An Ensemble Approach to Adaptation-Guided Retrieval

Vahid Jalali and David Leake
School of Informatics and Computing, Indiana University
Informatics West, 901 E. 10th Street
Bloomington, Indiana, 47408

Abstract

Instance-based learning methods predict the solution of a case from the solutions of similar cases. However, solutions can be generated from less similar cases as well, provided appropriate "case adaptation" rules are available to adjust the prior solutions to account for dissimilarities. In fact, case-based reasoning research on adaptation-guided retrieval (AGR) shows that it may be beneficial to base retrieval decisions primarily on the availability of suitable adaptation knowledge, rather than on similarity. This paper proposes a new method for adaptation-guided retrieval for numerical prediction (regression) tasks. The method, EAGR (*ensemble of adaptations*-guided retrieval) works by retrieving an ensemble of cases, with a case favored for retrieval if there exists an ensemble of adaptation rules suitable for adapting its solution to the current problem. The solution for the input problem is then calculated by applying each retrieved case's ensemble of adaptations to that case, and combining the generated values. The approach is evaluated on four sample domains compared to three baseline methods: k-NN, an adaptation-guided retrieval approach, and a previous approach using ensembles of adaptations without adaptation-guided retrieval. EAGR improves accuracy in the tested domains compared to the other methods.

Introduction

Case-based reasoning (CBR) (e.g., (Mantaras et al. 2005)) is a reasoning paradigm consisting of four steps: Retrieve; Reuse; Revise and Retain. Given an input problem, a new solution is generated by retrieving a prior case (the "base case") and revising its solution to fit the new problem (multiple base cases may be used as well). Revision is done by applying case adaptation rules to adjust the old solution for the differences between the input problem and the problem solved by the base case. Commonly, CBR systems retrieve the most similar case to the new problem to use as the base case. However, appropriate adaptation rules can enable generating quality solutions from cases that are not necessarily the most similar. Accordingly, some CBR systems aim their retrieval process primarily at retrieving cases for which they have suitable case adaptation knowledge. This method, adaptation-guided retrieval (AGR), can improve system performance (Smyth and Keane 1998).

When adaptation-guided retrieval is applied to regression (numerical prediction) tasks, it is common to base retrieval decisions on a single adaptation per difference (e.g. (Hanney and Keane 1996; McSherry 1998; McDonnell and Cunningham 2006). We generalize this with *Ensemble of Adaptations*-Guided Retrieval (EAGR) a method that selects cases to retrieve and adapt based on existence of ensemble of candidate adaptations for them. An important aspect of the approach is that it automatically generates adaptation rules from cases in the case base, providing a rich set of adaptations on which to draw. The method is tested in four sample domains, and results show that it increases accuracy over baseline methods on the test domains.

The paper is organized as follows: We first review existing work on adaptation-guided retrieval in case-based reasoning. Next we discuss instance-based methods for regression tasks. We then introduce *ensemble of Adaptations*-Guided Retrieval (EAGR) and compare it with other existing case-based regression methods. We follow with our experimental design and results, and close with a summary and discussion of future work.

Related Work

In this section we present a brief synopsis of adaptation-guided retrieval, a discussion of work on combining adaptation rule generation with AGR, and a discussion of previous work on using rule ensembles for case adaptation and prior instance-based regression methods, which EAGR extends by adding adaptation-guided retrieval.

Adaptation-Guided Retrieval

Adaptation-Guided Retrieval was introduced by Smyth and Keane in their seminal work questioning that similar experiences are always the best guide to future reasoning (Smyth and Keane 1998). They argue for using adaptation knowledge to determine whether a case can be easily modified to fit new circumstances, and favoring adaptable cases. They illustrate in the domain of autonomous robotic vehicle programming, showing that AGR can improve the problem-solving efficiency over using standard similarity for retrieving cases to adapt. Adaptation-guided retrieval has been applied to a number of other task domains, including medical diagnosis (Djebbar and Merouani 2012), failure diagnosis (Haouchine, Chebel-Morello, and Zerhouni 2008), case-based planning (Tonidandel and Rillo 2001), and example-based machine translation (Collins and Cunningham 1996).

