LEARNING BY BEING TOLD AND LEARNING FROM EXAMPLES: AN EXPERIMENTAL COMPARISON OF THE TWO METHODS OF KNOWLEDGE ACQUISITION

by

R. S. Michalski

International Journal of Policy Analysis and Information Systmes, Vol. 4, No. 2, 1980.

90-S

International Journal of Policy Analysis and Information Systems, Vol. 4, No. 2, 1980

Learning by Being Told and Learning from Examples: An Experimental Comparison of the Two Methods of Knowledge Acquisition in the Context of Developing an Expert System for Soybean Disease Diagnosis

R. S. Michalski[1] and R. L. Chilausky[1]

Received January 30, 1980

Current methods of machine acquisition of knowledge rely entirely on hand-crafted encoding of the knowledge of human experts. Such a process is very time and effort consuming, and therefore alternative methods are needed. This paper contrasts the method of knowledge acquisition by encoding decision rules of human experts with that of learning the rules (by means of an inductive program) from examples of decisions made by these experts. Both types of rules—the expert derived and the inductively derived—are expressed as special cases of *general variable-valued logic rules* (GVL_1 *rules*), which are an extension of the conventional condition–action rules. The problem is considered in the context of developing a knowledge base for an expert system PLANT for the diagnosis of crop diseases. The experiments (which involved testing both types of rules on several hundred cases of soybean diseases) have demonstrated the usefulness and practicality of the inductive method of knowledge acquisition for the limited problem under consideration. They also indicated how the method could be further improved.

KEY WORDS: Knowledge acquisition; knowledge representation; learning; inductive inference; expert systems; decision rules; pattern-directed inference; soybean disease diagnosis; agricultural consultation.

1. INTRODUCTION

It is a growing conviction that this decade will witness a rapid development of knowledge-based expert systems in many different fields. Two prerequisites—the strong social need for such systems and the technical feasibility—both exist now. In the area of medicine alone, the amount of

[1]University of Illinois, Urbana, Illinois.

diagnostic, therapeutic, and general knowledge is already so vast, and growing so rapidly, that no single physician can encompass it. Similar situations exist in many other areas. Expert systems can help significantly: they can provide an interactively accessible source of updated and well-organized knowledge on specific subjects, and can conduct a certain amount of inference to help a user with decision making.

In medical diagnosis, for example, an expert system can indicate the most likely problem for investigation, bring attention, when appropriate, to nontypical cases which are easy to overlook, alert a physician about the possibility of drug interaction, provide analysis of possible outcomes, and serve as an always ready source of expertise. Several experimental systems have already been developed in medicine. Among the most well-known are MYCIN for antimicrobial therapy advice (Shortliffe,[20] Davis,[4] INTERNIST for diagnosis in internal medicine (Myers and Pople[18]), and CASNET for glaucomas (Weiss *et al.*,[21]). Several other expert systems for medical and other applications are summarized by Feigenbaum.[8] Among them, earlier and most developed, is DENDRAL for determining molecular structures of complex organic chemicals from mass spectrograms and related data (Buchanan and Feigenbaum[3]). In geology a system PROSPECTOR was developed to provide consultation about potential mineral deposits (Duda *et al.*,[7]). Current aspects of research on expert systems are discussed in (Michie[16]).

An expert consulting system consists of a knowledge base and an inference mechanism, which matches the data provided by the users with rules in the knowledge base in order to compute conclusions. A knowledge base is a symbolic representation of factual and, as well, judgmental knowledge in the subject domain. In each of the above-mentioned consultation systems, the knowledge base was established by handcrafted encoding of the knowledge of human experts. Such encoding can be a very time consuming task, requiring close collaboration between experts of the subject domain and computer scientists trained as "knowledge engineers." This task can be simplified somewhat by special computer programs that facilitate the development, modification, and maintenance of the knowledge base (Davis,[4] Baskin and Levy[1]).

An attractive improvement of the process of constructing a knowledge base would be to use, whenever possible, an inductive program able to learn or refine decision rules on the basis of examples of expert decisions. It is generally much easier to collect and codify such examples than to formulate reliable and complete expert rules. The research on computer inductive inference is still at an early stage of development; however, it is already possible to obtain practical results, if the problem is sufficiently well defined and specialized. The papers by Buchanan and Feigenbaum,[3] Mitchell[17] and Dietterich and Michalski[6] describe some more recent work in this area.

This paper presents results from the application of an inductive program to the problem of learning the decision rules for the diagnosis of soybean diseases from examples. The rules produced are then compared with the decision rules obtained by directly representing the decision rules communicated by experts in plant pathology. In order to be able to relate both types of rules (inductively derived and expert derived) to each other and to use the same inference mechanism when applying them for diagnosis, a general format for representing decision rules was developed. This general format, the *general variable-valued logic rule* (*GVL_1 rule*), comprises either type of rule as a special case, and is an extension of the conventional format of condition–action rules.

2. THE FORMALISM USED FOR KNOWLEDGE REPRESENTATION

A formalism for knowledge representation should be not only equipped with adequate operators and data structures for representing many different aspects of human knowledge, but also well suited for implementing inference processes on this knowledge.

One of the most common methods for representing knowledge is to use condition-action rules or productions (e.g., Davis, Buchanan, and Shortliffe[5]). The condition part of such rules is typically a logical product of several conditions, and the action part describes a decision, an action, or an assignment of values to variables that is to be performed when a *situation* satisfies the condition part. A rule can have an associated "strength of implication"—a parameter indicating the *degree of confidence* in the correctness of the action when the condition part of the rule is fully satisfied.

Another method for representing expert knowledge is to use semantic nets (Brachman[2]) whose general form is a labeled graph with nodes representing various conceptual entities and links representing relationships among these entities. This way of representing knowledge is quite natural for many problems. The network representation has, however, several drawbacks. First, since everything is interconnected, it is difficult to modify, incrementally update, or extend the knowledge base. Also, it is difficult to represent nonbinary relationships. For example, it is difficult to represent a statement indicating that a certain logical product of concepts (associated with various nodes) implies some other concept, and that the "strength of the implication" is so and so. Such statements are, however, very common in human decision processes, and, therefore, rule representation is often preferable. In the study by Duda *et al.*[7] the initial representation of knowledge is in terms of rules, but in the final stage these rules are incorporated into a so-called *partitioned semantic net.*

An important advantage of the condition–action rules is that they can represent individual "chunks" or "modules" of human knowledge, which makes it relatively easy to comprehend the rules, to modify them, and to incrementally build up the knowledge base. Also, it is simpler to explain to a user the inference process conducted by the system by indicating the involved decision rules, than by showing a part of a network. In addition, the rule representation seems to be more convenient for inductive learning, and, in fact, there are already some quite advanced inductive programs working in this framework.

An accurate encapsulation of knowledge in the form of rules does, however, encounter a number of problems. Typically, an expert's knowledge is expressed in terms of imprecise concepts and involves many different operators with various shades of meaning. Also, much of this knowledge is accompanied by statements indicating varying degrees of credibility and varying levels of importance assigned to expressed conditions.

This paper uses the rule representation of knowledge. The knowledge involves descriptions of plant conditions indicating various soybean diseases. The format of the rules is based on the variable-valued logic calculus VL_1 (Michalski[11]). This calculus was developed for formally representing in a simple, compact, and self-explanatory way decision and inference processes involving many-valued variables. Commonly, the variables in such processes have problem dependent value sets, which can differ both in the scope and in the structure relating their elements. For example, "sex" is a 2-valued variable with no structure relating its possible values, while "height" or "temperature" of a human being varies within a certain range of possible values, and the values constitute a linearly ordered set.

A simple way of characterizing, e.g., a person is by a list of attribute–value pairs, which in VL_1 is written in the form

$$[\text{sex} = \text{male}][\text{height} = \text{medium}][\text{blood-type} = \text{O+}]$$

The form in brackets [ ] is called a *selector*, and represents a relational statement relating a variable to one or more values from its domain. A general form of a selector is

$$[x_i \# R] \tag{1}$$

where # stands for one of the relational symbols $=, \neq, \geq, >, \leq$, or $<$, and R (the *reference*) denotes a subset of the value set of variable x_i. A concatenation of selectors denotes a *term* (usually interpreted as the logical product). VL_1 does not include functions or relations. In many applications, however, descriptions using only variables are sufficient. (A richer language developed in the same spirit, which includes functions, relations, and some other forms, is VL_2 (Michalski[13]).)

An *event* e is defined as a list of values of an assumed set of variables. For example, assuming the variables sex, height, and blood-type, an event can be

$$e\text{: (male, 5 ft 11 in, A+)}$$

An event e is said to *satisfy* a selector $S:[x_i \# R]$ if the value of x_i in e is related by $\#$ to any element of R. For example, selector

$$[\text{albumin} = \text{low, high}] \quad \text{or, equivalently,} \quad [\text{albumin} \neq \text{medium}]$$

is satisfied by e if the value of albumin in e is low or high.

