Goals and Kinds of Explanations in Case-Based Reasoning

Thomas R. Roth-Berghofer[1], Jörg Cassens[2], and Frode Sørmo[2]

[1] Knowledge-Based Systems Group, Department of Computer Science,
University of Kaiserslautern, P.O. Box 3049, 67653 Kaiserslautern
and
Knowledge Management Department
German Research Center for Artificial Intelligence DFKI GmbH,
Erwin-Schrödinger-Straße 57, 67663 Kaiserslautern, Germany
thomas.roth-berghofer@dfki.uni-kl.de

[2] Norwegian University of Science and Technology (NTNU),
Department of Computer and Information Science (IDI),
7491 Trondheim, Norway
{frode.sormo|jorg.cassens}@idi.ntnu.no

Abstract. Research on explanation in Case-Based Reasoning (CBR) is a topic that gains momentum. In this context, fundamental issues on what are and to which end do we use explanations have to be reconsidered. This article presents a prelimenary outline of the combination of two recently proposed classifications of explanations based on the type of the explanation itself and user goals which should be fulfilled.

1 What do We Need Explanations for?

The more complex knowledge-based systems get, the more explanation capabilities the users expect when using such systems. This requirement was recognized early on in expert systems research and development [1,2]. Considerable results were produced, but research activity decreased together with the general decline of expert systems research in the nineties of the last century.

At the same time there was an increasing interest on this topic in Case-Based Reasoning (CBR) [3,4]. At the turn of the century, we find the issue discussed in the context of knowledge-based systems [5,6]. Recently, we can see a renewed focus in CBR on this track of research. ECCBR 2004 featured for example a workshop on Explanation in Case-Based Reasoning as well as a couple of papers on explanation at the main conference [7,8].

It is important to note that the term explanation can be interpreted in two different ways. One interpretation deals with explanations as part of the reasoning process itself. The other interpretation deals with usage aspects: making the reasoning process, its results, or the usage of the result transparent to the user. In this paper, we will focus on the second interpretation.

2 Views on Explanations

In this section, we outline two perspectives on explanation: The *Explanation Goals* focus on user needs and expectations towards explanations and help to understand *what* the system has to be able to explain and *when* to explain something. The *Kinds of Explanations* focus on different *types* of explanations, their *usefulness* for the user, and how they can be represented in the different *knowledge-containers* [9].

2.1 Explanation Goals

Sørmo and Cassens [10] suggest several explanation goals for Case-Based Reasoning systems (which are valid for knowledge-based systems in general). They also argue that those goals are indeed reachable because case-based reasoners are mostly made to perform limited tasks for a limited audience, thus allowing to make reasonable assumptions about the user's goals and the explanation context. The identified explanation goals are:

Justification: Explain why the answer is a good answer
This goal allows for a simplification of the explanation compared to the actual process the system goes through to find a solution. It will even allow a posteriori explanations formed after the solution is found, i.e., explanations that have nothing to do with how the reasoner came up with the answer.

Transparency: Explain how the system reached the answer
This goal is subtly different from the previous goal in that it seeks to impart an understanding of how the system found the answer. This allows the users to control the system's quality by examining the way it reasons and allows them to look for explanations for why the system has reached a surprising anomalous result.

Relevance: Explain why a question asked is relevant
In conversational systems, the user may wish to know why a question asked by the system is relevant to the task at hand. An explanation of this type would have to justify the strategy pursued by the system and why a question is relevant in this strategy.

Learning: Teach the user about the domain
In Intelligent Tutoring Systems, the goal often is not only to find a good solution to a problem, but to explain the solution to the user in a way that will increase his understanding of the domain. The goal can be to teach more general domain theory or tutor the user in solving problems similar to those solved by the system.

These user goals can guide us in identifying what the CBR system should be able to give as an explanation when starting from a user model. But how and where can the necessary knowledge be represented in the system?

2.2 Kinds of Explanations

Roth-Berghofer [11] looks at explanations from a knowledge-container perspective. He addresses the issue of what can naturally be explained by the four containers (vocabulary, similarity measures, adaptation knowledge, and case base).

One starting point is the work of Spieker [12] on the usefulness of explanations. According to Spieker, there are five useful kinds of explanations in the context of expert systems:

Conceptual Explanations are of the form 'What is ...?' or 'What is the meaning of ...?'. The goal of this kind of explanation is to build links between unknown and known concepts. Conceptual explanations can take different forms: Definition, Theoretical Proposition, Prototypical Example, or Functional Description.

Why-explanations provide causes or justifications for facts or the occurrence of events. Whereas the first concept is causal in nature and not symmetric, the latter only provides evidence for what has been asked for.

How-explanations are a special case of why-explanations, describing processes that lead to an event by providing a causal chain. How-questions ask for an explanation of the function of a device.

Purpose-explanations: The goal of *Purpose-explanations* is to describe the purpose of a fact or object. Typical questions are of the form 'What is ... for?' or 'What is the purpose of ...?'

Cognitive Explanations explain or predict the behavior of 'intelligent systems' on the basis of known goals, beliefs, constraints, and rationality assumptions. They are also a special case of why-explanations, distinguishing action and negative explanations.

Roth-Berghofer continues to identify the contribution of the different knowledge containers to the representation of these different kinds of explanations. We will not go into detail here and refer the reader to [11].

3 Exploring the Relations of Goals and Kinds

We propose a 2-step process to identify which explanations a CBR system should be able to give and to understand how to make the necessary knowledge accessible in the different knowledge containers:

1. Use the *Explanation Goals* perspective to identify user needs for explanations from a user model and system view which takes the usage situation into account.
2. Use the *Explanation Kinds* view to find useful prototypical explanations and assess the requirements for contents that have to be modeled into the system.

