Proceedings of the Fourteenth National Conference on Artificial Intelligence, AAAI Press, Menlo Park, CA, 1997.

Case-Based Similarity Assessment: Estimating Adaptability from Experience*

David B. Leake, Andrew Kinley, and David Wilson
Computer Science Department
Lindley Hall 215, Indiana University
Bloomington, IN 47405
{leake, kinley, wilson}@cs.indiana.edu

Abstract

Case-based problem-solving systems rely on *similarity assessment* to select stored cases whose solutions are easily *adaptable* to fit current problems. However, widely-used similarity assessment strategies, such as evaluation of semantic similarity, can be poor predictors of adaptability. As a result, systems may select cases that are difficult or impossible for them to adapt, even when easily adaptable cases are available in memory. This paper presents a new similarity assessment approach which couples similarity judgments directly to a case library containing the system's adaptation knowledge. It examines this approach in the context of a case-based planning system that learns both new plans and new adaptations. Empirical tests of alternative similarity assessment strategies show that this approach enables better case selection and increases the benefits accrued from learned adaptations.

Introduction

Case-based problem-solving solves new problems by retrieving and adapting the solutions for similar prior problems. In order for a case-based reasoning (CBR) system to generate solutions as effectively as possible, similarity criteria must reflect the *adaptability* of stored cases to the new situation (Birnbaum *et al.* 1991; Leake 1992a; Smyth & Keane 1996). However, it has recently been shown that widely-used similarity assessment criteria, such as semantic similarity of situation features, can be poor predictors of the actual difficulty of adapting prior cases to new needs (Smyth & Keane 1996). When similarity assessment picks an inappropriate case, solution generation will be unnecessarily expensive. If the difficulty of adaptation outstrips the system's adaptation abilities, the choice may even prevent the system from generating a solution. Thus developing more accurate similarity assessment is an important goal for CBR. In addition, because determining and tuning similarity criteria is a major part of the knowledge acquisition effort for CBR applications, it is desirable to learn useful similarity criteria. This paper presents a new case-based similarity assessment approach which addresses these problems.

Estimating Adaptability from Experience: Our similarity assessment approach exploits the case library of adaptation knowledge that is already present in CBR systems that use case-based reasoning to guide case adaptation (Berger 1995; Sycara 1988). The method is based on the observation that when adaptations themselves are performed by case-based reasoning, adaptation cases encapsulating derivational traces of prior adaptations can be used not only to guide adaptations but also to predict the difficulty of adapting similar problems in the future. Our method, RCR (Re-application Costs and Relevance), predicts the adaptability of a retrieved planning case by retrieving the adaptation cases that would be used to guide its adaptation, considering the number of steps that would have to be replayed, and scaling the estimated cost using a coarse-grained estimate of the difference between the current problems and those the adaptation cases previously addressed. The RCR method couples similarity judgments directly to a system's case adaptation experiences and to the knowledge it will use to adapt the problems in the planning case it selects.

Such a method is appealing for a number of reasons. It simplifies knowledge acquisition for similarity criteria, because similarity judgments are based on experiences with adaptations rather than *a priori* analysis that may be hard to connect to actual performance on specific problems. It provides a finer-grained method for estimating adaptation costs, reflecting knowledge of individual prior problems. It also provides a simple way to refine similarity criteria as new adaptations are learned—similarity criteria change naturally when new

 This work was supported in part by the National Science Foundation under Grant No. IRI-9409348. We thank an anonymous reviewer for helpful comments.

adaptations become available.

Issues to Address: Whether the RCR method can realize the desired benefits depends on four issues. The first is whether estimates of adaptability based on the prior cost of using an adaptation case are in fact a better predictor of adaptation costs in a new situation than standard similarity assessment methods used in CBR systems. The second is whether the finer-grained case-based cost predictions are appreciably more accurate than aggregate estimates (e.g., of the difficulty of adapting a broad problem class). The third is whether using RCR improves the ability to exploit learned adaptations. The fourth is whether the possible overhead of estimating adaptability in a case-based way will counterbalance any savings in adaptation cost. This paper examines the first three of these issues; the fourth is briefly summarized here and is addressed in Leake, Kinley, & Wilson (1997).

Task Domain and System

The RCR similarity method has been applied in the DIAL system, a case-based planner (Leake, Kinley, & Wilson 1996). DIAL's task domain is disaster response planning. Disaster response planning is the initial strategic planning used to determine how to assess damage, evacuate victims, etc., in response to natural and man-made disasters. Human disaster response planners appear to depend heavily on prior experiences when they address new problem situations, making disaster response planning a natural domain for case-based reasoning.

