In: Proceedings of the 2nd Conference "Professionelles Wissensmanagement", Luzern, Switzerland, 2003.

Evaluation of two Strategies for Case-Based Diagnosis handling Multiple Faults (Extended Abstract)

Martin Atzmueller, Joachim Baumeister, Frank Puppe

Department of Computer Science
University of Würzburg, 97074 Würzburg, Germany
email: {atzmueller, baumeister, puppe}@informatik.uni-wuerzburg.de

Introduction Case-based diagnosis handling multiple faults is still a challenging task. In our domain of sonography the examination considers several partially disjunctive subdomains, e.g. liver or kidney, which results in multiple faults, i.e. most cases contain multiple diagnoses. In this paper we present methods for handling multiple faults, embedded in the standard CBR cycle. The context of our work is to supplement a medical documentation and consultation system by CBR techniques enabling the extended retrieval of experiences for experience management. E.g. explanations for a query case based on the similarity to former cases and additional information contained in these can be retrieved. Similarly to other problem solving methods being able to handle multiple faults in medical domains, we can exploit independence assumptions about the domain to reduce the combinatorial search space. The major ideas are: decomposing complex cases into several simpler cases, finding solutions for the simple cases, and combining these solutions.

Case-Based Diagnosis with Multiple Faults As a similarity measure when comparing a query case with another case, we apply an adaptation of the *Hamming distance* with weights and partial similarities. This similarity knowledge is learned automatically from the case base. We say, that a case is *sufficiently similar* to another case, if the similarity between these two cases exceeds a given (and usually high) threshold. If we cannot retrieve sufficiently similar cases by standard CBR, then we rely on an adapted *Retrieve* and *Reuse* process, which is depicted in Figure 1: We generate a set of *candidate cases* according to a *candidate case generation strategy*. Below we will discuss two such strategies.
We create a *candidate case* by merging a set of subcases into one case. A candidate case consists of a set of findings, a set of diagnoses and a set of subcases from the case base as background information. When generating a candidate case, the set of findings is created by joining the findings of the subcases applying conflict resolution if necessary. The set of diagnoses is simply the union of the subcases' diagnoses. The generated candidate cases make up the candidate case base. From this case base we retrieve sufficiently similar cases to the query case.

Strategies for Candidate Case Generation We present two approaches for generating candidate cases. The first approach uses partitioning knowledge, to decompose cases into subcases. Partition class knowledge describes how to divide the set of diagnoses and

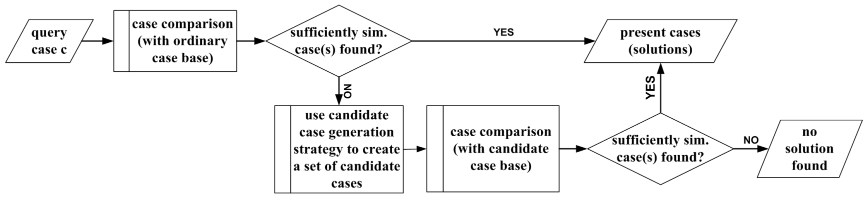

Figure 1: Process model for CBR handling multiple faults using candidate case creation

attributes of the domain into partially disjunctive subsets. These correspond to rather independent subdomains of the application domain. For example, in the medical domain of sonography, there are subsets corresponding to problem areas like *liver* or *kidney*.
Candidate case generation works as follows: The partitioning of the original case base into several 'partition class' case bases is precomputed once. The query case is decomposed into several subcases. Then we apply CBR for each query subcase using the respective 'partition class' case base and collect the most similar partial cases into result sets. These are recombined drawing one case from each set to generate a set of candidate cases.
The second approach uses learned diagnostic profiles to build set-covering models. These are used to generate hypotheses, i.e. sets of diagnoses, that represent an explanation for the query case. Given a hypothesis we combine cases to candidate cases such that the diagnoses of a candidate case have a high coverage w.r.t. the generated hypothesis.

Evaluation For the evaluation of the presented approaches we applied part of a real-world case base containing 744 cases, based on the knowledge-based documentation and consultation system for sonography SONOCONSULT. We performed 3 experiments E0, E1 and E2 in which we applied standard CBR and CBR combined with one of the two strategies for candidate case generation. We used leave-one-out cross-validation for evaluation. We say that a compare case c' *solves* a query case c, iff $sim(c, c') \geq \mathcal{T}_{CBR}$, i.e. if it is sufficiently similar. The more advanced strategies enable us to decrease the similarity threshold $\mathcal{T}_{CBR}$ without receiving a dramatically decreased accuracy of the solved cases. We present the results in the following table.

	used knowledge/ method	*threshold* $\mathcal{T}_{CBR}$	*solved cases*	*mean acc*
E0	no knowledge	0.78	33 (4%)	0.73
E1	set-covering strategy	0.54	443 (60%)	0.70
E2	partition class strategy	0.40	635 (85%)	0.78

E0 shows that the standard CBR method is performing poor for cases with multiple faults. E1 shows, that the set-covering approach performs acceptable since it can solve about 60% of all cases with a mean accuracy of 70%. Experiment E2 using partition class knowledge is even better solving about 85% of all cases with a mean accuracy of 78%, which demonstrates that this strategy can deal with the multiple fault problem quite well. These results are quite promising. Nevertheless, we see enhancements for the number of solved cases and accuracy when applying refined partition class knowledge or improved set-covering models using quality measures.

References For references we refer to our full paper, available in the CEUR workshop proceedings (`www.ceur-ws.org`) of the *2nd German Workshop on Experience Management (GWEM 2003)*, or at `http://ki.informatik.uni-wuerzburg.de/papers/`.