Feature-Centric Approaches to Case-Base Maintenance

Brian Schack
Advisor: Dr. David Leake

School of Informatics and Computing, Indiana University
Bloomington IN 47408, USA
leake@indiana.edu, schackb@indiana.edu

Abstract. Flexible case-base maintenance (FCBM) and adaptation-guided feature deletion (AGFD) extend case-base maintenance (CBM) research. FBCM examines superficial properties of cases and their features such as the size of a case or the rarity of a feature. Then it deletes either entire cases, or components of cases, measuring competence by the number of problems solved and measuring size by the number of case-feature pairs. AGFD harnesses the adaptation knowledge of a system in order to improve FCBM. It prioritizes cases for maintenance based on the recoverability of their deleted components. Experiments evaluated FCBM in three domains: film recommendations, Congressional bill sponsorship, and travel agency packages, and AGFD in a path-finding domain. The proposed research plan is to continue to explore and improve approaches to CBM and their relationship with each of the other phases of the CBR cycle.

Key words: Case-base maintenance, case-base compression, case deletion

1 Introduction and Context

The performance of a case-based reasoning system depends on the coverage of its case base and the quality of its cases. As the number of cases in the case base grows, storage constraints and retrieval costs [1, 2] necessitate limiting the size, and so case base maintenance [3] remains an active area of CBR research through methods such as competence-based case deletion [4], optimizing the trade-off between size and accuracy [5], deletion aimed at preserving diversity [6], and strategies for forgetting (e.g. [7]). On the flip side, other methods limit which cases to retain during problem solving [8] or the order to add cases from a candidate set [9, 10]. All of these strategies delete or add entire cases as indivisible units, therefore this research summary refers to them as *per-case maintenance strategies*.

Research on the maintenance of case contents has generally focused on quality improvement rather than case-base compression (e.g., [11]). The key novelty of *flexible case-base maintenance* [1] is that it removes components of case solutions. This research summary describes FCBM by contrasting it with per-case maintenance and builds on it with adaptation-guided feature deletion. Then, it concludes with the remaining research questions and the upcoming research plan.

[1] Earlier research [12] referred to *flexible case-base maintenance* as *flexible feature deletion*. This name change reflects feedback from the reviewers.

In Proceedings of the ICCBR 2016 Workshops. Atlanta, Georgia, United States of America

2 Flexible Case-Base Maintenance

Per-case strategies reflect two common implicit assumptions: (1) that all of the cases in the CBR system will have sufficiently uniform size that the size effects of deletion or addition do not depend on the chosen case, and (2) that the size of the internal contents of cases cannot be reduced. In domains for which each case must contain uniform knowledge, so that removal of any case information would severely impair the ability to use the cases, per-case strategies are the only appropriate choice. However, in some CBR domains, case contents are more flexible.

Flexible case-base maintenance questions the assumption of uniform case size in case-base maintenance. The assumption of uniform size means that, if cases are of different size, it is not possible, for example, to favor retention of smaller cases when those cases have comparable coverage. It also questions the assumption of maintenance only on a per-case basis, proposing that compression strategies can consider not only case deletion/addition but the deletion of components of particular cases. Rather than pre-determining a static set of features to be used throughout the life of the CBR system, the set of features to include in the case base could be adjusted based on requirements for storage, processing speed, and accuracy.

There need not be a requirement that all cases in the case base include the same set of features, just as there need not be uniform collections of components in the solution parts of cases, and the solutions need not be represented at the same level of granularity. FCBM is a new, more flexible maintenance approach in which selective compression can be done at the level of the contents of individual cases, by removing selected features from either indexing or solution information. Thus, this can be used to maintain both indexing features and features of a solution.

The motivation for adjusting case contents arises from domains in which cases are large and can be represented in multiple ways. For example, CBR has attracted interest for reasoning from imagery such as medical images (e.g., [13]). From any image, different features may be extracted, at different resolutions, and the amount of information required to represent different images might vary dramatically. In diagnostic domains, numerous features may carry information relevant to the diagnosis, with different pieces relevant to different degrees for different problems. When CBR is applied to design support, stored designs could selectively include different subsets of a full design or could include the design at different levels of detail. In a case-based planner generating highly complex plans, it is possible to retain the entire plan, or only key pieces, or to preserve full details for parts of the plans and high-level abstractions for others.

