Dispatching Cases versus Merging Case-Bases: When MCBR Matters*

David B. Leake
Computer Science Department, Lindley Hall 215
Indiana University
Bloomington, IN 47405, U.S.A.
leake@cs.indiana.edu

Raja Sooriamurthi
Kelley School of Business, Business 540D
Indiana University
Bloomington, IN 47405, U.S.A.
raja@indiana.edu

Abstract

Multi-case-base reasoning (MCBR) extends case-based reasoning to draw on multiple case bases that may address somewhat different tasks. In MCBR, an agent selectively supplements its own case-base as needed, by dispatching problems to external case-bases and using cross-case-base adaptation to adjust their solutions for inter-case-base differences. MCBR is often advocated as a means to facilitate handling large case-bases, or to enable use of distributed case sources. However, this raises an important question: When storage is not an issue, and the entire external case-base is available, is there any reason for MCBR? This paper answers that question with an experimental assessment of how MCBR affects the quality of solutions generated. It demonstrates that for a given local case-base and an external case-base for a task environment that is similar to, but different from, the local task environment, MCBR can improve accuracy compared to merging the case-bases into a single case-base. This improvement holds even if the cross-case-base adaptation method used by MCBR is also applied to the external cases before merging. The paper hypothesizes an explanation of this behavior in terms of the ability of MCBR to exploit the tradeoffs between similarity of problems and similarity of solution contexts. It provides experimental evidence to support this hypothesis, and also demonstrates that MCBR is a useful framework for selecting cases to add to a case-base.

Introduction

Multi-case-base reasoning (MCBR) addresses how case-based reasoning systems can supplement their own experiences by drawing on the experiences of other case-based reasoners: how they can make effective use of external case-bases which may have been generated for related, but possibly non-identical tasks (Leake & Sooriamurthi 2002b; 2002a). When the local case-base is sparse, as in the early phases of case-base development, MCBR effectively extends the system's case-base by importing cases. MCBR reasons about issues such as when to dispatch a problem for case retrieval from an external case-base, which case-base is most suitable for solving the current problem, and how to revise solutions in light of inter-case-base differences. Using MCBR to access external case sources as needed contrasts with "eager" merging, in which all cases from all sources are standardized and merged into a combined case-base.

MCBR has been advocated as a means to facilitate processing of large case-bases, by handling subsets of the case-base separately, and for exploiting distributed case sources which may not be available to merge. However, this raises an obvious question: When merging is practical, is there any reason to perform MCBR rather than simply merging cases into a single case-base?

This paper examines the question of when MCBR is preferable to case-base merging, focusing on an issue that has not been studied previously: The comparative effects of MCBR and eager merging on solution accuracy. The paper presents experiments demonstrating a surprising result: that even when the same cross-case-base adaptation strategy is applied to each externally-retrieved case by MCBR, and to each of the cases used for eager merging, MCBR can result in markedly higher accuracy when the local and external case-bases address slightly different tasks. It hypothesizes an explanation for this behavior—that MCBR enables making useful tradeoffs between similarity and expected solution quality when using cases from different sources. It then provides evidence for this hypothesis with an additional experiment showing that its predictions are borne out in another context, when the local and external case-bases reflect identical tasks but external cases have noisy solutions. It also demonstrates that MCBR processes can be used as an effective strategy for selectively adding cases to the case-base. These results show that MCBR is sometimes useful not only for space and case availability concerns, but for reasoning accuracy as well.

The MCBR Process

Figure 1 illustrates one possible MCBR architecture. When problems are input to the system, a dispatcher selects case-bases to query and a strategy for pursuing the query sequence. The system then selects returned cases to consider, performs cross-case-base adaptation for inter-case-base differences, and merges solutions if needed. These steps may be ordered in different ways, e.g., cross-case-base adaptation may precede or follow the solution merging step.

*This research is supported in part by NASA under award No NCC 2-1216 and by the National Science Foundation under award EIA-0202048. We thank the reviewers for their helpful comments.

