Combining Rules and Cases

K. Vanhoof, J. Surma

Faculty of Applied Economics

Limburg University Center

B-3590 Diepenbeek, Belgium

email: vanhoof@rsftew.luc.ac.be

phone: 00-32-11-268608

Abstract

The recent progress in case-based reasoning has shown that one of the most important challenges in developing future AI methods will be to combine and synergetically utilize general and case-based knowledge. In this paper an integration of rule-based and case-based reasoning is described. A typicality evaluator determines the applied method. The proposed method is tested and compared with alternative approaches. Experimental results show that the presented integration method can lead to an improvement in accuracy and comprehensibility.

Motivation

The recent progress in case-based reasoning has shown that one of the most important challenges in developing future AI methods will be to combine and synergetically utilize general and case-based knowledge [Aamodt 1995]. The presented approach has its origin in the Riesbeck and Schank's psychological consideration [Riesbeck & Schank 1989, pp.11]: " When an activity has been repeated often enough it becomes rule-like in nature. We do not reason from prior cases when well-established rules are available. [...] When the rule fails, the only alternative for its user is to create a case that captures that failure". Schank and Abelson proposed that our general knowledge and anomalies about situations is recorded and that these memory structures allow us to set up expectations. This is a specialized version of the standard schematic view of reasoning that says that general knowledge is resident in memory in chunks and that reasoning is a process of applying those chunks to new situations. What was new and different about Schank's approach is that it specified several particular types of chunks as being useful. Some chunks contain general knowledge, other chunks contain kowledge about specific cases. A basic problem remained: how do we retrieve and use both types of chunks. The theory of dynamic memory [Kolodner 1993] presented 'indexing' as the key to solve this problem. The hypothesis is that people label memory structures (both those holding general knowledge and those holding specific cases) according to their type and the way they differ from other structures. Translating this to our approach means thas cases are grouped by class label (type) and by an inter-class difference measure. When a memory structure, containing a specific case, is frequently used to understand another similar case, the memory structure becomes a memory structure that contains general knowledge. Memory structures containing specific instances are only stored in which expectations were violated or anomalies were encountered. Whether a general knowledge structure or a specific case structure will be applied to solve a new case depends on the frequency of similar cases solved. A new case with a lot of similar cases in a certain class has a high intra-class similarity. Summarising: the inter-class difference and the intra-class similarity of a case are labels used for retrieving the appropriate memory structure.

Combining cases and rules

We will call a case a typical case when there are a lot of similar cases with the same class label and there are few similar cases with a different class label. By using the theory of dynamic memory, the basic structure of a problem solver to classify a new case becomes :

IF a new case is a typical case
Then look for the appropriate memory structure containing general knowledge
Else look for the appropriate memory structure containing a specific case

This architecture contains three components: a typicality evaluator, a general knowledge component and a case-base component. The general knowledge component has been implemented with a rule set (derived with C4.5) and an inference engine. The case base component has been implemented with a cased-based classifier (1-NN). The main problem in the presented approach is to find a suitable function which gauges the typicality of a new case. In general this can be done by:

- *an expert* - the evaluation obtained from an expert is very valuable and gives an opportunity to obtain an additional explanation .
- *a statistical approach* (e.g. cluster analysis)
- *a heuristic approach* - there are a lot of informal approaches based on geometrical interpretations. The function can be based on weighting schemes as well, where exceptional cases can be determined to their performance and/or frequency of use for problem solving.

We took into consideration a heuristic approach based on Zhang's formalization of the family resemblance idea [Zhang 1992]. The similarity measure between cases X (x_1,x_2,..,x_m) and Y (y_1,y_2,...,y_m) is :

$$sim(X, Y) = 1 - \sqrt{\frac{1}{m}\sum_{i=1}^{m} dis(x_i, y_i)^2}$$

The metric used is the overlap metric. For nominal attributes, the distance is one when the attribute values (including unknowns) are different and zero otherwise. This similarity measure is reflexive and normalized to the range from 0 to 1. Intra-class similarity of a case is defined as a case's average similarity to other cases in the same class, and the inter-class similarity is defined as its average similarity to cases of all other classes.