Experiments evaluated eight CBM strategies and three hybrid strategies [14] across three domains: film recommendations, Congressional bill sponsorships, and travel agency packages. Results supported that, FBCM may outperform per-case maintenance at the same levels of compression for suitable domains where cases have varying sizes, contents are compressible, and reasoning requires different amounts of information.

3 Adaptation-Guided Feature Deletion

When flexible case-base maintenance removes components of cases, for flat feature representations, its process is restricted to deleting particular features from a feature vector.

However, for structured cases, FCBM includes a wider space of possible operations, not restricted to deleting individual features, or even limited to deletion per se. FCBM could compress cases through substructure deletion, substructure substitution, or substructure abstraction. These correspond to common operations for case adaptation, and an adaption chain can apply such operations successively [15]. If the adaptation component contains procedures for these, then adaptation-guided feature deletion can apply them directly. If the adaptation knowledge includes specific guidance on applicability, then the reuse of the adaptation knowledge makes it available to the maintenance phase.

Case abstraction research, because it aims to compact the case base by removing concrete cases subsumed by abstractions [16], can also be seen as in the spirit of replacing cases with more compact versions. For a recent example, ICARUS [17] retains abstract planning cases in a tree structure according to similarities, and this resembles feature bundling except constrained to a hierarchy. Exploiting adaptation knowledge for FCBM raises the key question of when an adaptation should be applied for FCBM. This depends crucially on four factors: compression, feature-centric recoverability, quality retention, and recovery cost.

- **Compression:** Compression simply refers to the change in case base size from applying an adaptation. However, because FCBM can change case sizes by feature deletion, this is measured not in terms of the number of cases, but in terms of the finest-grained subunit of cases that is meaningful to delete. For cases represented by feature vectors, this is a feature-value pair.
- **Feature-centric recoverability:** Feature-centric recoverability refers to the ability of the system to recover a competent solution to the problem of a given case from the FCBM-modified case and the remaining case base. Recovering the competence may not require recovering a solution identical to the original if multiple solutions are satisfactory.
- **Quality retention:** Quality retention refers to the quality of the solutions the system is able to generate, beyond simply generating a correct solution. For example, in a path planning domain, a deletion from a path would be recoverable if the system were still able to generate *some* path between the same endpoints. Quality retention might be measured by the ratio of the costs of old and new paths.
- **Recovery cost:** Recovery cost refers to the resources required to generate a new solution to the problem. For example, in a case-based planner able to draw on a generative planner when needed, all deletions might be recoverable, but some might be computationally expensive when done by reasoning from scratch. In those instances, other deletions might be more appropriate.

We note that feature-centric recoverability is closely related to the notion of *reachability*, defined by Smyth and Keane [4]. However, there is an important difference. Reachability refers to the ability to adapt other cases in the case base to cover the problem addressed by a candidate case. Feature-centric recoverability refers to the ability to adapt either other cases in the case base or the FCBM-revised case to restore the coverage that the candidate case initially provided.

4 Remaining Research Questions and Challenges

The FCBM approach raises a rich range of questions for fully exploiting its potential. Questions include how FCBM strategies should interact with the indexing and adaptation knowledge containers, how feature deletion can preserve case integrity, and how feature deletion should be reflected in case provenance and explanation.

- **Coupling feature deletion with index maintenance:** As case contents are deleted, the relevance of case indices may change. Consequently, this researcher's hypothesis is that FCBM needs accompanying index maintenance to assure that, as cases are compressed, the system still retrieves the most similar cases. I intend to evaluate this by measuring the performance of different combinations of FCBM and indexing strategies. Additionally, feature weight information might suggest features which could be deleted with limited harm.
- **Maintaining case integrity despite feature deletion:** Another question is the relationship of feature deletion to the cohesiveness of a case. From the early days of CBR, an argument for CBR has been that cases can implicitly capture interactions among case parts. Deleting portions of a case risks some of that cohesion, making it a concern to address in feature deletion strategies. I intend to formally define case cohesiveness and measure the impact of FCBM strategies on this metric.
- **Reflecting feature deletion in provenance and explanation:** Because feature deletion results in stored cases which differ from the cases originally captured, it (like case adaptation) may weaken the ability to justify proposed solutions by past experience. Likewise, changes from the original cases may make it difficult to apply provenance-based methods for predicting solution characteristics such as solution accuracy and trust (e.g., [18]). I intend to extend provenance-based explanations (when possible) to explain solutions adapted from cases restructured by FCBM.

