REPRESENTING AND ACQUIRING IMPRECISE AND CONTEXT-DEPENDENT CONCEPTS IN KNOWLEDGE-BASED SYSTEMS

by

F. Bergadano
S. Matwin
R. S. Michalski

Invited paper, International Symposium on Methodologies for Intelligent Systems,
Torino, 1988.

88-5

REPRESENTING AND ACQUIRING FLEXIBLE CONCEPTS IN KNOWLEDGE-BASE SYSTEMS

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

MLI 88-5
TR-12-88

A publication of *Machine Learning & Inference Laboratory*
Artificial Intelligence Center
George Mason University
Fairfax, VA 22030 USA
(703) 764-6259

Editor: R. S. Michalski
Assistant Editor: J. Zhang

MLI Reports replaces *ISG Reports* published until December, 1987 by the Artificial Intelligence Laboratory Department of Computer Science, University of Illinois Urbana, IL 61801.

Invited paper for the International Symposium on Methodologies for Intelligent Systems, Torino, October 1988

Representing and Acquiring Imprecise and Context-dependent Concepts in Knowledge-based Systems

F. Bergadano[1], S. Matwin[2], R. S. Michalski, J. Zhang

Artificial Intelligence Center
George Mason University,
Fairfax, VA 22030

ABSTRACT

The paper presents a method for representing and learning imprecise and context-dependent concepts. The concepts are learned from examples of varying typicality, and represented using a two-tiered approach. The first tier, the *base concept representation*, represents typical features of a concept. The second tier, the *inferential concept interpretation,* handles exceptional, nontypical, and context-dependent instances of the concept.

A learning algorithm and a *general description quality* measure are described. The measure is used for evaluating concept descriptions generated in the process of learning. Presented ideas are illustrated by an example of constructing a general description of a labor-management contract from specific examples.

INTRODUCTION

All scientific concepts are expected to be defined precisely, with their borderlines clearly delineated. The precision of concepts used in any scientific theory is traditionally viewed as a mark of quality of the theory. Theorems, rules, and scientific methods are also expected to bind concepts in well-defined, precise ways.

In contrast to such scientific standards and traditions, most concepts humans use are imprecisely defined, have flexible borders, and their meaning is often context-dependent. This can be explained by the fact that with the growth of complexity of a system, the precision of its description needs to decrease to maintain the description's complexity within manageable limits. The imprecision of a description does not necessarily lead, however, to the imprecision of its meaning, as

[1] University of Torino, Torino, Italy
[2] [illegible]

the meaning depends on the context in which a description is used. For example, an imprecise description "blond man with glasses" identifies precisely a person in a group, if this person is the only one in this group with non-dark hair and wearing glasses.

Despite numerous efforts, the problem of representing the imprecision and context-dependency of human concepts remains unsolved. The most known approach is based on the idea of fuzzy sets, introduced by Zadeh [e..g. 1,11]. This approach deals with the imprecision of concepts by defining a numeric degree of membership for each concept. Such a degree is subjective and introduces an additional complexity in the concept representation. Also, this approach does not provide adequate means for handling context-dependency of concepts.

This paper presents a different approach to this problem, based on the idea of *two-tiered concept representation* [10]. In this representation, the flexibility and context-dependency of concepts stem from an interaction between two components: a *base concept representation* (BCR) and an *inferential concept interpretation* (ICI). The BCR embodies the essential, typical features of a concept. The ICI matches the BCR with the concept instances using a body of context-dependent inference rules. Such a matching process may involve not only deductive, but approximate, analogical, and inductive forms of inference.

This report is a continuation of an earlier work on the topic of learning of two-tiered concept representations [12], and presents an algorithm and a comprehensive method for learning concepts using such a representation.

TWO-TIERED CONCEPT DESCRIPTIONS

The traditional work on concept representation assumes that the whole meaning of a concept resides in a single stored structure, e.g. a semantic network that captures all relevant properties of the concept (e.g. [2,3,4}). The process of recognizing a concept involves direct matching of the stored representation with perceived facts. Such a matching may include a simple deduction, or tracing of links in a network, but has not been assumed to involve any complex inferential processes.

