UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Representing Cases as Knowledge Sources that Apply Local Similarity Metrics

Permalink
https://escholarship.org/uc/item/7mp069vf

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 14(0)

Author
Skalak, David B.

Publication Date
1992

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Representing Cases as Knowledge Sources that Apply Local Similarity Metrics

David B. Skalak
Department of Computer Science
University of Massachusetts
Amherst, MA 01003
skalak@cs.umass.edu

Abstract [1]

A model of case-based reasoning is presented that relies on a procedural representation for cases. In an implementation of this model, cases are represented as knowledge sources in a blackboard architecture. Case knowledge sources define local neighborhoods of similarity and are triggered if a problem case falls within a neighborhood. This form of "local indexing" is a viable alternative where global similarity metrics are unavailable. Other features of this approach include the potential for fine-grained scheduling of case retrieval, a uniform representation for cases and other knowledge sources in hybrid systems that incorporate case-based reasoning and other reasoning methods, and a straightforward way to represent the actions generated by cases. This model of case-based reasoning has been implemented in a prototype system ("Broadway") that selects from a case base automobiles that meet a car buyer's requirements most closely and explains its selections.

Introduction

This paper addresses two fundamental problems of case-based reasoning (CBR): case representation and case similarity. Its central point is that a procedural, locally-indexed representation for cases provides several benefits. We use the term "locally indexed" to refer to a case retrieval technique that uses similarity metrics that are applicable only within a neighborhood of a case in the space of cases. Local indexing is distinguished from a global indexing method that relies on a single function to assess similarity throughout the case base. Metrics that are locally defined can be viewed as an attempt to approximate *piecewise* an ideal — but often difficult to construct — function that measures the similarity of a problem situation to any case, where each piece of the local metric is applicable only in a suitable area of the case.

The intuition behind local indexing is that each case is in the best position to map the topography of the case space in a neighborhood of itself. What then counts as similar depends on where the case is located in case space. Informally, the general perspective of this research is to impose problem-solving responsibility on the cases themselves by including in the case representation knowledge that is usually external to cases, including similarity metrics. In addition, this case-centric perspective regards cases as active entities, rather than as responsive to external procedures.

Seminal work in CBR has exploited the notion that how one assesses the similarity of a stored case depends on the problem situation, e.g., [Ashley, 1990], [Bareiss, 1989], [Kolodner, 1983], [Sycara, 1987]. The model we describe tries to exploit the complementary idea that cases of a particular sort possess features — independent of a problem situation — that will help determine how similarity will be assessed for cases of that variety. Informally, regardless of whether you're looking to buy a Cadillac or a Miata, if on the used car lot you encounter a pickup truck with huge tires whose body is six feet off the

---

[1] This research was supported in part by the National Science Foundation, contract IRI-890841, the Air Force Office of Sponsored Research under contract 90-0359, the Office of Naval Research under a University Research Initiative Grant, contract N00014-87-K-0238, and a grant from GTE Laboratories, Waltham, MA.

ground, your assessment will be governed by whether you want a vehicle with a six-foot high cab. Following this intuition, we try to build into each case several ways of assessing similarity that are useful for a case of that variety.

Briefly, the model of retrieval presented here regards a case as a procedural entity that is activated when a problem situation falls within a local neighborhood of the case. We claim that the following benefits flow from this approach:

- A procedural case representation yields a consistent knowledge representation for hybrid architectures that combine CBR with other reasoning methods.
- Local indexing is a useful alternative in situations where no available similarity metric can be applied uniformly across cases.
- Fine-grained scheduling for case retrieval is facilitated and permits focused control of problem-solving.
- Multiple perspectives on a case can be easily represented through local similarity functions.

Implementation

Our implementation of the model uses a blackboard architecture in which knowledge sources respond to changes on a global blackboard. The Broadway prototype uses GBB v.2.0, a toolkit for developing high-performance blackboard applications [Blackboard Technology Group, 1991]. The GBB Agenda Shell enables a user to define knowledge sources that may be triggered, checked for the fulfillment of their respective preconditions, and, if fulfilled, instantiated in knowledge source activation records with some execution rating. Knowledge source activations are then placed on an agenda of activations pending execution.

