Investigating different Methods for efficient Retrieval of Generalized Cases

Rainer Maximini Alexander Tartakovski
Ralph Bergmann
University of Hildesheim, Data- and Knowledge Management Group
PO-Box 101363, D-31113 Hildesheim, Germany
{r_maximi|tartakov|bergmann}@dwm.uni-hildesheim.de

In CBR applications, the traditional concept of a case is that of a point in the space spawned by the case's attributes. Driven by examinations of several new applications, like the retrieval for intellectual properties (IPs), we proposed the concept of *generalized cases* [BV99, Ber02]. A generalized case (gc) covers not only one point of the attribute space, but a whole subspace of it. It immediately provides solutions to a set of closely related problems rather than to one single problem only. The solutions a gc represents are very close to each other; basically they should be considered as (slight) variations of the same principle solution. In this paper we explore a more formal and systematic view on gcs by using constraints to express the dependencies between several attributes.

The similarity computation between a *point case* pc and a gc is very difficult because the following problem must be solved.

$$\sup_{\forall gc_i \in gc} \{sim(pc, gc_i)\}$$

In this equation the association is used that a gc is a possibly infinite set of pcs. Mougouie, Bergmann, and Vollrath [MB02] have analyzed methods from optimization theory to solve the similarity problem. However, there are several alternative approaches to solve this problem. This paper points out three radically different approaches to similarity assessment and retrieval. Briefly, the methods discussed are:

Mathematically Optimization

As described above, Bergmann and Mougouie [MB02] presented a solution to calculate the similarity between a pc and an gc that is represented through constraints over an n-dimensional real-valued vector space. They have shown that the difficulty of this calculation depends on whether the gc spawn a convex or nonconvex subspace which is defined by the constraints. For convex constraints and by usage of convex similarity measures, the Topkis-Veinott method can be easily applied to determine exactly the similarity between a pc and a gc. If the similarity measure is nonconvex or the gc contains also nonconvex constraints, the problem is more difficult. For this situation an algorithm is proposed that allows to incrementally compute sequences of upper and lower bounds for the similarity and assures the convergence of the algorithm. It allows to rank the gc in the case base

according to their similarity to the pc query without the exact computation of all similarity measures. The main problem of this concept is the query dependency and that it is only only analyzed for real valued attributes.

Sampling

The idea of the sampling converter[MA03] is to sample the spawn subspace for each gc, select reasonable pcs out of it and create a new case base with only pcs belonging to the current gc. For this case base well known retrieval engines can be used where only less modifications have to be made.

This is of course an approximative technique and the resulting quality mainly depends on the sampling quality and the number of result cases. The offline phase of this method takes a long time when the gc are converted in an intelligent way into pcs. The online phase is nearly the same as for retrieval with traditional cases. It has only to be granted that all retrieved cases originally belonged to different generalized or pcs.

Computer Graphics

The idea is to use methods and algorithms from computer graphics for the retrieval of gcs. Interesting are methods from the area of 3D real time calculating that are applied in 3D games. The known techniques for space dividing, removing of hidden surfaces or others can probably be useable for the retrieval of gcs.

Realtime 3D applications demand very high requirements on the efficiency of the used techniques. The amount of items grows continuously and requires more powerful hard- and software. Even if not all of the used techniques can be adapted to CBR purposes, there could be some methods which possibly can be used or are able to improve existing retrieval techniques.

Conclusion

The presented ideas have their own pro and cons depending on the kind of attributes and constraints. Our goal is to examine each idea and to combine the methods to a general one which hopefully handles each kind of attribute and constraint.

References

[Ber02] Ralph Bergmann. *Experience Management - Foundations, Development Methodology, and Internet-Based Applications*. Lecture Notes in Artificial Intelligence 2432. Springer Berlin, Heidelberg, New York, Hong Kong, London, Milan, paris, Tokyo, 2002.

[BV99] R. Bergmann and I. Vollrath. Generalized Cases: Representation and Steps Towards Efficient Similarity Assessment. In W. Burgard, Th. Christaller, and A. B. Cremers, editors, *KI-99: Advances in Artificial Intelligence.*, LNAI 1701. Springer, 1999.

[MA03] Rainer Maximini and Vera Austermann. Retrieval on attribute dependent generalized cases. Technical report, University of Hildesheim, 2003.

[MB02] B. Mougouie and R. Bergmann. Similarity Assessment for Generalizied Cases by Optimization Methods. In *Proceedings of the European Conference on Case-Based Reasoning (ECCBR-02)*. Springer., 2002.