The difficulty with the traditional representation is that it ignores the fact that our knowledge can be viewed as a combination of two components, recorded knowledge and inferential extension, i.e. knowledge that can be created from recorded knowledge by conducting inference. In view of this fact we propose that the meaning of a concept in any given situation is a result of an interplay between two parts: a simple memorized structure (called the base concept representation, or the BCR), and its dynamic modification (the inferential concept interpretation, or the ICI) created by applying a body of rules to the BCR. The base concept representation is a structure residing in memory that records both specific facts about the concept, and its general characteristics. The specific facts may include representative examples and counter-examples. The general characteristics are teacher-defined, or inferred by induction from examples or by analogy. They include typical, easily-defined, and possibly context-independent assertions about the concept. These characteristics tend to capture the principle, the ideal or intention behind a given concept. If this principle changes to reflect a deeper knowledge about the concept involved, the base concept representation is redefined. To see this, consider, for example, the changes of our understanding of concepts such as whale (from fish to mammal) or atom (from the smallest indivisible particle to the contemporary notion of dual wave-matter form).

The inferential concept interpretation is a process of assigning the meaning to a concept using the base concept representation and the context of discourse. This process involves the interpreter's background knowledge and relevant inference methods that allow one to recognize, extend or modify the concept meaning according to context. When an unknown entity is matched against a base concept representation, it may satisfy it directly, or it may satisfy some of its inferential extensions. The type of inference performed to match the description of the entity with the base concept representation determines the type of match. If the description of the entity strictly satisfies the BCR, i.e. it matches it directly, then we have a *strict match*. If it satisfies a probabilistic inferential extension, then we have a *flexible match*. If it satisfies an extension obtained deductively using (potentially chained) rules, then we have a *deductive match*. Sometimes, there is a trade-off between representing certain features of a concept explicitly in the BCR, or implicitly in the ICI. Moving parts of a representation from the BCR to the ICI allows us to reduce the memory necessary to represent a concept. It will, however, result in a higher cost of matching an event against the description. To summarize, we may say that inference allows us remember less and to know more, but at the price of execution time.

Let us illustrate the proposed approach by an example. Consider the concept of fish. Typical and general characteristics of fish are that they live in water and they swim. These and other typical physical properties of fish will be stored in the BCR. Suppose someone found an animal that matches many characteristics of fish, but does not swim. Suppose this animal appears to be sick. The ICI would involve background knowledge that sick animals may not be able to move, and that swimming is a form of moving. By deductive reasoning from these rules one concludes that lack of ability to swim should not be taken as negative evidence for the animal being a fish. To the contrary, the fact that the animal does not swim might even add to the confidence that it is a fish, once the animal was recognized as being sick.

The rules used in the above reasoning about sick fish would not be stored in the BCR for fish. They would be a part of methods for the ICI. These methods would be associated with the general concept of animal.

rather than with the concept of fish, because they apply to all animals. Thus, we postulate that the methods for inferentially interpreting a concept can be inherited from those applicable to a more general concept.

ACQUIRING TWO-TIERED CONCEPT DESCRIPTIONS: THE PRINCIPLES

In this paper we propose a method for acquiring two-tiered concept descriptions, that is based on a search that transforms descriptions initially given into descriptions of better quality. The initial descriptions are either given by the domain expert or supplied by an automated learning system. If necessary the disjunction of all the training examples can also be used as an initial description. A measure of quality of two-tiered concept descriptions is introduced and is used as a heuristic guiding the search.

The Base Concept Representation is expressed as a disjunctive normal form expression, and can be simplified by dropping or modifying some of its components. Learning is implemented as a best-first search in a space of two-tiered descriptions. In this approach different kinds of reasoning participate to the learning process (inductive, approximate, deductive) and the typicality of the given events is taken into account.

Since our objective is to obtain concept descriptions of good quality, the notion of quality should be introduced, and an operational definition usable in our system should be given. We will now discuss the important aspects of description quality.

Criteria for Determining the Quality of Concept Descriptions

The quality of a concept description is influenced by three basic characteristics: the accuracy, the comprehensibility, and the cost. This section discusses these three components, as well as a mechanism for combining them into a a single measure.

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 learning events [5,6,7]. Even if a description is incomplete and inconsistent, the number of positive and negative examples it covers provides important information for evaluating its quality. In this case, we can measure the degree of completeness and consistency of a given description. If the description is also sufficiently general and does not depend on the particular characteristics of the learning events, these measures can be a meaningful estimate of the accuracy of the description. In order to achieve completeness and consistency in presence of noise, one may generate overly complex and detailed descriptions. Such descriptions, however, may not perform well in future cases. This is the well-known phenomenon of overfitting [8,9].

Accuracy depends linearly on completeness and consistency of the description, as well as on the typicality of the events covered by the two parts of the description. In evaluating the accuracy of a two-tiered representation, we have to take into account the fact that degree of confidence in the results of inference decreases from deduction to induction [10]. These requirements are met by making completeness and consistency dependent on the typicality of the covered examples and on the way these examples are covered. We assume that an expert can provide typicality of examples at the time they are presented to the system responsible for building the initial description. The experts are usually able to determine the typicality of events in their area of expertise.