It is easy to see that if the *reference* of a selector has more than one element, the selector is equivalent to a disjunction of selectors with one-element references:

$$[x_i \# a, b, \ldots] \equiv [x_i \# a] \vee [x_i \# b] \vee \cdots$$

A selector with a reference consisting of more than one element denotes the so-called *internal disjunction* (disjunction of values of the same variable).

In discussions with experts who are trying to describe their decision processes, we observed that they often state a condition for a specific diagnosis as a sequence of observations or symptoms (which can be represented by a conjunction of appropriate selectors). However, these experts often also indicate that certain observations are more important than others. In our experiment, observations have ranged from very important to merely supportive or confirmatory.

Therefore, we extended here the concept of a selector as defined above (Michalski[11]) by adding to it a *weight.*

The weight of a selector expresses the *strength of evidence* provided by a satisfied selector to support a decision. Typically in our experiments the *weight* of selectors was not a constant, but varied with the values of the variable in the selector. Such a situation arises, for example, when one wants to express a statement: "A precondition for *Diaporthe stem canker* is high precipitation; and with the increase of precipitation the chances of the disease also increase."

To be able to express a range of conditions of this type, we defined the *weighted selector* as follows:

$$[x_i \# R\!:\!@fn] \tag{2}$$

where

1. fn is a *weight assigning function* defined on the values of x_i that satisfy the relation $x_i \# R$.

2. @ stands for an optional symbol indicating a general behavior of function fn. It can be $\uparrow$ ($\downarrow$) when fn is monotonically increasing (decreasing), or $\cup$ ($\cap$), when it has a maximum (minimum) around some mean and decreases (increases) with the distance from this mean. When fn is defined over the whole domain of x_i, then the part "$\# R$" is dropped, i.e., the selector has the form $[x_i : @fn]$.

If the function fn is the constant 1, then a selector is written simply $[x_i \# R_i]$, i.e., it reduces to the original form (1). Such a form is called an *unweighted selector.*

For example, [number of years crop repeated: $\uparrow YR_1$] is a weighted selector in which the weight assigning function YR_1 is monotonically growing with the number of years the crop was repeated in the same field. The function YR_1 may be defined, e.g.,

$$YR_1 : w = \begin{cases} 1.0, & \text{if the crop is repeated 3 or more years} \\ 0.8, & \text{if the crop is repeated 2 years} \\ 0.7, & \text{if the crop is repeated 1 year} \\ 0.2, & \text{if the crop has not been repeated} \end{cases}$$

which graphically is as shown in Fig. 1.

In medical, plant pathological, and other applications, many variables are numerical measurements, on an interval or stronger scale, whose values are known only approximately. In such cases the match between an event and a selector may be better described by a *degree* than by a YES–NO answer. Such a degree can be determined, e.g., from the error estimate of the measuring instrument. We will therefore assume that, in general, matching an event e with an unweighted selector S produces a value $v(S, e) \in [0, 1]$, called a *degree of confirmation.*

The degree of confirmation, $v(S^w, e)$, of a weighted selector S^w by event e, expresses the evidence provided by the selector to support a

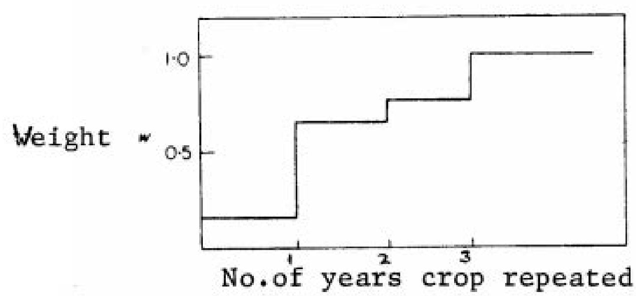

Fig. 1

decision. It can be computed by using functions such as:

$$P: \quad v(S^w, e) = v(S, e) + (1 - v(S, e))(1 - w) \tag{3}$$

or

$$N: \quad v(S^w, e) = (2v(S, e) - 1)w \tag{4}$$

Function P ("positive evidence") assumes that the role of weight w is mainly to affect the evidence when a selector is not satisfied (if $v(S, e) \approx 0$ then $v(S^w, e) \approx 1 - w$), that is, to weaken the negative effect of not satisfying a selector when the weight is small. Function N ("positive or negative evidence") assumes that the weight symmetrically affects the evidence, i.e., increases it when the sèlector is satisfied, and decreases it when the selector is not satisfied. The choice of the function can be determined experimentally.

Terms and GVL_1 Rules. A concatenation of weighted selectors is called a *term*. If a term is interpreted as a conjunction, then it can be used as a conjunctive condition in a decision rule. Conjunctive conditions, however, although very common, are generally not sufficient. Experts often predicate one condition upon another condition. This can be expressed as a *one-dimensional* or *bidirectional conditional statement*, respectively:

$$Term_1 \Rightarrow Term_2 \tag{5}$$

or

$$Term_1 \Leftrightarrow Term_2 \tag{6}$$

where $\Rightarrow$ and $\Leftrightarrow$ denote logical implication and equivalence, respectively.

Since $1 \Rightarrow Term$ is equivalent to *Term*, a conjunctive term can be viewed as a special case of a conditional statement.

In describing symptoms for a disease, plant pathologists often qualify whole groups of symptoms as, e.g., "significant," "confirmatory," etc., independently of an individual symptom's weight. In order to express such statements, the concept of a *linear expression* was developed. It is a linear function:

$$q_1 \cdot \mathrm{ET}_1 + q_2 \cdot \mathrm{ET}_2 + \cdots + q_k \cdot \mathrm{ET}_k \tag{7}$$

where:

1. ET_i, $i = 1, 2, \ldots, k$, are *extended terms*, defined as concatenations of conditional statements.
2. $q_1, q_2, \ldots, q_k \in [0, 1]$, $q_1 + q_2 + \cdots + q_k = 1$, are coefficients indicating the relative significance of extended terms.

3. $\cdot$ is the operator defined as the arithmetic multiplication of coefficient q_i by the degree of confirmation of ET_i.

A linear expression reduces to an extended term when it has only one coefficient.

Thus, the expression (7) provides a means for combining both logical and arithmetic operators in representing various aspects of expert knowledge.

The following is the general form of decision rules accepted in this study:

$$\bigvee_j \mathrm{LE}_j \overset{\alpha}{::>} \mathrm{CT} \tag{8}$$

where

1. $\bigvee$ denotes the disjunction operator.
2. LE_j, $j = 1, 2, \ldots$, are linear expressions.
3. CT is a term describing the decision to be assigned when the condition part (on the left of $::>$) is satisfied.
4. α denotes the *strength of implication*, $\alpha \in [0, 1]$ (is not listed when $\alpha = 1$).
5. $::>$ denotes the decision assignment operator. (It is used instead of the often used symbol $\rightarrow$ in order to avoid confusion with logical implication.)

Decision rules in form (8) are called *generalized VL_1 rules*, or in short, GVL_1 rules. GVL_1 rules in which each linear expression is a conjunctive term are called *disjunctive VL_1 rules*, or in short *DVL_1 rules*. Figure 2 gives a summary of introduced concepts.

Interpretation of GVL_1 Rules. A GLV_1 rule provides a very general structure for expressing different kinds of relationships of elementary observations (selectors) in support of a decision. It can reduce to various special cases, and by this can provide a best "fit" between the informal experts' knowledge and its formal representation. Weighted selectors, the concatenation of selectors, and the disjunction operator have no fixed interpretation. The interpretation is given by an *evaluation scheme*. For example, a scheme called (N, MIN, MAX) assumes that weighted selectors are evaluated according to function N (4), the concatenation is the *conjunction*, and the disjunction is the *maximum function*.

According to this scheme, e.g., the rule

$$0.9 \cdot ([x_1 = 3][x_3 \geq 2]) + 0.1 \cdot ([x_3 = 2 \ldots 4] \Rightarrow [x_5 = 0]) ::> [\text{decision} = \text{A}] \tag{9}$$

$\underline{\text{DVL}_1 \text{ rules}}$

(Unweighted) selector	S: $[x_i \# R]$
Term	T: [ ][ ][ ]..
DVL_1 rule (disjunctive variable-valued system one rule)	$T_1 \vee T_2 \vee \ldots ::> T \ldots$

$\underline{\text{GVL}_1 \text{ rules}}$

(Weighted) Selector	S^W: $[x_i \# R{:}@f_n]$
(Conjunctive) Term	CT: [ ][ ]...
Conditional Statement	CS: [ ][ ]..⇒[ ][ ]..
Extended Term	ET: $(CS_1)(CS_2)\ldots$
Linear expression	LE: $q_1 \cdot ET_1 + q_2 \cdot ET_2 + \ldots q_k \cdot ET_k$, where $q_1 + q_2 + \ldots + q_k = 1$
GVL_1 rule (generalized VL_1 rule)	$LE_1 \vee LE_2 \vee \ldots ::> CT$

Fig. 2. Structure of DVL_1 and GVL_1 rules.

can be interpreted:

> If x_1 is 3 and x_3 is greater than or equal to 2, then it constitutes 90% of the support of decision A. Additional 10% support is given when x_3 is between 2 and 4 and x_5 is 0, or when x_3 is not between 2 and 4 (in this case x_5 can have any value).

