LEARNING TWO-TIERED DESCRIPTIONS OF IMPRECISE CONCEPTS: A METHOD EMPLOYING EXAMPLES OF VARIED TYPICALITY AND AN OPTIMIZED BASE CONCEPT REPRESENTATION: PART I: PRINCIPLES AND METHODOLOGY

by

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

Submitted to the Machine Learning Journal.

LEARNING TWO-TIERED DESCRIPTIONS OF FLEXIBLE CONCEPTS

A Method Employing Examples of Varied Typicality and a Two-staged Construction of the Base Concept Representation

Part I: Principles and Methodology

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

Artificial Intelligence Center
George Mason University,
Fairfax, VA 22030

MLI 88-[illegible]
TR-14-88

November 1988

[1] On leave from Universita di Torino, Italy
[2] On leave from University of Ottawa, Canada

LEARNING TWO-TIERED DESCRIPTIONS OF FLEXIBLE CONCEPTS

Part I: Principles and Methodology

ABSTRACT

A method for learning flexible concepts is described, that is concepts that are imprecise and 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*, contains inference rules and metaknowledge that define allowable transformations of the concept under different contexts, and handle exceptional instances.

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 optimized through a heuristic search, employing a *description quality criterion.* The second tier is defined by an expert under the guidance from the system, which asks the expert to explain the context-dependent meaning or special cases of the concept. Alternatively, the second tier can be inherited from more general concepts. This part of the paper concentrates on basic ideas behind the method, and gives illustrative examples. Part II of the paper describes algorithms, their implementation, and experimental results.

ACKNOWLEDGEMENTS

The authors thank G. Thornburg, K. Kaufman, H. Ko, A. Giordana, and F. Oppacher for useful criticism.

This research was done in the Artificial Intelligence Center of George Mason University. Research 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. N00014-87-K-0874, in part by the Office of Naval Research under grant No. N00014-88-K-0226, and in part by the Office of Naval Research under grant No. 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 No. A2480.

1. INTRODUCTION

Most methods of machine learning research assume that concepts are precise entities, representable by a single symbolic description. In such a representation, the boundaries of a concept are well-defined and context-independent. All instances of a concept are assumed to be equally representative. If an instance satisfies the given concept description, then it belongs to the concept, otherwise it does not. Some of these assumptions are relieved in works that assume that the concepts are defined by a probability distribution (e.g., Cheeseman et al., 88) or set membership function (Zadeh, 74). However, once such a probability distribution or a set membership function is defined explicitly for a given concept, the concept again has a fixed meaning, independent of the context in which is used. Moreover, such concept representations remain inadequate for handling exceptional cases, or for capturing increases of knowledge about the properties of the concept.

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. That mean that the classification of instances of imprecise concepts typically involves logical, rather than probabilistic inference. For example, a human trying to resolve the problem if a given object is a desk or a table will evoke logical arguments rather than a probability value.

Examples of human concepts are usually not all equivalent. They may have different *degrees of typicality* in representing the concept. For example, a robin is conventionally viewed as a more typical bird than a penguin or an ostrich. Also, under different contexts the "bird" concept may apply to a live, flying bird, a picture or a sculpture, a chick hatching out of the egg, or even an airplane. Thus human concepts are *flexible*, as they adapt to the context in which they are used. It is clear that in order to handle such flexible concepts, machine learning systems need to employ richer concept representations than are currently used. Developing methods for acquiring flexible concepts and reasoning with them is thus an important goal in the new phase of machine learning research.

One problem that arises is how to represent such concepts. There have been a number of attempts to deal with representing imprecise concepts. Multiple-valued logic (e.g. Rine, 77) introduces additional truth-values, which represent different degrees of certainty with which an instance is believed to belong to a given concept.

The fuzzy set approach (Zadeh, 74) introduces a numerical degree of membership for examples in the concept. This set-membership function is described by people describing the concept, and thus is subjective. The fuzzy set approach allows one to express varying degree of membership of instances in the concept, for example, in the concept "tall person." It does not provide, however, appropriate mechanisms for expressing and handling the context-dependence of the concept meaning.

When the knowledge is incomplete rather than imprecise, it may be appropriate to make default assumptions about the world. In non-monotonic reasoning systems (e.g., Doyle, 79; McCarthy, 80; deKleer 86; Reiter, 87) these assumptions are revised if they can be refuted, i.e., if their negation can be proven. The non-monotonic approach to knowledge representation usually adds significant complexity and computational cost to knowledge representation system, and does not appropriately handle the context-dependency nor express degrees of uncertainty.

The starting point of 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 or ideal properties of the concept. The ICI makes the boundaries of the concept flexible by describing allowed modifications of the concept's features in different contexts. Consequently, learning any concept requires constructing the appropriate BCR and ICI.

In many cases, the ICI is common for a class of concepts: e.g., a method for matching an instance of a specific liver disease with descriptions of different liver diseases will usually be the same for every liver disease. The ICI may also be inherited from more general concepts. For example, many properties of trees are the same as those of plants. The process of inferential matching goes beyond the standard partial matching based on the probability that an instance covers a given concept. It may involve any type of reasoning: deductive, analogical, or inductive inference.

Early ideas on learning two-tiered concept representations were presented in (Michalski, 88a) and closely related earlier work (Michalski et al., 86; Michalski, 87). An intriguing result of that research was that a substantial reduction of the description complexity can be achieved using even a very simple version of such a representation, without affecting its performance.

This paper is an extension and continuation of these early ideas. Important advances are the introduction of a very general description quality measure, the use of a rule base for performing the ICI, and the development of a heuristic search procedure that explores the trade-offs between the BCR and the ICI.