Combining Adaptation Rule Generation with AGR

Adaptation-guided retrieval can be combined with methods to automatically generate adaptation rules (e.g., (d'Aquin et al. 2007; McDonnell and Cunningham 2006). A common "knowledge light" approach to adaptation rule generation for regression is the *Case Difference Heuristic* (Hanney and Keane 1996), which generates new adaptation rules by comparing pairs of cases in the case base, and ascribing their solution difference to the differences in their problem descriptions. For example, for real estate price prediction, if there are two cases for apartments in the case base whose descriptions differ only in that one apartment is 100 square feet larger than the other, and the larger apartment's rent is $20 more per month, this could suggest the rule that a 100 square foot size increase should increase the rent by $20. Note that this difference could also suggest other rules. How to determine the right rule is an interesting issue for AGR, but is beyond the scope of this paper; our focus is simply on making good use of whatever adaptation rules are available and alleviating some of the problems from varying rule quality through ensemble methods.

d'Aquin et al. (2007) apply the case difference heuristic to generate adaptations in a semi-automated decision support method for cancer treatment, and apply principles of knowledge discovery from databases for generating the final set of adaptations. McDonnell and Cunningham (2006) apply AGR to regression tasks, with rules generated by the case difference heuristic. Their AGR method ranks cases to adapt by both their similarity to the input problem and availability of adaptations that can address the differences between cases to adapt and the input query. They adjust the value of the highest-ranked cases by applying the retrieved adaptations and combine the adjusted values for generating the final values.

Case-based reasoning has itself been applied both to generation of new adaptations, and to selecting adaptable cases (Leake, Kinley, and Wilson 1997). Generating adaptation knowledge from cases in the case base can also be looked at from the perspective of shifting knowledge between CBR *knowledge containers* (Richter 1998). Shiu et al. 2001 have explored such shifting in a case base maintenance method that decreases the number of cases by transferring the case knowledge into adaptation knowledge, using fuzzy decision trees to generate adaptations from cases in the case base.

Ensembles for Case Adaptation

Previous research includes some work on using ensembles of adaptations, and has supported the approach for increasing accuracy for both classification (Arshadi and Jurisica 2005; Wiratunga, Craw, and Rowe 2002) and regression (Jalali and Leake 2013b) tasks. For example, Jalali and Leake (2013b) show that fixing a set of base cases, applying an ensemble of adaptations for adjusting their values, and combining the adjusted values for creating the final solution can increase the accuracy of a case-based regression system compared to other candidate methods. Our method is in accordance with that of Jalali and Leake in fixing a base case and applying an ensemble of adaptations for adjusting its value instead of applying a single adaptation per base case. However, instead of relying on the similarity assumption when selecting base cases, as in that work, our approach uses the new adaptation-guided retrieval method we present here.

Instance-Based Regression

Instance-Based Regression, done by k-Nearest Neighbors, (Aha, Kibler, and Albert 1991; Altman 1992) estimates the numeric target value for an input problem by combining the values of the top k nearest neighbor instances, often by averaging their values. If Q represents the input problem, and C_i's are the k cases most similar to the input query, k-NN calculates the value for Q as:

$$Val(Q) \equiv CombineVal(\bigcup_{i=1}^{k} C_i) \tag{1}$$

Adding Richer Adaptation: Case-Based Regression

Case-Based Regression (e.g. (McSherry 1998)) enriches instance-based regression by adding a case adaptation step before values are combined. If $AdjustedVal(C_i)$ represents the value of C_i after applying the selected adaptation, case-based regression estimates the value of Q as follows:

$$Val(Q) \equiv CombineVal(\bigcup_{i=1}^{k} AdjustedVal(C_i)) \tag{2}$$

where the cases C_i are the k most similar cases to Q, and their adjusted values are calculated by:

$$AdjustedVal(C) \equiv Val(C) + (Val(C_l) - Val(C_m)) \tag{3}$$

where C_l and C_m are selected so that their differences are most similar to the differences between Q and C among all pairs of cases in the case base (the case difference heuristic is used for generating adaptations). The generic case-based regression process is illustrated in Fig. 1.

