In: W. Dilger, M. Schlosser, J. Zeidler & A. Ittner (eds.), *Proc. of the 9th German Workshop on Machine Learning* (GI Special Interest Group 1.1.3), Chemnitzer Informatik-Berichte, University of Technology Chemnitz, pp. 6-11

An Evaluation of the INRECA CBR System

Klaus-Dieter Althoff, Karl-Heinz Weis jun.
Centre for Learning Systems and Applications (LSA)
Department of Computer Science
University of Kaiserslautern
D-67653 Kaiserslautern, Germany

We present an approach to systematically describing case-based reasoning systems by different kinds of criteria. One main requirement was the practical relevance of these criteria and their usability for real-life applications. We report on the results we achieved from a case study carried out in the INRECA[1] Esprit project.

Problem

In case-based reasoning (CBR) a huge number of systems (including commercial systems) has been developed so far and is steadily increasing. While this very dynamic situation corresponds to the situation in other computer science subfields like e.g. programming languages, the question arises whether a systematic description of CBR systems exists (again like for programming languages). The answer is „yes“: based on a number of case studies we developed a systematic description for CBR systems which includes a high number of different criteria like *technical criteria* dealing with the limitations and abilities of the CBR systems, *ergonomic criteria* concerning the consultation of the execution system and application development, *application domain criteria* dealing with concept structure, knowledge sources, and knowledge base characteristics as well as *application task criteria* like integration of reasoning strategies, decomposition methods, and task properties.

One main requirement was the practical relevance and applicability of these criteria. We used them for a systematic comparison (evaluation) of the INRECA research prototype system with commercial CBR tools as well as CBR-related research prototype systems (Althoff, 1996). However, the criteria should also be helpful for application development, the integration of different methods, and tool development.

The underlying intention of the criteria set can be best described as to give decision support for answering a number of questions (we therefore call these criteria *decision support criteria* as suggested in Althoff & Bartsch-Spörl, 1996). Important questions, which have been raised first by Cohen (1989) from the more general point of view of evaluating CBR research, are listed below. Though historically these questions have not been the starting point for our evaluation work, they are helpful for giving some structured overview of the used criteria. If we take e.g. the question „How is the CBR method to be evaluated an improvement, an alternative, or a complement to other methods?“, then we can derive the following sub-questions, which aim at characterising the associated system aspects (among others):

- Does the method account for more situations?
  Important aspects are the representation of general knowledge, the representation of structured cases, (fully) object-oriented case and knowledge representation, processing of symbolic and numeric data etc.

[1] Induction and reasoning from cases

- Does the method produce a wider variety of desired behaviour?
  Are for instance different retrieval strategies possible, can they be influenced by means of global or local parameters; can the different strategies be combined, or is it possible to improve the retrieval strategy through learning.
- Is the method more efficient in time or space?
  Interesting characteristics are the performance of the application system, the space requirements for main memory and disk, the time needed to generate the application system, the time necessary to load the case base etc.
- Does the method model human behaviour in a more useful way/detail?
  Is it possible to support a systematic acquisition of missing information during problem solving; can the similarity assessment mechanism be represented in a way that is understandable for the user/expert.

Approach

The general approach we took for evaluating the INRECA CBR system can be divided into three main tasks, namely *describing* CBR systems, domains and application tasks using a set of decision support criteria, *analysing* the underlying methods and domain/task characteristics by relating domain/task criteria with technical/ergonomic criteria with the additional use of a task-method decomposition hierarchy for CBR (Aamodt & Plaza, 1994), and *developing* abstracted algorithms based on the extracted methods and knowledge modelling techniques underlying the analysed CBR-related systems.

We view CBR system evaluation as a systematic organisation and presentation of information on the systems – not just a ranking –, using a mixture of qualitative criteria and formally defined experiments if running systems are available, combining technical, ergonomic, domain, and application task criteria, carrying out case studies, explaining and publishing the criteria (Althoff, Auriol et al. 1995; Althoff, 1996), etc.

We used quantitative, domain-dependent criteria for measuring the classification accuracy on noisy/incomplete data for training/query cases, retrieval speed, building speed for the case library etc. Qualitative, domain-dependent criteria were used for characterising a CBR system's ability to handle structured domains, to solve simple similarity problems, to handle weak theories, to handle open domains etc. Qualitative, domain-independent criteria were applied to describe a CBR system's capabilities for supporting data acquisition/maintenance and validation, the effectiveness of the execution system, the adaptivity of the development system etc.

