for reasoning, reducing the need to reason from distant plans requiring more adaptation. Adaptation learning should increase the availability of relevant adaptation knowledge, reducing the amount of effort required for each adaptation. Prior tests showed that as expected, the learning methods, used individually, each produced a marked improvement in the speed of case adaptation. Surprisingly, however, adding case learning to adaptation learning (the method that' performed best individually) produced only small additional speedup when compared with the best of the individual learning methods (adaptation learning) [Leake *et* al., 1996].

We hypothesized that the problem might be caused by a mismatch between the system's similarity assessment criteria and the system's changing case adaptation abilities. To facilitate adaptation, similarity criteria should reflect adaptability [Birnbaum *et* al., 1991; Leake, 1992a; Smyth and Keane, 1996]. Thus when new adaptations are learned, similarity criteria should be modified to reflect changed adaptation abilities, in order to select the cases that will be easiest to adapt. However, early versions of DIAL—like most other CBR systems—relied on static similarity criteria. As a result, when it learned both new plan cases and new adaptations, the new plan cases it selected as most similar might be more difficult to adapt than plans that appeared less similar, but that involved problems it had learned how to adapt.

To link similarity assessment directly to adaptation knowledge, we developed a simple similarity assessment method called RCR (for Re-application costs and relevance) [Leake *et* al., 1996]. RCR estimates the cost of performing adaptations by using simple case-based reasoning about the costs of previous adaptations. Such a method makes learning to refine similarity a natural side-effect of adaptation learning, but also has two potential drawbacks: either generating inaccurate similarity judgments (if the costs of the previous adaptations retrieved turn out to be poor predictors), or imposing excessive computational overhead, because of embedding another case-based reasoning process within the main CBR cycle. Consequently, we asked four questions:

1. Whether the linkage between similarity and adaptation knowledge provided by RCR similarity assessment can markedly decrease case adaptation effort when case learning and adaptation learning are used together.
2. How the *overall* planning efficiency of DIAL is affected by RCR and adaptation learning.
3. How the total planning cost breaks down into costs of RCR similarity assessment versus case adaptation.
4. How adaptation learning and case learning affect the range of problems that the system can solve.

After a synopsis of the learning methods investigated and how they are applied, this paper examines these four issues. It briefly addresses the first issue, which is considered in depth in Leake, Kinley, & Wilson [1997], and focuses primarily on the remaining three.

3 Task domain and basic processing sequence

DIAL's task domain is disaster response planning: the strategic planning used to guide damage assessment, evacuations, etc., in response to natural and man-made disasters such as earthquakes and chemical spills. Human disaster response planners appear to depend heavily on prior experiences when they address new problem situations [Rosenthal *et a/.*, 1989], making it a natural task domain for case-based reasoning. For example, when generating a response plan to bring help to an isolated area, a previously-generated plan for another isolated area may provide helpful information for planning emergency transportation.

DIAL generates disaster response plans for disasters reported in simple (1-2 line) news stories. The system includes a simple schema-based story understanding component that processes conceptual representations of news stories describing the initial events in a disaster, and a retrieval component that selects a prior response plan expected to be easily adaptable to the new disaster. Problems in the retrieved plan are detected by a simple evaluator for candidate response plans (based on the problem-detection process described in [Leake, 1992b], and supplemented by inputs from a human user).

When problems are found, a description of the problem in a pre-defined problem vocabulary is provided to the adaptation component. That component can either build up adaptations from scratch or by case-based reasoning starting from previous adaptations. During adaptation, DIAL learns by storing traces of its case adaptation process and of the memory search process used to find needed information. For example, if it performs a substitution to replace an unavailable object (e.g., supplies were previously delivered by the Red Cross but there is no Red Cross in the country where the new disaster occured), the stored memory search trace records the path it followed to find a substitution (e.g., moving from a memory node for Red Cross to the memory node for its abstraction of relief organizations, and then moving to specifications of that node). More complete descriptions of the system axe available in [Leake *et a/.*, 1996],