5 Conclusion

This research summary described two approaches which build on past ideas about CBM. FCBM removes components of cases, and adaptation-guided feature deletion uses adaptation knowledge. The proposed research plan is to continue to explore and improve upon approaches to CBM and their relationship with each of the other phases of the CBR cycle.

6 Acknowledgments

This research summary contains excerpts from two earlier papers by the same co-author [12, 19].

References

1. Francis, A., Ram, A.: Computational models of the utility problem and their application to a utility analysis of case-based reasoning. In: Proceedings of the Workshop on Knowledge Compilation and Speed-Up Learning. (1993)

2. Smyth, B., Cunningham, P.: The utility problem analysed: A case-based reasoning perspective. In: Proceedings of the Third European Workshop on Case-Based Reasoning, Berlin, Springer (1996) 392–399
3. Wilson, D., Leake, D.: Maintaining case-based reasoners: Dimensions and directions. Computational Intelligence **17**(2) (2001) 196–213
4. Smyth, B., Keane, M.: Remembering to forget: A competence-preserving case deletion policy for case-based reasoning systems. In: Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, San Mateo, Morgan Kaufmann (1995) 377–382
5. Lupiani, E., Craw, S., Massie, S., Juárez, J.M., Palma, J.T.: A multi-objective evolutionary algorithm fitness function for case-base maintenance. In: Case-Based Reasoning Research and Development - 21st International Conference, (ICCBR-13), Springer 218–232
6. Lieber, J.: A criterion of comparison between two case bases. In: Advances in Case-Based Reasoning. Volume 984 of Lecture Notes in Computer Science. Springer, Berlin (1995) 87–100
7. Ontañón, S., Plaza, E.: Collaborative case retention strategies for CBR agents. In: Case-Based Reasoning Research and Development: Proceedings of the Fifth International Conference on Case-Based Reasoning, ICCBR-03, Berlin, Springer-Verlag (2003)
8. Muñoz-Avila, H.: A case retention policy based on detrimental retrieval. In: Proceedings of ICCBR-99. (1999)
9. Zhu, J., Yang, Q.: Remembering to add: Competence-preserving case-addition policies for case base maintenance. In: Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, Morgan Kaufmann (1999) 234–241
10. Smyth, B., McKenna, E.: Building compact competent case-bases. In: Proceedings of the Third International Conference on Case-Based Reasoning, Berlin, Springer Verlag (1999) 329–342
11. Racine, K., Yang, Q.: Maintaining unstructured case bases. In: Case-Based Reasoning Research and Development, ICCBR 1997, Berlin, Springer (1997) 553–564
12. Leake, D., & Schack, B. (2015, September). Flexible Feature Deletion: Compacting Case Bases by Selectively Compressing Case Contents. In: International Conference on Case-Based Reasoning (pp. 212-227). Springer International Publishing.
13. Wilson, D., O'Sullivan, D.: Medical imagery in case-based reasoning. In Perner, P., ed.: Case-Based Reasoning on Images and Signals. Volume 73 of Studies in Computational Intelligence. Springer (2008) 389–418
14. Cummins, L., & Bridge, D. (2009, July). Maintenance by a committee of experts: The MACE approach to case-base maintenance. In *International Conference on Case-Based Reasoning* (pp. 120-134). Springer Berlin Heidelberg.
15. Fuchs, B., Lieber, J., Mille, A., & Napoli, A. (2014). Differential adaptation: An operational approach to adaptation for solving numerical problems with CBR. Knowledge-Based Systems, 68, 103-114.
16. Bergmann, R., Wilke, W.: On the role of abstraction in case-based reasoning. In: Advances in Case-Based Reasoning, Third European Workshop, EWCBR-96, Lausanne, Switzerland, November 14-16, 1996, Proceedings. (1996) 28–43
17. Menager, D. (2016, October). Episodic Memory in a Cognitive Model. In *International Conference on Case-Based Reasoning Doctoral Consortium*. Springer International Publishing.
18. Leake, D., Whitehead, M.: Case provenance: The value of remembering case sources. In: Case-Based Reasoning Research and Development: Proceedings of the Seventh International Conference on Case-Based Reasoning, ICCBR-07, Berlin, Springer-Verlag (2007) 194–208
19. Leake, D., & Schack, B. (2016, October). Adaptation-Guided Feature Deletion: Testing Recoverability to Guide Case Compression. In: International Conference on Case-Based Reasoning. Springer International Publishing.