

# Case-Base Maintenance Beyond Case Deletion

Brian Schack \*\*

Indiana University, Bloomington IN 47408, USA  
schackb@indiana.edu  
<http://bit.ly/schackbrian>

**Abstract.** *Case-base maintenance* strategies judiciously choose the most valuable cases to retain and the least valuable cases to delete in order to maintain a compact, competent case base. This research summary presents three case-base maintenance strategies which involve more than merely deleting cases: (1) *Flexible feature deletion* deletes components of cases instead of whole cases. (2) *Adaptation-guided feature deletion* prioritizes components for deletion according to their recoverability via adaptation knowledge. (3) *Expansion-contraction compression*, in addition to deleting cases, also adds cases in unexplored regions of the problem space. Evaluation of the strategies compared to standard case-base maintenance shows improved retention of competence and solution quality for suitable data sets compressed to the same sizes.

**Keywords:** artificial intelligence; case-based reasoning; swamping utility problem; case-base maintenance; flexible feature deletion; adaptation-guided feature deletion; expansion-contraction compression

## 1 Introduction

Case-based reasoning gradually builds up a case base from training data, knowledge engineered by human experts, and the retention phase. Each case retained in the case base can potentially, through adaptation, solve future problems. On the other hand, each retained case makes the case base larger. A larger case base requires more storage, more time to search through, more time to transmit over a network, and more time to manually review.

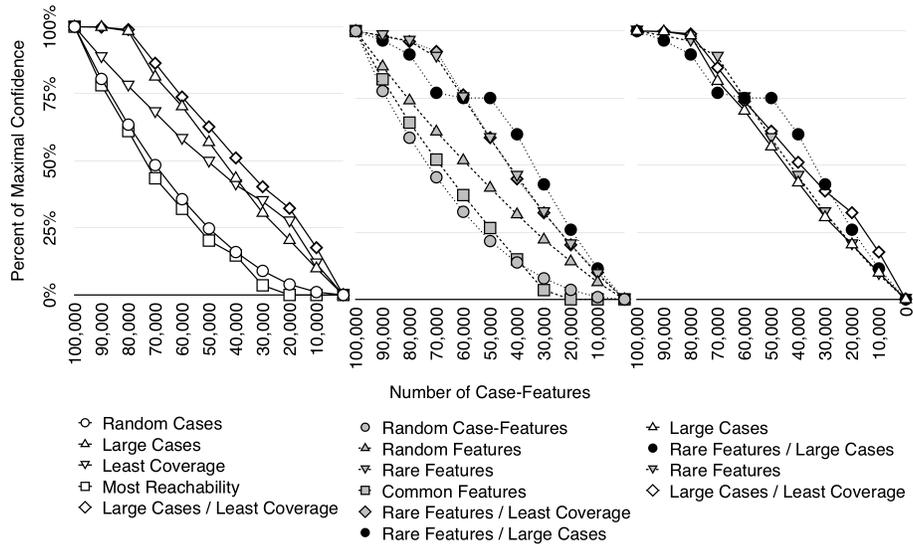
The *swamping utility problem* describes this trade-off between the competence, quality, and speed contribution of a case versus its storage, retrieval, and bandwidth cost [11]. Much research over the years has attempted to mitigate the utility problem [2]. *Case-base maintenance* strategies judiciously choose the most valuable cases to retain and the least valuable cases to delete in order to maintain a compact, competent case base. This research summary compares standard case-base maintenance strategies to three strategies which go beyond case deletion.

---

\*\* The author gratefully acknowledges the generous advice of his advisor Dr. David Leake at Indiana University Bloomington.

## 2 Flexible Feature Deletion

Case-base maintenance strategies, whether based on coverage and reachability or not, normally make two assumptions: (1) that all cases have a uniform storage cost and (2) that they must retain or delete whole cases. This research summary describes *flexible feature deletion*, a knowledge-light case-base maintenance strategy which, in contrast, subdivides variable-size cases for deletion of their components [4] [10].



**Fig. 1.** Competence retention of knowledge-light flexible feature deletion strategies for varying compression levels on the IMDb case base.

Cases can have varying storage cost when they contain varying amounts of information at varying levels of detail. The storage cost of both the problem and the solution can vary independently because a simple problem may have a complex solution and vice versa. This suggests balancing the competence contribution of a case against its storage cost.

A case-base maintenance strategy could delete an entire case, but it could also delete a single feature across all cases, or a single feature from a single case. Each of these alternatives presumably degrades problem-solving competence but not necessarily to the same degree. Compared to per-case strategies, flexible feature deletion could reduce the size of a case base with less reduction in the number of cases. It could also vary in the metric that it uses to order features for deletion. Each of the variations uses a knowledge-light metric like the size of a case, the rarity of a feature, or a hybrid of multiple metrics.

