LEARNING FLEXIBLE CONCEPTS THROUGH A SEARCH FOR SIMPLER BUT STILL ACCURATE DESCRIPTIONS

by

F. Bergadano
S. Matwin
R. S. Michalski
J. Zhang

Proceedings of the Fourth AAAI-Sponsored Knowledge Acquisition for Knowledge-Based Systems Workshop, Banff, Canada, pp. 1-10, October 1989.

Proceedings of the 4TH KNOWLEDGE ACQUISITION FOR KNOWLEDGE-BASED SYSTEMS WORKSHOP

Banff Conference Centre, Banff, Alberta, Canada
October 1-6, 1989

Sponsored by the American Association for Artificial Intelligence

Workshop Co-Chairs:
John H. Boose, Boeing Advanced Technology Center
Brian R. Gaines, University of Calgary

4th AAAI-Sponsored Knowledge Acquisition for Knowledge-Based Systems Workshop
Banff, Canada, October, 1989.

LEARNING FLEXIBLE CONCEPTS THROUGH A SEARCH FOR SIMPLER BUT STILL ACCURATE DESCRIPTIONS

Francesco Bergadano* , Stan Matwin** , Ryszard S. Michalski+, Jianping Zhang+

*Dipartimento di Informatica, Universita di Torino, Corso Svizzera 185, Torino,Italy
**Department of Computer Science, University of Ottawa, Ottawa, Ont. K1N 6N5 Canada
+Artificial Intelligence Center, George Mason University, Fairfax, VA 22030, USA

ABSTRACT

This paper describes a method for learning flexible concepts, which are defined as concepts that lack precise definition and are context-dependent. The method is based on a *two-tiered concept representation.* In such a representation the first tier, called the *Base Concept Representation,* describes typical properties of a concept in an explicit, comprehensible, and efficient form. The second tier, called the *Inferential Concept Interpretation*, defines allowable transformations of the concept under different contexts.

In the method, the first tier is created in two stages. In the first stage, a complete and consistent description of the concept is learned by applying the inductive learning methodology (AQ and INDUCE) to examples of varying typicality. In the second stage, so obtained description is improved through a heuristic search, employing a *description quality criterion.* During the search, the concept description is simplified by applying truncation operators. The second tier includes expert defined deductive rules and a flexible matching function. Experimental results provide evidence that two-tiered concept descriptions are not only simpler, but also have a better predictive power.

1. INTRODUCTION

Most methods of machine learning research assume that concepts are precise entities, representable by a single symbolic description. If an instance satisfies the given concept description, then it belongs to the concept, otherwise it does not. In contrast, most human concepts have a context-dependent meaning and lack precisely defined boundaries. The imprecision of the boundaries seems to have a logical rather than probabilistic character.

In order to handle such flexible concepts, machine learning systems need to employ richer concept representations than currently used. There have been numerous attempts to deal with representing of imprecise concepts, e.g. multiple-valued logic (e.g. Rine, 77), fuzzy logic (Zadeh, 74), and non-monotonic reasoning systems (e.g. Reiter, 87). Starting point of the research presented here is the idea of the *two-tiered concept representation* (Michalski et al. 86). In this representation the total meaning of a concept consists of two components, the *Base Concept Representation* (BCR) and the *Inferential Concept Interpretation* (ICI). The BCR defines the most typical properties of the concept. The ICI makes the boundaries of the concept flexible by describing allowed modifications of the concept's features in

This work was done while the first two authors were with the Artificial Intelligence Center, George Mason University.

different contexts.

Early ideas on learning two-tiered concept representations were presented in (Michalski 87) and closely related earlier work (Michalski et al. 86). An intriguing result of that research was that a concept description can be reduced to its most important components, without affecting its performance. This paper is an extension and continuation of these early ideas. Important novelties are the introduction of a general description quality measure, the use of both a rule base and a flexible matching function in the ICI, and the development of a heuristic search procedure for simplifying the concept descriptions. Simplifications are more selective. They also have a better heuristic justification in terms of their impact on the accuracy of the simplified descriptions, than the simple truncations described in (Michalski et al. 86). Empirical results discussed in sec. 5 show that simplification of concept descriptions by heuristic search, as proposed in this work, produces descriptions performing comparably well with the results of (Michalski et al. 86).