Adding Adaptation-Guided Retrieval

If the case difference heuristic is used for generating adaptations, then each pair of two arbitrary cases in the case base can give rise to an adaptation rule to apply to the base case. Therefore, in this case-based regression approach the target value is determined by triple of cases, one as the base case and two for generating the adaptation. One way of applying adaptation-guided retrieval is ranking the relevance of each triple of cases based on two criteria as proposed by McDonnell and Cunningham (2006): First, the similarity of the base case and the input query; second, the similarity of the differences addressed by the adaptation rule and the differences between the pair of input query and the base case.

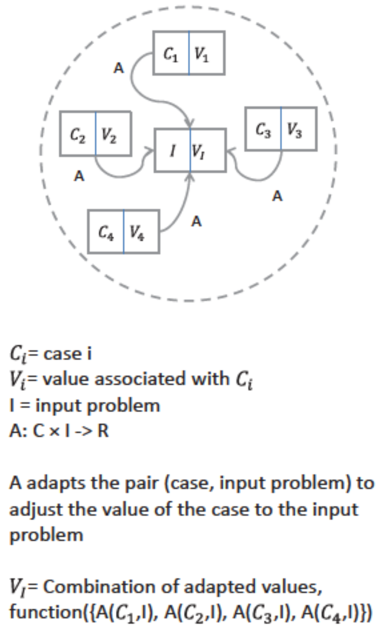

Figure 1: Illustration of the generic case-based regression process

McDonnell and Cunningham propose combining values derived by them for generating the final solution as follows:

$$Val(Q) \equiv CombineVal(\bigcup_{i=1}^{k} SuggestedVal(Triple_i)) \tag{4}$$

where $SuggestedVal$ is a function for calculating the value of the first case in $Triple_i$ by applying the adaptation derived from the other two cases in it according to Eq. 3. If Δ represents the difference vector between two cases and Sim represents the similarity of two cases or difference vectors, $Triple_i$'s are ordered triples of cases belonging to the case base and ranked by their descending scores calculated as:

$$\begin{aligned} &Score(Q, Triple(BC, C_l, C_m)) \equiv \\ &\quad Sim(Q, BC) + Sim(\Delta(Q, BC), \Delta(C_l, C_m)) \end{aligned} \tag{5}$$

The similarity assumption is not completely discarded in McDonnell and Cunningham's method, but given an appropriate adaptation it is possible that given two base cases, the one that is less similar to the input query will be preferred to the more similar. If $Sim(Q, BC)$ were removed from Eq. 5, the resulting method would be closer to the principles of adaptation-guided retrieval which we test in this paper.

Ensemble of Adaptations for Regression

EAR, a method introduced by Jalali and Leake (2013b), selects a fixed number of base cases by using a specific criterion (e.g. their similarity to the input query) and applying an ensemble of adaptations for adjusting their values. The final estimation is still calculated by combining the adjusted values of the selected base cases.

The value of an input query Q, using EAR is calculated as explained in Eq. 2. However, in contrast to case-based regression, the adjusted values of C_i's are calculated by applying an ensemble of adaptations as follows:

$$\begin{aligned} &AdjustedVal(C) \equiv \\ &\qquad Val(C) + CombineVal(\bigcup_{i=1}^{n} \Delta(Pair_i)) \end{aligned} \tag{6}$$

where $CombineVal$ plays the role explained in Eq. 1, though here—instead of taking a case as its input—it takes a difference vector as its input. $Pair_i$'s are ordered pairs of cases belonging to the case base; each of these pairs determines an adaptation rule, generated on the fly from the two cases, using the case difference heuristic. We call the pair of cases the *composing cases* of the adaptation rule. The pairs are ranked by their descending score calculated as follows, if $Pair_i = (C_l, C_m)$.

$$\begin{aligned} &Adaptation\_Relevance\_Score(Q, Pair_i, BC) \equiv \\ &\qquad Sim(\Delta(Q, BC), \Delta(C_l, C_m)) \end{aligned} \tag{7}$$

where BC denotes the base case selected in Eq. 2. Note that retrieval in EAR is not guided by adaptability considerations.