A correspondence between cases and blackboard knowledge sources can be exploited. A [case/knowledge source] is [similar/activated] when its [index/ precondition function] is [triggered/ satisfied], and therefore the [case's suggested action/ knowledge source function] should be executed. The current implementation fuses these corresponding aspects: cases are represented as knowledge sources. The precondition of a case knowledge source is a similarity predicate. If a problem situation satisfies this locally defined similarity predicate, the case knowledge source is activated. This incorporation of case similarity into knowledge source precondition functions yields one immediate by-product. The greater the similarity of a problem situation to a case, the higher the execution rating returned by the precondition of the knowledge source corresponding to the case, and the earlier the case's action is executed. The most similar cases execute their constituent actions first.

The Domain

The Broadway prototype addresses the almost century-old quandary "Which car should I buy?". For many people this question appears to invite anecdotal case-based reasoning. As in many common-sense areas, no strong domain theory is available to resolve the competing constraints involved in automobile purchase. However, this domain does not present the complexity of such classical blackboard applications as speech understanding [Erman et al., 1980] or sonar signal interpretation [Nii et al., 1982]. For our purposes the automobile domain is an interim vehicle for investigating the utility of procedural case representation and associated ideas of similarity, indexing and control.

Description of Control Flow

The skeletal flow of control for Broadway is given in general terms in Figure 1.

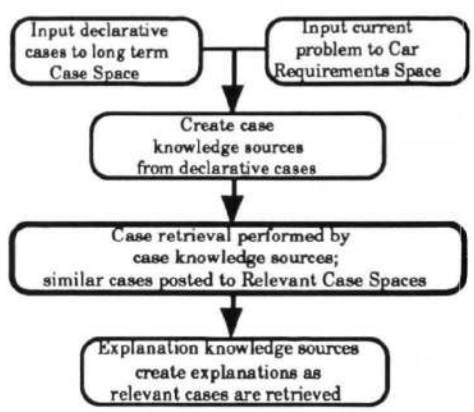

Figure 1: Flow of control in Broadway.

When the system is initialized, declarative, frame-based car case representations are loaded onto a blackboard space that is Broadway's long term case memory. Broadway's current case base consists of 93 cases that represent automobiles of a particular make and model, e.g., an Eagle Talon. Domain engineering has been aided by the 1991 annual automobile issue of Consumer Reports magazine [Consumer Reports, 1991]. Each declarative case then is used as the basis for creating several case knowledge sources that are activated when a similar problem case is posted to a problem space. See Figure 2 for a simplified example,

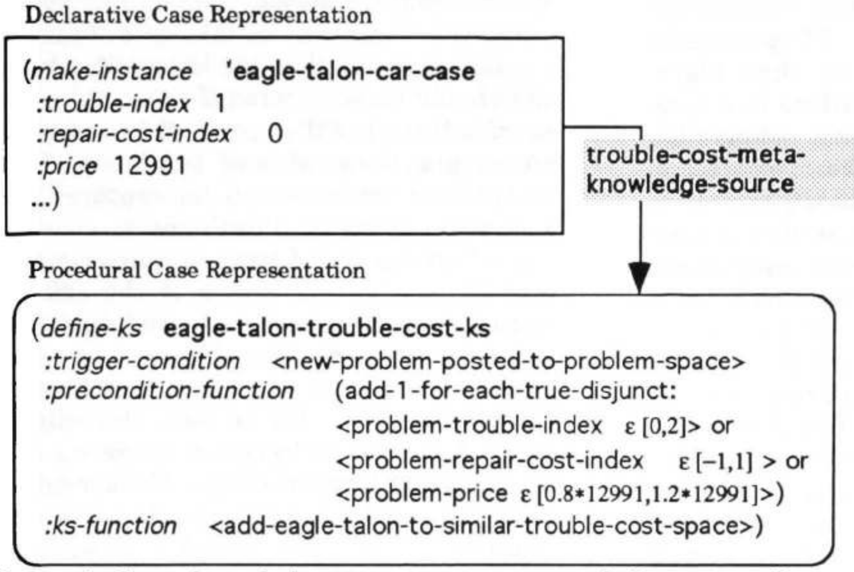

Figure 2: Case knowledge sources are created from case-frame representations by meta-knowledge-sources. The scale for the trouble-index and repair-cost-index is -2 (much worse than average) to 2 (much better than average).

which prescribes an interval of similarity for each feature and then adds the number of problem features that fall within those intervals to yield an execution rating.