The DIAL system starts with a small case library of simple disaster response plans for disasters such as earthquakes, floods, and chemical spills. It processes conceptual representations of news briefs about disasters and generates response plans by retrieving and adapting the response plan for a similar prior disaster. The system initially uses a rule-based approach to adaptation, using comparatively unguided methods, and supplants those general rules with specific *adaptation cases* as it gains experience. This paper focuses on how those cases can be used for similarity assessment.

Basic Processing Sequence

DIAL's basic processing sequence is as follows:

1. A story is input to the system.
2. A small set of potentially relevant cases from similar prior disasters is retrieved. This process uses coarse-grained semantic similarity criteria (based on distances between role-fillers in the schemas representing old and new disaster situations) to retrieve a user-defined number of prior cases.
3. The system determines correspondences between old and new disasters and does an initial mapping of the prior response plans to the new situation, identifying roles for which this initial simple adaptation fails.
4. Additional problems are identified by a combination of automatic stereotype-based problem detection (Leake 1992b) and user feedback (possibly rejecting system mappings).
5. The RCR similarity assessment process takes prior response plan cases and their problems as input, retrieves the adaptation cases that would be used to adapt the problems, and estimates the new adaptation cost based on the adaptation cases' previous cost and a coarse-grained estimate of their applicability to the new problem. This information is used to select the response plan case expected to be easiest to adapt.
6. Problems in the response plan suggested by the selected case are repaired by case adaptation using derivational analogy (Veloso & Carbonell 1994).
7. The resulting response plan case is stored for reuse, as are traces of the adaptations performed, packaged as adaptation cases.

This paper will focus on aspects of the system with a direct bearing on step 5, the actual case-based similarity assessment process. For this purpose, the crucial points are the information contained in adaptation cases, how it is used in similarity assessment, and how the similarity assessment process affects the benefits of adaptation learning. See Leake, Kinley, & Wilson (1996) for details on building the initial adaptation case library by storing traces of adaptations built from scratch.

Contents of Adaptation Cases

DIAL's adaptation cases represent adaptations as the combination of transformations (e.g., addition, deletion, substitution) plus the memory search for the knowledge needed to operationalize the transformations (e.g., to find what to add or substitute).

DIAL's adaptation cases package information about the *context* of an adaptation, the *derivation* of its solution, and the *effort* involved in the derivation process. The context information includes characteristics of the problem for which the adaptation was generated, such as the type of problem (in a problem vocabulary based on Leake, 1992b), the value being adapted, and the roles that value fills in the response plan.

The derivation records the primitive memory operations needed to find appropriate values in memory. This may include operations to extract role-fillers or other information to guide the memory search process. When adaptation cases are built, unsuccessful paths are trimmed from the stored adaptation trace. However, the record of effort also notes the actual effort expended to find the solution path.

Filtering and Selecting Cases

Given a new problem situation, DIAL first retrieves an initial pool of response-plan cases using static feature-based similarity measures. For example, in processing a story about a flood in West Virginia, DIAL selects three prior flood responses as being most relevant: one in Alaska, one in Georgia, and one in Oregon. DIAL performs an initial mapping of the new situation onto the prior situations to generate candidate response plans.

Problem detection: Any case-based adaptation system requires a method for determining problems requiring adaptation. Many methods exist, ranging from explanations of problems (Hammond 1989) to pattern-based verification (Leake 1992b). External information may also directly identify needed adaptations. For example, in the planner PLEXIS (Alterman 1988), adaptation to find a replacement object is triggered by failure to find that object during plan execution.

DIAL relies on a combination of pattern-based methods—which can detect potential problems at minimal cost (Leake 1992b)—and user feedback for information not available in the system's prior knowledge. In the disaster response domain, user input might reflect situation-specific information (e.g., that a road is impassable, or that a region is not under the jurisdiction of a particular agency). The DIAL model assumes that the planner's knowledge is incomplete and augmented by both case learning and adaptation learning.

For each candidate response plan, DIAL identifies problems detectable from its knowledge, and the user identifies and describes additional problems. For example, in the aftermath of the Oregon flood, the police forces maintained order. This response would not work for the new disaster in West Virginia, because the police force is not equipped to deal with the access problems posed by the flood. This results in a problem categorized as "lack of access" to the area. Re-applying the Oregon plan to the new situation will require adaptation to overcome this problem (as well as two additional problems that will not be discussed here).