4 Types of learning

Response plan case learning: DIAL begins its processing supplied with a small library of hand-coded disaster response plans, using a representation analogous to that used by CHEF [Hammond, 1989]. When new disasters are encountered, these response plan cases are reapplied by transformational analogy, changing components as needed to fit new constraints. The results are then stored for future reuse, adding to the case library. Because this process is a standard part of case-based planning systems, we will not discuss it further. When DIAL is unable to generate a suitable plan autonomously, its plan library can be augmented by user-generated plans, increasing the range of problems the system can solve autonomously, as described in the following paragraphs.

Adaptation case learning: As described in [Leake *et al,* 1996], DIAL'S initial case adaptation knowledge is a small set of abstract transformation rules and a library of domain-independent "weak methods" for memory search (e.g., the "local search" strategy to find related concepts by considering nearby nodes in memory). When presented with a new adaptation problem, DIAL first selects a transformation rule to apply and then performs memory search to find the information needed to operationalize the transformation rule and apply it to the problem at hand (for example, if a *substitution* transformation is selected, to find what to substitute). Once a successful adaptation has been generated, the system saves a trace of the steps used in solving the adaptation problem for future reuse. In this way, the system learns specific adaptation procedures starting from domain-independent adaptation methods when no specific knowledge is available. Adaptation cases may themselves be "adapted" in a simple way: If the derivation does not identify a solution, "local search" considers alternatives near the one suggested by the derivation, terminating its search after reaching a user-defined limit on the number of nodes visited. When the process terminates without finding an adaptation, the user can guide the system through the adaptation process to generate the new plan. Both the new adaptation and the resulting plan are stored for future use.

DIAL'S adaptation cases have two basic parts: indexing information and adaptation information. The indexing information includes a representation of the type of problem to adapt and information about the response plan for which the adaptation case was generated. The problem description information is similar in spirit to the problem vocabularies used to guide adaptation in other CBR systems (e.g., [Leake, 1992b]), and serves as an index to guide retrieval of adaptation cases to use for new adaptation problems. The problem vocabulary divides problems according to categories such as UNAVAILABLE-FILLER and LACK-OF-ACCESS. Each problem type is associated with a structure to be filled by a fixed range of descriptive information (e.g., the particular role, filler, and attempted action involved). To streamline access to relevant adaptation cases, stored adaptation cases are organized in memory by the types of problems they address.

The adaptation information packages a transformation type (e.g., substitute, add, delete) and a pointer to a memory search case containing the memory search steps used to find the information needed to apply the transformation. The memory search steps are described in terms of a vocabulary of standard memory operations, such as extracting a role-filler or moving up the abstraction hierarchy in memory.

The adaptation information is used both to guide future adaptations and to estimate their cost. Once an adaptation case has been retrieved, the cost of memory search dominates all other costs involved in adaptation. Consequently, the cost can be approximated by the memory search cost involved in replaying the stored memory search trace.

Similarity learning: The RCR similarity assessment method predicts the cost of adapting a problem in a case-based way, using learned adaptation knowledge. Given a new disaster situation and a candidate response plan with applicability problems, RCR first retrieves the adaptation cases most relevant to the current problem types, one for each problem to adapt, using the problem description as an index into the library of adaptation cases. It next estimates the cost to re-apply each of the adaptation cases retrieved, based on the length of its adaptation derivation.

Ideally, in similar future contexts, replaying the derivation will lead to an analogous result that applies to the new context, so that the length of the stored derivation suggests the re-application cost. However, differences between the old and new problems may prevent the prior derivation from being directly applicable, increasing the cost of adaptation. Consequently, the estimated cost is multiplied by a "dissimilarity" factor based in a simple way on the semantic similarity of old and new situations. To calculate the dissimilarity factor DIAL simply sums semantic distances between role-fillers in the problem descriptions, according to its memory hierarchy. The benefits of RCR compared to alternative methods are discussed in Leake, Kinley, and Wilson [1997].