As a simplified example, we look at a diagnostic system for engine failures. We have a mixed initiative dialogue system where the system can ask questions about the engine status and the user can voluntarily provide information he deems important.[1] The system can give detailed explanations on possible causes for the problems as well as advice on how to avoid future occurrences. It is supportive, e.g., the user should be enabled to understand similar situations in the future without having to rely on the system.

Learning goal fullfilled by a conceptual explanation (definition): During the symptom assesment, the system asks the user to fill in the specific gravity of the fuel. The user is not familiar with the term specific gravity so he asks the system to explain this.

User: *"What is the specific gravity?"*

System: *"It is a measure of the density of a liquid relative to the density of water, with water having a specific gravity of 1.0."*

Relevance goal fullfilled by a causal why-explanation: The user can further on request to know why the system asked this specific question. The system should give explanations which set the question asked into context with the symptoms already known.

User: *"Why is the density of the fuel interesting?"*

System: *"Air/Fuel ratio is based on pounds of AIR to pounds of FUEL. Changing the specific gravity of your fuel can have a profound affect on the A/F ratio of your engine since carburetors are metered by volume."*

Justification goal fullfilled by a theoretical conceptual explanation: The system finally presents a solution to the user which requires him to change the spark lead time. In this situation, the user might want to know why this is a solution.

User: *"How is this adjustment related to the problem of too much CO_2 in the exhaust gases?"*

System: *"Different fuels and engine setups require different spark lead times. When using a wrong timing, the combustion process is suboptimal, and this will lead to high CO_2 values."*

Transparency goal fullfilled by a how explanation: The user might not be fully satisfied and tries to understand how the system came up with this answer.

User: *"How did you come up with this answer?"*

System: *"Looking at my case base we can see that the most relevant cases had problems with the spark lead time as well, with only the exception of those cases where the problem can be explained by the wrong type of fuel."*

---

[1] Please, keep in mind that there are other possible human computer interfaces than natural language dialogue systems as depicted here for easier understanding. For example, one possibility is to provide the user with pre-formulated queries from which he or she can choose the most appropriate one.

4 Conclusions and Future Research Directions

We have outlined a unified view on explanations in CBR which takes both the goals of the user and the type of an explanation into account. Both perspectives are to a certain degree independent from each other.

The next step in our fellow work is to analyze the relation between these two perspectives in more detail and to map the unified view onto the different knowledge containers. Further on, we want to integrate an explanation goals view with methods for the analysis of workplace situations like activity theory (as proposed, e.g., by Cassens [13]) and integrate the explanation kind perspective with existing design methodologies (such as INRECA [14] and SIAM [15]).

Our goal is to utilize this analytical framework in the system design process in order to identify requirements for the contents of the knowledge containers with regard to the intended usage situation and the anticipated user goals.

References

1. Buchanan, B.G., Shortliffe, E.H.: Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Addison Wesley (1984)
2. Swartout, W., Smoliar, S.: On Making Expert Systems More Like Experts. Expert Systems **4** (1987) 196–207
3. Leake, D.B., ed.: Case-Based Reasoning: Experiences, Lessons, & Future Directions. AAAI Press/MIT Press, Menlo Park (1996)
4. Schank, R.C., Kass, A., Riesbeck, C.K., eds.: Inside Case-Based Explanation. Lawrence Erlbaum Associates, Hillsdale, New Jersey (1994)
5. Gregor, S., Benbasat, I.: Explanations From Intelligent Systems: Theoretical Foundations and Implications for Practice. MIS Quarterly **23** (1999) 497–530
6. Swartout, W.R., Moore, J.D.: Explanation in second generation expert systems. In David, J., Krivine, J., Simmons, R., eds.: Second Generation Expert Systems. Springer Verlag, Berlin (1993) 543–585
7. Gervás, P., Gupta, K.M., eds.: Proceedings of the ECCBR 2004 Workshops. Number 142-04 in Technical Report, Madrid, Departamento de Sistemas Informáticos y Programación, Universidad Complutense Madrid (2004)
8. Funk, P., Calero, P.A.G., eds.: Advances in Case-Based Reasoning: Proceedings ECCBR 2004. Number 3155 in LNAI, Madrid, Springer (2004)
9. Richter, M.M.: The knowledge contained in similarity measures. Invited Talk at ICCBR 1995, Sesimbra, Portugal (1995)
10. Sørmo, F., Cassens, J.: Explanation goals in case-based reasoning. [7] 165–174
11. Roth-Berghofer, T.: Explanations and Case-Based Reasoning: Foundational Issues. [8] 389–403
12. Spieker, P.: Natürlichsprachliche Erklärungen in technischen Expertensystemen. Dissertation, University of Kaiserslautern (1991)
13. Cassens, J.: Knowing what to explain and when. [7] 97–104
14. Bergmann, R., Althoff, K.D., Breen, S., Göker, M., Manago, M., Traphöner, R., Wess, S.: Developing Industrial Case-Based Resoning Applications: The INRECA Methodology. Second edn. LNAI 1612. Springer-Verlag, Berlin (2003)
15. Roth-Berghofer, T.R.: Knowledge Maintenance of Case-Based Reasoning Systems – The SIAM Methodology. Volume 262 of Dissertationen zur KI. Akademische Verlagsgesellschaft Aka GmbH / IOS Press, Berlin, Germany (2003)