These procedural representations of cases assume the usual CBR tasks of retrieving similar cases and manipulating them. Next, the user inputs a problem case, which is a specification of the features that the user desires in a car, presented in the same form as the declarative case frames. Cars that partially or completely meet the user's specifications are then retrieved through the activation of car case knowledge sources whose preconditions recognize the similarity of that case to the problem requirements. In Broadway, the action taken by a case upon activation is to post the corresponding declarative case to a relevant case space on the blackboard. This posting in turn triggers explanation knowledge sources, which we do not describe further here except to say that stereotypical explanation patterns are applied to explain to the user why these cars are appropriate recommendations. In general, the action taken by each case knowledge source depends on the application, for example, to suggest a repair to a plan case (e.g., CHEF [Hammond, 1989]) or to supply an argument fragment (e.g., HYPO [Ashley, 1990], CABARET [Rissland and Skalak, 1991]).

Procedural Cases

Procedural representations are created by several meta-knowledge-sources, which encode the knowledge to create a case knowledge source for a given perspective on a case. For example (Fig. 2), Broadway applies a meta-knowledge-source that uses a modified Manhattan metric to create a similarity neighborhood based on several case features related to the perspective of economy: purchase price, repair frequency and repair cost. Each neighborhood is local partly because the size of the interval of similarity for each case feature depends on its value in that case. To take a simple example, the interval surrounding the price of an expensive car is larger than that around the price of a cheap car. We have also experimented with metrics that are local in that they are applicable only to certain types of cases, e.g., to sports cars. Perspectives are reflected in metrics and include any means of evaluating cases that corresponds to a particular way of assessing similarity [Ashley, 1989], [Kolodner, 1989] including, e.g., dimensions [Rissland, Valcarce and Ashley, 1984], [Ashley, 1990], signatures of feature values [Samuel, 1967], or the reasoning or explanation captured in a case [Barletta and Mark, 1988], [Branting, 1991]. To assess case similarity, Broadway currently uses two perspectives: a perspective based on figures for reliability and cost provided by Consumer Reports and a perspective based on comparison of the signatures of features of cases, resulting in 186 case knowledge sources.

Some Advantages and Shortcomings of this Representation

Local Neighborhoods and Indexing. Local similarity neighborhoods provide the framework to tailor the measurement of similarity to each case and to the region of the case space where the case resides. To take an example from a classical AI program, imagine a case-based system that plays checkers and stores board positions in a case library. [Samuel, 1967] applied polynomial evaluation functions to evaluate checkerboard positions, and noted improved performance when the game was divided into opening, middle-game, and end-game phases, with a different set of evaluation function coefficients used for each phase. As many as six game phases were used. In a CBR system that used similarity metrics based on these evaluation functions, the appropriate metric would be determined by the game phase, reflected in the location in case space of the cases under consideration.

The primary advantage of localized metrics stems from the practical and theoretical difficulties of capturing in a global approach to similarity the nuances of case similarity across all cases in the case base. Case-based retrieval mechanisms generally rely on a system-wide method or metric to compute case similarity. As a practical matter, it may be difficult to reflect in a single metric or global evaluation function the important differences among all cases and account for interactions between related features, but still avoid implicit comparison of features that are incommensurate. Also, entirely different means of assessing similarity may be required for different types of cases. We speculate that since local metrics need only work in a neighborhood of a case, spurious feature interactions or inappropriate comparisons may more likely be avoided. Since CBR is often useful in poorly understood, "weak theory" domains, a globally applicable similarity function may be hard to come by, as it implicitly would reflect strong knowledge about the domain that holds across features, cases, perspectives and contexts.

A second advantage to local metrics is the facility with which multiple views of a case can be captured (see, e.g., [Rissland, Valcarce and Ashley, 1984], [Ashley, 1989], [Kolodner, 1989]). Different perspectives are reflected in distinct similarity metrics that capture the varying importance to be accorded features when reasoning from diverse vantage points.

Additionally, exceptional cases may have unusual features that are known in advance and should be considered if relevant to one's specifications, but that are hard to incorporate into a global calculus of similarity. These exceptional features can be captured in a similarity metric local to the case.

A minor benefit of local similarity metrics may be in the observance of the software engineering principles of modularization and data encapsulation. Unusual or salient aspects of a case that are important to determining similarity can usefully be encoded in a metric local to the case.

Scheduling Granularity. Case retrieval may be scheduled at a fine level of granularity in this model. Case-based retrieval has sometimes been modeled and implemented as a monolithic action, "Search the case base for relevant cases and return them." See, e.g., [Rissland, Kolodner and Waltz, 1989] for a description of the classical control flow of CBR. A retrieval mechanism that is both large-grained and uninterruptible will potentially consume computing resources that may be applied more efficiently than to additional search of a large case base [Veloso and Carbonell, 1991]. For example, the cases initially retrieved may suggest a modification of the current case probe [Owens, 1989]. However, the current implementation of the model in Broadway does not schedule case knowledge source preconditions, and so does not exhibit this benefit. An extension to the blackboard control shell would be required to realize this advantage.