Note that because RCR focuses on the difficulty of adapting problems, a response plan that requires several simple adaptations could be chosen over a response plan that requires a single difficult adaptation. Because this similarity learning method focuses on finding the cases that are easiest to adapt (those with the least important differences), it differs from learning methods such as Prodigy/Analogy's [Veloso, 1994] "foot-print" similarity metric that are aimed at learning situations with the most relevant similarities. RCR is in the spirit of Smyth and Keane's [1996] adaptation-guided retrieval, but learns about the difficulty of adaptations from experience rather than using static criteria to estimate adaptation cost.

Thus DIAL'S learning mechanisms include response plan learning, by CBR/transformational analogy; adaptation learning, by CBR/derivational analogy applied to traces of internal processing or user adaptations; and similarity learning, by CBR/transformational analogy-applied to previous adaptations. The combination of methods allows different lessons to be drawn from a single episode and reapplied independently in new contexts.

5 Effects of Individual and Combined Learning Strategies

To answer the questions listed in section 2, we performed a series of tests. These tests compared DIAL'S performance under five conditions: No learning of either cases or adaptations (NL); case learning, of plan cases only—the standard learning of case-based planners (CL); adaptation learning, in which only adaptation cases are stored (AL); learning of both response plan cases and adaptation cases (AL+CL); and learning of both response plan cases and adaptation cases, using the RCR method to base similarity assessment during plan retrieval on learned adaptation cases (AL+CL+RCR). Each condition except the last used traditional semantic similarity for case retrieval, with ties broken by a simple count of the number of problems in a plan case requiring adaptation.

The initial memory for the trials included nodes for 1264 concepts and an initial case library containing 5 response plans for earthquake, air quality, flood, and fire disasters. During testing, DIAL processed conceptual representations of 18 news stories (7 floods, 5 earthquakes, 4 forest fires, and 2 industrial air quality problems). Generating response plans for these disasters required generating 119 adaptations, each of which was stored as a new adaptation case. These experiments extended the trials reported in [Leake *et al.*, 1996], which processed 5 stories, resulting in 30 adaptation cases.

Test runs were divided into two sets. Processing of the first third of the adaptation problems was treated as a learning phase to build up initial knowledge sources, and statistics were gathered on the remaining two thirds of the adaptations.

Effects of linking similarity and adaptation knowledge on adaptation efficiency: The measure used for adaptation efficiency was memory search effort, calculated by two machine-independent measures: the number of memory nodes visited, and the number of primitive memory search operations performed. Figure 1 shows that both case learning and adaptation learning individually provide large efficiency increases over no learning (as expected), while adaptation efficiency with AL+CL provides smaller gains over adaptation learning alone. The results for the first four cases are consistent with those of [Leake *et al.*, 1996].

The fifth result, for AL+CL+RCR, suggests the potential benefits of directly linking similarity judgments to learned adaptation knowledge. The tests do not,

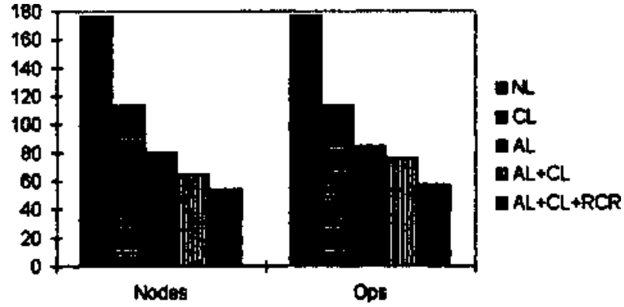

Figure 1: Average adaptation costs.