We assume that typical cases have bigger intra-class similarity than inter-class similarity; the opposite holds for exception cases. So the final algorithm is simple:

IF intra-class similarity(case) > inter-class similarity(case)
THEN apply rule set
ELSE apply case-base

For a small data set the learning set should be used for determining typicality, generating rules and retrieving cases.

When the data set is large it is more reasonable to split the given set of cases into two disjoined subsets: Exception_Case_Base and Standard_Case_Base, and then to generate rules from the total dataset with a separate class (atypical) for the atypical cases. This has the advantage that you don't need a typicality evaluation during the classification. The basic algorithm slightly changes:

IF rule classification is atypical
THEN apply case-base
ELSE apply rule base

The first algorithm has been applied to a small marketing set and the latter to some standard datasets.

Experiments

A small marketing data set

Product-mix decisions in large retailing firms have become of growing concern to both manufacturers and retail management [Montgomery 1975, Swinnen 1983, McLaughlin & Rao 1988,1990]. As a consequence of the increasing buying power of distributors and the growing centralisation of buying decisions, less and less buyers decide on what products will get a chance to prove themselves on the shelves of large stores. For suppliers, it is therefore important not only to study consumer behaviour but also to find out more about what retail purchasing people want. For retail buyers as well, a better understanding of the decision-making process

may be a first step towards improvement. The database for the present study, consists of the data [Montgomery 1975] used to explore the relationship between 18 variables and a supermarket buyer's decision (from three different supermarkets) to accept or reject a new product.

The data of the first two supermarkets have been used to determine the typicality of the unseen cases, to induce the rules and to populate the case base. The dataset of the third supermarket have been used to test different methods. Table 1 summarizes the results The hold-out error rate is the error rate on the data set of the third supermarket .

Statistical methods	Holdout error rate
Linear disciminant analysis	18 %
Probit	14 %
Gate keeper	9 %
Disfre	9 %
AI methods	**Holdout error rate**
Decision tree	14 %
Pruned decision tree	14 %
Pruned rule sets	14 %
Cases and rules	9 %
Rule sets	14 %
Case-based	7 %

table 1: Holdout error rate of different methods

Experiments on standard datasets: Voting and Led display

The experiments were performed in order to compare the second integrated approach with the rule-based (rules from C4.5 algorithm) and the nearest-neighbour approach. The tests were based on three databases: the LED Display (with seven attributes) symbolic-valued artificial database with 10% amount of noise, the U.S. Congressional Voting 1984, symbolic valued database with unknown values, and a medical database Nurses (Surma 1994, a real database with numeric and symbolic attributes). Table 2 briefly characterises the domain and the experiments.

	LED Display	Voting	Nurses
Train size (no of records)	200	300	115
Tert size (no of records)	500	135	51
No of attributes	7	16	3
No of classes	10	2	5

table 2: Database Characteristics

The results of the accuracy comparison are shown in Table 3. The results for Nearest-Neighbour (1-NN) were obtained thanks to the Inducer utility from the MLC++ Machine Learning Library (Kohavi et al. 1994). In all experiments the rules were generated and tested thanks to Quinlan's C4.5 Machine Learning programs (Quinlan 1993). All the classifiers were tested on randomly drawn and separate training and test sets.

	LED display	Voting	Nurses
1 - NN	69.3 %	94.1 %	56.9 %
C4.5	69.1 %	95.6 %	60.8 %

Combined	69.7 %	97 %	58.8 %

table 3: Average classification accuracies

In general the combined approach showed a slightly higher accuracy than the methods on their own. However the number of rules decreased roughly with 30%, which enhances the comprehensability [Surma&Vanhoof 1995].

Also the quality of the derived rules seems to increase. In order to verify this hypothesis a simple experiment was done on the medical data set. The dataset consists of 39 atypical and 76 typical cases. From all 115 cases 8 rules (set R1) were generated (using C4.5). Next, from 76 typical cases six rules (set R2) were generated as wel. The rules were mixed and given to an expert for examination. The expert's task was to evaluate each rule in terms of a nonsense, wrong, tolerable or good rule. The results of this subjective evaluation are presented in table 4.

Rule set	Nonsense no of rules	Wrong no of rules	Tolerable no of rules	Good no of rules
R1 R2	3 0	1 2	2 2	2 2

table 4: Quality evaluation of derived rules

The rules from the R2 set are much closer to the expert opinion than the rules from R1. Of course this is only an experiment, but the result opens promising avenues for further evaluations.