The *general description quality* (GDQ) measure takes into consideration the accuracy, the simplicity of the total description (BCR+ICI), the computational cost and the comprehensibility. The introduction of such a general evaluation measure allows us to redefine the concept of learning (Bergadano et al., 88a). Namely the learning activity is seen as the multistep process of improving the initial concept description (e.g., a set of instances of the concept, or an incomplete or inconsistent description) in terms of the above mentioned description quality measure.

To demonstrate these ideas we have built a system and then applied it to selected problems. Two problems in particular have been investigated: the problem of learning the concept of a "chair" and the problem of learning the concept of a labor contract.

2. TWO-TIERED CONCEPT REPRESENTATION

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. Collins and Quinlan, 72, Minsky, 75, Sowa, 84). The process of recognizing a concept involves simple matching of the stored representation with the perceived facts. Such matching may include comparing features of concept descriptions, or tracing links in networks of concepts, but has not been assumed to involve any complex inferential processes.

On the other hand, 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. This view leads us to the proposition that the meaning we assign to a concept in any given situation is a result of an interplay between two parts. The first part represents what the agent knows, or remembers. The second part represents what the agent can infer from his knowledge, using rules of inference.

The same principle of two-tiered representation seems to apply to individual concepts. In order to investigate the consequences of this conjecture, Michalski (87) has proposed a two-tiered knowledge representation. A concept description is split into two-parts: a Base Concept Representation (BCR) and an Inferential Concept Interpretation (ICI). The BCR defines the concept explicitly, by giving a description of the concept in terms of either the attributes observed in the example, or in terms constructively learned during concept formation. The prototypical instances of the concept are classified by matching with the BCR. Characteristics of the concept represented in the BCR tend to capture the principle, the ideal or the intention behind the concept.

Anomalies, exceptions and context-dependencies are covered by a reasoning process that uses information contained in the ICI. The ICI deals with exceptions by inferring that they are instances of the concept (concept extending), or that they ought to be excluded from the description supplied by the BCR (concept shrinking). The ICI is used in the process of assigning the meaning to a concept using the BCR and the context. This process involves the background knowledge and relevant inference methods that allow the recognition, extension, or modification of the concept meaning according to context. When an unknown entity is matched against the BCR, it may satisfy it directly, or it may satisfy some of its inferential extensions. During the process of interpreting the ICI, one may use a probabilistic inference based on a simple distance measure (so called flexible matching (Michalski et al. 86)), analogical reasoning, inductive reasoning, or deductive reasoning to classify "special" uses of concepts.

Let us illustrate the idea of two-tiered representation with the concept of chair. A two-tiered representation of the chair concept could have the following form:

BCR: A piece of furniture.
Purpose: to seat one person.
Structure: seat, four legs, and a backrest.
(A picture of a typical chair, or a description of the relationship among the parts may be included).

ICI: no-of legs may vary from 1 to 4
the shape, the size, the color and the material of all components can vary as long as the function defined in the BCR is preserved
(chair without the backrest) ---> (stool rather than chair)
(chair with arm-rests) ---> (chair specializes to armchair)

(context = museum exhibit) --> (chair is not used for seating any more)
(context = capital punishment) --> (specializes to electric_chair)
(context = toys) --> (dimensions can be much smaller, but other structural properties are preserved. Does not serve for sitting by normal persons, but by correspondingly small dolls)
(a part of the chair is broken) --> (a broken chair)

This simple example illustrates several important features of the two-tiered representation. If recognition time is important, only BCR will be used to match an example. If more time can be allocated, or if a more precise classification is required for a given event, ICI is used. When interpreting the ICI, one relies on background and general knowledge, and on the context in which the concept operates. Contexts can have hierarchical organization. Finally, ICI rules may chain, although it is not shown in this simple example.

Some systems that generate and use two-tiered representations have been described in the literature (Michalski et al., 86; Bergadano et al., 88b; Bergadano, Giordana, to appear). Also, the work on pruning ID3 trees in order to avoid overfitting (e.g. Cestnik et al., 87) can be viewed as related to this approach. However, pruning the trees results in the loss of coverage of concept instances. The approach presented here does not experience this problem, since the instances removed from the BCR are covered by the ICI.

Two-tiered concept descriptions are usually simpler, easier to understand and more efficient to use than the conventional ones. They also exhibit performance improvement on a testing set. In the systems developed so far, the ICI includes only a flexible matching function. More importantly, in their quality evaluation measures, these systems do not take into account the inferentially covered parts of concept descriptions. Improvement in quality is therefore measured only by the improvement in the first tier.

3. THE LEARNING SYSTEM: A GENERAL OVERVIEW.

3.1. The General Architecture.

The learning system presented here produces two-tiered concept descriptions by performing inference on examples obtained from the source. Fig. 1 presents the general architecture of such a system.

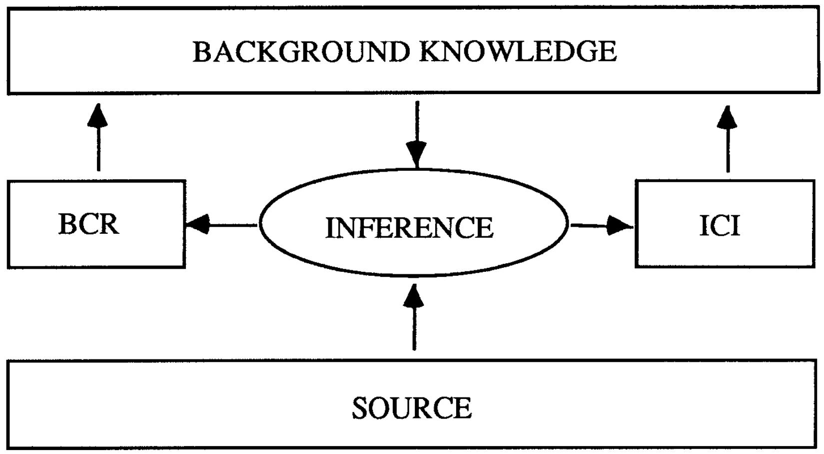


Fig. 1. General architecture of a system for learning two-tiered concept descriptions.