Ensemble of Adaptations-Guided Retrieval

In AGR, a single base case is normally retrieved, and in AGR for regression, a single adaptation is normally applied. Especially when adaptation rules are being generated automatically from the case base, their quality is not guaranteed (e.g. due to variation in the context of the cases (Jalali and Leake 2013a) or due to varying confidence in the value of the composing cases (Jalali and Leake 2013c)). Therefore, we hypothesize that applying an ensemble of adaptations can make it possible to generate better solutions.

Our ensemble method builds on the ideas introduced by McDonnell and Cunningham (2006) and Jalali and Leake (2013b), extending Jalali and Leake's method by having the AGR process retrieve base cases for which an *ensemble* of appropriate adaptation exists. For example, if an ensemble of four adaptations will be applied to adjust the values of base cases, in the retrieval process, cases are favored based on the availability of four quality adaptations for them.

EAGR, like EAR, uses Eq. 2 and Eq. 6 for calculating the value of an input query Q. However, in contrast to EAR, selection of base cases is done by ranking each candidate base case according to the expected quality of the best ensemble of r adaptations available to adapt it. In particular, given a candidate base case BC, pairs of cases (corresponding to possible adaptations) are selected and ranked according to Eq. 8. The top-ranked k base cases are retained. Then, for each retained base case BC, the top k adaptations retrieved in the first step are combined and applied to BC to adjust its value.

Algorithm 1 Ensemble Adaptation-Guided Retrieval's basic algorithm

Input:
Q: input problem
k: number of base cases to adapt to solve query
r: number of rules to be applied per base case
CB: case base
R: set of existing adaptations
Output: Estimated solution value for Q

$(CasesToAdapt, AdaptationsToApply) \leftarrow$ BaseCaseSelection(Q,k,r,CB)
for c in $CasesToAdapt$ **do**
  $ValEstimate(c) \leftarrow$ Val(c) + CombineAdaptations($AdaptationsToApply$)
end for
return CombineVals($\cup_{c \in CasesToAdapt} ValEstimate(c)$)

$$Case\_Adaptability\_Score(Q, BC) \equiv \sum_{i=1}^{k} Sim(\Delta(Q, BC), \Delta(Pair_i)) \tag{8}$$

Here, for each base case BC, $Pair_1$ to $Pair_k$ are composing cases of adaptations that best represent the differences between Q and BC (compared to other pairs of cases in the case base).

To estimate the target value of an input query, the EAGR method ranks cases based on the availability of an ensemble of appropriate adaptations that addresses the differences between them and the input problem. For each case to adapt, its value is adjusted by applying an ensemble of adaptations to it and the final solution is then calculated by combining the adjusted values. Algorithm 1 summarizes the entire process.

The most naive way of applying EAGR requires considering each case as a candidate base case and for each case all pairs of cases in the case base can be considered as potential adaptations to be applied. This gives rise to considering $O(n^2)$ rules per base case which consequently ends up to $O(n^3)$ comparisons for determining the target value of an input problem. However, in practice the time complexity of EAGR is less than $O(n^3)$ because at any point if the differences between the pair of input query-current base case and the ensemble of considered adaptations exceeds the level of difference observed for the current top k base cases, then that rule can be discarded. Another way to make application of EAGR to bigger case bases feasible is exploiting a competence-based case deletion strategy to compact the case base, or using case prototypes in solution building instead of examining all cases in the case base. In addition, another possible remedy for the time complexity issue is reusing case base maintenance lessons (e.g., (Smyth 1998) to maintain and condense the adaptation knowledge, and threfore reducing the number of cases that needs to be considered for base case selection.