Similarly, a rule

$$[x_2 \neq 3][x_3 = 1, 3] \vee [x_4 < 4] ::> [\text{decision} = A] \tag{10}$$

can be interpreted:

> Decision A is taken if x_2 is not 3 and x_3 is 1 or 3, or if x_4 is smaller than 4.

Some other interpretation schemes are discussed in (Michalski[14]). The GVL_1 rules provide a general format of which expert derived rules and inductively derived rules are special cases. The form of expert derived rules is like that of rule (9) while the form of inductively derived rules is like that of rule (10) (i.e., a DVL_1 form).

3. DESCRIPTION SPACE

In the case study the diagnosis of soybean diseases was selected as being representative of the problems one faces in the diagnosis of plant diseases in general. The task was to develop a knowledge base which contained sufficient information to diagnose the following 15 soybean diseases:

D1: *Diaporthe stem canker*
D2: *Charcoal rot*
D3: *Rhizoctonia root rot*
D4: *Phytophthora root rot*
D5: *Brown stem rot*
D6: *Powdery mildew*
D7: *Downy mildew*
D8: *Brown spot*
D9: *Bacterial blight*
D10: *Bacterial pustule*
D11: *Purple seed stain*
D12: *Anthracnose*
D13: *Phyllostica leaf spot*
D14: *Alternaria leaf spot*
D15: *Frog eye leaf spot*

A description space for diagnosing the selected soybean diseases was developed in collaboration with an expert in soybean pathology. The variables used were 35 plant and environmental descriptors and one decision variable (specifying diagnosis). The intent in selecting the particular descriptors and their associated values was to provide a description space that was sufficient to describe the diseases of soybeans in terms of macrosymptoms, i.e., those symptoms that could be clearly observed with no sophisticated mechanical assistance. The reason was that an extension service agent, a farmer, or even a layman should be able to make the required observations.

A descriptor is a function that associates with the plant or its environment a specific value from the set called the *value set* (*domain*) of the descriptor. For example, the descriptor "Time of Occurrence" (TOC) specifies for the diseased plant the time of occurrence of the disease in the

Table I. Plant Descriptors Used in the Experiment

Name of descriptor	Number of values	Variable
1. Environmental descriptors		
1.1 Time of occurrence	(7)	(x_1)
1.2 Plant stand	(2)	(x_2)
1.3 Precipitation	(3)	(x_3)
1.4 Temperature	(3)	(x_4)
1.5 Occurrence of hail	(2)	(x_5)
1.6 Number years crop repeated	(10)	(x_6)
1.7 Damaged area	(4)	(x_7)
2. Plant global descriptors		
2.1 Severity	(3)	(x_8)
2.2 Seed treatment	(3)	(x_9)
2.3 Seed germination	(3)	(x_{10})
2.4 Plant height	(2)	(x_{11})
3. Plant local descriptors		
3.1 Condition of leaves	(2)	(x_{12})
3.1.1 Leafspots—halos	(3)	(x_{13})
3.1.2 Leafspots—margin	(3)	(x_{14})
3.1.3 Leafspot size	(3)	(x_{15})
3.1.4 Leaf shredding or shot holing	(2)	(x_{16})
3.1.5 Leaf malformation	(2)	(x_{17})
3.1.6 Leaf mildew growth	(3)	(x_{18})
3.2 Condition of stem	(2)	(x_{19})
3.2.1 Presence of lodging	(2)	(x_{20})
3.2.2 Stem cankers	(4)	(x_{21})
3.2.3 Canker lesion color	(4)	(x_{22})
3.2.4 Fruiting pod on stem	(2)	(x_{23})
3.2.5 External decay	(3)	(x_{24})
3.2.6 Mycelium on stem	(2)	(x_{25})
3.2.7 Internal discoloration	(3)	(x_{26})
3.2.8 Sclerotia—internal or external	(2)	(x_{27})
3.3 Conditions of fruits—pods	(4)	(x_{28})
3.3.1 Fruit spots	(5)	(x_{29})
3.4 Condition of seed	(2)	(x_{30})
3.4.1 Mold growth	(2)	(x_{31})
3.4.2 Seed discoloration	(2)	(x_{32})
3.4.3 Seed size	(2)	(x_{33})
3.4.4 Seed shriveling	(2)	(x_{34})
3.5 Condition of roots	(3)	(x_{35})

field. The descriptor "Condition of Roots" (COR) assigns a value describing the state of the roots of the plant. The domains of these descriptors in the knowledge base were:

Domain(TOC) = {April, May, June, July, August, September, October}

Domain(COR) = {Normal, Rotted, Galls or Cysts Present}

Table I lists all 35 descriptors that were used. The number in parentheses following each descriptor indicates the number of possible values the descriptor can take. In addition, there is a decision variable which specifies the diagnosis of a disease from the assumed set of soybean diseases.

Individual diseased plants were described in terms of the above 35 descriptors. Thus, the theoretical size of the description space (i.e., the set of all possible sequences of values of descriptors) was:

$$7 \times 2 \times 3 \times \cdots \times 2 \times 2 \times 3 = \text{approx. } 3 \times 10^{15} \text{ events}$$

Owing to relationships among some of the variables (see Fig. 3) the actual size of the event space was somewhat smaller.

4. EXPERT DERIVED DECISION RULES ("LEARNING BY BEING TOLD")

The knowledge base of a computer expert system can store (at least at the present time) only symbolic descriptions of the diseases. These descriptions are formal expressions involving names of observable symptoms, manifestations, and characteristics indicative of each disease. Such a knowledge base is vastly different from the knowledge base of a human expert, which typically contains a multiplicity of sensory recordings (primarily mental images, as indicated by experts), memories of experiences, teachings from scholarly and less scholarly texts, opinions of others, etc. It involves complex models of structures, physiology, behavior, etc., of plants, and of causal interrelationships between plants, pathogens, and the environment. When an expert is asked to describe the symptoms of a disease, he skims from this vast knowledge structure some "most important" simple characteristics which he turns into a verbal form.

A "knowledge engineer" then transforms these verbal statements—typically loosely structured, of varied precision and uncontrolled scope—into precise, well-defined formal expressions, involving a restricted vocabulary of concepts.

In this experiment it took several conferences with a plant pathologist and approximately 45 hours of time to develop formal descriptions of the 15 soybean diseases listed in Sec. 3. These descriptions were expressed as

GVL_1 rules. Each rule characterizes a single disease, and involves a two-component linear expression:

$$q_s \cdot ET_s + q_c \cdot ET_c :: > [\text{disease} = d] \tag{11}$$

where ET_s and ET_c are extended terms characterizing significant and confirmatory conditions for each disease, respectively.

For example, the following is an example of the rule describing *Diapothre stem canker*:

$$\begin{aligned}
&q_s \cdot ([\text{time} = \text{August} \ldots \text{September}] \\
&\quad [\text{precipitation: } \uparrow P][\text{fruiting bodies} = \text{present}] \\
&\quad [\text{stem cankers} = \text{above second node}][\text{fruit pods} = \text{absent}]) \\
&\qquad\qquad\qquad\qquad + \\
&q_c \cdot ([\text{temperature} \geq \text{n}][\text{canker lesion color} = \text{brown}] \\
&\quad [\#\text{ years crop repeated: } \uparrow YR_1]) \\
&\qquad\qquad :: > [\text{Diagnosis} = \textit{Diaporthe stem canker}]
\end{aligned}$$

The weight assigning function YR_1 is defined as in Fig. 1, and the function[2] P:

$$P: w = \begin{cases} 1.0, & \text{if precipitation} > \text{n} \\ 0.7, & \text{if precipitation} = \text{n} \\ 0.3, & \text{otherwise} \end{cases} \qquad (\text{n} = \text{normal})$$

The complete listing of the expert-derived decision rules and the weight assigning functions is given in Appendix A. The interpretation of the rules is delayed until Sec. 6, where expert derived and inductively derived rules are compared.

5. INDUCTIVELY DERIVED RULES ("LEARNING FROM EXAMPLES")

5.1. The Framework for the Experiment

It is obvious that if an expert could provide a complete, precise, and definite characterization of a disease, then it would be the simplest way of introducing knowledge into a consulting system. Practice shows, however, that the above premise is rarely satisfied, and that the formalization of expert knowledge is a tedious, time consuming and error prone effort. This process

[2]Function P can be defined as a DVL_1 expression:

$$1[\text{precipitation} > \text{n}] \vee 0.7[\text{precipitation} = \text{n}] \vee 0.3$$

where the concatenation is interpreted as the minimum function and disjunction as the maximum function (Michalski[11]).