The core of the system is an inference engine that learns concept descriptions in a two-tiered format, consisting of the Base Concept Representation (BCR) and the Inferential Concept Interpretation (ICI). Different types of inference are involved in obtaining a two-tiered representation.

The BCR describes typical and easy-to-define meaning of the concept. Inductive inference is appropriate to perform this task. The inductive inference performed here is knowledge intensive. First the background knowledge can be used as the semantic criterion to restrict the search space. For example, we could determine that one attribute is more important than the others through the deductive inference performed on the background knowledge. Second, the background knowledge can be used to guide the constructive induction. The knowledge intensive method can help the system to learn the BCR from a small set of typical examples.

The ICI is used mainly to handle the special and rare cases. It is obtained from exceptions and nontypical events. Their number is usually limited, and therefore deductive or analogical inference is used to acquire some of the deductive rules. One of the methods to deductively obtain the rules can be outlined as follows: the background knowledge is used to produce the explanations of those exceptions and the ICI rules are generated from the explanations. The explanations can also be obtained from human experts.

Concepts in a given domain usually form a hierarchy, ordered by the relation of one concept being a specialization of another concept. Some ICI rules of a concept may then be inherited from a higher-level concept.

Moreover, as indicated with the arrows in Fig. 1, both parts of the two-tiered description obtained by the system can contribute to the improvement and enrichment of the background knowledge.

3.2. Architecture of the Experimental System.

An experimental version of the system implements the architecture presented in Fig. 1. Table 1 specifies the input, output, and the function of the system.

Input:

Examples obtained from the source.

Output

BCR and ICI for the concept.

(The BCR is an efficient and comprehensible representation of typical instances of the concept. The ICI provides inferential means for assigning instances to classes, and interpreting concept exceptions.)

Function;

Initial concept description is provided by an inductive learner
This description becomes the root of a tree-like search space
Non-root nodes of the search space are obtained by simplification of the ancestor nodes
Search is guided by the General Description Quality measure.

Table 1. A specification of the input, output and the function of the system

Fig. 2 shows the design of the experimental version of the system. There are two differences between the general architecture presented in Fig. 1, and the specific experimental system presented in Fig. 2 that implements it. First, the experimental system relies on an inductive learning system, such as AQ15 (Michalski et al., 86) or INDUCE (Hoff et al., 82), to learn a concept description from examples. This description is treated as the BCR of the initial two-tiered representation, whose ICI is empty. Second, the experimental system relies on the expert to provide rules that explain special events.

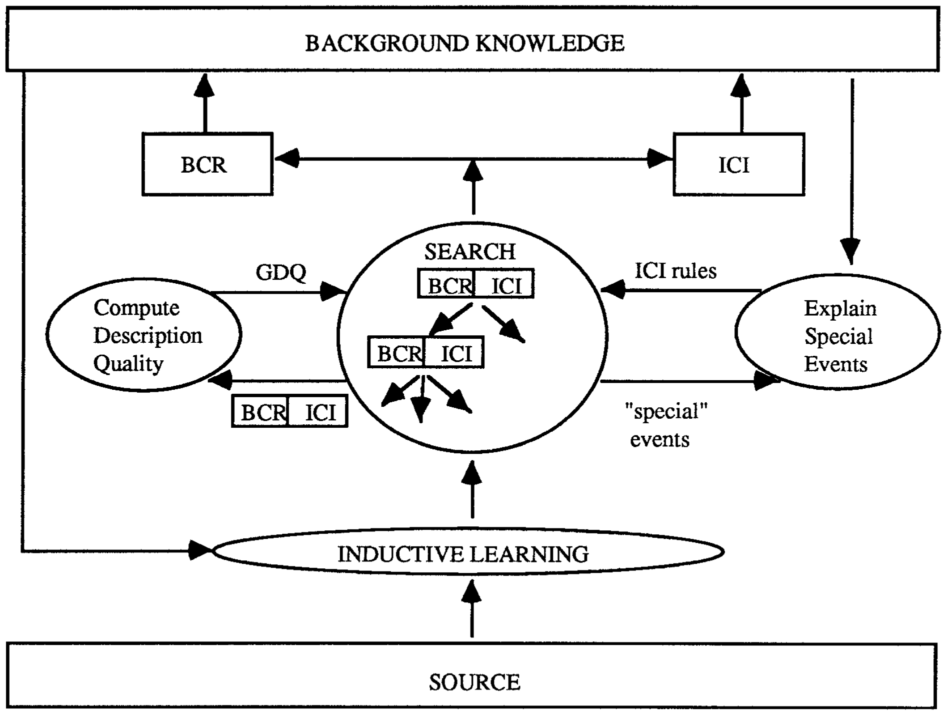


Fig. 2. Design of the experimental version of the system

The BCR is represented in disjunctive normal form described in VL_1 or VL_2 notation (Michalski, 83). The ICI consists of two parts: a flexible matching function and deductive rules. The system learns the BCR and the deductive rules. The flexible matching function is predefined. Corresponding to the three kinds of knowledge, three types of matching are defined:

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

The system attempts to build a better quality two-tiered representation of the concept under consideration, using a heuristic search procedure. The search space is a potentially infinite set of all possible two-tiered descriptions of a given concept. The search procedure searches only a finite fragment of the space. Only those descriptions that can be obtained from the initial description by applying search operations are involved in the search. The search process is driven by the General Description Quality (GDQ) of the descriptions obtained at each step of the search.