Experimental Design

We conducted experiments on four sample domains from the UCI repository (Frank and Asuncion 2010): MPG, with 7 features and 392 cases, Auto Price, with 13 features and 195 cases, Housing, with 13 features and 506 cases, and Hardware, with 7 features and 209 cases. Each data set is cleaned by removing cases with unknown feature values. For each input feature, values are standardized by subtracting that feature value's mean from each individual feature value and dividing the result by the standard deviation of that feature. Target values are not standardized. Experiments assess the performance of four candidate methods: kNN, AGR, EAR and EAGR. The retrieval process for AGR was as described in Equation 5, except that our approach does not consider the similarity of the base case to the input problem; consequently we omitted $Sim(Q, BC)$ from Equation 5 in our implementation of AGR, so every case in the case base has an equal chance to be selected as a base case, if it has an equally good ensemble of adaptations.

In our experiments we use two versions of AGR based on the number of unique base cases used for building the solution. We considered two conditions, limiting the number of base cases to 5 and to 20, respectively; we refer to the corresponding AGR methods as AGR5 and AGR20. For other methods the number of base cases is always limited to 5.

Ten fold cross validation is used for evaluating the performance of each method. Parameters used in each method are turned by using Leave-One-Out testing on the training data. The version of EAR used in the experiments is EAR5 introduced in (Jalali and Leake 2013b).

Experimental Results

Comparisons of the Candidate Methods

Table 1 shows the Mean Absolute Error for the candidate methods in the test domains. The results show that EAGR outperformed the other methods for all tests. We note that performance of AGR20 was always superior to AGR 5—additional cases always improved performance—and that the level of improvement was related to the case base size. This shows that the adjusted values of single base cases by applying a single adaptation might not be as accurate as they can be, but given enough of them it is possible to estimate a target value more accurately.

	Domain			
Approach	MPG	Auto	Housing	hardware
kNN	2.07	1.59	2.46	36.33
AGR5	3.12	1.67	2.72	39.39
AGR20	2.22	1.57	2.13	32.73
EAR	1.93	1.52	2.1	32.53
EAGR	**1.9**	**1.5**	**1.87**	**29.2**

Table 1: Mean Absolute Error of kNN, AGR, EAR and EAGR for the sample domains

Effect of the Number of Used Base Cases on the Performance of AGR

To assess the effect of increasing the number of base cases on the performance of AGR we conducted experiments and measured MAE in the Auto domain by using different number of base cases for building the solutions. The results are reported in Fig. 2. Lines in the graph are for visibility only. Increasing the number of base cases initially decreases MAE, until increasing base cases begins to degrade performance. We hypothesize that this results when the number of base cases considered outstrips the system's ability to find relevant adaptations.

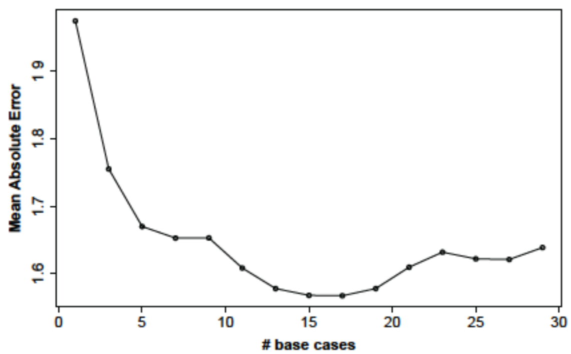

Figure 2: MAE of AGR when different numbers of base cases are used in the Auto domain

Percent of Improvement of AGR, EAR and EAGR over k-NN

Figure. 3 depicts the MAE of AGR20, EAR and EAGR methods for the test domains compared to that of k-NN. It can be seen EAGR shows the greatest improvement (ranging between 6% to 24% in the test domains) over k-NN. The performance of AGR20 and EAR in the housing and hardware domains are very close however, for the mpg and auto domains EAR outperforms AGR20.

Statistical Significance

To assess the significance of the results achieved by EAGR compared to other methods, one side paired t-test with 95% confidence interval is used. The null hypothesis is always MAE of EAGR being less than those of other methods.