Consistent Hybrid Representation. In a hybrid system cases may be represented consistently with knowledge sources from other reasoning paradigms. A uniform representation supports the use of CBR as a component in a hybrid architecture where cases and other sources of expertise respond uniformly and cooperatively to progress and failure in problem-solving.

Shortcomings in the Implementation of the Model. A primary disadvantage of this implementation is that identifying a similarity rating with a knowledge source execution rating reduces a complex assessment to an information-losing numeric scale. Previous research on analogy and on case retrieval has cast doubt that similarity can be captured so simply, e.g., [Ashley, 1990], [Carbonell, 1986], [Falkenhainer et al., 1989], [Gentner, 1983], [Holyoak and Thagard, 1989]. On the other hand, on a serial computer, cases must be individually retrieved in some order, which implicitly ranks cases ordinally.

Secondly, it is not at all clear that every case should be proceduralized. Creating knowledge sources for rarely referenced cases or perspectives incurs computational overhead without apparent benefit. This problem will have to be addressed if case bases are to be scaled up to realistic levels, possibly consisting of tens or hundreds of thousands of cases [Schank, 1991]. Reserving procedural representation only for prototypical cases may present one way to deal with knowledge source proliferation.

Related Work

This project benefits from a long history of thought about the relative benefits of representing knowledge declaratively or procedurally, including [Anderson, 1983] (ACT), [Bobrow and Winograd, 1977] (frame-driven dialog), [Minsky, 1975] (procedural attachment), and [Schank and Abelson, 1977] (scripts). A more dynamic approach to cases was inspired by PANDEMONIUM [Selfridge, 1959]. Several systems have also used a blackboard architecture to combine CBR with other reasoning methods, but all have used a declarative representation for cases (FIRST [Daube and Hayes-Roth, 1988], PROLEXS [Oskamp et al., 1989] and ABISS [Rissland, et al. 1991].) The memory-based approach of [Stanfill and Waltz, 1986], which represents both rules and stored experience within the MBR paradigm, suggested the search for a uniform knowledge representation for hybrid systems with a CBR component.

Summary

Our approach tests the fit of cases and knowledge sources, reflecting an alternative model of case-based retrieval. This procedural, locally-indexed approach is characterized by similarity metrics that are local to cases, fine scheduling granularity for case retrieval, case-generated actions that are incorporated within cases themselves, and the uniform representation of knowledge in hybrid systems.

Acknowledgments

I thank Edwina Rissland, who also suggested the Consumer Reports automobile domain, Dan Corkill, Jamie Callan, Claire Cardie, Jody Daniels, Kevin Gallagher, Ellen Riloff, Zack Rubinstein, Dan Suthers and Richard Weiss. Thanks also to Blackboard Technology Group, Inc., Amherst, MA, for providing GBB v.2.0.

References

Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard.

Ashley, K. D. (1989). *Toward a Computational Theory of Arguing with Precedents: Accommodating Multiple Interpretations of Cases*. Vancouver, BC: ACM, 93-102.

Ashley, K. D. (1990). *Modelling Legal Argument: Reasoning with Cases and Hypotheticals*. Cambridge, MA: M.I.T.

Bareiss, E. R. (1989). *Exemplar-Based Knowledge Acquisition*. Academic Press.

Barletta, R. and Mark, W. (1988). Explanation-Based Indexing of Cases. *Proceedings, Case-Based Reasoning Workshop*. Clearwater Beach, FL: Morgan Kaufmann, 50-60.

Blackboard Technology Group (1991). *GBB Reference Manual, Version 2.0*. Amherst, MA: Blackboard Technology Group, Inc.

Bobrow, D. G. and Winograd, T. (1977). An Overview of KRL, a Knowledge

Representation Language. *Cognitive Science*, 1, 3-46.

Branting, L. K. (1991). *Integrating Rules and Precedents for Classification and Explanation: Automating Legal Analysis.* Ph.D. Thesis, AI Laboratory, Univ. Texas.