To demonstrate the methodology of learning two-tiered concept descriptions, we have implemented a system and then applied it to selected problems. Experimental evidence shows that by applying a two-tiered concept representation one can significantly decrease the complexity of a description, improve its comprehensibility and at the same time obtain better accuracy in recognizing new instances of the concept.

2. CONCEPT REPRESENTATION

The Base Concept Representation (BCR) is a disjunctive normal form which is called a cover; it defines the most typical properties of a concept in a simple and efficient form. A cover is a disjunction of complexes and a complex is a conjunction of selectors (Michalski 83). The Inferential Concept Interpretation includes a flexible matching function and a set of deductive rules.

The flexible matching function F is predefined and fixed during learning and testing. The value of F for an event e and a cover c is defined as the probabilistic sum of the values of F for its complexes, e.g. if c consists of a disjunction of two complexes $cpx1$ and $cpx2$, then:

$$F(e, c) = F(e, cpx1) + F(e, cpx2) - F(e, cpx1) * F(e, cpx2)$$

F of event e and a complex cpx is defined as the average of the Fs for a conjunction of its constituent selectors, weighed by the proportion of positive examples covered by the complex:

$$F(e, cpx) = (\Sigma_i F(e, sel_i) / n) * \#cpxpos/(\#cpxpos + \#cpxneg)$$

where n is the number of the selectors in cpx, and $\#cpxpos$ ($\#cpxneg$) is the number of positive (negative) examples covered by cpx, respectively.

F of an event e and a selector sel is defined by the degree of match between the event and the selector weighted by the coverage of positive and negative examples of the selector:

$$F(e, sel) = DegMatch(e, sel) * (1+(\#selpos/\#pos - \#selneg/\#neg))/2$$

where $\#selpos$ ($\#selneg$) is the number of positive (negative) examples covered by the selector, respectively. $\#pos$ ($\#neg$) is the number of positive (negative) examples, respectively. The degree of match depends on the type of the selector. It is either 1 (match) or 0 (no match) for nominal selectors. In case of a linear selector, *DegMatch* inversely depends on the distance of the event from the selector, weighed by the cardinality of the range of the selector.

The ICI also includes a set of deductive rules, allowing the system to recognize transformed or special cases. In fact, the flexible matching is most useful to cover instances that are close to the typical case. For example, flexible matching could allow us to recognize a sequoia as a tree, although it does not match the typical size requirements, while deductive reasoning would be required to recognize a tree without leaves (in the winter time) or to include in the concept of tree some metaphorical meaning (e.g. a genealogical tree or a search tree). The deductive rules in the ICI are expressed as Horn clauses. Rules in the ICI may chain, but simpler deductions are preferred, in order to make the classification easy to understand for the users. Although the implementation supports recursion, non-recursive rules should be used when possible, and the number of rule activations should be limited.

An event can then be covered by a two-tiered description through the following three types of matching:

1. *Strict matching*: the event matches the BCR exactly
2. *Flexible matching*: the event matches the BCR through a flexible matching function
3. *Deductive matching*: the event matches the concept through deductive reasoning by using the ICI rules.

In sequel, we use the term *inferential matching* when an event can be matched either flexibly, or deductively.

3. QUALITY OF CONCEPT DESCRIPTIONS

Our objective is to obtain concept descriptions of good quality, so the notion of quality has to be introduced, and an operational definition usable in our system has to be given. In the presented method, the quality of a concept description is influenced by three basic characteristics: the accuracy, the comprehensibility, and the cost.

The accuracy represents the description's ability to produce correct classifications. A common way to prefer more accurate descriptions is to require that they be complete and consistent with respect to the training events (Mitchell 77, Michalski 80). Even if a description is incomplete and inconsistent, the number of positive and negative examples it covers provides important information for evaluating its quality.

Accuracy of a concept description depends on the typicality of the examples it covers. A good description should cover the typical examples explicitly, and the non-typical ones implicitly. Moreover, covering a typical *negative* example in the BCR is far worse for the accuracy of the description than covering a non-typical negative example. Generally, descriptions in which the typical events are covered by the BCR, and non-typical and exceptional ones are covered by the ICI, are preferred.

The comprehensibility of the acquired knowledge is related to subjective and domain dependent criteria. An important requirement of an AI system is that knowledge has to be explicit and easily understandable by human experts. This is important for improving or modifying the knowledge, and for communicating with experts.