Table 2 shows the statistical significance of EAGR compared to the results reported for the candidate methods. It can be seen that the gain over k-NN, AGR5 and AGR20 is often significant except for the Auto domain. Also, for Auto and MPG domains the difference between EAR and EAGR is not statistically significant.

Conclusion and Future Work

We have introduced EAGR, an adaptation-guided retrieval method for regression tasks which selects cases to adapt for which it can find an ensemble of appropriate adaptations.

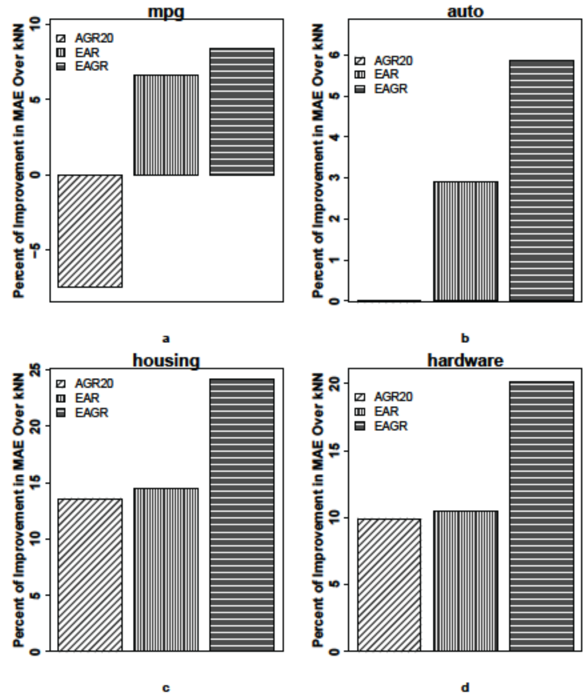

Figure 3: MAE of AGR, EAR and EAGR compared to kNN for the sample domains

	Domain			
Approach	MPG	Auto	Housing	hardware
kNN	p<.03	**p<.3**	p<.01	p<.01
AGR5	p<.01	**p<.08**	p<.01	p<.01
AGR20	p<.01	**p<.2**	p<.01	p<.01
EAR	**p<.4**	**p<.5**	p<.01	p<.05

Table 2: Statistical significance of performance of EAGR compared to other candidate methods in the test domains

EAGR uses the case difference heuristic as a knowledge-light method for generating adaptations by comparing pairs of cases in the case base. In the retrieval stage, EAGR ranks cases based on the existence of an ensemble of appropriate adaptations that can address differences between them and the input query. The values of top ranked cases are then adjusted by applying the ensemble of adaptations and the final estimation is calculated by combining those adjusted values. We tested EAGR on four sample domains and compared its performance with that of three candidate methods: kNN, AGR and EAR. In our tests, EAGR outperformed these methods.

Because for AGR and EAGR, each case in the case base has the chance of being used in building the final solution, these methods require more computational resources than k-NN, in which no adaptation is required, and than EAR in which adaptations are only retrieved for a limited number of cases. Future work includes increasing the efficiency of EAGR, e.g., by applying case base maintenance techniques to shrink the case base size or the set of adaptation rules before applying EAGR, by limiting the number of cases to be considered as candidate base cases by exploiting techniques such as clustering, or by using indexing techniques to improve the efficiency of EAGR's retrieval. Another future direction for our work is extending the ideas introduced in EAGR to domains with symbolic features/solutions where cases to adapt are selected based on availability of an ensemble of adaptations to apply.

References

Aha, D.; Kibler, D.; and Albert, M. 1991. Instance-based learning algorithms. *Machine Learning* 6(1):37–66.

Altman, N. S. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* 46(3):pp. 175–185.

Arshadi, N., and Jurisica, I. 2005. An ensemble of case-based classifiers for high-dimensional biological domains. In *ICCBR*, 21–34.

Collins, B., and Cunningham, P. 1996. Adaptation-guided retrieval in ebmt: A case-based approach to machine translation. In Smith, I., and Faltings, B., eds., *Advances in Case-Based Reasoning*, volume 1168 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 91–104.