Related work

The mixed paradigm involving case- based and traditional rule- based reasoning was included in the original CBR systems. The CHEF contained a rule- based sub module to support Case- Based Reasoning (Hammond 1988), and the CASEY used cases to supplement rule- based mechanism (Koton 1988).

The advanced researches in this field were made in the 1990s. Rissland and Skalak described a system CABARET that integrates reasoning with rules and reasoning with previous cases (Rissland & Skalak 1991). This integration was performed via a collection of control heuristics. Golding and Rosenblum proposed the architecture for combining Rule- Based and Case- Based Reasoning for the task of pronouncing surnames (Golding & Rosenblaum 1991). The central idea of their approach is to apply the rules to a target problem to get a first approximation to the answer, but if the problem is judged to be compellingly similar to a known exception of the rules, then the solution is rather based on the exception than on the rules. A good example of combined reasoning in framework for problem solving in knowledge rich environment is the CREEK system (Aamodt 1991), where first, the system is attempting to solve the problem by case- based reasoning, and if no acceptable match is found, rule based reasoning is attempted.

The selected researches in inductive learning were based on integration with case- based approach. Utgoff showed how the updating costs of incremental decision tree algorithms can be significantly decreased by saving specific instances (Utgoff 1989). Cardie proposed a method for using decision trees to specify the features to be included in k- nearest neighbour retrieval (Cardie 1993). The classical applying the decision tree as an index for case- based retrieval is implemented in commercial CBR tool ReMind, described and evaluated by Barletta (Barletta 1994).

Recently in Europe a number of systems have been constructed exploring various approaches of case and rule integration, including Cabata (Lenz 1994), BUBE/CcC+ (Bamberger & Goos 1993), and IKBALS (Zeleznikow et al., 1993). Malek and Rialle developed a computer aided medical diagnosis system for neuropathy diseases where the domain knowledge is represented in prototype cases, and non- typical cases. During a phase prototypes matched to the presented case are extracted, if no prototype is matched then the group of non-typical cases is retrieved (Malek & Rialle 1994). Armengol and Plaza presented the general knowledge modelling framework for solving the purification task of proteins. This task can be solved in different ways, especially by combining case- based retrieval and by using domain knowledge in the form of prototypes that are generated by an inductive method (Armengol & Plaza 1994).

The four possible levels of integration between the Induction and Case- Based Reasoning method were esta-

blished and tested in the INRECA project (Manago et al., 1993), (Auriol et al., 1994). In terms of stand-alone, cooperative, workbench and seamless level of integration, the approach presented in this paper is cooperative. This means that both methods are kept separated but they cooperate.
The few researches in multistrategy learning are very interesting from the rule and case integration point of view. Especially Matwin's and Plante's prototypical and marginal examples (Matwin & Plante 1994) are very close to the presented approach. That idea is based on Micalski Two- Tired representation concept (Michalski 1990).

Conclusion

The approach presented in this paper is one of the possible ways of integration. The described integration seems to be valuable for two reasons. Firstly, the outcomes of the accuracy comparisons are acceptable. Secondly, a better comprehensibility is given to the end user.

References

Aamodt, A. 1995. Knowledge Acquisition and Learning by Experience - The Role of Case- Specific Knowledge. Kodratoff, Y., Tecuci, G. eds. *On Integration of Knowledge Acquisition and Machine Learning.* Academic Press (in press).

Aamodt, A. 1991. A Knowledge- Intensive, Integrated Approach to Problem Solving and Sustained Learning. A Doctoral Dissertation - University of Trondheim.

Armengol, E.; Plaza, E; 1994. Integrating Induction in a Case- Based Reasoner. In Proceedings of the Second European Conference on Case- Based Reasoning,243-252. AcknoSoft Press, Paris.

Auriol, E.; Manago, M.; Althoff, K.D.; Wess, S.; Dittrich, S. 1994. Integration Induction and Case- Based Reasoning: Methodological Approach and First Evaluation. In Proceedings of the Second European Conference on Case- Based Reasoning,145-156. AcknoSoft Press, Paris.

Bamberger, S.K.; Goos, K. 1993. Integraton of Case- Based Reasoning and Inductive Learning Methods. In Proceedings of the First European Conference on Case- Based Reasoning,296-300. SEKI Report SR-93-12. University of Keisers-lautern.