3.3. General Description Quality.

The notion of general description quality of concepts (Bergadano et. al. 88a) is formalized and discussed in more detail in Part II. In order to understand how it influences the search, it suffices to say here that the quality of a concept description depends on the accuracy, the comprehensibility, and the cost of the description. Covering the typical instances of the concept by the BCR, and the less typical ones by the ICI, also improves the quality of a description.

The accuracy of a concept description reflects the degree to which the description relates to the concept it describes. In the case of concept learning form examples, accuracy depends on the completeness and consistency of the description with regard to learning examples. It also depends on the typicality of the examples it covers, and the justification that can be constructed for the description. If the description can be plausibly justified in terms of the domain knowledge, the confidence in its correctness will increase.

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. Therefore 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 ensure.

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.

Both the BCR and the ICI are parts of the concept description and they are used together in concept recognition. They influence each other. It is not necessarily true that if a BCR performs well with one ICI, then it also performs well with a different ICI. For example, the experiment described in Part II showed that a BCR performed poorly with an empty ICI, but it performed well with a flexible matching function. Furthermore, in order to learn a better two-tiered concept description, the distribution between the BCR and the ICI should be adjusted during learning. Therefore they should not be learned separately and they should be related in some way during learning. Most of the current learning systems learned the BCR and the ICI independently, and flexible matching is only applied in the performance element. In our approach, when computing GDQ of a concept description, both the BCR and the ICI are considered. When we compute the completeness and consistency, the type of matching (strict matching, deductive matching and flexible matching) of an event is taken into account. The comprehensibility and cost of the ICI are also considered.

3.4. Learning the Base Concept Representation.

An elementary search operation may either specialize or generalize a description. The heuristics used at a given step of the search (Part II) decide which operation is applied. In the experimental version of the system, generalization is implemented as selector truncation, and specialization is implemented as complex truncation. Another operation, referent modification, simplifies the range of a selector, and may behave either as a generalization, or a specialization, depending on the selector relation (see table II). For instance, if the selector is

[size = 1..5, 7]

then referent modification giving

[size = 1..7]

is a generalization, since the cover is extended. On the other hand, if the selector is

[size <> 1..5, 7]

then the same referent modification represents a specialization, since the cover shrinks. Table II summarizes the implementation of generalization and specialization operators in the existing system.

generalization:	selector truncation	referent modification
specialization:	complex truncation	referent modification

Table 2. Implementation of the search operators

3.5. Learning the Inferential Concept Interpretation.

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. The coverage by the concept description is arranged by building ICI rules that will cover them.

On the other hand, when a description obtained from BCRd is more general than BCRd, some new events, previously not covered by BCRd, may have been added. These events could be positive as well as negative. If negative events have been added as a result of generalization, they will have to be excluded from the set of events covered by the concept description by means of the ICI rules. There are two types of rules: rules that cover a positive example otherwise left out of the BCR, and rules that eliminate a negative example from the BCR. Rules of the first type are referred to as extending rules, and rules of the second type are called shrinking rules.

In order to exclude or cover an event by the ICI part of a concept description, one has to obtain rules that will match the event, and perform the action necessary for the exclusion or coverage of the event. These rules, or their chains that ultimately lead to a conclusion regarding the membership of an event in the concept, are treated as an explanation of the event. The rules used in an explanation can be obtained in several ways.

First, suppose that the system is given an instance of a broken chair. More precisely, the description of the instance matches neither the BCR nor the ICI of the chair, because one of the four legs is shorter than the other three. Suppose that the ICI of the leg concept will deduce from this description that the shorter leg is broken. The chair concept is a specialization of the higher level concept furniture. Suppose further that the ICI for furniture has the following rule: if some part of an object that otherwise is an instance of furniture is broken, then the object still is an instance of furniture. The following rule may now be inherited by the ICI of the chair concept: if some part of an object, that otherwise matches the BCR of chair, is broken, than the object is a chair.

Second, an analytic learning system, such as described in (Mitchell et al., 86; DeJong 86), can be applied to obtain an explanation. In these systems, a typical event is explained by deductively inferring it from the underlying background theory. The result is an operational rule for the concept. This rule is then generalized, e.g. using techniques described in (Mitchell et al. 86; Prieditis and Mostow 87). In our approach, only nontypical events will be subject to explanation. The purpose of the explanation here is to justify the special character of the event explained, rather that to operationalize the proof of its membership in the concept.

Finally, explanation of an event may be obtained from an expert, as is the case in knowledge acquisition for expert systems. The experimental system described here uses this method in its initial version.

3.6. A Summary of the Design of the Experimental System.

Once the ICI rules associated with the new description are known, its GDQ is computed. The GDQ value directs the best-first search. A heuristic, described in sec. 5, stops the search process. When the search process stops, the selected node defines the BCR and the ICI of a good concept description. As in the general architecture of Fig. 1, if these give rise to interesting generalizations that impact the general knowledge of the system, this knowledge will be modified. On the other hand, the general knowledge is used in the process of explaining the special cases to be covered by the ICI. Moreover, the general knowledge is necessary to perform constructive induction in the inductive learning phase that produces the initial description.