The cost captures the properties of a description related to its storage and use. Other things being equal, descriptions which are easier to store and easier to use for recognizing new examples are preferred. When considering the cost of a description, two characteristics are of primary importance. The first one is the cost of measuring the values of variables occurring in the description. In some application domains, e.g. in medicine, this may be a very important consideration. The second one is the computational cost of evaluating the description. Again, certain applications in real-time environment, e.g. speech or image recognition, may impose constraints on the evaluation time of a description.

These criteria need to be combined into a single evaluation procedure that can be used to compare different concept descriptions. A possible solution is to have an algebraic formula that, given the numeric evaluations of single criteria, produces a number that represents their combined value. Nevertheless, the meaning of this final number is hard to understand for a human expert. Secondly, this may force the system to evaluate all the criteria, even if it would be sufficient to compare two given descriptions on the basis of the most important one, if one is much better than the other. In order to overcome some of these problems, we use a lexicografic evaluation functional (LEF) (Michalski 83) that combines the above mentioned criteria. A precise quality measure based on these ideas is given in (Bergadano et al. 88) and has been implemented in the system described in this paper.

4. IMPROVING THE QUALITY OF A CONCEPT DESCRIPTION

Learning two-tiered concept descriptions is performed in two stages. In the first stage, a complete and consistent concept description is obtained from an inductive learning system. In our approach, we have relied on AQ15 (Michalski et al. 86), and INDUCE (Hoff et al. 82) to obtain such descriptions. The description generated in this stage, together with an empty ICI, forms the initial two-tiered description for the second stage. The second stage improves the initial description generated in the first stage. This process is guided by the description quality measure discussed in the previous section, and is implemented as a best first search, i.e., the descriptions of better quality are considered first. According to the nature of the quality measure, descriptions can be improved mainly by increasing their accuracy or by decreasing their complexity. For this reason the operators in the search simplify the BCR of a given description by dropping (truncating) some of its components or by modifying the argument of some predicate. This does not always result in a loss of accuracy, especially when measured on a testing set of new examples, since simpler features might be more stable and depend less on the set of training examples. The search process is described as follows:

Search space:	a tree structure, in which the nodes are two-tiered descriptions (BCR + ICI) of a given concept.
Operators:	selector truncation, complex truncation, referent modification.
Search strategy:	controlled by the quality measure.

The goal of this procedure is not necessarily to find an optimal solution, i.e. the description with the highest quality, because this would require a combinatorial search. On the contrary, the system tries to improve the given concept description by expanding a limited number of nodes in the search tree and is guided by a heuristic measure. In the current implementation these heuristics are based on the coverage of the individual selectors and complexes.

The operators in the search correspond to generalizations or specializations of the BCR of a given description. In particular, selector truncation is a generalization operator, making the new BCR cover more positive and negative examples, while complex truncation is a specialization operator, making the set of examples covered by the modified the BCR smaller. Referent modification can be either a specialization or a generalization operator, depending on the type of modification that is being used and on the type of selector involved.

After a search operation is applied on the BCR of a description d, referred to as BCRd, a new BCR may be either more specialized or more general than BCRd. If the description is more specialized, some positive events previously covered by BCRd may not be covered any more. On the other hand, when a description obtained from BCRd is more general than BCRd, some new negative events, previously not covered by BCRd, may have been added. These events are treated as exceptions. Two types ICI rules, rules that cover a positive events otherwise left out of the BCR and rules that eliminate a negative event from the BCR, are used to handle these positive and negative examples. These rules are learned from the explanations for the exceptions. In the system, these explanations are provided by human experts.

5. EXPERIMENTS

We have run experiments with the system in three different domains: labor management contracts, congressional voting record, and breast cancer reoccurrence. For the labor management application, we have collected real data on 57 labor-management contracts negotiated in Canada in the last 18 months. Each contract is described by sixteen attributes, belonging to two main groups: issues related to salaries (e.g. pay increases in each year of contract, cost of living allowance, etc.), and issues related to fringe benefits (e.g. different kinds of pension contributions, holidays, vacation, dental insurance, etc.). We have run two experiments. In each experiment we were dealing with two descriptions: a description of a contract, and a description of a contract proposal deemed unacceptable by one of the parties involved in the negotiation. We have evaluated concept descriptions both on the training set (18 positive and 9 negative examples) and on a testing set (19 positive and 11 negative examples) of cases not previously seen by the system. The results of the two experiments are given in Tables 1 and 2.