For example, in an e-commerce application, a local merchant asked to estimate the price for a particular home theater system might first seek product information from its own case-base, and, if no sufficiently similar cases were available there, might dispatch the queries to other suppliers. It might then have to adjust the predicted prices based on systematic inter-case-base differences (e.g., to convert prices from Euros to dollars, if cases from a European supplier were being retrieved for American use).

The general MCBR framework may treat the local case-base in a special way, or may treat all case-bases, local and external, according to a uniform set of dispatching criteria. For example, if the local case-base has a low access cost, it might be favored by a cost-based dispatching strategy, but if it contains noisy cases, it might not be favored by a quality-oriented dispatching strategy. In what follows, for simplicity we will assume that there is a single fixed local case-base and a single fixed external case-base. However, we believe that the results will apply to the more general model as well.

We have explored a dispatching method in which problems are dispatched to an external case-base if their distance from the most similar local case exceeds a threshold, as well as methods for automatically setting this threshold (Leake & Sooriamurthi 2002b). The dispatching method treats the external case-base and its similarity/retrieval method as a "black box," analyzing system performance on initial problems and selecting the dispatching threshold that gives the best performance on that sample.

New Motivations for MCBR: Two Hypotheses

A number of potential benefits have been advanced for using MCBR (Leake & Sooriamurthi 2002b). These include *increasing efficiency*—dividing up the case-base may increase retrieval speed—*augmenting coverage when needed*, and *exploiting distributed case information available on a per-case basis.* If external case-bases become commercial knowledge resources, cases may be available individually, but without access to the case-base as a whole.

Although there are compelling arguments for these advantages, they apply primarily when constraints on processing speed, storage, or case access prevent merging the different case-bases. If those constraints do not apply, it would be possible to simply import all external cases as a group, perform cross-case-base adaptation on the entire group, and then to reason from the single resulting case-base by the normal CBR process. However, this paper proposes two new motivations that can make MCBR useful even when it is practical to perform eager case-base merging:

1. **Increasing solution quality:** If two case-bases are merged, all the cases in the resulting case-base are treated uniformly for future retrievals, and the criterion for choosing between their cases is simply their similarity to new problems. In MCBR, however, considerations beyond similarity—such as the quality of cross-case-base adaptation—can be traded off against similarity. This prompts the *MCBR quality hypothesis*: MCBR—lazy retrieval and cross-case-base adaptation—should sometimes provide improved solution quality compared to cross-case-base adapting all external cases and merging them with the local case-base.
2. **Guiding selective case addition:** When a CBR system has insufficient competence, additional cases may be needed. Research on case discovery examines how to select cases to add (McSherry 2000) and to identify cases to fill competence gaps in the case-base (e.g., (McKenna & Smyth 2001)). The MCBR process of dispatching cases to external case-bases may be seen as another strategy for case-base building, in which cases are added to fill only those competence gaps that affect the current performance of the CBR system, given the characteristics of local case-base, external case-base, and cross-case-base adaptation. This prompts the *MCBR case-base building hypothesis*: MCBR is a useful framework for guiding case addition.

These hypotheses are examined in the following sections.

Solution Quality Effects for Merging vs. Dispatching

The success of MCBR depends on the availability of external case-bases to fill the local case-base's competence gaps. When such external case-bases are available, merging all case-bases may seem to be a good alternative to MCBR's lazy merging: Eager merging avoids the overhead and potential for error of case dispatching. However, the *MCBR quality hypothesis* suggests that MCBR may be advantageous even when merging is possible. To explore this hypothesis, we performed experiments on the quality effects of eager vs. lazy merging.

Experimental Design

Our experiments compared lazy and eager case-base merging for the task of predicting median housing prices, using a publicly-available data set from the Delve group.[1] This data set includes 22,784 cases from the 1990 U.S. census, divided by states. In the experiments, a randomly-selected subset of the cases from one state was used as the "local" case-base, and the complete case-base for another state was used as the "external" case-base. The reason for including only a subset of the local cases was to simulate the circumstances when MCBR is expected to be useful: when the local case-base has incomplete coverage, so additional cases are needed.