Carbonell, J. G. (1986). Derivational Analogy: A Theory of Reconstructive Problem Solving and Expertise Acquisition. In R. S. Michalski, J. G. Carbonell & T. M. Mitchell (Ed.), *Machine Learning: An Artificial Intelligence Approach*, 371-392. Los Altos, CA: Morgan Kaufmann.

Consumer Reports. (1991). *Annual Auto Issue, The 1991 Cars*, 56(4), 206-294.

Daube, F. and Hayes-Roth, B. (1988). FIRST: A Case-Based Redesign System in the BB1 Blackboard Architecture. *Proceedings, Case-Based Reasoning Workshop.* AAAI-88. St. Paul, MN.

Erman, L. D., Hayes-Roth, F., Lesser, V. R. and Reddy, D. R. (1980). The HEARSAY-II Speech Understanding System: Integrating Knowledge to Resolve Uncertainty. *Computing Surveys*, 12.

Falkenhainer, B., Forbus, K. and Gentner, D. (1989). The Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence*, 41, 1-63.

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, 7, 155-170.

Hammond, K. J. (1989). *Case-Based Planning: Viewing Planning as a Memory Task.* Boston, MA: Academic Press.

Holyoak, K. J. and Thagard, P. (1989). Analogical Mapping by Constraint Satisfaction. *Cognitive Science*, 13, 295-355.

Kolodner, J. L. (1983). Maintaining Organization in a Dynamic Long-Term Memory. *Cognitive Science*, 7(4), 243-280.

Kolodner, J. L. (1989). Judging Which is the "Best" Case for a Case-Based Reasoner. *Proceedings, CBR Workshop.* Pensacola Beach, FL: Morgan Kaufmann, 77-81.

Minsky, M. (1975). A Framework for the Representation of Knowledge. In P. Winston (Ed.), *The Psychology of Computer Vision* New York: McGraw-Hill.

Nii, H. P., Feigenbaum, E. A., Anton, J. J. and Rockmore, A. J. (1982). Signal-to-Symbol Transformation: HASP/SIAP Case Study. *AI Magazine*, 3(2), 23-35.

Oskamp, A., Walker, R. F., Schrickx, J. A. and Berg, P.H.v.d. (1989). PROLEXS, Divide and Rule: A Legal Application. *Proceedings, ICAIL-89*, Boston, MA: ACM.

Owens, C. (1989). Integrating Feature Extraction and Memory Search. *Proceedings, The 11th Annual Conference of the Cognitive Science Society.* Ann Arbor, MI: Lawrence Erlbaum.

Rissland, E. L., Basu, C., Daniels, J. J., McCarthy, J., Rubinstein, Z. B. and Skalak, D. B. (1991). A Blackboard-based Architecture for CBR: An Initial Report. *Proceedings, CBR Workshop.* Washington, DC: Morgan Kaufmann, 77-92.

Rissland, E. L., Kolodner, J. and Waltz, D. (1989). Case-Based Reasoning, Introduction. *Proceedings, CBR Workshop.* Pensacola Beach, FL: Morgan Kaufmann.

Rissland, E. L. and Skalak, D. B. (1991). CABARET: Rule Interpretation in a Hybrid Architecture. *International Journal of Man-Machine Studies*, 34, 839-887.

Rissland, E. L., Valcarce, E. M. and Ashley, K. D. (1984). Explaining and Arguing with Examples. *Proceedings, AAAI-84*, Austin, TX. AAAI.

Samuel, A. L. (1967). Some Studies in Machine Learning using the Game of Checkers II — Recent Progress. *IBM J. Research and Development*, 11, 601-617.

Schank, R. (1991). Where's the AI? *AI Magazine*, 12(4), 38-49.

Schank, R. and Abelson, R. (1977). *Scripts, Plans, Goals, and Understanding.* Hillsdale, NJ: Lawrence Erlbaum.

Selfridge, O. G. (1959). Pandemonium: A Paradigm for Learning. *Proceedings of the Symposium on the Mechanization of Thought Processes*, 511-529.

Stanfill, C. and Waltz, D. (1986). Memory-Based Reasoning. *Communications of the ACM*, 12(12), 1213-28.

Sycara, K. P. (1987). *Resolving Adversarial Conflicts: An Approach Integrating Case-Based and Analytic Methods.* Ph.D. Thesis, School of Information and Computer Science, Georgia Institute of Technology.

Veloso, M. M. and Carbonell, J. G. (1991). Variable-Precision Case Retrieval in Analogical Problem Solving. *Proceedings, Case-Based Reasoning Workshop, May 1991.* Morgan Kaufmann, San Mateo, CA.