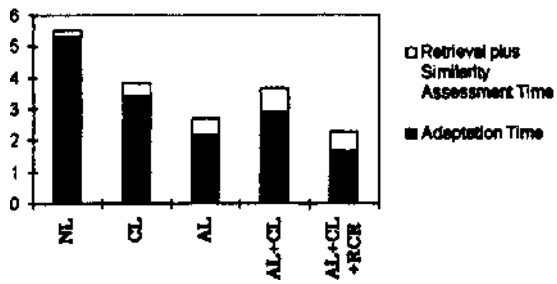

Figure 2: Overall processing costs.

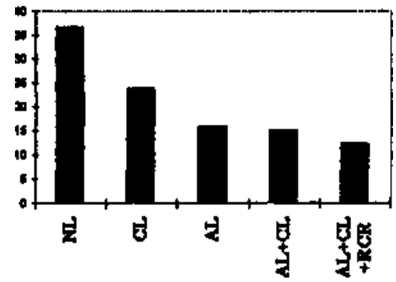

Figure 3: Failure rates for the adaptation process.

however, address another crucial question: whether the combination of AL+CL+RCR improves overall planning performance. RCR similarity assessment involves retrieving adaptation cases applicable to all the problems in a set of candidate cases, possibly imposing considerable overhead on the retrieval/similarity assessment process. This led us to examine the overall efficiency of generating response plans, measured by execution time.

Overall Planning Efficiency and Cost Breakdown: Figure 2, illustrating execution time in CPU seconds, shows that AL+CL+RCR in fact provided some improvement over the other conditions in terms of overall planning time. The light bands at the top of the bars show the portion of the execution time for the retrieval/similarity assessment process. Interestingly, although AL+CL decreased the machine-independent measures of adaptation effort compared to AL (as shown by figure 1), it resulted in a noticeably worse total execution time than AL. This is partially due to increased retrieval time of retrieving from growing sets of plan and adaptation cases. However, it also appears that the machine-independent measures do not completely capture the factors affecting adaptation time.

Effects on the Range of Problems the System Can Solve: DIAL's initial domain theory is incomplete, but its ability to store and reuse user-provided solutions (both disaster response plans and adaptations) enables it to augment its knowledge. Consequently, its learning affects not only efficiency, but also the range of problems that it can solve. Figure 3 shows the percentage of the trial problems the system is unable to solve autonomously after the learning phase on the test set of problems. The combination of case learning and adaptation learning performed better than either method alone, and the improved case selection of AL+CL+RCR increased the proportion of problems that the system could solve compared to AL+CL based on static semantic similarity criteria. However, these differences are small and possibly insignificant.

6 Lessons and Directions

These results suggest the importance of adjusting similarity criteria—and hence the selection of learned cases—to keep pace with adaptation learning. By enabling more effective use of two types of learned knowledge, integrating different types of learning improved both the speed of processing and the range of problems that DIAL could solve. These pilot experiments raise many issues for future study: the relative importance of adaptation and case learning in different domains, the learning curve for each type of knowledge, the utility of case-based similarity assessment methods (like RCR) as the number of adaptation cases increases, whether it is possible to partially alleviate the utility problem by retaining only a subset of the many adaptation cases that are generated, and the tradeoffs and utility of alternative methods for retrieving and applying adaptation knowledge during case selection. However, the results support the potential value, of coordinating different types of learning in CBR and the need for further investigation.

7 Conclusion

Case-based reasoning exploits multiple knowledge sources. Consequently, it provides an opportunity for multistrategy learning to refine each of those knowledge sources. Our studies of multistrategy learning in the case-based planner DIAL provide—to our knowledge—the first empirical demonstrations of the complementary roles that can be played by these multiple learning processes. However, they also show that the learning strategies must be coordinated to realize their potential benefit. Similarity criteria for selecting cases must change as adaptation knowledge is learned; neither coverage of the case library, nor case adaptation abilities, can be judged in isolation from the other knowledge sources. Likewise, in developing a combined method, how one type of learning affects the efficiency of one component of the CBR process is secondary to the efficiency effects of the learning on the CBR process as a whole. The close coupling of multiple processes and knowledge sources in CBR complicates the application of learning to each one, but also provides a new motivation for combined learning: Combined learning can enable a CBR system to better exploit the relationships between multiple types of knowledge.