Djebbar, A., and Merouani, H. F. 2012. Applying bn in cbr adaptation-guided retrieval for medical diagnosis. *differences* 5(2).

d'Aquin, M.; Badra, F.; Lafrogne, S.; Lieber, J.; Napoli, A.; and Szathmary, L. 2007. Case base mining for adaptation knowledge acquisition. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI-07)*, 750–755. San Mateo: Morgan Kaufmann.

Hanney, K., and Keane, M. 1996. Learning adaptation rules from a case-base. In *Proceedings of the Third European Workshop on Case-Based Reasoning*, 179–192. Berlin: Springer Verlag.

Haouchine, M. K.; Chebel-Morello, B.; and Zerhouni, N. 2008. Adaptation-guided retrieval for a diagnostic and repair help system dedicated to a pallets transfer. In Schaaf, M., ed., *ECCBR Workshops*, 33–42.

Jalali, V., and Leake, D. 2013a. A context-aware approach to selecting adaptations for case-based reasoning. In Brézillon, P.; Blackburn, P.; and Dapoigny, R., eds., *Modeling and Using Context*, volume 8175 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 101–114.

Jalali, V., and Leake, D. 2013b. Extending case adaptation with automatically-generated ensembles of adaptation rules. In *Case-Based Reasoning Research and Development, ICCBR 2013*, 188–202. Berlin: Springer-Verlag.

Jalali, V., and Leake, D. 2013c. On deriving adaptation rule confidence from the rule generation process. In *Case-Based Reasoning Research and Development, ICCBR 2013*, 179–187. Berlin: Springer-Verlag.

Leake, D.; Kinley, A.; and Wilson, D. 1997. A case study of case-based CBR. In *Proceedings of the Second International Conference on Case-Based Reasoning*, 371–382. Berlin: Springer Verlag.

Mantaras, R.; McSherry, D.; Bridge, D.; Leake, D.; Smyth, B.; Craw, S.; Faltings, B.; Maher, M.; Cox, M.; Forbus, K.; Keane, M.; Aamodt, A.; and Watson, I. 2005. Retrieval, reuse, revision, and retention in CBR. *Knowledge Engineering Review* 20(3).

McDonnell, N., and Cunningham, P. 2006. A knowledge-light approach to regression using case-based reasoning. In *Proceedings of the 8th European conference on Case-Based Reasoning*, ECCBR'06, 91–105. Berlin: Springer-Verlag.

McSherry, D. 1998. An adaptation heuristic for case-based estimation. In *Proceedings of the 4th European Workshop on Advances in Case-Based Reasoning*, EWCBR '98, 184–195. London, UK, UK: Springer-Verlag.

Richter, M. 1998. Introduction. In Lenz, M.; Bartsch-Spörl, B.; Burkhard, H.-D.; and Wess, S., eds., *CBR Technology: From Foundations to Applications*. Berlin: Springer. chapter 1, 1–15.

Shiu, S.; Yeung, D.; Sun, C.; and Wang, X. 2001. Transferring case knowledge to adaptation knowledge: An approach for case-base maintenance. *Computational Intelligence* 17(2):295–314.

Smyth, B., and Keane, M. 1998. Adaptation-guided retrieval: Questioning the similarity assumption in reasoning. *Artificial Intelligence* 102(2):249–293.

Smyth, B. 1998. Case-base maintenance. In *Proceedings of the Eleventh International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems.*

Tonidandel, F., and Rillo, M. 2001. An accurate adaptation-guided similarity metric for case-based planning. In *Proceedings of the 4th International Conference on Case-Based Reasoning: Case-Based Reasoning Research and Development*, ICCBR '01, 531–545. London, UK, UK: Springer-Verlag.

Wiratunga, N.; Craw, S.; and Rowe, R. 2002. Learning to adapt for case-based design. In *Proceedings of the 6th European Conference on Advances in Case-Based Reasoning*, ECCBR '02, 421–435. London, UK, UK: Springer-Verlag.