Table II. An Example of Learning Event (Completed Questionnaire Describing a Diseased Plant)

- Environmental descriptors
  - Time of occurrence = July
  - Plant stand = normal
  - Precipitation = above normal
  - Temperature = normal
  - Occurrence of hail = no
  - Number years crop repeated = 4
  - Damaged area = whole fields
- Plant global descriptors
  - Severity = potentially severe
  - Seed treatment = none
  - Seed germination = less than 80%
  - Plant height = normal
- Plant local descriptors
  - Condition of leaves = abnormal
    - Leafspots—halos = without yellow halos
    - Leafspots—margin = without watersoaked margin
    - Leafspot size = greater than $\frac{1}{8}''$
    - Leaf shredding or shot holding = present
    - Leaf malformation = absent
    - Leaf mildew growth = absent
  - Condition of stem = abnormal
    - Presence of lodging = no
    - Stem cankers = above the second node
    - Canker lesion color = brown
    - Fruiting bodies on stem = present
    - External decay = absent
    - Mycelium on stem = absent
    - Internal discoloration of stem = none
    - Sclerotia—internal or external = absent
  - Conditions of fruits—pods = normal
    - Fruit spots = absent
  - Condition of seed = normal
    - Mold growth = absent
    - Seed discoloration = absent
    - Seed size = normal
    - Seed shriveling = absent
  - Condition of roots = normal

Diagnosis:

Diaporthe stem canker( ) *Charcoal rot*( ) *Rhizoctonia root rot*( ) *Phytophthora root rot*( ) *Brown stem root rot*( ) *Powdery mildew*( ) *Downy mildew*( ) <u>*Brown spot*(×)</u> *Bacterial blight*( ) *Bacterial pustule*( ) *Purpose seed stain*( ) *Anthracnose*( ) *Phyllosticta leaf spot*( ) *Alternaria leaf spot*( ) *Frog eye leaf spot*( )

could be simplified if the experts specified directly only the set of descriptors to be used (the description space), the semantic relationships among the descriptors, and examples (expressed in terms of these descriptions) of correct decisions in various situations (which they can easily provide using their standard professional skills). This method requires application of an inductive learning program able to formulate general rules on the basis of exemplary decisions. Such inductive learning is generally a problem of immense complexity. Fortunately, the relative simplicity of the rules for soybean diagnosis permits the current state of the art in computer induction to offer solutions.

To generate diagnostic rules from examples in this experiment we used the inductive program AQ11 (Michalski and Larson[15]). The examples of expert decisions were represented as pairs (*events, disease*), where the *event* was a list of values of 35 descriptors (Table I) characterizing a single diseased plant, and *disease* was the diagnosis provided by an expert. The events were specified in the form of questionnaires completed by plant pathologists. Table II is an example of a completed questionnaire that describes a case of *brown spot.* All available events (630) were partitioned into a learning set and testing set (Table III).

The learning events were used for the actual development of the rules; testing events were used to test the inductively developed rules and also the expert derived rules. In addition to learning events, the input to AQ11 also included some underlying problem knowledge. This knowledge consisted of the following:

1. Definitions of the domains of the descriptors used (the sizes of the domains and their structure). Three types of descriptors were distinguished: *nominal* (whose domain has no structure), *linear* (whose domain is a linearly ordered set), and *structured* (whose domain has values representing concepts of different degree of generality). Domains of structured descriptors can be represented as labeled directed graphs (usually trees) in which the parent node represents a concept that is more general than concepts represented by descendent nodes (Michalski and Larson[15]). Two structured descriptors were used: DAMAGED AREA and LEAFSPOTS—HALOS (Fig. 3).

2. A description of relationships among variables. They were relations stating that if some part of a plant is healthy then all the descriptors that specify the particular conditions of that part do not apply. For example,

$$
\begin{aligned}
[\text{leaves} = \text{normal}] \Rightarrow\ & [\text{leafspots halos} = *][\text{leafspots margin} = *] \\
& [\text{leafspot size} = *][\text{leaf shredding} = *] \\
& [\text{leaf malformation} = *][\text{leaf mildew growth} = *]
\end{aligned}
$$

where * denotes "does not apply." Table IV gives the rules used.

Table III. Events-Available for Learning and Testing

Disease	Learning events	Testing events	Available events
Diaporthe stem canker	10	10	20
Charcoal rot	10	10	20
Rhizoctonia root rot	10	10	20
Phytophtora root rot	40	48	88
Brown stem rot	20	24	44
Powdery mildew	10	10	20
Downy mildew	10	10	20
Brown spot	40	52	92
Bacterial pustule	10	10	20
Bacterial blight	10	10	20
Purple seed stain	10	10	20
Anthracnose	20	24	44
Phyllosticta leaf spot	10	10	20
Alternaria leaf spot	40	51	91
Frog eye leaf spot	40	51	91
Total	290	340	630

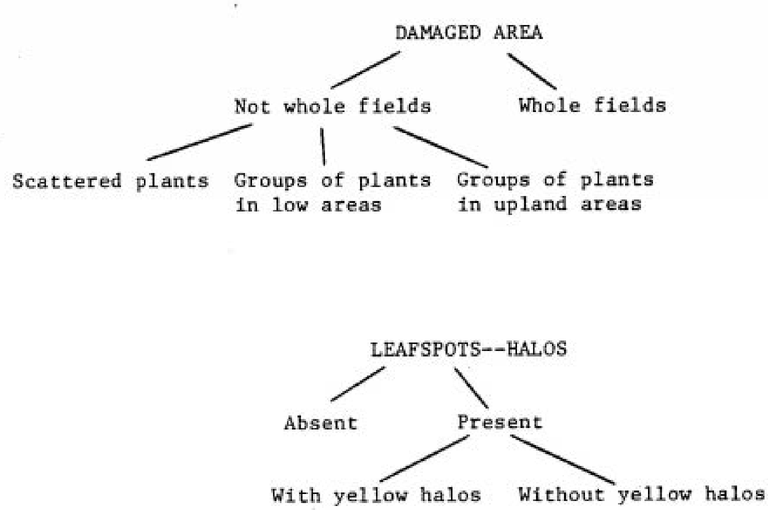

Fig. 3. Generalization structures of structured descriptors DAMAGED AREA and LEAFSPOTS—HALOS.

Table IV. Rules Describing Relationships Among Variables

1. [leaves = normal]⇒	[leafspots halos = *][leafspots margin = *] [leafspot size = *][leaf shredding = *] [leaf malformation = *][leaf mildew growth = *]
2. [leafspots halos = absent]⇒	[leafspots margin = *][leafspot size = *]
3. [stem = normal]⇒	[presence of lodging = *][stem cankers = *] [canker lesion color = *][fruiting bodies on stem = *] [external decay of stem = *][mycelium on stem = *] [internal discoloration = *] [sclerotia internal or external = *]
4. [fruit pods = normal]⇒	[fruit spots = *]
5. [seed = normal]⇒	[seed mold growth = *][seed discoloration = *] [seed size = *][seed shriveling = *]

5.2. A Brief Description of the Learning Algorithm

Given examples and the underlying problem knowledge, program AQ11 determines the DVL_1 rules characterizing the decision process. The rules are optimized according to user selected criteria, which require the rules to have the minimum number of selectors, the minimum number of descriptors, the minimum number of terms, etc., or a combination of these. The program can work in either unistep or multistep (incremental) modes of learning. Each step of the incremental learning mode starts with an initial set of rules (hypotheses), and a set of learning events. Some events may contradict the rules, e.g., an event of decision class i (here disease i) may not satisfy the rule for decision i, but satisfy the rule for decision j, or may not satisfy any rule. Execution of the step produces a new set of rules that are consistent with the events. If there are no initial rules, the step generates rules describing the events. The basic block of the learning algorithm is a *covering algorithm*, which generates a *cover* $C(E_1/E_0)$ *of one event set* (E_1) *against another event set* (E_0). Such a cover can be interpreted as a DVL_1 expression that is satisfied by every event in E_1 and not satisfied by any event in E_0 (or in $E_0\backslash E_1$, if E_0 and E_1 intersect). The covering algorithm is based on the effective use of "negative events" (i.e., those in E_0), and is especially efficient when the negative events are expressed as a cover. The major idea behind the covering algorithm is to generate the cover in steps, each step producing one conjunctive term of the cover. Each term is determined by focusing attention on one specially selected event from E_1, generating a set of all terms (a *star*) which cover this event and do not cover any event in E_0, and then selecting the "best term" from the star according to the assumed criteria. The covering algorithm (called A^q) is described in detail in (Michalski[10,12]).

Let us go back now to the learning algorithm. Suppose that a set of initial hypotheses is $V=\{V_i\}$, $i=1,\ldots,m$, and a family of event sets ("facts") which these hypotheses are supposed to describe is $F=\{F_i\}$. Suppose that for any i, V_i describes correctly only a part of the events from F_i.

The problem is to produce a new set of hypotheses, $V^1=\{V_i^1\}$, where each V_i^1 describes all events from set F_i, and does not describe events from other event sets F_j, $j\neq i$.

