German Workshop on Experience Management (GWEM 2002)

Towards the Fast Monitoring of Case Properties

Ioannis Iglezakis

DaimlerChrysler AG, Research & Technology, RIC/AM,
P.O. Box 2360, 89013 Ulm, Germany
ioannis.iglezakis@daimlerchrysler.com

Abstract. Applying case properties over a case base can indicate conflicts in or between cases. Monitor these results is important to trigger the maintenance of a case–based reasoning system. However, the complexity of applying a single case property over a case base is $\mathcal{O}(n^2)$.
This short paper describes how appropriate sampling methods can be used to reduce the effort of triggering the maintenance for case–based reasoning systems.

1 Introduction

Case–based reasoning is a process rather than an algorithm. The classical case–based reasoning cycle presented by Aamodt and Plaza [AP94] supports this process with the *retrieve*, *reuse*, *revise*, and *retain* step. In contrast the extended case–based reasoning cycle by Reinartz et al. [RIR01] contains two additional steps. These steps are *review* and *restore*. The review step consists of the two substeps *measuring* and *monitoring*. The measuring substep determines the quality of a case base by calculating the case properties between cases. The result of this calculation is controlled by the monitoring substep and is used to trigger the restore step.

The measuring substep within the extended case–based reasoning cycle calculates with the help of the case properties the quality of a case base. Case properties are disjunct measures who indicate possible conflicts within or between cases. The application range of case properties are diagnostic domains like help desks for which a single fault assumption holds. A formal definition of the case properties *correctness*, *consistency*, *incoherence*, *minimality*, and *uniqueness* can be found in Reinartz et al. [RIR01]. To calculate these case properties over the whole case base each pair of cases is applied to the case property. Therefore, the complexity of this algorithm is $\mathcal{O}(n(n-1)/2) = \mathcal{O}(n^2)$.

The monitoring substep within the review step collects the values of each single case property p_t at the time t and triggers the restore step if a threshold τ is reached. There are three different kinds of possible triggers. The first trigger uses the value of a case property and starts the restore step if $p_t \leq \tau$. The second trigger uses the difference of a case property value at the point at time t and t' (with $t' > t$) and maintains if $p_{t'} - p_t \leq \tau$. Finally, the third trigger builds an gradient of the case property values at the point at time t and t' and maintains if $(p_{t'} - p_t)/(t' - t) \leq \tau$. All these triggers are using the results of applying the case properties over the case base. Hence, a fast approximation of the case properties can reduce the expenses for maintaining the case base.

2 Sampling

Sampling methods are used to speed up the analysis of basic characteristics from a collection of units selected of a finite collection (population).

There are various sampling methods like *random sampling*, *systematic sampling*, *stratified sampling*, *cluster sampling*, and *two–stage cluster sampling*. According to the complexity only random sampling, systematic sampling, and stratified sampling are applicable. The reason why is that the complexity of the case properties is $\mathcal{O}(n^2)$ (cf. section 1) and with more complex sampling methods like cluster sampling or two–stage cluster sampling the compound complexity of the sampling method and case properties calculation for the sample is higher than the calculation of the case properties over the whole population. For example, random sampling selects randomly a collection of cases (sample) for examination from a finite collection (population). Each member of the sample is drawn entirely by chance. Furthermore, each member of the population has a known, but not necessary equal chance of being added into the sample at each draw. The complexity of this algorithm is $\mathcal{O}(n)$ and the overall complexity is reduced. Detailed descriptions about various sampling methods can be found in Scheaffer et al. [SIO96].

A possible evaluation scenario would be to draw a sample over a case base and apply a case property over this sample. The result can be compared to the total result which is received by applying the same case property over the whole case base. Furthermore, different sample sizes can be used to find the sample size for which the total result can be predicted as accurate as possible in respect to the aim of decreasing the total time of applying the case property over the sample.

3 Conclusions and Further Work

By using sampling methods, it is possible to decrease the effort of calculating the case properties for a case base. The aim of reducing this effort is a fast determination if a case base must be maintained. This is part of the monitoring substep in the extended case–based reasoning cycle. This short paper introduced sampling methods as a promissing method to trigger the restore step in the extended case–based reasoning process.

First examinations with various sampling methods showed that it is possible to reduce the effort of triggering the maintenance. Further tasks are to evaluate the use of these methods in more detail and eventually to use regression analysis to predict the value of a case property more accurate. For example: With a 5%, 10%, 15%, and 20% sample, it is possible to calculate a regression curve for which it is feasible to predict the total result.

References

[AP94] Agnar Aamodt and Enric Plaza. Case–based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 7(1):39–59, 1994.

[RIR01] Thomas Reinartz, Ioannis Iglezakis, and Thomas Roth–Berghofer. Review and restore for case–base maintenance. *Computational Intelligence: special issue on maintaining CBR systems*, 17(2):214–234, 2001.

[SIO96] Richard L. Scheaffer, William Mendenhall III, and Lyman Ott. *Elementary Survey Sampling*. Duxbury Press, 1996.