The Knowledge Engineering Review, Vol. 00:0, 1–4. © 2005, Cambridge University Press
DOI: 10.1017/S000000000000000 Printed in the United Kingdom

Representation in case-based reasoning

RALPH BERGMANN[1], JANET KOLODNER[2] and ENRIC PLAZA[3]

[1] *University of Trier, Business Information Systems II, 54286 Trier, Germany*
E-mail: bergmann@uni-trier.de
[2] *Georgia Institute of Technology, Atlanta, Georgia 30032-0280, USA*
E-mail: janet.kolodner@cc.gatech.edu
[3] *IIIA - Institut d'Investigació en Intelligència Artificial, Campus UAB, 08193 Bellaterra, Catalonia, Spain*
E-mail: enric@iiia.csic.es

Abstract

A case in case-based reasoning is a contextualized piece of experience, which can be represented in various forms. Traditional approaches can be classified into three main categories: *feature vector representations*, *structured representations*, and *textual representations*. More sophisticated approaches make use of hierarchical representations or generalized cases. For particular tasks such as design and planning highly specific representations have been developed.

1 Introduction

Originally, the notion of case was that of a "problem solving episode", based on the cognitive science distinction between semantic and episodic memory. According to (Kolodner, 1993) a case is a "contextualized piece of knowledge representing an experience that teaches a lesson fundamental to achieving the goals of the reasoner". The experience a case represents can be structured in various ways. Very often it is only subdivided into a problem and a solution description, but additional knowledge might be necessary depending on the kind of intended reuse. For example, (Kolodner, 1993) proposes a comprehensive case structure consisting of the following five parts: (i) a situation and its goal, (ii) the solution and, sometimes, means of deriving it, (iii) the result of carrying it out, (iv) explanations of results, and (v) lessons that can be learned from the experience.

Case representation in case-based reasoning (CBR) makes use of familiar knowledge representation formalisms from AI to represent the experience contained in the cases for reasoning purposes. A large variety of representation formalisms have been proposed. However, three major types of case representation have arisen: feature vector (or *propositional*) cases, structured (or *relational*) cases, and textual (or *semi-structured*) cases. In addition, specialized representations of cases for specific tasks have been used, such as plans used as cases in planning.

Feature-vector approaches represent a case as a vector of attribute-value pairs, similar to the propositional representations used in Machine Learning (ML), that support *k*-nearest neighbor matching and instance-based learning (Aha *et al.*, 1991). The structured approach to case representation originates from the episodic memory notion of cases. The representational structure itself is usually developed around a frame-based formalism. Since frames can be seen as a subset of first-order logic, cases represented as frames or some frame-like structure are examples of what ML calls relational representations. Such representations arise from gathering together clusters of relations that occur together in elementary objects. Cases, from this perspective, are clusters of relations between the kinds of elementary objects that comprise them. Frame representations also strongly resemble object-oriented case representations (Bergmann, 2002, Chapter 3), that have a different origin and are often used in industrial applications. Finally, textual case representations (Lenz & Burkhard, 1996) go in the opposite direction by imposing only a weak structure on the cases. This allows easy exploitation of the experience captured in documents such as bug reports or FAQ's. Text fields are usually represented as sets of linguistic items (e.g. words reduced by some stemming algorithm) similar to vector representations in information retrieval. However, the available (weak) structure of the text allows the introduction of more semantics for a better interpretation of the reusability of a case.

Representation of cases and the way similarity is assessed during retrieval are strongly related to each other. Distance-based similarity metrics are easy to apply to feature vectors, while techniques related to Information Retrieval (IR) can easily be applied to textual representations. Frame-based cases often require knowledge-intensive indexing and matching algorithms.

2 Basic approaches to case representation

2.1 Content of a case

In a representative paper on case representation, "Improving human decision making through case-based decision aiding," Janet Kolodner (1991) focuses on the role cases can play in helping people make decisions and on the content cases need to have to play that role. Kolodner does not try to comment on the form that a case should take but rather focuses on the kinds of things that should be represented in a case so that it can be productively used for reasoning. Kolodner recommends that cases include a problem situation description, the solution that was proposed (often including how that solution was derived), and the outcome, including the state of the world after the solution was carried out, how close that was to what was expected, and explanations, if necessary and available, of why it might not have worked as well as expected.

2.2 Feature vector representation

The PROTOS system (Porter *et al.*, 1990) uses a feature vector approach for domains with weak or intractable theories. A category is extensionally represented as a collection of cases called *exemplars*. A new case is classified into a category if a match can be found between an exemplar and the new case. This matching process is knowledge intensive and tries to build an explanation that connects the features of the new case with an exemplar. Since each explanation is a path constructed inside a semantic net, retrieval is the process of explaining the (similarity) relation between a new case and an exemplar. Unlike most early CBR systems that use feature vector representations, PROTOS already uses a knowledge-intensive similarity measure.