Table 1 Labor Management Contracts (First Experiment): performance of the descriptions derived by AQ15 (11 complexes and 28 selectors).

		Correct	Incorrect	No_Match
Strict Match				
	Training Set	100%	0%	0%
	Testing Set	80%	17%	3%
Flexible Match				
	Training Set	100%	0%	0%
	Testing Set	80%	4-15%	3%

Table 2 Labor Management Contracts (Second Experiment): performance of the descriptions obtained using the search procedure. In the case of inferential matching, expert rules were used in combination with flexible matching.

		Correct	Incorrect	No_Match
Strict Match				
	Training Set	63%	0%	37%
	Testing Set	43%	3%	54%
Flexible Match				
	Training Set	85%	0%	15%
	Testing Set	83%	13%	4%
Inferential Match				
	Training Set	96%	0%	4%
	Testing Set	90%	10%	0%

In the first experiment, we have used the descriptions learned by AQ15 (Initial Description), given in Fig. 1, and in the second experiment we have used the description generated by the system, given in Fig. 2.

[duration ≠ 1] & [wage_incr_yr2 ≠ 3.0] & [holidays ≠ 10] v
[wage_incr_yr1 ≠ 2.0% v 2.5% v 2.8% v 3.0% v 4.0% v 4.5%] v
[wage_incr_yr1 ≠ 2.0% v 2.5% v 2.8% v 4.0%] & [wage_incr_yr2 ≠ 2.0% v 4.0%] v
[wage_incr_yr1 ≠ 2.0% v 2.5%v 3.0%v 4.0%v 4.5%] & [holidays ≠ 9] v
[wage_incr_yr1 ≠ 2.0%] & [vacation = above_average] ::> acceptable contract

[wage_incr_yr1=2.0% v2.5% v 4.0%]&[holidays=10] &[vacation=below_average v average]v
[wage_incr_yr1 = 2.0% v 2.5% v 3.0%v 4.0% v 4.5%] &[wage_incr_yr2 = 2.0% v 4.0%] &
    [holidays = 10] & [vacation = below_average v average] v
[duration = 1] & [wage_incr_yr1 = 2.0% v 2.5% v 2.8%v 4.0%] & [holidays = 9] &
    [vacation = below_average v average] v
[wage_incr_yr1=2.0%v2.5% v4.0%]&[wage_incr_yr2=3.0%]&
    [vacation=below_average v average]v
[duration = 1] & [wage_incr_yr1=2.0%v2.5% v4.0%]&[vacation=below_average v average] v
[wage_incr_yr1 = 2.0%] & [wage_incr_yr2 = 3.0%] ::> unacceptable contract

---

Fig. 1. Labor Management Contracts: Descriptions Generated by AQ15

[wage_incr_yr2 > 3.0%] v
[wage_incr_yr1 > 4.5%] v
[holidays > 9] v
[vacation = above_average] ::> acceptable contract

[wage_incr_yr1 = 2.0% v 2.5% v 4.0%] & [holidays = 10] v
[wage_incr_yr2 = 2.0% v 4.0%] & [vacation = below_average v average] v
[holidays = 9] v
[duration = 1] & [wage_incr_yr1 = 2.0% v 2.5% v 4.0%] v
[wage_incr_yr2 = 3.0%] ::> unacceptable contract

Fig. 2. Labor Management Contracts: Improved Descriptions.

The improved descriptions are simpler than the ones generated by AQ15, and they represent the most important characteristics of the labor management contracts: a contract is acceptable when it offers a significant wage increase (the first two complexes in Fig. 2), or it offers many holiday days, or the vacation is above average. The training events that were not correctly classified by the description, as it was modified step by step during the search, were analyzed by a domain expert, who provided deductive rules allowing the system to classify almost all the training events (one of them could not be explained by the expert). Following is one of those deductive rules obtained for the ICI of the contract concept.

[wage_incr_yr1 < 3.1%] & [wage_incr_yr2 < wage_incr_yr1] ::> unacceptable contract

The rule addresses the case of a contract with a low wage increase in year one of the contract, and an even lower increase in the second year. In those circumstances, the holiday and vacation offered do not matter: the contract is deemed unacceptable by the union.