For analytical purposes we associated qualitative domain and application task criteria with CBR systems/methods. In addition, we analysed „CBR applications“ by associating technical criteria with a concrete domain/task. We also used the task-method decomposition model for CBR suggested by Aamodt and Plaza (1994) to classify CBR methods within this model.

We developed different algorithms to extend the task-method decomposition model for CBR with additional task-execution methods. These algorithms have been abstracted from the analysed systems. We developed specialised CBR systems/system components as well as real-life applications to achieve additional feedback for the correspondence of technical/ergonomic criteria and domain/application task criteria.

Realisation

We compared INRECA with five industrial CBR tools, namely CBR EXPRESS/CBR 2 (Inference, USA), ESTEEM (Esteem Software, USA), KATE tools (AcknoSoft, France), ReMind (Cognitive Systems, USA) and S^3-CASE (tecInno, Germany), and about twenty CBR-related research prototype systems (CREEK: Aamodt, 1991; PROTOS, Bareiss, 1989;

CASEY: Koton, 1989; CcC+: Goos, 1995; PATDEX/1: Stadler & Wess, 1989; PATDEX/2: Wess, 1993; MOLTKE: Althoff, 1992; MoCAS/1: Pews & Weiler, 1992; MoCAS/2: Pews, 1994; KoDIAG: Rahmel & von Wangenheim, 1994; S^3+/INRECA: Priebisch, 1995; AWACS: Faupel, 1992; KALES: De la Ossa, 1992; CHEF: Hammond, 1989; PRODIGY/ANALOGY: Veloso, 1992; FABEL: Voß, Bartsch-Spörl et al., 1996; GenRule: Althoff, 1992; iMAKE: Rehbold, 1991; WheelNet: Weis, 1994; Weiß, 1993 a.o.).

We compared INRECA with (additional) related work: based on a list of 26 topics about 300 papers have been collected and classified into four different categories according to their respective interestingness with a special focus on the period of 1991-1995 (Althoff, 1996).

In the scope of the INRECA project we developed four real-life applications: help-desk for robot diagnosis, maintenance system for supporting troubleshooting of aircraft engines, assessing wind risk factors for Irish forests, and reuse of object-oriented software (Lenz, Auriol et al., 1996; Auriol, Manago & Guyot-Dorel, 1996; Bergmann & Eisenecker, 1995; Althoff, 1996).

We developed the model-based case adaptation system MoCAS (Pews & Wess, 1993; Pews, 1994; Bergmann, Pews & Wilke, 1994), the integrated S^3+/INRECA system (Priebisch, 1995; Althoff, 1996), and the seamless integration level for induction and CBR in INRECA (Wess, 1995; Althoff, Bergmann et al., 1996).

We abstracted 40 algorithms from the analysed CBR-related research systems as additional task-execution methods for the task-method decomposition model of Aamodt and Plaza (1994) and explicitly related these algorithms to the INRECA system (Weis, 1995; Althoff, 1996).

Advantages and Results

Besides enabling the evaluation of the INRECA research prototype system and giving an systematic overview of CBR-related methods for diagnosis and decision support, our approach has two additional advantages, namely giving decision support for developing applications of CBR technology by building a (case-based) information system on CBR applications/systems (Althoff & Bartsch-Spörl, 1996; Meissonnier, 1996) as well as by going first concrete steps towards building an analytic framework for relating case-based problem solving and learning methods with domain/task characteristics as suggested by Althoff and Aamodt (1996).

Our evaluation work resulted in a review of industrial CBR tools and a documented example how to evaluate CBR systems and applications, where the latter also can be viewed as a systematically organised documentation of CBR-related methods for decision support and diagnosis. Further important results are an improved version of the final INRECA system, namely the seamless integration of induction and CBR, as well as improved versions of related commercial CBR tools (KATE tools version 4 and later; CBR-Works).

Scientific Grounding

For an introductory view on evaluation we would like to cite Paul Cohen and Adele Howe (1988a), who wrote an article on how evaluation guides AI research:

„Evaluation means making observations of all aspects of one's research. Often, we think of evaluation only in terms of how well AI (in our case: CBR) systems perform, yet it is vital to all stages of research, from early conceptualisation to retrospective analyses of series of programmes. Indeed, some of the most informative observations are not performance measures but, rather, describe why we are doing the research, why our tasks are particularly illustrative, why our views and methods are a step forward, how completely they are implemented by our programmes, how these programmes work, whether their performance is likely to increase or has reached a limit (and why), and what problems we encounter at each stage of our research. To the research community, these observations are more important than performance measures because they tell us how research should proceed."