The solution consists of four major steps:

Step 1. The first step isolates those facts which are not consistent with the given hypotheses. For each hypothesis, two sets are created: F^+, a set of events which should be covered by the hypothesis, but are not; and F^-, a set of events which are covered by the hypothesis, but should not be covered. (An event is said to be *covered* by a hypothesis if the event satisfies the DVL_1 formula representing the hypothesis.) Specifically, this step determines, for each i, $i=1,2,\ldots,m$, the sets of "exceptions"[3]:

$$F_i^+ = F_i \setminus \tilde{V}_i$$

$$F_{ij}^- = \tilde{V}_i \cap F_j, \qquad j=1,2,\ldots,m; \quad j\neq i$$

Thus, F_i^+ denotes events which should be covered by V_i but are not, and F_{ij}^- denotes events F_j, $j\neq i$, which are covered by V_i, but should not be covered (Fig. 4).

Step 2. This step determines, for each i, a generalized formula V_i^- describing all exception events (the union of sets F_{ij}^-, $j=1,2,\ldots,m$, $j\neq i$). This is done by generating, for given i and each j, a cover of F_{ij}^- against the events in the sets $(\tilde{V}_k \cup F_k^+)$, $k=1,2,\ldots,m$:

$$V_{ij}^- = C\left(F_{ij}^- \Big/ \bigcup_{k=1}^{m} (\tilde{V}_k \cup F_k^+)\right)$$

and then taking the logical union of V_{ij}^-:

$$V_i^- = \bigvee_{\substack{j=1\\ j\neq i}}^{m} V_{ij}^-.$$

The reason for this step is that it is computationally more efficient to use formulas V_i^- than the union of F_{ij}^-, $i=1,2,\ldots,m$, $j\neq i$.

Step 3. New "correct" hypotheses are obtained by "subtracting" from each V_i the formula V_i^- and "adding" to it the set F_i^+. To do this

[3] $\tilde{V}_i$ denotes the set of events covered for formula V_i.

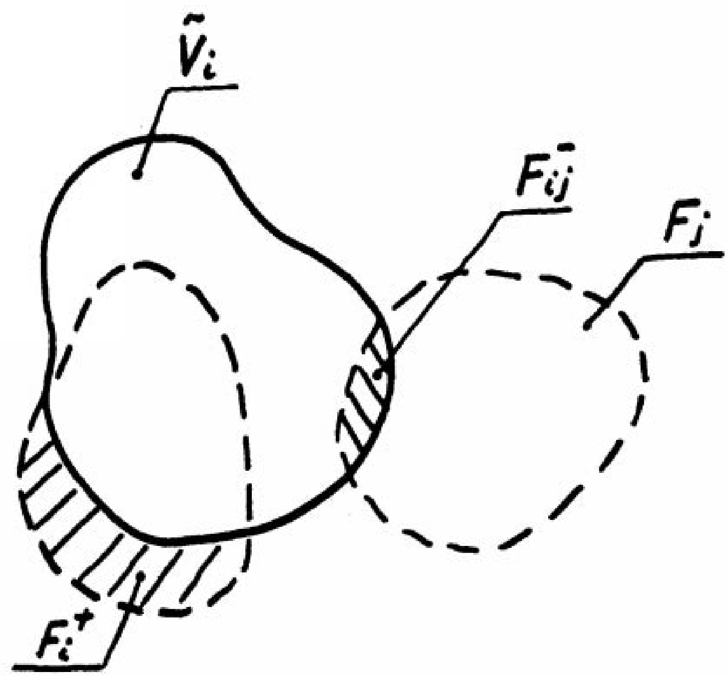

Fig. 4. An illustration of sets of "exceptions" F_i^+ and F_{ij}^-.

directly, however, is difficult. Again, an advantage is taken of the efficiency of the covering algorithm. (The point is that directly simplifying a union of terms is difficult; but generating a cover of an event set against a formula is easier.) The new hypotheses, denoted V_i^1, $i = 1, 2, \ldots, m$, are thus determined as covers:

$$V_i^1 = C\left(F_i \Big/ \bigcup_{\substack{k=1 \\ k \neq i}}^{m} [(\tilde{V}_k \backslash \tilde{V}_k^-) \cup F_k^+]\right)$$

Step 4. Rules representing the underlying problem knowledge (the rules describing relationships among variables, rules for generalization by *climbing the generalization trees* of structured variables, and some "cleaning procedures") are applied.

The obtained rules V_i^1, $i = 1, 2, \ldots, m$, become new hypotheses. If the sets of facts F_i are now enlarged with new facts contradicting these hypotheses, a new learning step is repeated.

5.3. The Inductively Derived Rules

Program AQ11 produces decision rules in the form of disjunctive VL_1 rules, in which each selector has weight $w = 1$. The following is an example of an inductively derived decision rule describing *Phytophthora root rot*:

NEW IND TOTAL

[plant stand < n][precipitation ≥ n]
[temperature ≤ n][stem = abn]
[plant height = abn][leaves = abn]
[leaf malformation = abs] (24, 6, 24)

∨

[time = April . . . Aug][plant stand = abn]
[damaged area = low areas][plant height = abn]
[leaves = abn][stem = abn]
[external decay ≠ firm & dry] (16, 16, 34)
:: > [Diagnosis = *Phytophthora root rot*]

The triplet of numbers next to each term in the rule indicates the performance of that term in covering the learning events. The first element of the triplet, labeled NEW, indicates the number of new events covered by this term (those which were not covered by previously generated terms; for the first term in the rule, NEW is equal to the total number of events covered (TOTAL)); the second, labeled IND ("independent"), is the number of events that only this term covers; the third, labeled TOTAL, is the number of events that this term covers totally. This triplet provides information about the relative importance of each term in a given decision rule. The complete set of inductively derived decision rules is given in Appendix B. (AQ11, written in PL/1, took approximately 4 minutes and 30 seconds on an IBM 360/75 to generate all the rules.)

In the experiment, program ESEL (Michalski and Larson[15]) was used to select the learning events from the available population of events. This program selects the most representative events from each disease class using a "distance" measuring technique. Namely, it selects events that are most distant from each other in the class. This method of selecting the learning events biased the testing set in some sense since the testing events were those that were not selected by the ESEL. To eliminate this bias one could acquire a distinct set of testing events or select learning events totally randomly. The point of this study was, however, not to test the learning method using a teacher that randomly selects examples, but a "good" teacher that selects representative learning examples. The program ESEL was such a teacher.

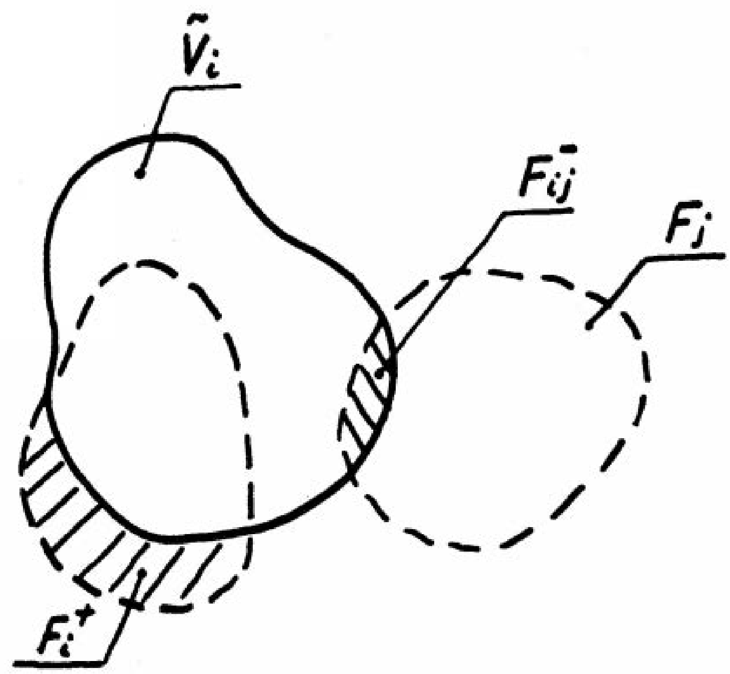

Fig. 4. An illustration of sets of "exceptions" F_i^+ and F_{ij}^-.

directly, however, is difficult. Again, an advantage is taken of the efficiency of the covering algorithm. (The point is that directly simplifying a union of terms is difficult; but generating a cover of an event set against a formula is easier.) The new hypotheses, denoted V_i^1, $i = 1, 2, \ldots, m$, are thus determined as covers:

$$V_i^1 = C\left(F_i \Big/ \bigcup_{\substack{k=1 \\ k \neq i}}^{m} [(\tilde{V}_k \backslash \tilde{V}_k^-) \cup F_k^+]\right)$$

Step 4. Rules representing the underlying problem knowledge (the rules describing relationships among variables, rules for generalization by *climbing the generalization trees* of structured variables, and some "cleaning procedures") are applied.

The obtained rules V_i^1, $i = 1, 2, \ldots, m$, become new hypotheses. If the sets of facts F_i are now enlarged with new facts contradicting these hypotheses, a new learning step is repeated.

6. COMPARISON OF EXPERT DERIVED AND INDUCTIVELY DERIVED RULES

6.1. Comparison of Forms of the Rules

Expert derived and inductively derived rules clearly differ in form. Each expert derived rule consists of two extended terms representing conditions with different strength to support the decision (significant and confirmatory conditions). Some of the selectors have a weight assigning function. Some rules use conditional terms. Noticeably, rules do not include disjunction: experts characterized each disease by only one sequence of characteristics; no subcategories within any disease were indicated.