Here all cases use the same representation scheme, but the price for a given set of property features will change based on differences in the housing markets in different states. Intuitively, this prediction task can be seen as related to what a real estate appraiser might do after moving to a new area, when it is necessary to reason from a combination of local and non-local experience. The state case-bases used (with their abbreviations and sizes) are Alabama (AL, 470 cases), Florida (FL, 752 cases), Indiana (IN, 590 cases), Illinois (IL, 1308 cases), Kentucky (KY, 471 cases), Mississippi (MS, 324 cases), and Ohio (OH, 1051 cases). In the experiments, sparse versions of the IN case-base were used as the local case-base. (An asterisk will designate the sparse version of a case-base, e.g., IN* is the sparse version of the Indiana

[1] http://www.cs.toronto.edu/~delve/data/census-house/desc.html.

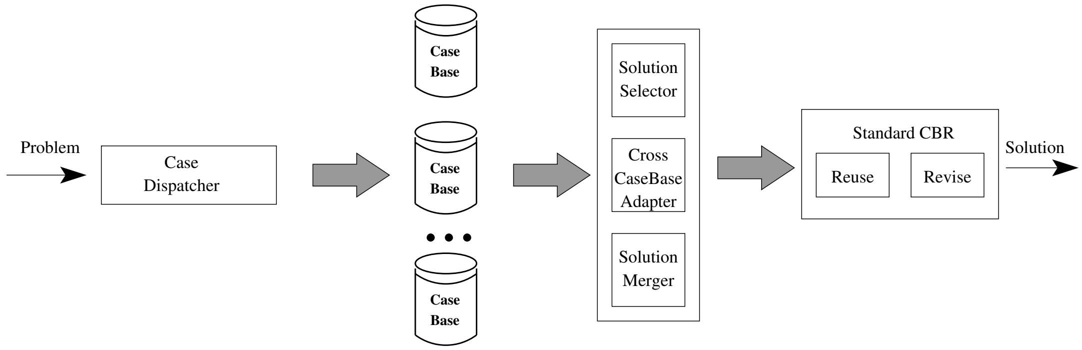

Figure 1: Multi-case-base reasoning framework for drawing on a set of case-bases (Leake & Sooriamurthi 2002b)

case-base). Complete versions of the other case-bases were taken as the external case-bases.

For each combination of sparse local case-base and external case-base, prediction quality was compared for lazy and eager merging. Both conditions used test problems from the original IN case-base, with leave-one-out cross validation. The system automatically selected one of four cross-case-base adaptation strategies—identity, linear interpolation, or two alternative case-based methods—expected to maximize performance, using the inductive selection algorithm described in (Leake & Sooriamurthi 2002a), applied to the first 30 problems in the problem stream.

For the eager merging condition, the merged case-base was generated by taking each case from the external case-base, applying cross-case-base adaptation, and adding the cross-case-base adapted version to the local case-base. In subsequent tests, normal CBR was done on the resulting case-base, with prices predicted by averaging prices of the 3 nearest neighbors. For MCBR, a dispatching threshold was selected by the learning method in (Leake & Sooriamurthi 2002a); when the distance between an input problem and the most similar case in the local case-base exceeds the threshold, the problem is dispatched to the external case-base. The external case for the most similar problem is then retrieved. If it is more similar than the most similar local case, the solutions of the three nearest neighbors in the external case-base are averaged, and the result is cross-case-base adapted to yield a solution.

Experimental Results on MCBR vs. Eager Merging

Figure 2 graphs the performance of MCBR vs. eager merging for IN* as the local case-base (here IN* contains 1% of the original IN cases), and OH, IL, KY, respectively, as external case-bases. The bars show the average performance for 10 versions of the local case-base, all containing the same number of randomly-selected cases. The first bar shows the prediction performance of the local case-base in isolation. Next, each pair of bars represents the merged and dispatched behavior for MCBR with a given local case-base/external case-base combination. In all trials, including tests done with other case-base combinations, the performance of MCBR with case-dispatch is at least as good as that of the merged case-base behavior and often noticeably surpasses it (e.g., for [IN*,KY], [AL*,MS]). Thus under some circumstances, eager merging can be detrimental to the performance of the CBR system. The following section hypothesizes an explanation for that behavior.