How evaluation of knowledge-based systems can be viewed from the perspective of current social and economical science is described in Kirchhoff (1993). Here a *system* is specified as an ordered set of elements which are connected to one another. Internal relations are called *structure of the system*, the external relations are called *behaviour*. According to Kirchhoff (1993), the evaluation needs to begin with the knowledge base, which represents the interpretation of the knowledge engineer of the real world expert knowledge. For that purpose a kind of so-called content validation is proposed. Then empirical evaluation should continue with the directly implemented expert knowledge. The validity of all the system directly depends on two items they call *objectivity* and *reliability*, which means that results should be independent of the examiner and that a certain degree of exactness should be measured. Further Kirchhoff (1993) states that the developed evaluation methodology, which bases on classical expert systems, may be – from the point of view of real expert problem solving methodology – used also by developers of CBR or induction-based systems.

Both views underline that evaluation is necessary. However, how could we come up with a methodology for evaluating CBR systems? On the level the methodology is reported in Kirchhoff (1993) the suggestion that the reported methodology can be also applied to systems based on CBR and/or induction is somehow understandable. However, we do not think that this is the right way of generalising. A methodology needs to be more focused and specialised. Here we agree with Cohen and Howe (1988b) who argue that methods known from the, as they say, behavioural sciences are inappropriate for AI in general and, thus, CBR in particular. In Cohen and Howe (1988a) they suggest that we must develop our own evaluation criteria, experiment designs, and analytic tools according to the specific area in which we want to achieve results. Explorations along these lines have been reported for areas like machine learning (Langley, 1987), distributed systems (Decker, Durfee & Lesser, 1988), and knowledge-based systems (Geissman & Schultz, 1988; Gaschnig, Klahr et al., 1983; Rothenberg, Paul et al., 1987). Although these papers address different facets of evaluation, a common theme is that we must develop our own evaluation methods appropriate to the practise in our field (in our case: CBR).

In INRECA we focused on CBR and related methods – not on knowledge-based systems in general. In addition, we restricted the scope of our applications to analytic tasks. We believe that this was of central importance and allowed us to identify commonalties and differences with other approaches in the INRECA and MOLTKE system families (Althoff, 1992; Wess, 1995) or with approaches that clearly relate to INRECA because of the used methods (Althoff, 1996).

Future Activities

The results of the evaluation work in the INRECA project contribute to current activities on analysing CBR systems and applications (Althoff & Bartsch-Spörl, 1996; Althoff & Aamodt, 1996; Althoff, Richter & Wilke, 1996) and on developing practical guidelines for users of CBR technology (Althoff, Bergmann et al., 1995; Manago, Althoff et al., 1995; Wilke, Bergmann & Althoff, 1996).

Acknowledgement

Funding for INRECA has been partially provided by the Commission of the European Union (Esprit contract no. 6322) and to which the authors are greatly indebted. The partners of INRECA are AcknoSoft (prime contractor, France), tecInno (Germany), Interactive Multimedia Systems (Ireland), the University of Kaiserslautern (Germany). From the standpoint of developing a methodology for CBR applications partial funding has also been provided by the Stiftung Rheinland-Pfalz für Innovation (WiMo: „Knowledge Acquisition and Modelling for Case-Based Learning"). We would like to thank *Michael M. Richter* who initiated our evaluation work and *Stefan Wess* who contributed to our work with his competent and practical view on case-based reasoning. The authors are indebted to *Harald Holz*, *Carsten Priebisch*, *Alexandre Meissonnier*, and *Wolfgang Wilke* who heavily supported us in putting the evaluation criteria to use. *Eric Auriol, Ralph Bergmann, Roy Johnston, Michel Manago, Ralph Traphöner,* and *Wolfgang Wilke* gave us valuable feedback from a number of different applications they are familiar with. We also would like to thank *Agnar Aamodt* and *Brigitte Bartsch-Spörl* for critical feedback and helpful comments on our evaluation work.

References

Aamodt, A. (1991). *A Knowledge-Intensive, Integrated Approach to Problem Solving and Sustained Learning*. Ph. D. Thesis, University of Trondheim, Department of Computer Science and Telematics

Aamodt, A. & Plaza, E. (1994). Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches. *AI Communications Vol. 7*, *No. 1*, 39-59

Althoff, K.-D. (1992). *Eine fallbasierte Lernkomponente als integrierter Bestandteil der MOLTKE-Werkbank zur Diagnose technischer Systeme*. Doctoral Dissertation, University of Kaiserslautern; also: Sankt Augustin, Germany: infix Verlag

Althoff, K.-D. (1996). *Evaluating Case-Based Reasoning Systems: The INRECA Case Study*. Habilitationsschrift, University of Kaiserslautern (submitted)

