Retrieval of Cases Imperfectly Described and Explained: A Quantitative Approach

Carlos Bento **Ernesto Costa**
Laboratório de Informática e Sistemas
Universidade de Coimbra
Quinta da Boavista, lote 1, 1°, 3000 Coimbra - PORTUGAL
bento@uc.pt ernesto@moebius.uc.pt

The power of a Case-Based Reasoning (CBR) System is greatly determined by its capability to retrieve the relevant cases for prediction of the new outcome. The retrieval process involves indexing cases. The combination of nearest neighbour and knowledge-guided techniques for case indexing led to the development of hybrid systems [Cain *et al.*, 1991] joining CBR [Hammond 1986; Kolodner 82] and Explanation-Based Learning (EBL) techniques [Mitchell *et al.*, 1986].

We propose a CBR+EBL similarity metric for cases imperfectly described and explained. A case is represented by a past situation, an outcome and a set of explanations of why the situation had such an outcome. It is assumed that cases are imperfectly described due to an incorrect delimitation of the past situation. We represent three types of imperfections in explanations: broken explanations (proof trees with a gap between them and the case's outcome); partial explanations (proof trees that omit some branches); incomplete set of explanations (some outcome facts are not end of a proof tree).

Our similarity metric involves the concepts of matching situation, strong, weak, and undetermined situation snippet. A matching situation is a situation (represented by a set of facts) that is obtained by going down in one or more explanation trees in order to gather the maximum number of facts that match the new situation. A past or matching situation is seen as composed by a set of situation pieces called situation snippets. A situation snippet is a set of facts that are the leaves of a proof tree. Depending on the proof tree being complete, omit some branches or have a gap the situation snippet is called strong, weak, or undetermined.

The proposed similarity metric is composed of three terms:

$$\kappa \sum_{i=1}^{k} \text{relev}(f_i, SS_u)\, \text{sim}(f_i, f_i') + \lambda \sum_{i=1}^{r} \text{relev}(f_i, SS_w)\, \text{sim}(f_i, f_i') + \mu \sum_{i=1}^{t} \text{relev}(f_i, SS_s)\, \text{sim}(f_i, f_i')$$

with constants k, r, t, respectively, the number of occurrences of the matching situation facts in undetermined, weak and strong situation snippets; f_i a fact in the matching situation; f_i' a fact in the new situation; SS_u, SS_w and SS_s ,respectively, undetermined weak and strong situation snippets; constants κ, λ, and μ , respectively, represent the weight assigned to the facts belonging to undetermined, weak, and strong situation snippets;

$$\text{relev}(x, SS_{type}) = \frac{1}{\text{“cardinal of the } SS_{type} \text{ set to which x belongs”}} \; ; \quad \text{and} \quad \text{sim}(x,y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$$

Some important points about this metric are: (1) It discriminates between facts relevant to broken, partial and complete explanations; (2) It assigns relevance to each matching fact, function of its situation snippet size; (3) It takes into account a matching fact as many times as the number of times it occurs in the situation snippets.

This metric has been used in an expert system for culpability assignment in car accidents implemented in PROLOG.

References

[Cain *et al.*, 1991] Timothy Cain, M. J. Pazzani, and Glenn Silverstein, "Using Domain Knowledge to Influence Similarity Judgments", in Proceedings of a Case-Based Reasoning Workshop, Morgan-Kaufmann, 1991.

[Hammond 1986] K. Hammond., "Case-Based Planning: An Integrated Theory of Planning, Learning and Memory", Ph. D. Dissertation, Yale University, 1986.

[Kolodner 1982] J. Kolodner, "The Role of Experience in the Development of Expertise", in Proceedings of the National Conference on Artificial Intelligence, Pittsburgh, P.A., Morgan-Kaufmann, 1982.

[Mitchell *et al.*, 1986] T. Mitchell, R. Keller, and S. Kedar-Cabelli, "Explanation-Based Learning: A Unifying View", Machine Learning, vol. 1 (1), 1986.