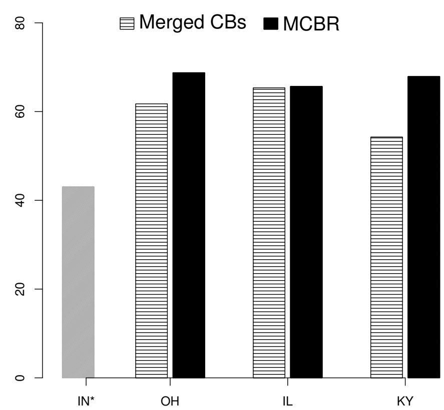

Figure 2: Performance of eager merging vs. MCBR.

Why MCBR May Increase Accuracy

Initially, it appears surprising that MCBR could improve quality compared to normal CBR after merging all available cases. At the very least, MCBR introduces a new potential source of error, the decision of whether to process a case in the local or (potentially less suitable) external case-base. However, we believe that the benefit of MCBR can be explained by the added control MCBR gives the case selection process, enabling using external cases only when they are expected to be beneficial.

The value of using a case from a case-base with somewhat different characteristics depends not only on the similarity of the problem that case solves to the new problem, but also on (1) the level of inter-case-base differences—the extent to which the solutions for similar problems may differ in the local and external case-bases, and (2) the ability of cross-case-base adaptation procedures to compensate for inter-case-base differences. Even if the external case is very similar to the input problem, using that case may de-

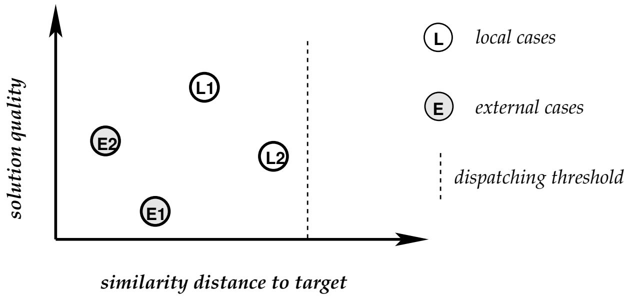

Figure 3: Why case-base merging may decrease accuracy.

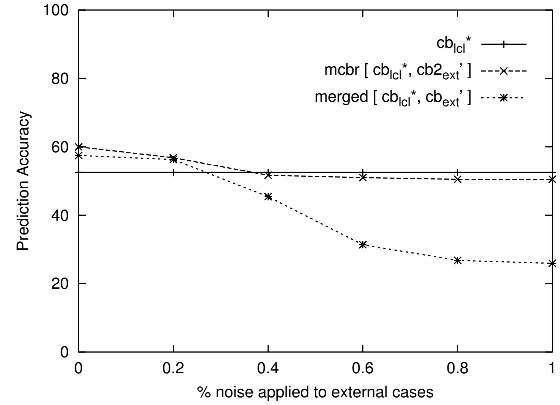

Figure 4: Comparative performance of case acquisition strategies: local CB (5% IN) and external noisy case-base (10% IN with added noise).

grade performance if it suggests a less-reliable solution than a more dissimilar case from the local case-base.