In contrast to the above, the inductively derived rules have a simple structure of a disjunction of conjunctive terms. The necessity of using disjunction in each disease description (in order to completely characterize all given cases of the disease) is an interesting property of the rules. In some cases it was due merely to the limitation of the program. A case in point is rule D3 (see Appendix B) where one term has a selector [leaves = normal], and another has a selector [leaf malformation = absent]. If the program could recognize that the second selector is the special case of the first, then it would merge the two terms into one. The current program does not also have the ability to create conditional terms, so none are present, although they might have been useful.

In addition it seems that if the program had appropriate *constructive generalization rules* (Michalski[13]), i.e., an ability to generate new, more general descriptors (as functions of the given ones), some rules could have been simpler. It is interesting to notice, however, that despite these limitations, the inductively derived rules, when presented to an expert, were judged quite favorably and in only a few cases did the expert indicate irrelevancy or lack of some selector, or low significance of a term (where several terms were present).

6.2. Comparison of the Performance of the Rules on Testing Data

The final stage of the experiment involved a comparison of the performance of both types of rules in diagnosing diseased plants, for which the correct diagnosis was known (the "correct diagnosis" means the diagnosis made by a plant pathologist).

In order to use the rules for diagnosis, an *evaluation scheme* must be specified, which gives an interpretation to the selector and the terms, specifies the coefficients q_i in linear expressions, and interprets the disjunction operator. Several evaluation schemes were tried experimentally (Michalski[14]). Here we will present results obtained by applying the best performing scheme for each type of rules.

Evaluation scheme for expert derived rules
(P, AVG, (0.9, 0.1), N/A)

- Evaluation of selectors (P):
  Unweighted selectors $[x_i \neq R]$

$$v(S) = \begin{cases} 1, & \text{if the selector is satisfied} \\ 0, & \text{otherwise} \end{cases}$$

  Weighted selectors $[x_i : @f_n]$:

$$v(S^w) = \text{ value of } f_n \text{ for given value of } x_i$$

- Evaluation of terms (the average function):

$$v(T) = \sum_{i \in I} v(CS_i)/|I|$$

  $|I|$ = the number of conditional statements CS_i in the extended term (or selectors in the conjunctive term)

- Evaluation of linear expressions:
  The condition part of each rule is in the form

$$q_s \cdot ET_s + q_c \cdot ET_c$$

  where ET_s and ET_c are extended terms corresponding to *significant* and *confirmatory* conditions, respectively (in two rules ET_c was empty). The coefficients were $q_s = 0.9$ and $q_c = 0.1$.

- Evaluation of disjunction—not applicable.

Evaluation scheme for inductively derived rules
(N, AVG, N/A, PSum)

- Evaluation of selectors (N):

$$v(s) = \begin{cases} 1, & \text{if satisfied} \\ -1, & \text{otherwise} \end{cases}$$

- Evaluation of terms (the average of degrees of confirmation of selectors).
- Evaluation of linear expressions—not applicable.
- Evaluation of disjunction—as the probabilistic sum (PSum):
  For $F = T_1 \vee T_2$

$$v(F) = v(T_1) + v(T_2) - v(T_1)v(T_2)$$

  When there are more terms, the rule is multiply applied.

Using the above evaluation schemes, the expert derived and inductively derived rules were applied to 340 testing cases of diseased plants with known diagnosis (testing events).

A rule $C :: > D$ is said to pass the *acceptability criterion* for an event e, if the degree of confirmation $v(C, e)$ is greater than the *threshold*, and is either the maximum of those of other rules, or differs from the maximum by not more than the *margin of uncertainty*. The *threshold* (determined experimentally) was 0.65 for expert derived rules and 0.8 for inductively derived rules. The *margin of uncertainty* was 0.2 in both cases. If there was more than one rule that passed the *acceptability criterion*, the diagnoses indicated by these rules were *alternative diagnoses*.

Tables V and VI present confusion matrices illustrating the performance of both types of rules on testing events. The labels are defined as follows:

Correct diagnosis: The diagnosis assigned by a plant pathologist.

Indecision ratio: The average number of alternative diagnoses per case of the given disease.

Ties: The number of testing cases of the disease which were not uniquely diagnosed.

Maximum # of altern: The maximum number of alternative diagnoses in diagnosing testing cases of the disease.

Test cases: The total number of testing cases of the given disease.

Assigned decision: Each column under this label is headed by a symbolic name of disease (D*i*), and contains percentages of diagnoses indicating this disease for all testing cases of the disease associated with the given row. Thus, the percent of correctly assigned diagnoses are on the diagonal of confusion matrix. Table VII gives a comparison of the overall performance of the two sets of rules.

The label "% 1st choice correct" specifies the percentage of cases in which the rule with the highest degree of confirmation indicated the correct disease. The label "% correct" specifies the percentage of cases in which a single correct diagnosis was assigned, or a set of alternative diagnoses was assigned that included the correct one. Both inductively derived and expert derived rules scored high and approximately even from the viewpoint of avoiding errors (% correct), but the inductively derived rules clearly performed better than expert derived rules in terms of frequency with which the "1st choice diagnosis" was correct. In addition, inductively derived rules had a smaller *indecision ratio* and larger *threshold*. This would seem to indicate that these rules were "cleaner" than expert derived rules, i.e., they involved less nonessential information.

Table V. Confusion Matrix Summarizing the Diagnosis of 340 Testing Events Using Expert-Derived Rules

Correct diagnosis	Indecision ratio	Ties	Maximum # of altern	Test cases	Assigned decision														
					D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
Diaporthe stem canker (D1)	1.8	7	3	10	100				40							40			
Charcoal rot (D2)	1.0	0	1	10		100													
Rhizoctonia root rot (D3)	0.9	0	1	10			90												
Phytophthora root rot (D4)	1.4	18	2	48		27	8	100	2										
Brown stem rot (D5)	0.96	2	3	24					87			4						4	
Powdery mildew (D6)	1.0	0	1	10						100									
Downy mildew (D7)	3.4	10	5	10						.	80	100	30				30	70	30
Septoria brown spot (D8)	4.9	52	8	52	37						40	100		38		37	90	44	100
Bacterial blight (D9)	2.7	9	4	10							50		100	90			30		
Bacterial pustule (D10)	3.2	9	5	10							10	70	50	100	30		30	20	10
Purple seed stain (D11)	2.1	8	5	10					20		10		10		80			60	30
Anthracnose (D12)	2.1	21	4	24	50			4	4						54	96			
Phyllosticta leaf spot (D13)	4.1	10	6	10							20	100		50			90	80	70
Alternaria leaf spot (D14)	3.1	51	5	51							39	100	20				8	94	69
Frog eye leaf spot (D15)	4.2	51	6	51				4	39		63	100				4	6	100	100

Table VI. Confusion Matrix Summarizing the Diagnosis of 340 Testing Events Using Inductively Derived Rules

Correct diagnosis	Indecision ratio	Ties	Maximum # of altern	Test cases	Assigned decision														
					D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
Diaporthe stem canker (D1)	2.7	10	3	10	100											100			70
Charcoal rot (D2)	1.0	0	1	10		100													
Rhizoctonia root rot (D3)	2.0	10	2	10			100	100											
Phytophthora root rot (D4)	1.0	0	1	48				100											
Brown stem rot (D5)	1.3	3	5	24				8	100			4						8	8
Powdery mildew (D6)	1.0	0	1	10						100									
Downy mildew (D7)	4.1	10	5	10							100	90					30	90	100
Septoria brown spot (D8)	4.0	52	5	52				10				100					88	100	100
Bacterial blight (D9)	3.2	10	4	10									100	50			10	80	80
Bacterial pustule (D10)	1.6	4	4	10								20		100	10		30		
Purple seed stain (D11)	2.8	7	4	10								40	10		100	10		60	60
Anthracnose (D12)	1.1	2	3	24				8								100			4
Phyllosticta leaf spot (D13)	3.9	10	4	10								100					100	100	90
Alternia leaf spot (D14)	3.2	51	4	51								100					22	100	100
Frog eye leaf spot (D15)	3.9	51	5	51				2				100				82	4	100	100

Table VII. The Summary of Performance of the Rules

Type	% 1st choice correct	% correct	% not diagnosed	Indecision ratio	Threshold
Inductively derived	97.6	100.0	—	2.64	0.80
Expert derived	71.8	96.9	2.1	2.90	0.65

7. CONCLUSION

The paper discussed the relative merits of knowledge acquisition through interviewing experts and formally representing their decision rules, and through inductively inferring the rules from examples of these experts' decisions. The acquisition of knowledge through representation is a commonly used and powerful method, but is also time consuming, prone to omissions, and requires high technical skills from the knowledge engineer encoding the rules (who has to be able to consolidate sometimes conflicting opinions of experts). The inductive method presented in the paper required less effort and produced decision rules whose overall performance was somewhat better than expert derived rules. The latter result was contrary to the expectations of the authors, and therefore the experiment was repeated several times, introducing modifications to expert derived rules and trying different rule evaluation schemes. The results have consistently followed the same pattern. A reason for the somewhat poorer performance of expert derived rules may be the insufficiently precise encoding of the decision rules of experts. It is likely that further interaction with experts and a refinement of the knowledge representation method will lead to better rules. Another possible reason is that experts are trained in making diagnoses, and not in explaining the process of diagnosis. These two functions are different. This would mean that examples of expert decisions represent more reliable information than experts' descriptions of the diagnostic procedures. This would provide an additional argument for the inductive method of knowledge acquisition.