The architecture presented above involves both analytic and empirical learning, and may support learning in a multi-concept environment. Moreover, its constructive induction feature implements a feedback between the previously learned knowledge and the future learning. Such an architecture satisfies therefore the basic tenets of the constructive learning approach as understood in (Michalski and Ko, 88). Given the encouraging results with the experimental version of the system, a more complex design of an integrated learning system, that does not rely on an inductive learner for its input, is considered.

4. EXAMPLES

This section presents some examples of two-tiered representations obtained using the experimental system. 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 VL_1. Consequently, the learning program AQ15 is used to obtain the initial concept description. The application, discussed in Part II, is the natural extension of this example.

The second example, learning the concept of a chair from structural examples, requires the use of structure-based language, such as VL_2, since relations between objects of given types have to be represented. Consequently, the learning program INDUCE is used to produce the initial concept description.

4.1. Learning two-tiered Description of Labor-management Contracts.

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 some job security, may be accepted during a deep recession. On the other hand, given an exceptionally good economic environment 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 the 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 (gwi), cost of living allowance (cola), job security (job_sec), retirement age (ret_age), extent of pension (p_ext), pension for overtime work (p_ovt), 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 Table. 3. 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).

The examples were submitted to AQ15, which produced a discriminant description of a concept of an acceptable contract (see Fig. 3). The quality of this description, according to the specific quality measure defined in Part II, is 0.989. This description is modified by the system in the following way:

1. Referent modification is performed on selectors. Close interval operation is applied on the first selector of the first two complexes, so that the complex [gwi <> 0..4 v 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.

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

Table 3. Cases of acceptable and unacceptable contracts.

2. The algorithm, described above, gives a modified description, shown in Fig. 5. 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.3 is truncated. The two negative examples are explained as exceptions. The reasoning about this exception is the following:

[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. 3. The initial concept description for the labor-management contract

---

[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. 4. The improved concept description obtained by the system (BCR only)

even if the three attributes occurring in the rule of Fig. 4 (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] --> exceptional(pension)
exceptional(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.4. 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.

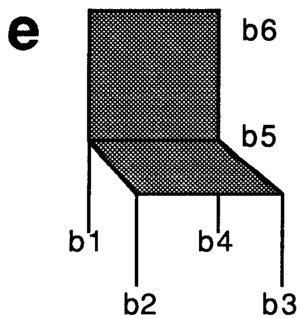
4.2. Learning a two-tiered description of the *chair* concept.

The instances of visual concepts present a high degree of variability, are affected by noise and are subject to modifications related to context. For this reason visual concepts can be better represented through a two-tiered scheme, allowing the system to capture the stable characteristics and reason about the special cases in a unified framework.

As in other pattern analysis problems, instances of visual concepts can be segmented into elementary components, and each component can be described by a list of attributes. Moreover, there can be spatial or functional relationships among the components, and attributes related to the whole instance can be present. It is important to point out that functional attributes have an important role in the representation of the instances, since they are the basis for the learning of a classification knowledge which is truly meaningful and understandable in a given domain. Human learning also seems to interleave the acquisition of concept descriptions and the detection of functional elements in the examples, without a well-defined separation or a clear ordering of the two activities in time.

Below we will give a simple example, where a two-tiered representation of the concept of "chair" is learned from examples in a partially automated way. Fig. 5 gives an example of how a concept instance is described to the system. First of all, the example is divided into components (b1-b6) and the instance is said to contain them. Then every component is described through its shape attribute and its space relationships to the other components (ontop, attached from above). Moreover the predicate person_can_sit_on is used to specify a functional property of the component b5.

Some other examples have been described in a similar way. These examples are given, in a pictorial form, in Fig. 5. There are 7 positive (1 thru 7) and 8 negative examples (8 thru 15) of the "chair" concept. The reader should realize that some of the information contained in the drawings is actually lost in the symbolic description and cannot possibly be obtained in the learned knowledge.



contains(e,b1,b2,b3,b4,b5,b6),
type(b1)=line, type(b2)=line,
type(b3)=line, type(b4)=line,
ontop(b1 & b2 & b3 & b4, floor),
type(b5)=rectangle,
person_can_sit_on(b5),
ontop(b5, b1 & b2 & b3 & b4),
type(b6)=rectangle,
attached_from_above(b6,b5)

Fig. 5 Symbolic representation of a chair

These examples were given as input to INDUCE, together with the following constructive rule, as part of the background knowledge:

[type(x)=line] & [ontop(x,floor)] => [type(x)=leg]

and the following discriminant descriptions were obtained, for the "chair" concept:

1: [type(b1)=leg] & [type(b2)=leg] & [type(b3)=leg] &
[person_can_sit_on(b4)]&[ontop(b4,b1Ÿb2Ÿb3)]&
attached_from_above(b5,b4)] & [type(b4)πcircle]
events covered: 1,3,5,6 - events uniquely covered: 1,5,6

2: [person_can_sit_on(b2)] & [ontop(b2,b1)] &
[type(b1)πline] & [type(b2)=square]
events covered: 7 - events uniquely covered: 7

3: [type(b1)=semicircle]
events covered: 3,4 - events uniquely covered: 4

4: [attached_from_above(b2,b1)]&[ontop(b3,b2)]&[type(b3)=rectangle]
events covered: 2 - events uniquely covered: 2

These four conjunctive descriptions (or complexes) form a disjunctive normal form expression (or cover), that is consistent and complete with respect to the examples given in Fig. 6. In particular, complex 1 covers the positive examples 1,3,5 and 6, among which examples 1,5