Domains with large cases and multiple representations call for the application of flexible feature deletion. For example, cases based on medical imagery may have various resolutions and a large number of features of which only some are relevant to the diagnosis.

### 3 Adaptation-Guided Feature Deletion

The adaptation-guided feature deletion case-base maintenance strategy builds on flexible feature deletion. Whereas flexible feature deletion orders the features according to a knowledge-light metric, *adaptation-guided feature deletion* integrates additional knowledge from the solution transformation container about the recoverability of features [5] [9]. Similar to how reachability measures the ability of adaptation knowledge applied to other cases to restore the solution to a case considered for deletion, *recoverability* measures the ability of adaptation knowledge applied to other features to restore a feature considered for deletion.

A solution with recovered features may either match exactly the original uncompressed solution, or it may solve the same problem in a different way. Compression to smaller sizes can increase the time required for recovery and decrease the quality of the recovered solution until adaptation knowledge can no longer recover any solution at all. Therefore, in order to preserve problem-solving competence, adaptation-guided feature deletion deletes features in order from most recoverable to least.

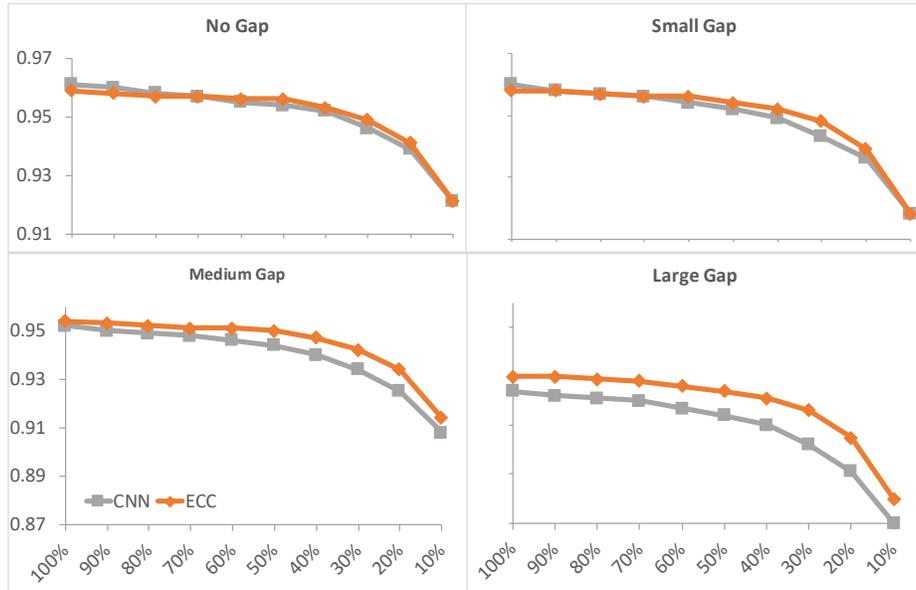
In addition to deleting features, adaptation-guided feature deletion can also replace them with a smaller substitution or abstraction. Occasionally, this reorganization can make case contents more accessible to an adaptation rule of limited power. Even though case-base compression normally reduces competence, compression under these circumstances, termed *creative destruction*, can improve competence instead.

### 4 Expansion-Contraction Compression

By the representativeness assumption, maintenance strategies predict that future problems will follow the same distribution as the current case base, and this works reasonably well for mature case bases in stable domains. But the representativeness assumption may apply less accurately during early case base growth, to dynamically changing domains, or in cross-domain transfer learning.

In these situations, case-base maintenance strategies optimizing for assumed representativeness may instead cause overfitting. *Overfitting* means that a statistical model or a machine learning algorithm makes predictions based on peculiarities in the training data not reflected in the testing data thereby improving performance on the training data and sacrificing performance on the testing data [1].

The overfitting problem has received significant attention in the context of artificial neural networks [3]. Among several potential mitigations, neural networks may employ *data augmentation* which perturbs training data in order to supplement it with additional instances [12]. For example, cropping images with-



**Fig. 2.** Absolute quality for condensed nearest neighbor and expansion-contraction compression on House data set for four different gap sizes in the training data.

out obscuring their subjects or other minor deformations which maintain overall cohesion.

Case-based reasoning does not normally apply data augmentation, but the solution transformation container provides a natural source for such perturbations. *Expansion-contraction compression* explores unseen regions of the problem space using adaptation knowledge to generate ghost cases and then exploits the ghost cases to broaden the range of cases available for competence-based deletion [6].