Despite the high performance of the inductively derived rules, they are by no means completely satisfactory. Some of the limitations were indicated in Sec. 6.1. The current program does not have sufficient ability to use the knowledge of the relationships among descriptors, and therefore it generated in some cases too many conjunctive terms. Also it possesses a very limited repertoire of syntactic forms in which it can express rules. The program does not have the ability to generate a layered system of rules, with intermediate variables. In the problem under consideration, there was no need for such facility, but more advanced inductive problems will require it.

Surprisingly, the inductively derived rules were viewed generally quite favorably by experts—with a few exceptions. This observation and previous remarks suggest that a procedure in which an expert would edit inductively derived rules, in combination with an improved inductive program, could lead to an attractive new method of knowledge acquisition.

The major conclusion of the paper is that the inductive method for introducing knowledge to expert systems can be both useful and practical, if the problem domain is sufficiently simple. The currently growing interest in computer induction will likely widen the scope of problems to which the inductive method is applicable.

APPENDIX A: EXPERT DERIVED RULES FOR 15 SOYBEAN DISEASES

q_s: the coefficient associated with significant conditions; q_c: the coefficient associated with confirmatory conditions. Abbreviations used: n, normal; abn, abnormal; p, present; abs, absent.

D1: $q_s \cdot$ ([time = Aug . . . Sep][precipitation: $\uparrow P$])
[stem cankers = above second node][fruiting bodies = p]
[fruit pods = n])
+
$q_c \cdot$ ([temperature $\geq$ n][canker lesion color = brown]
[# years crop repeated: $\uparrow YR_1$]
::> [Diagnosis = *Diaporthe stem canker*]

D2: $q_s \cdot$ ([time = Jul . . . Aug][precipitation $\leq$ n][temperature $\geq$ n]
[plant growth = abn][leaves = abn][stem = abn][sclerotia = p]
[roots = rotted][internal discoloration = black])
+
$q_c \cdot$ ([damaged area = upland areas][severity = severe][seed size < n]
[# years crop repeated: $\uparrow YR_2$]
::> [Diagnosis = *Charcoal rot*]

D3: $q_s \cdot$ ([time = May . . . Jun][plant stand < n][temperature < n]
[precipitation < n][leaves = abn][stem = abn]
[canker lesion color = brown][roots = rotted]
([occurrence of hail = no] $\Rightarrow$ [stem cankers = below soil line, at or slightly above soil line])
([occurrence of hail = yes] $\Rightarrow$ [stem cankers = above second node]))
+
$q_c \cdot$ ([fruiting bodies = abs][external decay = firm & dry]
[mycelium = abs])
::> [Diagnosis = *Rhizoctonia root rot*]
+

D4: $q_s \cdot$ ([time: $\cap\, TO$][plant stand < n]
([time = Apr . . . Jun] ⇒ [precipitation = n])
([time = Jul . . . Aug] ⇒ [precipitation = above n])
([time = Apr] ⇒ [temperature = above n])
([time = May . . . Aug] ⇒ [temperature = n])
[damaged areas = low areas]
[plant growth = abn][leaves = abn][stem = abn]
[stem cankers = at or slightly above soil line]
([time = May . . . Aug] ⇒ [canker lesion color = dark brown or black])
[roots = rotted])
+
$q_c \cdot$ ([# years crop repeated ≥ 2])
:: > [Diagnosis = *Phytophthora root rot*]

D5: $q_s \cdot$ ([time = Jul . . . Sep][precipitation > n][temperature ≤ n]
[leaves = abn][stem = abn]
[internal discoloration = brown][lodging = p])
+
$q_c \cdot$ ([seed size < n][# years crop repeated: ↑ YR_3]
:: > [Diagnosis = *Brown stem rot*]

D6: $q_s \cdot$ ([leaves = abn][leaf mildew growth = upper leaf surface])
+
$q_c \cdot$ [time = Aug . . . Sep]
:: > [Diagnosis = *Powdery mildew*]

D7: $q_s \cdot$ ([time = Jun . . . Aug][precipitation ≥ n]
[damaged areas = whole fields]
[leaves = abn][leafspots halos = no yellow halos]
[leaf mildew growth = lower leaf surface]
([time = Sep . . . Oct] ⇒ [see = abn])[mold growth on seed = p])
:: > [Diagnosis = *Downy mildew*]

D8 $q_c \cdot$ ([leaves = abn][leafspots halos = p]
[leafspots watersoaked margin = abs][leafspot size > $\frac{1}{8}''$])
+
D9: $q_c \cdot$ ([time = Apr . . . Jun, Aug . . . Sep]
([time = Apr . . . Jun] ⇒ [precipitation ≥ n])
([time = Aug . . . Sep] ⇒ [precipitation > n])
([time ≠ Aug] ⇒ [temperature = n])
([time = Aug] ⇒ [temperature < n][leaves = abn]
[leafspots halos = with yellow halos]
[leafspots watersoaked margin = p]
[leafspot size < $\frac{1}{8}''$][leaf shredding = p])
:: > [Diagnosis = *Bacterial blight*]

D10: $q_s \cdot$ ([time = Jun . . . Aug][precipitation $\geq$ n][leaves = abn]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = abs]
[leafspot size $< \frac{1}{8}''$][leaf shredding = p])
+
$q_c \cdot$ [# years crop repeated $\geq$ 1]
:: > [Diagnosis = *Bacterial pustule*]

D11: $q_s \cdot$ ([time = Sep . . . Oct][seed = abn][seed discoloration = p]
[seed size = smaller than n])
+
$q_c \cdot$ ([time = Aug . . . Sep][precipitation $\geq$ n][leaves = abn]
:: > [Diagnosis = *Purple seed strain*]

D12: $q_s \cdot$ ([time = Aug . . . Oct][precipitation $\geq$ n][stem = abn]
[canker lesion colour = brown][fruiting bodies = p]
([time = Sep . . . Oct] $\Rightarrow$ [seed = abn])
[fruit spots = abs, brown spots with black specks])
+
$q_c \cdot$ [damaged area = whole fields]
:: > [Diagnosis = *Anthracnose*]

D13: $q_s \cdot$ ([time = Apr . . . Jul][precipitation $\geq$ n][leaves = abn]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = abs]
[leafspot size $> \frac{1}{8}''$][leaf shredding = p])
+
$q_c \cdot$ ([damaged area = whole fields][time $\neq$ Jun] $\Rightarrow$ [temperature = n])
([time = Jun] $\Rightarrow$ [temperature < n])
:: > [Diagnosis = *Phyllosticta leaf spot*]

D14: $q_s \cdot$ ([time = Jul . . . Oct][leaves = abn]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = abs][leafspot size $> \frac{1}{8}''$]
[leaf shredding = abs])
+
$q_c \cdot$ (([time = Sep . . . Oct] $\Rightarrow$ [fruit pods = diseased])
([fruit pods = diseased] $\Rightarrow$ [fruit spots = colored spots])
([seed = abn] $\Rightarrow$ [seed discoloration = p]))
:: > [Diagnosis = *Alternaria leaf spot*]

D15: $q_s \cdot$ ([time = Jul . . . Sep][precipitation $\geq$ n][leaves = abn]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = abs]

[leafspot size > $\frac{1}{8}''$])

+

$q_c \cdot$ (([time = Sep] ⇒ [fruit spots = colored spots])
[stem canker = above second node][canker lesion color = tan]
[fruiting bodies = abs])
::> [Diagnosis = *Frog eye leaf spot*]

Definition of weight assigning functions:

$$P: \begin{cases} 1.0, & \text{if precipitation} = \text{above normal} \\ 0.7, & \text{if precipitation} = \text{normal} \\ 0.3, & \text{otherwise} \end{cases}$$

$$YR_1: \begin{cases} 1.0, & \text{if \# years crop repeated} \geq 3 \\ 0.8, & \text{if \# years crop repeated} = 2 \\ 0.7, & \text{if \# years crop repeated} = 1 \\ 0.2, & \text{if crop not repeated} \end{cases}$$

$$YR_2: \begin{cases} 1.0, & \text{if \# years crop repeated} \geq 2 \\ 0.6, & \text{if \# years crop repeated} = 1 \\ 0.2, & \text{if crop not repeated} \end{cases}$$

$$YR_3: \begin{cases} 1.0, & \text{if \# years crop repeated} \geq 2 \\ 0.5, & \text{if \# years crop repeated} = 1 \\ 0.1, & \text{if crop not repeated} \end{cases}$$

$$TO: \begin{cases} 1.0, & \text{if time of occurrence} = \text{May} \ldots \text{Jul} \\ 0.7, & \text{if time of occurrence} = \text{Apr, Aug} \\ 0.3, & \text{otherwise} \end{cases}$$

APPENDIX B: INDUCTIVELY DERIVED RULES FOR 15 SOYBEAN DISEASES

Abbreviations used: n, normal; abn, abnormal; p, present; abs, absent.