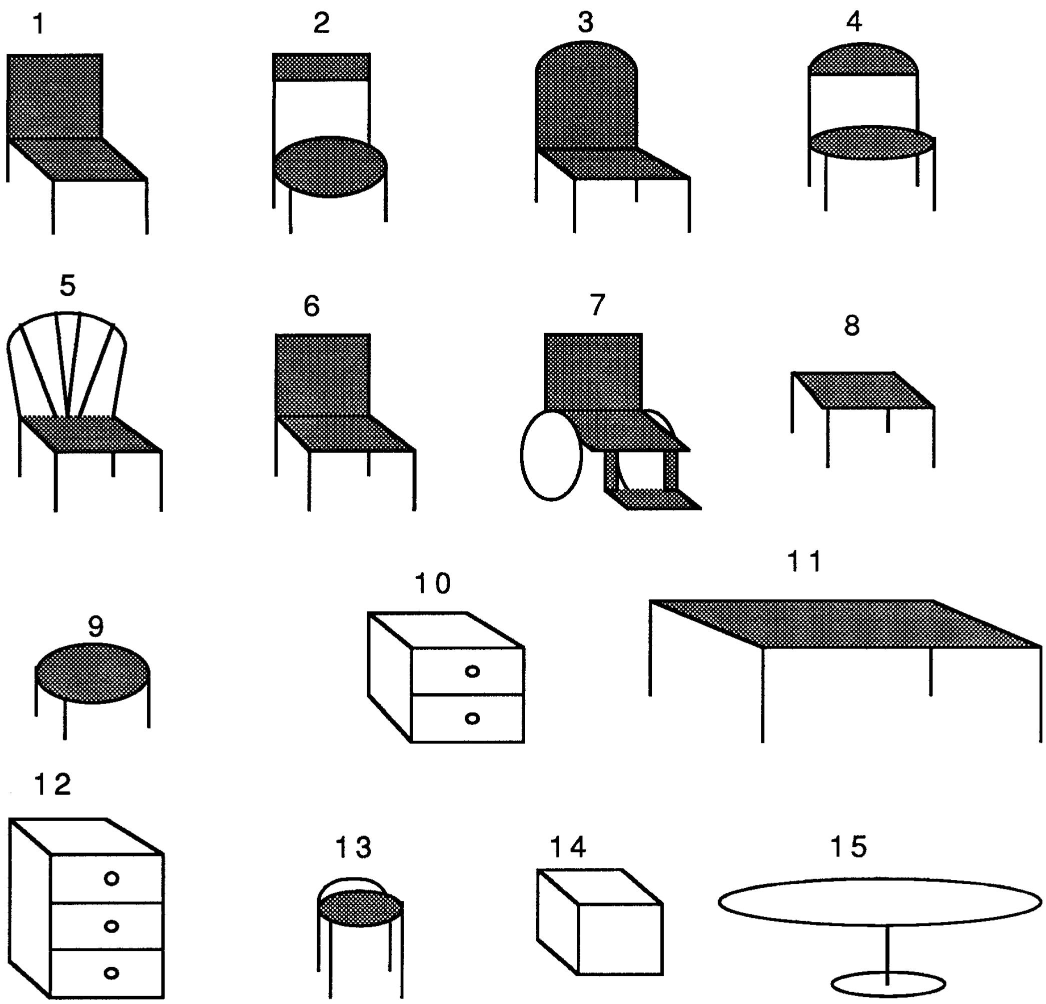


Fig. 6. Visual representation of examples used to obtain a two-tiered description of the chair concept.

and 6 are uniquely covered by this complex; complex 2 covers uniquely positive example 7; complex 3 covers positive examples 3 and 4, and covers uniquely example 4; complex 4 covers uniquely positive example 2. The induction process has effectively used the available background knowledge and the complexes that were produced also contain some functional description, but they still do not capture the most relevant characteristics of the concept of a chair. The reason is that the given examples contain a lot of information which was not really important, and an inductive learning system uses this information if it helps discriminating among the concepts. For example the presence of a component whose shape is semicircle actually covers examples 3 and 4 and no counterexamples in Fig. 6, but is by no means a plausible description of a chair. This is a fundamental problem in many applications of inductive learning, because the acquired knowledge will cover the given events but sometimes will not be understood and accepted by experts. This is often the case when only a limited number of examples is available.

In the following we will show how a two-tiered description of the concept "chair" can be automatically acquired from the description generated by INDUCE in the system described in sec. 3. The obtained concept description will then have two parts: the BCR and the ICI. The BCR for the "chair" concept will not necessarily be consistent nor complete, and some exceptional examples and counterexamples will need to be taken into account by a set of rules in the ICI.

Complex 1 is the most important one, since it covers four positive examples out of seven, and covers uniquely three of them. Since INDUCE only generates perfectly consistent complexes, complex 1 does not cover any negative examples, but this may change when the two-tiered description is constructed, since some inconsistency might be introduced during the truncation procedure. The other complexes cover uniquely only one example each. Moreover if a semantic measure of comprehensibility is available, complex 1 could be evaluated as closer to our intuitive idea of what a chair is, since it contains the requirement that there have to be at least three legs, on top of which there is something a person can sit on.