The combination of the modified description and inferential matching (rules plus flexible matching) produces the best results (90% correct classifications). The power of the improved description is due to a combination of all three types of matching (strict, flexible and deductive), and all three contribute to the quality measure of a description as computed during the learning process. This represents a new feature of the system, since inferential matching is usually introduced only after the learning phase is completed (Bergadano, Giordana, Saitta 88; Michalski et al. 86).

The second application was concerned with the U.S. Congress voting record. We have relied on the same data set as used by (Lebowitz 87) in the experiments on conceptual clustering. The data represents the 1981 voting record for 100 selected representatives. The data set was split randomly into a training and testing set, with voting records of democrats entered as positive examples, and voting records of republicans entered as the negative ones. The goal was to obtain discriminating descriptions of democrat and republican congresspersons. Again, we ran two experiments as in the first application domain. The results are shown in tables 3 and 4.

Table 3. Congress Voting Record (First Experiment): performance of the descriptions derived by AQ15 (10 complexes and 32 selectors).

		Correct	Incorrect	No_Match
Strict Match				
	Training Set	100%	0%	0%
	Testing Set	86%	14%	0%
Flexible Match				
	Training Set	100%	0%	0%
	Testing Set	86%	14%	0%

Table 4. Congress Voting Record (Second Experiment): performance of the descriptions obtained with the search procedure.

		Correct	Incorrect	No_Match
Strict Match				
	Training Set	84%	0%	16%
	Testing Set	73%	4%	23%
Flexible Match				
	Training Set	100%	0%	0%
	Testing Set	92%	8%	0%
Inferential Match				
	Training Set	96%	4%	0%
	Testing Set	92%	8%	0%

The learning set contained 31 positive and 20 negative examples, and the test set contained 29 positive and 20 negative examples. The obtained results are comparable with the other application: the description was simplified and the prediction power was increased. In fact, the AQ description contained 32 selectors compared to 21 in the improved description and the recognition rate moved from 86% to 92%. In the Congress application, though, the use of the deductive rules did not cause any performance increase. This can be explained by an already high percentage of correct classifications on the test set (92%). Nevertheless, the result must not be underestimated: although performance is the same, deductive matching is preferred over flexible matching. In fact, the examples that are correctly matched through some ICI rule, are actually explained by relevant domain knowledge, and not only on the basis of a knowledge-independent distance measure. This is important both to accuracy and comprehensibility of the modified description when evaluated with the deductive rules. The quality measure used in the system actually reflects this idea and scores higher when inferential matching is used.

Finally, we have also experimented with the same breast cancer reoccurrence data as (Michalski et al. 86). Our system had a distinct disadvantage on this data, since in the absence of typicality information all examples were assumed to be equally typical. Moreover, no deductive rules were available The following results were obtained on the testing set:

Table 5. Breast Cancer Reoccurrence (Michalski et al. 86).

	AQ15	best complex	two-tiered
# of complexes	22	2	12
# of selectors	102	7	52
performance	66%	68%	70.5%

All these experiments show how a two-tiered learning scheme allows a system to learn concepts from a small number of examples, and produces simpler but still accurate descriptions. The concept meaning is now divided into a cover, generated by the search, a flexible matching procedure, defined a-priori, and a set of deductive rules given by the expert on the basis of misclassified examples selected automatically during the search process.

6. CONCLUSION

In this paper, a system that is able to learn two-tiered concept descriptions has been described. Learning is viewed as a state space search guided by a measure of quality that is applied to concept descriptions. The operators in the search modify a given description by removing some of its components or by simplifying the referent of the selectors. The goal of the search process is to obtain simpler but still accurate descriptions. In this way, the comprehensibility and the predictive power of the acquired knowledge are improved. Some of the motivations behind the system come from previous work (Michalski et al, 1986), that produced some preliminary results in which the flexible matching function was applied during the recognition process. In the system presented here, a different flexible matching function is used. It is augmented by a set of deductive rules, describing symbolically possible transformations of the concept description. Selector truncation and referent modification are applied automatically during a search process.