RCR similarity assessment: The RCR method uses adaptation cases to estimate adaptability. For each problem identified, DIAL retrieves the single most relevant adaptation case from memory. In the flood example, based on the system's problem description DIAL finds an adaptation case that solved a previous lack of access problem similar to the one in West Virginia. The retrieved adaptation case solved a lack of access problem during a flood in Afghanistan by replacing the Afghan police with the better-equipped Afghan army. The cost to solve the original adaptation was very high—500 operations—because in its early runs the system had no applicable adaptation cases and had to try many alternatives. However, the successful memory search process that was eventually found only required 7 primitive operations.

After retrieving an adaptation case for each problem, DIAL estimates the adaptation cost for the response-plan cases, based on the length of the adaptation derivation for each problem. Ideally, in similar future contexts, replaying an adaptation will lead to an analogous result that applies to the new context. Consequently, the length of the stored derivation in an adaptation case suggests its re-application cost.

However, derivations that are used in new contexts are unlikely to apply directly. Consequently, the derivation length is multiplied by a "dissimilarity" factor reflecting the difference between the description of the current problem and the description of the problem that the derivation was generated to address. To calculate the dissimilarity DIAL simply sums semantic distances between role-fillers in the problem descriptions, according to its memory hierarchy.

When no adaptation case is retrieved for a particular problem type, the cost is estimated as the average cost for adapting problems of the current problem class, or—if no problems of that problem class have been encountered—the cost is estimated as the average cost of all adaptations generated up to that time. 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.

For the problem involving lack of access by the police, the reapplication cost is 7, multiplied by the dissimilarity factor of 1, for a total estimated adaptation cost of 7. Dissimilarity is small for this example because the lack of access problems in West Virginia and Afghanistan affect the same role in the same component of the response plans (the actor role for maintaining order), for the same type of disaster. In this example, the dissimilarity only arises from the difference of the two locations.

Finally, DIAL selects the response-plan case with the lowest estimated adaptation cost and begins adaptation. The estimated total adaptation costs for the candidate response-plan cases were: 547 for Alaska, 528 for Georgia, and 520 for Oregon. Consequently, DIAL chooses the Oregon plan as the easiest to adapt. In the West Virginia flood context, DIAL re-plays the derivation that originally led to the Afghan army. With a small amount of extra memory search from the result of following this derivation, DIAL finds a similar replacement, the national guard. A new adaptation case is created including the information that the total cost to find the national guard was 22, along with the successful path and its cost of 7.

Comparison of Similarity Metrics

To study the effects of case-based similarity assessment, we have compared DIAL's performance applying five alternative similarity assessment strategies, ranging from simple semantic similarity to methods aimed at providing increasingly close estimates of adaptability. Each one calculates a measure of the difference between response plan cases; cases with the lowest difference values are considered most similar.

- **Semantic similarity:** The difference is the sum of the semantic distance of features in the description of the new situation and the situation addressed by the retrieved response-plan case. This method, although commonly used in CBR, does not explicitly consider case adaptation effort. Thus it provides a baseline for judging the effects of considering adaptability during similarity assessment.

- **Number of problems:** The difference is a simple count of the number of problems that would need to be repaired to apply the candidate response-plan case. This is a crude approximation of difficulty that treats all problems as if equally hard to adapt.

- **Problem class averages:** The difference is the sum of the average adaptation costs for each type of problem to be adapted. This corresponds closely to the similarity assessment method used in adaptation-guided retrieval (Smyth & Keane 1996), another method for basing case selection on adaptability.

- **Actual prior costs of retrieved adaptations:** The difference is the sum of the actual prior adaptation costs of the retrieved adaptation cases. This is similar to the previous method but is intended to give a finer-grained estimate by using costs for the most similar specific prior cases rather than averages for broad problem classes.

- **Re-application costs and relevance (RCR) of retrieved adaptations:** As previously described, the difference is the sum of the prior costs of adaptation cases selected to adapt problems in the prior plan, each multiplied by a "dissimilarity" factor comparing the current problem to the problem for which the adaptation case was generated.

	Mean Ops	Std Dev	Max Ops	Min Ops
Semantic Similarity	65	81	316	3
Number of Problems	52	97	552	3
Problem class averages	47	93	552	3
Actual Prior Cost	48	94	552	1
RCR	36	30	110	1

Table 1: How similarity criteria affect adaptation effort.