The description generated by INDUCE, with an empty ICI is then used as a root node in the search process. The system will try to build a better quality two-tiered description of the concept "chair" from the initial description through the heuristic search discussed in sec. 3. The two-tiered description obtained by the system is shown in Fig. 7. It is obtained as follows. First, the last selector [type(b4)πcircle] of the complex 1 is truncated since it will improve the GDQ of the description most. In this way examples 2 and 4 will also be covered by the complex, which will now cover all the positive examples except for example 7, but one counterexample (event 13) will also be covered. Nevertheless, the overall quality of the complex is increased, after truncation of its last selector, also because it will then be possible to truncate complexes 3 and 4, since they will not uniquely cover any positive example. This will improve the simplicity of the description. Event 13 is then taken into account by the following

ICI rules, that are introduced by a human user:

$\exists y$ [person_can_sit_on(y)] & [attached_from_above(x,y)] & [big(x)] => [backrest(x)]
$\neg \exists x$ [backrest(x)] => ˇ chair

these rules will allow the system to reason about "special" counterexamples such as event 13, by understanding that, in a chair, the backrest has to be sufficiently large. At this point, complex 2 can also be truncated, since it only covers one example, and the following two-tiered representation for the "chair" concept will be produced:

BCR: $\exists x,z\ \exists(\geq 3) y$ [person_can_sit_on(x)] &
[type(y)=leg] & [ontop(x,y)] & [attached_from_above(z,x)]

ICI: $\exists y$ [person_can_sit_on(y)] & [attached_from_above(x,y)] &
[big(x)] => [type(x)=backrest]
$\neg \exists x$ [type(x)=backrest] => $\neg$ chair
$\exists(2) x$ [type(x)=wheel] => Irrelevant($\exists(\geq 3) y$ [type(y)=leg])

Fig 8. Two-tiered representation of the "chair" concept

The last rule in the ICI has been introduced in order to cover the positive example that was covered by complex 2, before the truncation. In fact example 7 represents an exceptional condition and should not be included in the BCR; on the contrary, the rule in the ICI captures the relevant information that was present in the example. Event 7 is still a chair, but it is a special kind of chair, where legs have been replaced by two wheels. By using this rule as a rewriting rule, in conjunction with the BCR, event 7 will be covered as a positive example. BCR and ICI, if taken together, are a complete and consistent representation of the concept.

The description is now more intuitive, and the BCR seems to capture our natural understanding of the concept of a chair, while the ICI deals with special cases in an explicit way. The quality of the two tiered description would be higher because its accuracy is the same as the one of the description generated by INDUCE, 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.

5. RELATED WORK.

The research presented here is different in several ways from the recent work in machine learning that investigates the effects of simplifying concept descriptions, e..g. (Fisher and Schlimmer, 88; Iba et al., 88). First, the method described here does not experience any loss of coverage as a result of description modification. This is a major difference between experimental results reported in Part II, and the findings of both (Iba et al., 88) and (Fisher and Schlimmer, 88). The reason is that in our approach events that lose their strict cover as the result of BCR simplification, become then covered by the ICI. Moreover, unlike (Fisher and Schlimmer, 88) and (Iba et al., 88), this approach to concept description simplification takes into account the typicality of events covered by the simplified description, thus preventing loss of coverage of typical events.

The experiments of (Fisher and Schlimmer, 88) in truncating the ID3's decision trees are based on a statistical attribute dependence measure that determines the attributes to be pruned. Because of its statistical character, there is a loss of predictive power when simplifying descriptions learned on small training sets. As the experiments described in Part II indicate, the approach presented here does not seem to suffer from this problem.

The system developed by (Iba et al. 88) uses a trade-off measure that is similar to the GDQ measure proposed in this paper. The GDQ measure, defined in detail in Part II, considers more factors. Besides taking into account the typicality of the instances covered by the description, it considers the type of matching between an instance and a description. Moreover, the simplicity measured by the GDQ depends not only on the number of disjuncts in the description, as in (Iba et al. 88), but also on the different syntactic features of the terms in the description.

An important difference between the approach presented here and pruning of decision trees (Cestnik et al., 87) 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. Moreover, pruning decision trees involves only generalization of the concept description, while the above method performs both generalization and specialization of the description.

Truncation of the BCR, obtained inductively from a small learning set does not affect predictive power if an adequate typicality measure is available. The existence of an adequate ICI further alleviates the problems resulting from induction with few examples.

The problem of defining and using the typicality of examples has been considered in the past both in machine learning and cognitive science. Negative examples of low typicality are referred to as *near misses* in Winston's system (Winston, 1975). Such examples, labeled by the user as near misses, are used in Winston's system to delineate the borders of a concept. Michalski and Larson, (78) introduced the idea of an outstanding representative of a concept. The concept of prototypical examples has been also studied by Smith and Medin (81) and by Roch and Mervis (75). Prototypical examples are fully specified instantiations of a given concept. In the method described here such prototypical examples do not need to exist, prototypical properties of a concept are learned automatically from examples of different typicality.

To summarize, there are four major differences between the work presented here and related research described in the literature. First, the above method does not experience a loss of coverage although it still yields a simpler description with improved predictive power. Second, it simplifies the description by performing both generalization and specialization. Third, any part of the description may be truncated in the simplification process. Finally, the method takes into account the typicality of the examples.

6. CONCLUSIONS AND FUTURE WORK.

The paper describes a method of learning two-tiered concept descriptions. The method is based on transforming an initial Base Concept Representation. The transformed BCR covers groups of events characterized by high typicality. It is also syntactically simpler, and therefore more comprehensible. A more complete coverage of the events from the learning set by the whole two-tiered description is achieved through inference. The method presented in this paper relies not only on the probabilistic inference, implemented as flexible matching in (Michalski et al., 86). It uses also a rule base for deductive inference. Deductive inference has the additional advantage of explaining why a given event is to be included (or excluded) from the cover.