2.3 Frame-based representation

Frame based representations have been (partially) formalized by description logics. The notion of "cases as terms" (Plaza, 1995) argues that viewing structured cases as terms in feature logics (a particular brand of description logics) helps to better understand several aspects of case-based reasoning. Domain knowledge can be integrated using a sort hierarchy and the issue of composite cases (cases that group together other objects or sub-cases) is understood by the fact that a sub-term is also a term. Finally, the notion of similarity between two cases is linked to the concepts of subsumption and anti-unification of terms. This notion also bridges relational cases in CBR with relational learning in inductive ML where subsumption and anti-unification are basic building blocks.

2.4 Object-oriented representation

Object-oriented case representations (Bergmann, 2002, Chapter 3) have an expressiveness similar to frame representations, but have a different origin. They make use of the data modeling approach of the object-oriented paradigm, including *is-a* and *part-of* relations as well as the inheritance principle. Cases are represented as collections of objects, each of which is described by a set of attribute-value pairs. The structure of an object is described by an object class. Several object-oriented case representation languages have been developed. CASUEL (Manago *et al.*, 1994) is an early example in plain ASCII, while more recent languages are XML compatible. Object-oriented representations are particularly suitable for complex domains in which cases with different structures occur. An impressive example is described by Göker & Roth-Berghofer (1999).

2.5 Textual representation

Textual case representations, as described by Lenz & Burkhard (1996), decompose the text that constitutes a case into information entities (IEs). An IE is a word or a phrase contained in the text that is relevant to determine the reusability of the episode captured in the case. Given a vocabulary of the relevant words or phrases, text cases can be mined for IEs, allowing case acquisition to be automated. The set of cases that form the case base is organized in the form of a case retrieval net (CRN), which is a directed graph with nodes representing cases and their IEs. These nodes are linked according to their similarity. Hence knowledge about similarity is encoded into the strength of the links between the nodes in the CRN. Case retrieval is similar to activation propagation in a neural network: the IEs that occur in the current problem are activated and this initial activation is propagated through the case retrieval net according to the similarity-based link strength.

3 Advanced approaches

3.1 Hierarchical case representation

The previously discussed approaches typically represent cases at a single level of abstraction. However, in recent CBR publications, the use of multiple representations at different levels of abstraction has been investigated (Bergmann & Wilke, 1996). The basic idea behind these approaches is to represent a case at multiple levels of detail, possibly using multiple vocabularies. When a new problem must be solved, similar cases at appropriate levels of abstraction are retrieved from the case base and the solutions from these cases are combined and refined.

JULIA (Hinrichs, 1992; Kolodner, 1993) provides one such hybrid representation for cases. It begins with a hierarchical frame-based representation that distinguishes between different parts of a complex case. Connecting the parts of a case, constraints show how they are related to each other. An advantage of this representation is that it allows a whole case or its parts to be accessed and used by the case-based reasoner, and the constraints can be used to guide adaptation.

3.2 Generalized cases

In the previous discussion, a case is regarded as a single experience, e.g. representing a point in the case space. In contrast, a generalized case covers a subspace of the representation space. A single generalized case immediately provides solutions to a set of closely related problems rather than to a single problem only. Kolodner (1983), Zito-Wolf & Alterman (1992), and Bergmann & Vollrath (1999) discuss this issue from different perspectives. In Zito-Wolf & Alterman's work, for example, a single case contains a variety of alternative plans that are available for achieving a common goal. From a technical point of view, generalized cases are represented by introducing variables in a traditional case representation approach. Allowed assignments to these variables can be further restricted by introducing constraints in the case representation.

3.3 Cases in case-based design

Specific case representations have been developed for particular tasks like planning or design. In design, for instance, the DRAMA system (Leake & Wilson, 1999) uses CMaps (Concept Maps) to represent aircraft design cases. A CMap is a semi-structured case containing both a structural aspect (similar to semantic nets to describe aircraft components) and a textual aspect (to document design decisions and rationale). The FABEL project developed CBR support systems for architect firms in the design of buildings (Gebhardt *et al.*, 1997). The representation of cases is based on CAD objects, which are described by coordinates, aspect, granularity, and size. On top of these objects, a FABEL case consists of a problem, a plan (a solution), and a collection of solution paths or explanations. The complexity of this specialized case representation gave rise to new or adapted techniques for retrieval and reuse (Gebhardt *et al.*, 1997). Hinrichs' hierarchical representations in JULIA (Hinrichs, 1992; Kolodner, 1993) were also developed to support case-based design.