Althoff, K.-D. & Aamodt, A. (1996). Zur Analyse fallbasierter Problemlöse- und Lernmethoden in Abhängigkeit von Charakteristika gegebener Aufgabenstellungen und Anwendungsdomänen. *KI 1/96, special issue on case-based reasoning* (edited by B. Bartsch-Spörl and S. Wess), 8-15

Althoff, K.-D., Auriol, E., Barletta, R. and Manago, M. (1995). *A Review of Industrial Case-Based Reasoning Tools*. AI Intelligence, Oxford, UK

Althoff, K.-D. & Bartsch-Spörl, B. (1996). Decision Support for Case-Based Applications. In: *Wirtschaftsinformatik 1/96*, special issue on case-based decision support (edited by D. Ehrenberg), 8-16

Althoff, K.-D., Bergmann, R., Maurer, M., Richter, M. M., Traphöner, R. and Wilke, W. (1995). *Wissensmodellierung und -akquisition für Case-Based Learning*. Project proposal, Stiftung Rheinland-Pfalz für Innovation

Althoff, K.-D., Bergmann, R., Wess, S., Manago, M. , Auriol, E., Larichev, O. I., Bolotov, A. & Gurov, S. I. (1996). *Integration of Induction and Case-Based Reasoning for Critical Decision Support Tasks in Medical Domains: The Inreca Approach*. Centre for Learning Systems and Applications, LSA-96-93E, University of Kaiserslautern

Althoff, K.-D., Richter, M. M. & Wilke, W. (1996). Case-Based Reasoning: A New Technology for Experienced Based Construction of Knowledge Systems. (*forthcoming*)

Auriol, E., Manago, M. & Guiot-Dorel, J. (1996). Cassiopée: Fehlerdiagnose von CFM 56-3 Triebwerken für Boeing 737 Flugzeuge. *KI 1/96, special issue on case-based reasoning* (edited by B. Bartsch-Spörl and S. Wess), 47-53

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

Bergmann, R. & Eisenecker, U. (1995). Fallbasiertes Schließen zur Unterstützung der Wiederverwendung objektorientierter Software: Eine Fallstudie. In: M. M. Richter & F. Maurer (eds.), *Expertensysteme 95*, infix Verlag

Bergmann, R., Pews, G. & Wilke, W. (1994). Explanation-Based Similarity: A Unifying Approach for Integrating Domain Knowledge into Case-Based Reasoning for Diagnosis and Planning Tasks. In: S. Wess, K.-D. Althoff & M. M. Richter (eds.), *Topics in Case-Based Reasoning*, Springer Verlag, 182-196

Cohen, P. R. (1989). Evaluation and Case-Based Reasoning. In: K. Hammond (ed.), *Proc. of the 2nd darpa Workshop on Case-Based Reasoning*, Morgan Kaufmann, 168-172

Cohen, P. R. & Howe, A. E. (1988a). How Evaluation Guides AI Research. *AI Magazine 9*, No. 4, Winter 1988, 35-43

Cohen, P. R. & Howe, A. E. (1988b). Toward AI Research Methodology: Case Studies in Evaluation. *IEEE Transactions on Systems, Man, and Cybernetics*

De la Ossa, A. (1992). *Knowledge Adaptation: Analogical Transfer of Empirical Knowledge Across Domains for Case-Based Problem Solving*. Doctoral Dissertation, University of Kaiserslautern

Decker, K. S., Durfee, E. H. & Lesser, V. R. (1988). Evaluating Research in Cooperative Distributed Problem Solving. *Dept. of Computer and Information Science, TR 8889*, University of Masschusetts

Faupel, B. (1992). *Ein modellbasiertes Akquisitionssystem für technische Diagnosesysteme*. Doctoral Dissertation, RWTH Aachen

Gaschnig, J., Klahr, P., Pople, H., Shortliffe, E. & Terry, A. (1983). Evaluation of Expert Systems: Issues and Case Studies. In: *Hayes-Roth, Waterman & Lenat (1983)*, 241-282

Geissman, J. R. & Schultz, R. D. (1988). Verification and Validation. *AI Expert 3*, 2, 26-33

Goos, K. (1995). *Fallbasierte Klassifikation in der Diagnostik*. Doctoral Dissertation, Universität Würzburg

Hammond, K. (1989). *Case-Based Planning*. Academic Press

Hayes-Roth, F., Waterman, D. A. & Lenat, D. B. (1983). *Building Expert Systems*. Reading, Massachusetts: Addison Wesley