In order to achieve this objective, transformations of BCR are implemented as truncations of the cover. The cover is provided by AQ15 in a standard, disjunctive normal form. The truncations either specialize the description (complex truncation), or generalize it (selector truncation). A search process, guided by a heuristic quality measure, is used to obtain a "good" description. The measure takes into account not only the explicit part of the description, but also the implicit one. The algorithms and heuristics are discussed in more detail in Part II of this paper.

The experiments that we have performed on real data indicate that the system is capable of learning good quality descriptions that are easy to comprehend and efficient to evaluate. In Part II, we describe an application in which a two-tiered description of a labor-management example has been learned from real data. The results indicate that recognition rate of two-tiered concept descriptions on the testing set is better than the performance of the normal, inductively learned descriptions on this set. This provides the evidence about the adequate predictive power of a two-tiered concept description of good quality. On the other hand, the BCR of the two-tiered description is much simpler than the cover learned by AQ15.

A number of problems remain to be addressed in the future. First, an integrated system that learns two-tiered descriptions needs to be designed and build. Such a system will have to satisfy a number of design goals: it will have to be incremental, learn descriptions of good quality, exhibit good predictive power, and be efficient. The incremental behavior of the system would involve remembering examples that are neither strictly matched nor close (in the sense of the measure used by flexible matching) to the BCR. Then an explanation of such an example using the existing background and context knowledge, and the existing deductive ICI rules, is attempted. If no explanation is obtained automatically or from an expert, the example is stored. When the ICI grows significantly, there will be a repeated attempt to explain such an example.

Second, the problem of automatic acquisition of the ICI has to be investigated and tackled. The methods developed in Explanation-Based Learning will provide a good starting point. It has to be observed, however, that since the events to be explained are usually exceptions, the knowledge necessary to explain them may be lacking from the system.

Third, the system does not address the problems of dynamically emerging hierarchies of concepts. In the existing version the system only learns one concept at a time, and concepts do not change or split as new examples become available.

Finally, the system should be able to self-reorganize. The distribution of knowledge between the BCR and the ICI will be determined by the performance of the system on large testing sets. If it turns out, e.g., that some ICI rules are used very often, then these rules could be compiled into explicit BCR assertions. The representation obtained in this manner will be faster, but it will occupy more memory. It seems, therefore, that in concept representation one can trade one parameter against the other, within certain limits. This interesting research problem merits further investigation.

REFERENCES

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

Bergadano, F. , Giordana, A. , Saitta, L., "Automated Concept Acquisition in Noisy Environments", IEEE Transactions on PAMI, Vol 10, NO. 4, pp. 555-578, July1988.

Bergadano, F., Giordana, A., "Pattern Classification: An Approximate Reasoning Framework", International Journal of Intelligent Systems, (To appear).

Cheeseman, P., Kelly, J., Self, M., Stutz, J., Taylor, W., Freeman, D., "AutoClass: A Bayesian Classification System", Procs. of the Fifth Int'l. Conf. On Machine Learning, Ann Arbor, pp. 54-64 (1988).

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

Clark, K., "Negation as Failure", pp. 293-322 in Logic and Data Bases, ed. Gallaire, H. and Minker, J., Plenum Press, New York (1978).

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.

de Kleer, J., "An Assumption Based TMS", Artificial Intelligence 28, pp. 127-162 (1986).

DeJong, G., Mooney, R., "Explanation-based Learning: An Alternative View", Machine Learning, vol. 1, No. 2, 1986.

Doyle, J., "A Truth Maintenance System", Artificial Intelligence 12, pp. 231-272 (1979).

Fisher, D. H., Schlimmer, J. C., "Concept Simplification and Prediction Accuracy", Procs. of the Fifth 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 from 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).

McCarthy, J., "Circumscription: a form of Non-monotonic Reasoning", Artificial Intelligence 13, 27-39 (1980).

Michalski, R.S., Larson, J. B., "Selection of Most Representative Training Examples and Incremental Generation of VL_1 Hypotheses: the Underlying Methodology and the Description of Programs ESEL and AQ11", TR 867, Dept. of Computer Sci., UIUC, 1978.

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, Hong, J.,I., Lavrac, "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), 1988.

Michalski, R. S. and Ko, H., "On the Nature of Explanation, or Why Dod the Wine Bottle Shatter", AAAI Symposium: Explanation-Based Learning, Stanford University, pp. 12-16, 1988a.

Michalski, R. S., "How to Learn Imprecise Cocepts: A Method Employing a Two-Tiered Representation for Learning", Procs. of the Fourth International Workshop on Machine Learning, Irvine, CA, pp. 50-58,1987.

Minsky, M., "A Framework for Representing Knowledge", in The Psychology of Computer Vision, ed. P. Winston, (1975) Mitchell, T. M., Keller, R. & Kedar-Cabelli, S., "Explanation-based Generalization: a unifying view," Machine Learning Journal, 1, 11-46 1986.

Prieditis, A. E. and Mostow, J., "PROLEARN: Towards A Prolog Interpreter that Learns," Proc. of IJCAI 87, Milan, pp 494-498 1987.

Reiter, R., "Non-monotonic Reasoning", pp. 147-186 in Annual Review of Computer Science, (1987).

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

Smith, E. E., Medin, D. L., "Categories and Concepts", Harvard University Press, 1981.

Rosch, E., and Mervis, C. B., "Family Resemblances: Studies in the Internal Structure of Categories", Cognitive Psychology, vol. 7, pp. 573-605, 1975.

Sowa, J. F., "Conceptual Structures", Addison Wesley, 1984.

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