Results

The DIAL system was seeded with 5 initial disaster response plans and 1264 memory nodes. In each trial, DIAL processed 20 novel disaster stories and generated response plans for them. A human user oversaw processing by identifying problems and evaluating suggested solutions to those problems. DIAL's memory grows during problem solving, and after one trial of the experiment the memory included up to 125 adaptation cases and 1693 total memory nodes.

In our first experiment, trials were conducted using each of the five similarity methods. For these tests, the four most promising cases according to semantic similarity were provided as input to RCR. Efficiency with each similarity metric was measured by the number of primitive memory operations that were executed to perform adaptations. Table 1 shows the average number of operations required to adapt the selected response plan until a successful response plan is generated.

The table shows that using the other similarity metrics decreases the overall adaptation effort compared to using semantic similarity. As expected, RCR shows the greatest decrease in cost. An interesting result was that using either problem class averages or actual prior cost gave almost identical performance.

Adaptation cost with RCR is not only low but also relatively stable, having both a low standard deviation and a smaller range between the maximum and minimum costs than the other methods. This suggests that RCR is choosing plans which contain primarily adaptations of moderate to low difficulty.

The accuracy of similarity assessment in selecting adaptable cases is also shown by examining the improvements in average adaptation cost over the course of a complete system run. Figure 1 shows the cost trend during learning for three of the similarity assessment methods. In all conditions, after an initial unstable period there is a downward trend in cost as-

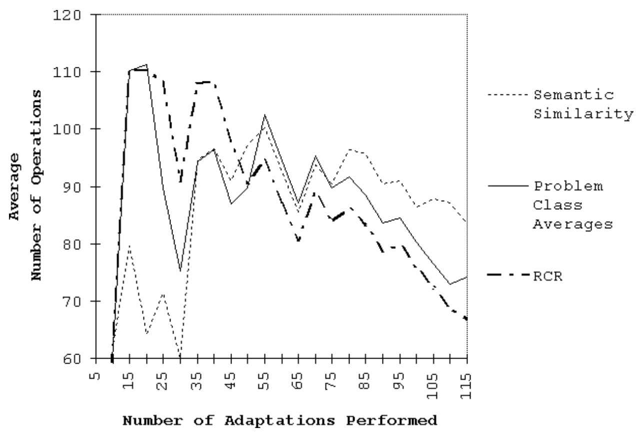

Figure 1: Trends in adaptation time with learning when cases are selected by semantic similarity, problem class average costs, and RCR.

sociated with the learning of adaptation cases, but it is clear that using RCR consistently averages lower costs after the initial period. Thus RCR enables better use of learned adaptations. These results suggest that even this simple case-based similarity assessment method is sufficient to achieve noticeable benefits in accuracy of case selection. Additional tests are needed to determine how to achieve the best possible case selection.

An obvious additional question is the impact of this process on the overall performance of the system. This can be measured in two ways. One is in overall processing time: does the savings in adaptation cost counterbalance additional overhead from the added cost of retrieving more adaptation cases in order to do similarity assessment? A second is how selection of better cases during retrieval increases the range of problems the system can solve. Leake, Kinley, & Wilson (1997) report tests in which the savings in adaptation cost did counterbalance increased similarity assessment effort and the better case retrieval increased the range of problems the system could solve.

Relationship to Prior Research on Linking Similarity and Adaptability

Similarity assessment for case-based reasoning systems often relies on semantic similarity or other criteria that may not reflect the difficulty of adaptation. Ideally, case selection should reflect anticipated usefulness as directly as possible (Kolodner 1988).

One approach is to refine similarity criteria to reflect the most relevant similarities. For example, explanation-based indexing (Barletta & Mark 1988) and the Prodigy/Analogy (Veloso & Carbonell 1994) system's "foot-print" similarity metric focus attention on goal-relevant features, in order to retrieve cases that refer to the prior problem situations with the most relevant similarities; other approaches do failure-driven learning to refine similarity criteria after detecting retrieval of a case that is needlessly difficult to adapt (Birnbaum *et al.* 1991; Fox & Leake 1995). All these approaches are worthwhile, but do not directly take adaptability into account.