Kirchhoff, S. (1993). *Abbildungsqualität von wissensbasierten Systemen – Eine Methodologie zur Evaluierung*. Bergisch Gladbach, Köln: Josef Eul Verlag

Koton, P. (1989). *Evaluated case-based problem solving*. Ph.D. thesis, Laboratory of computer science, MIT/LCS/TR-441

Langley, P. (1987). Research papers in machine learning. *Machine Learning 2*, 3, 195-198

Lenz, M., Auriol, E., Burkhard, H.-D., Manago, M. & Pirk, P. (1996). Fallbasierte Diagnostik und Entscheidungsunterstützung. *KI 1/96, special issue on case-based reasoning* (edited by B. Bartsch-Spörl and S. Wess), 16-21

Manago, M., Althoff, K.-D., Auriol, E., Bergmann, R., Breen, S., Richter, M. M., Stehr, J., Traphöner, R. and Wilke, W. (1995). *INRECA II: Information and Knowledge Reengineering for Reasoning from Cases*. Project proposal, Esprit IV

Meissonnier, A. (1996). *Ein fallbasiertes Informationssystem für fallbasierte Anwendungen und Systeme auf der Basis von Inreca*. Diploma Thesis, University of Kaiserslautern (forthcoming)

Pews, G. & Weiler, F. (1992). *MoCAS - ein Ansatz zur falladaptiven Diagnostik*. Project thesis, University of Kaiserslautern

Pews, G. & Wess, S. (1993). Combining Case-Based and Model-based Approaches for Diagnostic Applications in Technical Domains. In: M. M. Richter, S. Wess, K.-D. Althoff & F. Maurer (eds.), *Proc. of the First European Workshop on Case-Based Reasoning*, Seki-Report SR-93-12, University of Kaiserslautern, 325-328

Pews, G. (1994). *MoCAS/2 - Fallbasierte Diagnose in technischen Domänen*. Diploma Thesis, University of Kaiserslautern

Priebisch, C. (1995). *Integration von fallbasierter und kontextbasierter Diagnose am Beispiel der Systeme INRECA und* S^3. Diploma Thesis, University of Kaiserslautern

Rahmel, J. & von Wangenheim, A. (1994). KoDIAG: a Connectionist Expert System. *International Symposium on Integrating Knowledge and Neural Heuristics,* Pensacola

Rehbold, R. (1991). *Integration modellbasierten Wissens in technische Diagnostik-Expertensysteme*. Doctoral Dissertation, University of Kaiserslautern

Rothenberg, J., Paul, J., Kameny, I., Kipps, J. R. & Swenson, M. (1987). Evaluating Expert System Tools: A Framework and Methodology. *Technical report, R-3542*, DARPA, Rand Corporation

Stadler, M. & Wess, S. (1989). *PATDEX: Konzept und Implementierung eines fallbasierten, analogieorientierten Inferenzmechanismus zur Diagnose eines CNC-Bearbeitungszentrums*. Project thesis, University of Kaiserslautern

Veloso, M. M. (1992). *Learning by Analogy and Reasoning in General Problem Solving*. Ph.D. thesis, Carnegie-Mellon-University

Voß, A., Bartsch-Spörl, B., Hovestadt, L., Jantke, K. P., Petersohn, U. & Strube, G. (1996). FABEL. To appear in *KI*

Weis, K.-H. jun. (1994). Aufbau topologischer Weltmodelle mit neuronalen Netzen. In: *Proc. Seminar „artificial life & real robots",* University of Kaiserslautern

Weis, K.-H. jun. (1995). *Bewertung integrierter fallbasierter Systeme*. Diploma Thesis, University of Kaiserslautern

Weiß, S. (1993). *Modellbasiertes Lernen zur Wissensakquisition für technische Diagnosesysteme*. Doctoral Dissertation, Carl-Hanser Verlag

Wess, S. (1993). PATDEX - ein Ansatz zur wissensbasierten und inkrementellen Verbesserung von Ähnlichkeitsbewertungen in der fallbasierten Diagnostik. In: F. Puppe & A. Günter (eds.), *Expertensysteme 93*, Springer Verlag, 42-55

Wess, S. (1995). *Fallbasiertes Schließen in wissensbasierten Systemen zur Entscheidungsunterstützung und Diagnostik*. Doctoral Dissertation, University of Kaiserslautern; also: Sankt Augustin: infix Verlag

Wilke, W., Bergmann, R. & Althoff, K.-D. (1996). Fallbasiertes Schließen zur Kreditwürdigkeitsprüfung. To appear in: *KI 3/96 – special issue on AI in financial business* (edited by S. Kirn & C. Weinhardt)