References

[Barletta, 1994] R. Barletta. A hybrid indexing and retrieval strategy for advisory CBR systems built with ReMind. In *Proceedings of the Second European Workshop on Case-Based Reasoning,* pages 49-58, Chantilly, France, 1994.

[Birnbaum *et al.,* 1991] L. Birnbaum, G. Collins, M. Brand, M. Freed, B. Krulwich, and L. Pryor. A model-based approach to the construction of adaptive case-based planning systems. In R. Bareiss, editor, *Proceedings of the DARPA Case-Based Reasoning Workshop,* pages 215-224, San Mateo, 1991. Morgan Kaufmann.

[Carbonell, 1986] J. Carbonell. Derivational analogy: A theory of reconstructive problem solving and expertise acquisition. In R. Michalski, J. Carbonell, and T. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach,* volume 2, pages 371-392. Morgan Kaufmann, Los Altos, CA, 1986.

[Hammond, 1989] K. Hammond. *Case-Based Planning: Viewing Planning as a Memory Task.* Academic Press, San Diego, 1989.

[Hanney, 1997] K. Hanney. Learning adaptation rules from cases. Master's thesis, Trinity College, Dublin, 1997.

[Kolodner, 1993] J. Kolodner. *Case-Based Reasoning.* Morgan Kaufmann, San Mateo, CA, 1993.

[Leake *et al.,* 1996] D. Leake, A. Kinley, and D. Wilson. Acquiring case adaptation knowledge: A hybrid approach. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence,* pages 684-689, Menlo Park, CA, 1996. AAAI Press.

[Leake *et al,* 1997] D. Leake, A. Kinley, and D. Wilson. Case-based similarity assessment: Estimating adaptability from experience. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence.* AAAI Press, 1997.

[Leake, 1992a] D. Leake. Constructive similarity assessment: Using stored cases to define new situations. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society,* pages 313-318, Hillsdale, NJ, 1992. Lawrence Erlbaum.

[Leake, 1992b] D. Leake. *Evaluating Explanations: A Content Theory.* Lawrence Erlbaum, Hillsdale, NJ, 1992.

[Mark *et al.,* 1996] William Mark, Evangelos Simoudis, and David Hinkle. Case-based reasoning: Expectations and results. In D. Leake, editor, *Case-Based Reasoning: Experiences, Lessons, and Future Directions,* pages 269-294. AAAI Press, Menlo Park, CA, 1996.

[Michalski and Tecuci, 1994] R. Michalski and G. Tecuci, editors. *Machine Learning: A Multistrategy Approach.* Morgan Kaufmann, San Mateo, CA, 1994.

[Richter, 1995] Michael Richter. The knowledge contained in similarity measures. Invited talk, the First International Conference on Case-Based Reasoning, Sesimbra, Portugal., October 1995.

[Rosenthal *et al,* 1989] U. Rosenthal, M. Charles, and P. Hart, editors. *Coping with crises: The management of disasters, riots, and terrorism.* C.C. Thomas, Springfield, IL, 1989.

[Smyth and Keane, 1996] B. Smyth and M. Keane. Design a la Deja Vu: Reducing the adaptation overhead. In D. Leake, editor, *Case-Based Reasoning: Experiences, Lessons, and Future Directions.* AAAI Press, Menlo Park, CA, 1996.

[Sycara, 1988] K. Sycara. Using case-based reasoning for plan adaptation and repair. In J. Kolodner, editor, *Proceedings of the DARPA Case-Based Reasoning Workshop,* pages 425-434, San Mateo, CA, 1988. Morgan Kaufmann.

[Veloso, 1994] M. Veloso. *Planning and Learning by Analogical Reasoning.* Springer Verlag, Berlin, 1994.