Figure 3 illustrates a potential bad result of eager merging. In the figure, the x-axis measures the similarity distance of cases to the given target problem. The y-axis measures the quality of the solution associated with a case. Here, if the local and external case-bases are merged, the two external cases (E1 and E2) will be most similar to the target problem. E2, the closest case, will be used to solve the target, resulting in a comparatively poor solution. However, if cases in the external case-base systematically produce worse solutions for a given similarity level, MCBR can take that into account. By learning a sufficiently high dispatching threshold (thereby preferring to solve problems using local cases), MCBR can use only those external cases that are so much more similar that their similarity counterbalances their lower expected solution quality for a given similarity level. In the figure, the dotted line on the x-axis indicates a possible dispatching threshold to avoid the previous problem. In MCBR, because the distance from L1 to the target problem is less than the dispatching threshold, L1's (higher quality) solution is used. To illustrate using the real estate appraisal domain, a realtor may generate a better estimate pricing a new house based on a somewhat dissimilar house that is nearby, instead of a more similar house in a different area. However, using cases from another area might be worthwhile if local cases are too dissimilar (e.g., pricing an apartment when the only local cases are for houses).

Thus with the right dispatching criteria, MCBR can improve solution quality compared to eager merging, by taking into account the quality of the external case-base and cross-case-base adaptation. Eager merging lacks that capability.

MCBR to Guide Use of Noisy Case-Bases

If the explanation in the previous section is correct, MCBR should be beneficial when there are differences between the expected quality of solutions from external and local case-bases. To examine this hypothesis, we conducted an experiment in which the difference in expected quality comes from noise in the external case-base, rather than from differences in the tasks that the local and external case-bases address. This experiment also tested the ability of our learning strategy for case dispatching to automatically adjust to changing case-base characteristics.

In this experiment, we took two random samples of the IN case-base, treating one as the local and the other as the external case-base. To study the effects of case-base quality, varying amounts of uniformly-distributed noise were probabilistically added to the solution part of the cases in the external case-base. Our previous experiment on case-base merging vs. MCBR was repeated on this combination of local and external case-bases. Figure 4 graphs our results. We see that when a noisy case-base is merged, performance of the combined case-base drops. However, the dispatching learning strategy associated with MCBR is able to respond to the noisiness of the external case-base and its reduced contribution to forming an effective solution. Hence as noise increases in the external case-base, the dispatching learning strategy effectively shuts down case-dispatch, handles input problems locally, and thereby avoids the performance degradation that results when the case-bases are merged. This shows that when an external case-base is noisy, MCBR can automatically self-adjust the whole system to compensate for decreased effectiveness of the external case-base, again providing accuracy benefits compared to eager merging. This is consistent with our prediction.

Using MCBR to Guide Effective Case Addition

Although merging an entire external case-base may be detrimental when solutions in the external case-base are unreliable, it is clear that case additions may be beneficial, even from a case-base that is less reliable than the local one. For example, a new case from an external case-base may fill a crucial competence gap. We hypothesized that, given an original case-base and another case-base with less reliable solutions, MCBR is a good strategy for selecting which cases to import from the external case-base. To explore this hypothesis, we performed an experiment in which a very sparse (1%) version of the Indiana case-base (IN*) was the local case-base, and FL was the external case-base. We then compared (1) performance using only IN*, (2) performance using MCBR to select cases from FL to add, and adding the cross-case-base adapted versions of those cases to the local case-base; and, (3) performance selecting random cases from FL to add, and adding cross-case-base adapted versions of those cases. In conditions (2) and (3), all tests added one new case for each problem processed, and the same cross-case-base adaptation strategy was applied to all added cases. The test simulated conditions in the early phases of using a CBR system, when system competence is low and case addition is needed to increase competence.

The resulting competence was measured as the percentage of problems successfully solved. Tests were run 10 times, and results averaged. In these tests, on average IN* solved 43% of the problems correctly. IN* with random (cross-case-base adapted) additions from FL increased performance by 19% over the baseline (solving 51% of the problems correctly), while adding cases determined by MCBR gave a 30% increase over the baseline (solving 56% of the problems correctly). Thus MCBR may provide useful guidance for case addition during initial system use.

Relationships to Previous Research

The general idea of multi-case-base reasoning relates to research in distributed CBR, and in heterogeneous and distributed databases; some of these relationships are described in (Leake & Sooriamurthi 2002a). Previous CBR research has included research on problem dispatching (McGinty & Smyth 2001) and on exploiting solutions generated by different agents (Ontañón & Plaza 2001). Ontañón and Plaza's approach differs from that of MCBR, however, in focusing on when a sufficiently good solution has been obtained, rather than on deciding which case-base to query.