The research presented here is related to recent and important work in machine learning that investigates the effects of simplifying concept descriptions (e.g. Fisher, Schlimmer 88; Iba et al. 88). An advantage of the presented method is that it does not experience any major loss of coverage as a result of description modification. Other relevant work is concerned with the problem of pruning decision trees (Cestnik, Kononenko, Bratko 87). An important difference is lack of constraints on the part of the representation that is truncated when learning a two-tiered concept description. In post-pruning of decision trees, only paths ending in leaves may be truncated, which may improve the efficiency at the expense of the description quality.

The system was tested on three different domain applications. The experimental results that we obtained confirm the hypothesis that two-tiered descriptions can be more accurate and easier to understand. The ICI used in the experiments included a flexible matching function and a set of logical rules. The performance of the descriptions produced by the search process on the test set is influenced by the use of the inferential matching. This is due to the fact that ICI is used during learning, in order to choose and modify the best descriptions. This property represents an important difference between the presented system and previous approaches, that tend to apply flexible matching only after the learning process is completed.

Acknowledgements

This research was done in the Artificial Intelligence Center of George Mason University. The activitie the Center are supported in part by the Defence Advanced Research Projects Agency under grant N00014-85-K-0878, administered by the Office of Naval Research, and in part by the Office of Na Research under grant No. N00014-88-K-0226. The first author was supported in part by the Italian Minis of Education, and the second author was supported in part by the NSERC grant number A2480. The auth thank Zbig Koperczak for his help with acquiring the contract data and A. Giordana for useful criticism.

References

Bergadano, F., Matwin, S., Michalski, R. S., Zhang, J., "Measuring Quality of Concept Descriptio Proc. Third European Working Sessions on Learning, pp. 1-14, Glasgow, 1988a.

Bergadano, F. , Giordana, A. , Saitta, L., "Automated Concept Acquisition in Noisy Environments", IE Transactions on PAMI, 10, 4, pp. 555-577, 1988.

Cestnik, B., Kononenko, I., Bratko, I., "ASSISTANT 86: A Knowledge-elicitation Tool for Sophistica Users", Procs. of the 2nd European Workshop on Learning, pp. 31-45 (1987).

Fisher, D. H., Schlimmer, J. C., "Concept Simplification and Prediction Accuracy", Procs. of the Fi Int'l. Conf. On Machine Learning, Ann Arbor, pp. 22-28 , 1988.

Hoff, B., Michalski, R. S., Stepp, R., "INDUCE-2 - A Program for Learning Structural Descriptions fr Examples", TR 82-5, ISG, University of Illinois, 1982.

Iba, W., Wogulis, J., Langley, P., "Trading off Simplicity and Coverage in Incremental Concept Learning", Procs. of the Fifth Int'l. Conf. On Machine Learning, Ann Arbor, pp. 73-79 (1988).

Lebowitz, M., "Experiments with Incremental Concept Formation: UNIMEM", Machine Learning, vol. 2, No. 2, pp. 103-138, 1987.

Michalski, R. S., "Pattern Recognition as Rule-guided Inductive Inference", IEEE Transactions on PAMI, vol. 2, no. 2,3,4, pp. 349-361, 1980.

Michalski, R.S., "A Theory and Methodology of Inductive Learning", Chapter in the book "Machine Learning, an Artificial Intelligence Approach", Michalski, R. S., Carbonell, J. G., Mitchell, T. M. (Eds.), Tioga Pub. Co., Palo Alto, Ca, 1983.

Michalski, R. S., Mozetic, I., Hong, J.,Lavrac, N., "The Multi-purpose Incremental Learning System AQ15 and its Testing Application to Three Medical Domains", Proc. 5th AAAI, pp. 1041-1045, 1986.

Michalski, R. S., "Two-Tiered Concept Meaning, Inferential Matching and Conceptual Cohesiveness" , Chapter in the Book "Similarity and Analogy", Stella Vosniadou and A. Orton, (Eds), 1987.

Mitchell, T. M., "Version Spaces: An Approach to Concept Learning", Ph. D. dissertation, Stanford University, December 1977.

Reiter, R., "Non-monotonic Reasoning", in Annual Review of Computer Science, (1987).

Rine, D. C., ed., "Computer Science and Multiple-valued Logic. Theory and Applications", North Holland, 1977.

Zadeh, L. A., "Fuzzy Logic and its Applications to Approximate Reasoning", Information Processing, North Holland, pp. 591-594, 1974.