Barletta, R. 1994. A Hybrid Indexing and Retrieval Strategy for Advisory Case- Based Reasoning Systems Built with ReMind. In Proceedings of the Second European Conference on Case- Based Reasoning,49-58. AcknoSoft Press, Paris.

Cardie, C. 1993. Using Decision Trees to Improve Case- Based Learning. In Proceedings of the Tenth International Conference on Machine Learning, 25-32. Morgan Kaufmann.

Golding, A.R.; Rosenbloom, P.S. 1991. Improving Rule- Based System through Case- Based Reasoning. In Proceedings of the 1991 National Conference on AI,22-27. The MIT Press.

Hammond, K.J. 1988. Explaining and Repairing Plans That Fails. *Artificial Intelligence* 45: 173-228.

Koton, P.A. 1988. Reasoning about Evidence in Causal Explanations. In Proceedings AAAI-88, 256-261. Morgan Kaufmann, Los Altos.

Lenz, M. 1993. Cabata - A hybrid Case- Based Reasoning system. In Proceedings of the First European Conference on Case- Based Reasoning, 204-209. SEKI Report SR-93-12. University of Keiserslautern.

Malek, M.; Rialle, V. 1994. A Case- Based Reasoning System Applied to Neuropathy Diagnosis. In Proceedings of the Second European Conference on Case- Based Reasoning,329-336. AcknoSoft Press, Paris.

Manago, M.; Althoff, K.D.; Auriol, E.; Traphoner, R.; Wess, S.; Conruyt, N.; Maurer, F. 1993. Induction and Reasoning from Cases. In Proceedings of the first European Conference on Case- Based Reasoning,204-209. SEKI Report SR-93-12. University of Keiserslautern.

Matwin, S.; Plante, B. 1994. Theory Revision by Analyzing Explanations and Prototypes. In Michalski, R., Tecuci, G. eds. Machine Learning vol. 4, Morgan Kaufmann, San Mateo.

McLaughlin, E.; Rao, V. 1988. An Exploratory Modeling of the Decision Process of New Product Selection by Supermarket Buyers. *Agribusiness* 4(2): 177-185.

McLaughlin, E. ; Rao, V. 1990. The Strategic Role of Supermarket Buyer Intermediaries in New Product Selection : Implications for Systemwide Efficiency. *American Journal of Agricultural Economics*, May 1990: 358-370.

Michalski, R.S. 1990. Learning Flexible Concept: Fundamental Ideas and a Method Bases on Two- Tired Representation. In Kodratoff, Y., Michalski, R. (eds.) Machine Learning vol.3, Morgan Kaufmann, San Mateo.

Montgomery, D. 1975. New Product Distribution : An Analysis of Supermarket Buyer Decisions. *Journal of Marketing Research* XII: 255-264.

Quinlan, J.R. 1993. C4.5 Programs for Machine Learning. Morgan Kaufmann.

Quinlan, J.R. 1990. Learning Flexible Concept: Fundamental Ideas and a Method Bases on Two- Tired Representation. In Kodratoff, Y., Michalski, R. (eds.) Machine Learning vol. 3, Morgan Kaufmann, San Mateo.

Riesbeck, C.K.; Schank, R.C. 1991
Inside Case- Based Reasoning. Lawr Erlbaum, Hillsdale 1989.

Rissland, E.L.; Skalak, D.B. 1991. CABARET: rule integration in a hybrid architecture. International Journal of Man- Machine Studies 34:839-887.

Surma, J.; Vanhoof, K. 1995. Integrating cases and rules for the classification task
Case-based reasoning research and development.
Springer Verlag

Swinnen, G. 1983. Decisions on Product-Mix Changes in Supermarket Chains, doctoral dissertation, UFSIA, Antwerp University.

Zeleznikow, J.; Hunter, D.; Voss, G. 1993. Integrating Rule Based and Case- Based Reasoning with Information Retrieval: The IKBALS project. In Proceed. of the First European Conference on Case- Based Reasoning,314-346. SEKI Report SR-93-12. University of Keiserslautern.

Zhang,J. 1992. Selecting typical instances in instance-based learning, In Proceedings of the 9th International Conference on Machine Learning, 470-479. Morgan Kaufmann.