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. Since a black box classifier will not be accepted by experts verifying the knowledge of a performance element, knowledge acquired automatically should be easy to understand, should contain the descriptors most frequently used by experts, and should not be syntactically too complex. In practice, only the last feature is easy to measure.

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. Examples of such approaches are multiplication, weighted sum, maximum/minimum, t-norm/t-conorm [9]. Although these approaches are often appropriate, some of them may present disadvantages. Firstly, they usually combine a set of heterogeneous evaluations into a single number, and the meaning of this final number is hard to understand for a human expert. Secondly, they 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 lexicographic evaluation functional (LEF) [7] that combines the above mentioned criteria.

The criteria discussed above can also be applied to two-tiered descriptions [13]. The accuracy of the acquired knowledge does not only depend on the explicit information, but also on the implicit reasoning abilities. Inferential Concept Interpretation also affects cost, since it allows the performance system to use a simpler BCR, and reason about special details only in exceptional cases. Finally, the comprehensibility of a two-tiered representation must be carefully evaluated, since one of its implied goals is to state a clear and simple concept description in the BCR and to account for meaningful special cases through a reasoning process.

Matching an event with a two-tiered concept description

Completeness and consistency of a two-tiered concept description are not only computed on the basis of the BCR, but also take into account the ICI and the typicality of the covered events. More precisely, an event can be covered by a two-tiered description through the following three types of matching:

1. *Strict matching*: the event matches the BCR exactly, in which case we say that the event is S-covered,
2. *Flexible matching*: the event matches the BCR through a flexible matching function, and we say the event is F-covered.
3. *Deductive matching*: the event matches the concept through deductive reasoning by using the ICI Rules, and we say the event is D-covered.

These three sets are made mutually exclusive: if an event is S-covered, then it is not D-covered or F-covered, and if an event is D-covered, then it is not F-covered. In general, descriptions that cover many typical positive events are most preferred. Completeness is therefore proportional to the typicality of the events covered. Moreover, if negative events are covered, the consistency of the description is inversely proportional to the typicality of the negative events covered. Finally, the quality measure prefers descriptions in which typical events are S-covered, borderline examples are F-covered and exceptions or special cases are D-covered.

ACQUIRING TWO-TIERED DESCRIPTIONS: AN EXAMPLE

This section will briefly describe method of learning two-tiered representations by presenting an example. Learning the concept of a labor-management contract provides a suitable and interesting application of learning a two-tiered concept representation. The nature of the domain is such that the example can be described using an attribute-based language, such as VL1. Consequently, the learning program AQ15 is used to obtain the initial concept description

Labor-management contracts usually show a number of typical characteristics. Among those are: general wage increase, job security, and pensions. Meeting all those demands by the management would result in an "ideal" contract, from the labor point of view. In practice, labor demands are usually scaled down during negotiation, which results in a contract.

It is not uncommon, however, to see contracts that exhibit very nontypical characteristics. These exceptional contracts may be explained by the context and background knowledge. For instance, a contract which is highly unsatisfactory in both wages and pension areas, but offers minimal job security, may be accepted during a deep recession. On the other hand, in an exceptionally good economic environment and in an industry where labor supply is scarce, any contract proposal that is not highly satisfactory in all three areas may be unacceptable. Furthermore, a contract that is recognized as acceptable by the first tier representation may actually be evaluated negatively in the second tier: it may imply micro-economic consequences overwriting initial values of its attributes. For all these reasons, a two-tiered representation seems appropriate when learning the concept of a contract.

The example space is divided, from the labor point of view, into acceptable and unacceptable contracts. Both are agreements negotiated between a trade union and management of an organization. The former have furthermore been ratified by the general union membership, while the latter have been rejected.

In sequel we present a set of simple examples: they describe specific contracts, where only some selected characteristics of a contract are given. Those characteristics pertain to seven chosen attributes of a contract: general wage increase, cost of living allowance (cola), job security (job_sec), retirement age (ret_age), extent of pension (p_ext), pension for overtime (p_ovt) work, and fringe benefits. The following is therefore an example of an acceptable contract:

gwi	cola	job_sec	ret_age	p_ext	p_ovt	fringes
2%	inflation+.5%	some	60	part	false	maintained

A number of other examples have been described in the same way, and are shown in Fig. 1. The typicality of the last two positive and the last two negative examples was 0.5; the typicality of all the remaining examples was 1.0 (the highest value of typicality).