3.4 Cases in case-based planning

Case representations for case-based planning (see Cox *et al.* this issue) are sometimes more specialized than other case representations due to the specific structure of problems and solutions in planning. By developing the first case-based planner (CHEF), Hammond (1989) proposed a case-based approach that uses a plan-like representation of cases. In planning, a problem is typically described by an initial state and a goal state. A solution is a totally or partially ordered sequence of actions. In principle, a planning case contains such a problem description and its related plan. The representation formalisms used in general AI planning significantly influence the case representations used. States are usually represented as sets of propositions in a logic language. The solution plan consists of a set of actions described, for instance, by operator terms, together with an ordering relation on them. A crucial step in case-based planning is the reuse phase, which requires strong plan adaptation capabilities. For plan adaptation, derivational analogy has demonstrated several advantages over transformational adaptation approaches. However, this method requires that additional information about successful and/or failed planning decisions be recorded as part of the case representation. Carbonell (1986) was the first to propose storing a complete derivational trace of the decision process in a planning case. Each derivational trace is augmented with an explicit justification structure explaining each individual decision.

References

Aha, D, Kibler, D & Albert, M, 1991. "Instance-based learning algorithms," *Machine Learning* 6:37 – 66.

Bergmann, R, 2002. *Experience Management: Foundations, Development Methodology, and Internet-Based Applications*, Berlin: Springer.

Bergmann, R & Wilke, W, 1996. "On the role of abstraction in case-based reasoning," in Smith, I & Faltings, B (eds.) *Proceedings of the 3rd European Workshop on Case-Based Reasoning*, Berlin: Springer, pp. 28–43.

Bergmann, R & Vollrath, I, 1999. "Generalized cases: Representation and steps towards efficient similarity assessment," in Burgard, W, Christaller, T & Cremers, AB (eds.) *KI-99: Advances in Artificial Intelligence*, Berlin: Springer, pp. 195–206.

Carbonell, JG, 1986. "Derivational analogy: A theory of constructive problem solving and expertise acquisition," in Michalski, RS, Carbonell, JG & Mitchell, TM (eds.) *Machine Learning: An Artificial Intelligence Approach (Volume II)*, Los Altos, CA: Morgan Kaufmann, pp. 371 – 392.

Gebhardt, F, Voß, A, Gräther, W & Schmidt-Belz, B, 1997. *Reasoning with Complex Cases*, Kluwer Academic Publishers.

Göker, M & Roth-Berghofer, T, 1999. "The development and utilization of the case-based help-desk support system HOMER," *Engineering Applications of Artificial Intelligence* 12(6):665–680.

Hammond, KJ, 1989. *Case-Based Planning: Viewing Planning as a Memory Task,* Boston: Academic Press.

Hinrichs, TR, 1992. *Problem Solving in Open Worlds: A Case Study in Design*, Northvale, New Jersey: Erlbaum.

Kolodner, J, 1983. "Maintaining organization in a long term dynamic memory," *Cognitive Science* 7:243–280.

Kolodner, J, 1991. "Improving human decision making through case-based decision aiding," *Artificial Intelligence Magazine* 12(2):52–68.

Kolodner, J, 1993. *Case-Based Reasoning*, San Mateo, California: Morgan Kaufmann.

Leake, D & Wilson, DC, 1999. "Combining CBR with interactive knowledge acquisition," in Althoff, KD, Bergmann, R & Branting, LK (eds.) *Proceedings of the 3rd International Conference on Case-Based Reasoning*, Berlin: Springer, pp. 203–217.

Lenz, M & Burkhard, HD, 1996. "Case retrieval nets: Basic ideas and extensions," in Görz, G & Hölldobler, S (eds.) *KI-96: Advances in Artificial Intelligence*, Berlin: Springer, pp. 227–239.

Manago, M, Bergmann, R, Wess, S & Traphöner, R, 1994. *CASUEL: A Common Case Representation Language*, ESPRIT Project INRECA, No. 6322, Deliverable D1, University of Kaiserslautern, Kaiserslautern Germany.

Plaza, E, 1995. "Cases as terms: A feature term approach to the structured representation of cases," in Veloso, M & Aamodt, A (eds.) *Proceedings of the 1st International Conference on Case-Based Reasoning*, Berlin: Springer, pp. 265–276.

Porter, BW, Bareiss, R & Holte, RC, 1990. "Concept learning and heuristic classification in weak-theory domains," *Artificial Intelligence* 45:229–264.

Zito-Wolf, R & Alterman, R, 1992. "Multicases: A case-based representation for procedural knowledge," in *Proceedings of the 14th Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum, pp. 331–336.