## 5 Future Work

Expansion-contraction compression explores the problem space arbitrarily — not necessarily in the direction of unrepresentative regions. Therefore, similar to [8], work in progress involves designing a case-base maintenance strategy which models the competence holes in a case base and targets expansion-contraction compression to fill the competence holes located between nearby competence groups by using adaptation knowledge to discover new cases. Evaluation will compare this strategy to untargeted expansion-contraction compression on multiple standard machine learning data sets and measure the retention of competence and solution quality for case bases compressed to the same number of cases.

This research summary anticipates a dissertation consisting of six chapters. The first chapter will describe the case-based reasoning cycle and motivate case-

based maintenance in terms of the swamping utility problem. The second chapter will explain the uniform storage and indivisibility assumptions and evaluate flexible feature deletion. The third chapter will define recoverability and evaluate adaptation-guided feature deletion. The fourth chapter will explain the representativeness assumption and overfitting problem and evaluate expansion-contraction compression. The fifth chapter will explain competence groups and competence holes and evaluate targeted expansion-contraction compression. The sixth chapter will envision future work and conclude by restating the key contributions.

## References

1. Dietterich, T. (1995). Overfitting and Undercomputing in Machine Learning. *ACM Computing Surveys*, 27(3), 326-327. <https://doi.org/10.1145/212094.212114>
2. Juarez, J. M., Craw, S., Lopez-Delgado, J. R., & Campos, M. (2018). Maintenance of Case Bases: Current Algorithms After Fifty Years. *International Joint Conference on Artificial Intelligence*, 27, 5457-5463. <https://doi.org/10.24963/ijcai.2018/770>
3. Lawrence, S., Giles, C. L., & Tsoi, A. C. (1997). Lessons in Neural Network Training: Overfitting May Be Harder Than Expected. *Proceedings of the 14th National Conference on Artificial Intelligence*, 540-545. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.38.6468\&rep=rep1\&type=pdf>
4. Leake, D., & Schack, B. (2015). Flexible Feature Deletion: Compacting Case Bases by Selectively Compressing Case Contents. *Case-Based Reasoning Research and Development*, 212-227. [https://doi.org/10.1007/978-3-319-24586-7\\_15](https://doi.org/10.1007/978-3-319-24586-7_15)
5. Leake, D., & Schack, B. (2016). Adaptation-Guided Feature Deletion: Testing Recoverability to Guide Case Compression. *Case-Based Reasoning Research and Development*, 9969, 234-248. [https://doi.org/10.1007/978-3-319-47096-2\\_16](https://doi.org/10.1007/978-3-319-47096-2_16)
6. Leake, D., & Schack, B. (2018). Exploration vs. Exploitation in Case-Base Maintenance: Leveraging Competence-Based Deletion with Ghost Cases. *Case-Based Reasoning Research and Development*, 11156, 202-218. [https://doi.org/10.1007/978-3-030-01081-2\\_14](https://doi.org/10.1007/978-3-030-01081-2_14)
7. Mántaras, R. L. De, McSherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S., ... Watson, I. (2005). Retrieval, Reuse, Revision, and Retention in Case-Based Reasoning. *The Knowledge Engineering Review*, 20(3), 215-240. <https://doi.org/10.1017/S0269888906000646>
8. McKenna, E., & Smyth, B. (2002). Competence-Guided Case Discovery. *Research and Development in Intelligent Systems XVIII*, 97-108. [https://doi.org/10.1007/978-1-4471-0119-2\\_8](https://doi.org/10.1007/978-1-4471-0119-2_8)
9. Schack, B. (2016). Feature-Centric Approaches to Case-Base Maintenance. *Proceedings of the ICCBR 2016 Workshops*, 287-291. Retrieved from <https://pdfs.semanticscholar.org/9946/c00d9f03e5b7829b227d8501b4f7439d132d.pdf>
10. Schack, B., & Summers, R. (2017). Flexible Feature Deletion. *International Conference on Case-Based Reasoning Video Competition* (p. 2). United States. Retrieved from <http://sce.carleton.ca/~mfloyd/ICCBVC2017/#nominees>
11. Smyth, B., & Cunningham, P. (1996). The Utility Problem Analysed: A Case-Based Reasoning Perspective. *Advances in Case-Based Reasoning*, (November), 392-399. <https://doi.org/10.1007/BFb0020625>
12. Wong, S. C., Gatt, A., Stamatescu, V., & McDonnell, M. D. (2016). Understanding Data Augmentation for Classification: When to Warp? *Digital Image Computing: Techniques and Applications*. Retrieved from <https://arxiv.org/pdf/1609.08764.pdf>