Adaptability is the basis for similarity assessment in Smyth and Keane's (1996) Déjà Vu, a case-based design system. Both DIAL and Déjà Vu assume that an initial filtering stage retrieves a subset of the case library for fine-grained consideration. In Déjà Vu, this step checks all stored cases to select those for which the system has adaptation rules that potentially can adapt the features of the retrieved case to the new situation. DIAL's filtering stage currently uses semantic similarity to retrieve a smaller set of candidates to be considered by RCR. Because of the potentially high cost of retrieving adaptation cases to judge adaptability (due to the proliferation of adaptation cases), it is desirable to restrict the set of cases considered by RCR. The selective prefiltering of RCR trades off the risk of missing an adaptable case against the ability to spend more time examining each retrieved case, enabling more accurate estimates of its adaptability. Because adaptations in the DIAL domain are potentially extremely expensive, this tradeoff appears to be worthwhile, but its success depends on the filtering process retrieving at least one comparatively adaptable case. Filtering by semantic similarity seems reasonable, in light of evidence in prior CBR systems that semantic similarity can provide general estimates of similarity. However, RCR is not bound to filtering being done by semantic similarity. For example, it might be worthwhile to filter by explanation-based indexing methods rather than semantic similarity of the disaster situations.

In the second step of estimating adaptation effort, Déjà Vu's method relies on coarser-grained estimates of adaptation cost: each of the system's adaptation rules is associated with a single predefined cost. DIAL's process uses the costs of the most similar prior adaptations, which involves more costly retrievals but which also increases accuracy. The systems also differ in problem detection: Déjà Vu assumes a perfect domain theory and does not require any human intervention to detect problems.

An additional benefit of case-based similarity assessment is that it links similarity criteria directly to adaptation experience, automatically updating similarity criteria as adaptations are learned. This has the potential to be especially valuable given recent attention to systems that address the knowledge acquisition problem for adaptation by learning new adaptations (Leake, Kinley, & Wilson 1996; Hanney & Keane 1996).

Conclusions

Our new similarity assessment approach, Re-application Costs and Relevance (RCR), couples similarity judgments directly to a case library containing the system's adaptation knowledge. Its case-based approach simplifies knowledge acquisition for similarity criteria, because similarity is learned from experiences with adaptations rather than based on *a priori* analysis. RCR provides fine-grained estimates of adaptation costs, reflecting knowledge of individual prior problems, and provides a natural way to refine similarity criteria as new adaptations are learned.

In tests with the DIAL system, use of RCR for similarity assessment decreases the effort that must be expended on case adaptation. We expected case-based similarity assessment to increase accuracy of case selection compared to using problem class averages, but when case-based estimates consider only the cost of the most similar prior adaptation, we found no improvement. Substantial improvement does occur when the relevance of the prior adaptation case is taken into account by RCR, even using a very simple relevance criterion for adaptation cases. It is plausible that tuning the relevance criterion would further improve performance.

Tests also show that after an initial period establishing a case library of adaptation cases, RCR enables more effective adaptation than other similarity methods as adaptations are learned. A concern is increased similarity assessment cost, but preliminary tests suggest this is counterbalanced by savings in adaptation costs. We are now conducting additional tests and extending DIAL to a new domain for large-scale experiments to further investigate the benefits of RCR similarity assessment.

References

Alterman, R. 1988. Adaptive planning. *Cognitive Science* 12:393–422.

Barletta, R., and Mark, W. 1988. Explanation-based indexing of cases. In Kolodner, J., ed., *Proceedings of a Workshop on Case-Based Reasoning*, 50–60. Palo Alto: DARPA.

Berger, J. 1995. *ROENTGEN: A Case-based Radiation Therapy Planner.* Ph.D. Dissertation, The University of Chicago. Computer Science Department.

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

Fox, S., and Leake, D. 1995. Using introspective reasoning to refine indexing. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, 391–397. San Francisco, CA: Morgan Kaufmann.

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

Hanney, K., and Keane, M. 1996. Learning adaptation rules from a case-base. In *Proceedings of the Third European Workshop on Case-Based Reasoning.* Berlin: Springer Verlag.

Kolodner, J. 1988. Retrieving events from a case memory: A parallel implementation. In Kolodner, J., ed., *Proceedings of a Workshop on Case-Based Reasoning*, 233–249. Palo Alto: DARPA.

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

Leake, D.; Kinley, A.; and Wilson, D. 1997. Learning to integrate multiple knowledge sources for case-based reasoning. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence.* Morgan Kaufmann. In press.

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

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

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

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

Veloso, M., and Carbonell, J. 1994. Case-based reasoning in PRODIGY. In Michalski, R., and Tecuci, G., eds., *Machine Learning: A Multistrategy Approach.* Morgan Kaufmann. chapter 20, 523–548.