acceptable-events

#	gwi	cola	job_sec	p_age	p_ext	p_ovt	fringes
1	2	ilpos	some	60	part	false	main
2	10	inf	good	65	none	false	main
3	2	inf	some	60	part	false	main
4	1	zero	good	65	full	false	main
5	10	zero	good	65	full	true	inc
6	3	inf	good	64	part	true	loss
7	6	zero	good	65	full	true	loss
8	14	ilpos	good	62	part	false	inc
9	4	ilpos	good	64	full	false	main
10	3	zero	some	58	full	false	inc
11	5	zero	some	60	full	true	loss
12	2	inf	good	61	part	false	main
13	11	zero	good	65	none	false	main
14	14	zero	good	65	none	false	main
15	11	ilpos	good	65	none	false	main
16	15	zero	some	65	none	false	loss
17	15	zero	some	65	none	false	inc
18	12	zero	some	61	part	false	loss
19	10	zero	none	65	full	false	loss
20	11	zero	none	65	full	false	loss
21	15	ilneg	none	58	part	false	inc
22	14	inf	none	61	part	true	loss
23	10	inf	none	60	full	false	main
24	7	ilpos	none	55	full	true	loss
25	6	ilpos	none	55	full	true	main

unacceptable-events

#	gwi	cola	job_sec	p_age	p_ext	p_ovt	fringes
1	1..8	zero	none	60..65	none	false	loss
2	14	zero	none	58..65	none	false	loss
3	11	zero	none	65	none	false	loss
4	10	inf	none	64	none	true	main
5	2	inf	some	62	none	false	loss
6	0..4	zero	some	65	none	false	main
7	4	zero	good	60	none	false	inc
8	6	inf	some	65	none	false	inc
9	2	zero	good	65	none	false	main
10	14	zero	none	65	none	false	main
11	2...7	inf	none	60..65	full	false	main
12	0	ilneg	none	60	part	false	main
13	0	zero	none	65	part	false	main
14	3	zero	none	65	full	false	loss
15	2	zero	some	64..65	part	false	loss

Figure 1. Example instances of acceptable and unacceptable contracts.

The examples were submitted to AQ15, which produced a discriminant description of a concept of an acceptable contract (see Fig. 2).

[gwi <> 0..4 v 6] [job_sec = some v good] v
[gwi <> 0..2 v 7] [p_ext = full v part] v
[job_sec = some v good] [p_ext = full v part] [fringes = incr v maint] v
[gwi <> 10] [p_ovt = true]

Fig. 2 The initial concept description for the labor-management contract

The quality of this description, according to the specific quality measure defined in Appendix 1, is 0.989. This description is modified by the system in the following way:

1. Reference modification is performed on selectors. The close interval operation is applied on the first selector of the first two complexes, so that the complex [gwi <> 0..4, 6] becomes [gwi > 6], and the complex [gwi <> 0 v 2..7] becomes [gwi > 7].

In this example, completeness and inconsistency are left unchanged by range modification, and simplicity (and comprehensibility) are improved.

[gwi > 6] [job_sec = some v good] v
[gwi > 7] [p_ext = full v part] v
[job_sec = some v good] [p_ext = full v part]

Fig. 3 The improved concept description obtained by the system (BCR only)

2. The system gives a modified description, shown in Fig. 3. The quality of this description is 0.976. This represents a small deterioration of quality, since two negative examples are covered by the modified description. Moreover, two positive examples are lost because the last complex of the description shown in Fig. 2 is truncated. The two negative examples are explained as exceptions. The reasoning about this exception is the following: even if the three attributes occurring in the rule of Fig. 3 (gwi, job_sec, p_ext) have acceptable values, when all the other attributes have the worst values possible, the contract is not acceptable. This shrinking rule may be expressed as follows:

(gwi = very_low) ^ (cola = zero) ^ (p_age = very_high) ^ (p_ovt = false) ^ (fringes = loss)
--> unacceptable_contract

Is there a good extending explanation rule for the two events $e_{24}+$, $e_{25}+$ covered uniquely by complex 4? These events may be covered by applying the following reasoning: if the pension offer is extremely good, and the state of the economy is good, the value of the gwi attribute does not matter anymore. The following rules convey this reasoning:

(p_age = very_low) ^ (p_ext = full) ^ (p_ovt = true) --> extremely_good(pension)

extremely_good(pension) ^ good_economy --> irrelevant(gwi).

Therefore, complex 4 is truncated and the above rules are added to the ICI. The final BCR obtained in this example is shown in Fig. 3. The BCR obtained represents a concept of acceptable contract which has high values of two of the three areas important for the union. The quality of this description is 1.0: it is complete and consistent, and the partition of events between the BCR and the ICI corresponds to the typicality of these events.