A significant body of CBR research studies case-base maintenance (e.g., (Leake *et al.* 2001)). MCBR can be seen as performing a form of "lazy" case-base maintenance: Cases are imported, and converted to local requirements by cross-case-base adaptation, as needed. A perhaps-surprising result from our experiments is that for accuracy, this lazy maintenance, done as-needed, can provide better results than eager maintenance.

Future Directions

As described previously, inter-case-base differences may have a strong effect on the expected usefulness of cases with given similarity levels. Although our experiments show that threshold-based dispatching can help address this when local and external case-bases have comparable similarity metrics, we anticipate that better performance could be achieved by a richer model of the relationships between the similarity criteria of two case-bases. We note that the MCBR dispatching process can be seen as imposing a new similarity measure on the space, based on expected usefulness of cases from different case-bases. Combining learned dispatching criteria with the case-bases' own similarity metrics is a convenient "knowledge light" way to reflect utility differences between cases in different case-bases. However, how additional knowledge could be used, how to decide when to use MCBR, rather than merging, and how to refine the dispatching threshold over time are all interesting open questions.

Our tests suggest that MCBR may also have value for selective case addition. This potentially includes updating legacy case-bases, by treating any already-updated cases as the "local" case-base and the legacy case-base as the "external" case-base, and automatically importing and cross-case-base adapting the legacy cases needed during processing.

Conclusion

Multi-case-base reasoning (MCBR) enables an agent to selectively augment its own case-base as needed, drawing on external case-bases and adjusting their solutions for inter-case-base differences. MCBR is often motivated by the desire to support experience-sharing without requiring access to a complete external case-base, and without the storage required to merge complete case-bases. Although its advantages are clear when those constraints are significant, this appeared to limit its usefulness to situations when those constraints hold. This paper shows that MCBR may have two additional benefits: Increasing solution accuracy compared to eager merging, and guiding case-base building. For a given local case-base and external case-base for a similar but different task, MCBR can improve accuracy compared to merging both case-bases into a single case-base, even if the same cross-case-base adaptation process is applied to both cases-bases. The paper hypothesizes an explanation of this behavior in terms of a tradeoff between similarity of problems and similarity of solution contexts, and demonstrates that the expected benefits also apply when using local and external case-bases developed for the same task environment, but with differing levels of reliability. These results make a case for the broader usefulness of MCBR.

References

Leake, D., and Sooriamurthi, R. 2002a. Automatically selecting strategies for multi-case-base reasoning. In Craw, S., and Preece, A., eds., *Proceedings of ECCBR-2002*, 204–219. Berlin: Springer Verlag.

Leake, D., and Sooriamurthi, R. 2002b. Managing multiple case-bases: Dimensions and issues. In *Proceedings of FLAIRS-2002*, 106–110. Menlo Park: AAAI Press.

Leake, D.; Smyth, B.; Yang, Q.; and Wilson, D., eds. 2001. *Maintaining Case-Based Reasoning Systems*. Blackwell. Special issue of *Computational Intelligence*, 17(2), 2001.

McKenna, E., and Smyth, B. 2001. Competence-guided case discovery. In *Research and Development in Intelligent Systems XVIII: Proceedings of ES 2001*. Berlin: Springer-Verlag.

McGinty, L., and Smyth, B. 2001. Collaborative case-based reasoning: Applications in personalised route planning. In *Proceedings of ICCBR-2001*. Berlin: Springer Verlag.

McSherry, D. 2000. Automating case selection in the construction of a case library. *Knowledge-Based Systems* 13(2-3):133–140.

Ontañón, S., and Plaza, E. 2001. Learning when to collaborate among learning agents. In *Machine Learning: ECML 2001*, 395–405. Berlin: Springer-Verlag.