The description is now more intuitive, and the BCR seems to capture our natural understanding of the concept of an acceptable contract, while the ICI deals with special cases in an explicit way. The quality of the two tiered description is higher because its accuracy is the same as that of the description generated by AQ, but its comprehensibility is significantly better. This learning method seems to be appropriate in applications where a limited number of examples is available, and general and domain knowledge is needed in order to capture the most relevant aspects of the learning events.

CONCLUSION AND FUTURE WORK

The paper presents a method of two-tiered concept representation. The method deals with the problem of a flexible, context-dependent knowledge representation that involves both a frame-like approach and the inferential interpretation. Different types of inference are used to interpret the concepts that do not match the base representation.

The paper discusses a system for automatic learning of two-tiered concept representations from examples with varying typicality. The system is based on a notion of quality of two-tiered descriptions, that depends on such characteristics of descriptions as their accuracy, simplicity, and cost. The novel and important feature of the measure is that both tiers contribute to the quality of the learned descriptions.

The paper reports results of preliminary work. An experimental system implementing the ideas presented here is currently being developed. It relies on the concept description produced by an inductive learner AQ15 [12] as the initial BCR, and transforms it into a better quality concept representation involving a highly simplified BCR and a powerful ICI. Early experimentation with the system indicates that the system will meet its objectives of learning concept descriptions that are characterized by high predictive power and good comprehensibility.

A number of interesting research problems remain. First, an integrated system that learns two-tiered descriptions, uncoupled from an inductive learning program, should be designed and developed. Second, a method of automatic generation of ICI rules from exceptions and anomalous examples should be investigated. Such a method will rely on background and high-level knowledge to produce explanations of nontypical events and then generalize them into powerful rules to be included in the ICI. Third, the ICI rules themselves may be subject to generalization by induction. If such an generalization is constructive, the system possessing the necessary minimum knowledge could in future learn automatically hierarchies of interesting concepts.

ACKNOWLEDGEMENTS

This research was done in the Artificial Intelligence Center of George Mason University. The activities of the Center are supported in part by the Defence Advanced Research Projects Agency under grant administered by the Office of Naval Research No. N00014-85-K-0878, in part by the Office of Naval Research under grant N00014-88-K-0226, and in part by the Office of Naval Research under grant N00014-88-K-0397. The first author was supported in part by the Italian Ministry of Education (ASSI) project, and the second author was supported in part by the NSERC grant number A2480. The authors thank Joan Elliott from the GMU Center for Artificial Intelligence for useful comments and criticism.

REFERENCES

1. Zadeh, L. A., "Fuzzy Logic and its Applications to Approximate Reasoning", Information Processing, North Holland, pp. 591-594, 1974.
2. Collins, A. M. and Quillian, M. R., "Experiments on Semantic Memory and Language Comprehension" in Cognition, Learning and Memory, L. W. Gregg ed., John Wiley, 1972.
3. Minsky, M., "A Framework for Representing Knowledge", in The Psychology of Computer Vision, ed. P. Winston, (1975)
4. Sowa, J. F., "Conceptual Structures", Addison Wesley, 1984.
5. Michalski, R. S., "AQVAL/1--Computer Implementation of a Variable-Valued Logic System VL1 and Examples of its Application to Pattern Recognition", Proc. of the 1rst Inbternational Joint Conf. on Pattern Recognition, Washington, D.C., pp. 3-17, 1973.
6. Mitchell, T. M., "Version Spaces: An Approach to Concept Learning", Ph. D. dissertation, Stanford University, December 1978.
7. Michalski, R. S., "Pattern Recognition as Rule-guided Inductive Inference", IEEE Transactions on PAMI, vol. 2, no. 2,3,4, pp. 349-361, 1980.
8. Watanabe, S. , "Knowing and Guessing - a Formal and Quantitative Study", Wiley Pub. Co., 1969.
9. Sturt, E., "Computerized Construction in Fortran of a Discriminant Function for Categorical Data", Applied Statistics, vol. 30, pp. 213-222, 1981.

10. 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.
11. Zadeh, L. A., "Fuzzy-algorithmic Approach to the Definition of Complex or Imprecise Concepts", International Journal of Man-Machine Studies, vol. 8, 1976.
12. Michalski, R. S., "How to Learn Imprecise Concepts: A Method for Employing a Two-Tiered Knowledge Representation in Learning, Procs. of the 4th International Workshop on Machine Learning, Irvine, CA, pp. 50-58, 1987.
13. Bergadano, F. Matwin, S., Michalski, R.S., Zhang, J., "Measuring Quality of Concept Descriptions", Research Report, Artificial Intelligence Center, George Mason University, May, 1988.