NEW IND TOTAL

D1: [time = Jul . . . Oct][precipitation > n]
[leaf malformation = abs]
[stem = abn][stem cankers = above second node]
[external decay = firm & dry][fruit pods = n] (10, 10, 10)
::> [Diagnosis = *Diaporthe stem canker*]

D2: [leaf malformation = abs][stem = abn]
[internal discoloration = black] (10, 10, 10)
::> [Diagnosis = *Charcoal rot*]

D3: [leaves = n][stem = abn][stem cankers = below soil line]
[canker lesion color = brown] (9, 9, 9)
$\vee$
[leaf malformation = abs][stem = abn]
[stem cankers = below soil line]
[canker lesion color = brown] (1, 1, 1)
::> [Diagnosis = *Rhizoctonia root rot*]

D4: [plant stand > n][precipitation $\geq$ n][temperature $\leq$ n]
[plant height = abn][leaves = abn]
[leaf malformation = abs]
[stem = abn] (24, 6, 24)
$\vee$
[time = Apr . . . Aug][plant stand = abn]
[damaged area = low areas]
[plant height = abn][leaves = abn][stem = abn]
[external decay = abs, soft and watery] (16, 16, 16)
::> [Diagnosis = *Phytophthora root rot*]

D5: [leaf malformation = abs][stem = abn]
[internal discoloration = brown] (13, 13, 13)
$\vee$
[leaves = n][stem = abn][internal discoloration = brown]
(7, 7, 7)
::> [Diagnosis = *Brown stem rot*]

D6: [leaves = abn][leaf malformation = abs]
[leaf mildew growth = on upper leaf surface][roots = n] (10, 10, 10)
::> [Diagnosis = *Powdery mildew*]

D7: [leafspots halos = p]
[leaf mildew growth = on lower leaf surface]
[stem = n][seed mold growth = p] (10, 10, 10)
::> [Diagnosis = *Downy mildew*]

D8: [precipitation $\geq$ n][# years crop repeated > 1]
[damaged area $\neq$ whole fields][leaves = abn]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = abs][leafspot size > $\frac{1}{8}''$]
[leaf malformation = abs][roots = n] (19, 2, 19)
$\vee$
[precipitation > n][leaves = abn]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = abs][leafspot size > $\frac{1}{8}''$]

[root = n] (15, 11, 30)

$\vee$

[time = Apr . . . Jun][damaged area ≠ whole fields]
[leaves = abn]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = abs][leafspot size > $\frac{1}{8}''$]
[leaf shredding = abs][leaf malformation = abs]
[roots = n] (6, 6, 12)
:: > [Diagnosis = *Brown spot*]

D9: [time = Jun . . . Sep][temperature ≥ n][leaves = abn]
[leafspots halos = p][leafspots watersoaked margin = p]
[leafspot size < $\frac{1}{8}''$][fruit pods = n][roots = n] (10, 10, 10)
:: > [Diagnosis = *Bacterial blight*]

D10: [leaves = abn][leafspots halos = with yellow halos]
[leafspots watersoaked margin = abs][leafspot size < $\frac{1}{8}''$]
[stem = n][fruit pods = n] (7, 6, 7)

$\vee$

[leafspots halos = p][leafspot size < $\frac{1}{8}''$][stem = n]
[roots = rotted] (2, 2, 2)

$\vee$

[time = May][precipitation = n][leaves = abn]
[leafspots halos = with yellow halos] (1, 1, 2)
:: > [Diagnosis = *Bacterial pustule*]

D11: [plant stand = n][precipitation > n][severity = minor]
[plant height = n][leafspots halos = no yellow halos]
[seed = abn]
[seed discoloration = p][seed size = n] (5, 5, 5)

$\vee$

[leaves = n][seed = abn][seed size = n] (5, 5, 5)
:: > [Diagnosis = *Purple seed stain*]

D12: [precipitation > n][leaf malformation = abs][stem = abn]
[stem cankers = at or slightly above soil line, above second node]
[seed = abn][roots = n] (10, 8, 10)

$\vee$

[time = Aug . . . Oct][precipitation > n][leaves = n]
[stem cankers = above second node]
[fruit pods = diseased]
[fruit spots = brown spots with black specks] (5, 5, 5)

∨

[temperature > n][leafspots halos = abs]
[leaf malformation = abs]
[stem = abn][external decay = firm and dry] (5, 5, 7)
:: > [Diagnosis = *Anthracnose*]

D13: [time = Jun . . . Jul][precipitation ≤ n][severity = minor]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = abs][stem = n]
[roots = n] (6, 5, 6)

∨

[precipitation < n][leaves = abn]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = abs][roots = n] (3, 3, 4)

∨

[plant stand < n][precipitation = n][occurrence of hail = no]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = abs][stem = n]
[roots = n] (1, 1, 1)
:: > [Diagnosis = *Phyllosticta leaf spot*]

D14: [time = Aug][precipitation > n][seed treatment = none]
[leaves = abn][leafspots halos = no yellow halos]
[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}''$]
[leaf mildew growth = abs][stem = n][fruit pods = n] (8, 5, 8)

∨

[time = Sep . . . Oct][precipitation > n]
[damaged area = scattered plants, low areas, whole fields]
[seed germination ≥ 80%][leaves = abn]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}''$]
[stem = n] (13, 4, 13)

∨

[time = Aug . . . Oct]
[damaged area = scattered plants, low areas]
[seed germination < 80%][plant height = n][leaves = abn]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}''$]
[leaf mildew growth = abs][stem = n] (7, 3, 10)

∨

[time = Oct][seed germination < 90%][leaves = abn]
[leafspots halos = no yellow halos]

[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}$"]
[leaf mildew growth = abs][stem = n] (4, 2, 7)

∨

[time = Aug . . . Oct]
[damaged area = upland areas, whole fields]
[seed treatment = none, other][seed germination ≥ 80%]
[leaves = abn][leafspots halos = no yellow halos]
[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}$"]
[leaf mildew growth = abs][stem = n][fruit pods = n] (3, 3, 3)

∨

[occurrence of hail = no][damaged area = scattered plants]
[severity = potentially severe][seed germination ≥ 80%]
[leaves = abn][leafspots halos = no yellow halos]
[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}$"]
[leaf mildew growth = abs][stem = n] (3, 3, 11)

∨

[time = Aug . . . Oct][temperature = n]
[seed treatment = fungicide]
[seed germination = 80–89%][leaves = abn]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}$"]
[leaf mildew growth = abs][stem = n][fruit pods = n] (1, 1, 6)

∨

[time = Sep . . . Oct][leaves = abn]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}$"]
[leaf shredding = p] (1, 1, 1)
:: > [Diagnosis = *Alternaria leaf spot*]

D15: [precipitation ≥ n]
[damaged area = low areas, upland areas, whole fields]
[leaves = abn][leafspots halos = no yellow halos]
[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}$"]
[leaf shredding = abs][leaf mildew growth = abs]
[stem = abn][roots = n]

∨

[time = Jul . . Sep][precipitation ≥ n][temperature = n]
[occurrence of hail = no]
[damaged area = low areas, whole fields]
[seed treatment = fungicide][leaves = abn]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}$"]
[leaf shredding = abs][leaf malformation = abs][roots] (7, 5, 8)

∨

[time = Aug . . . Sep][precipitation ≥ n]
[damaged area = low areas, upland areas][severity = minor]
[leaves = abn][leafspots halos = no yellow halos]
[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}''$]
[leaf shredding = abs][leaf mildew growth = abs]
[seed = n][roots = n] (8, 4, 20)

∨

[time = Jul . . . Aug][precipitation > n]
[# years crop repeated ≥ 1]
[damaged area = scattered plants]
[seed treatment = none, other]
[leaves = abn][leafspots halos = no yellow halos]
[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}''$]
[leaf shredding = abs][leaf mildew growth = abs]
[roots = n] (4, 3, 8)

∨

[precipitation > n][# years crop repeated ≤ 2]
[damaged area = scattered plants, upland areas]
[severity = potentially severe][seed germination < 80%]
[leaves = abn][leafspots halos = no yellow halos]
[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}''$]
[leaf mildew growth = abs][roots = n] (4, 3, 9)

∨

[time = Jul][occurrence of hail = yes][leaves = abn]
[leafspots halos = no yellow halos]
[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}''$]
[leaf mildew growth = abs][stem = n] (2, 2, 4)

∨

[plant stand = n][precipitation ≥ n]
[# years crop repeated = 2]
[leaves = abn][leafspots halos = no yellow halos]
[leafspots watersoaked margin = p][leafspot size > $\frac{1}{8}''$]
[leaf shredding = abs][leaf mildew growth = abs]
[seed = n][roots = n] (2, 2, 5)

::> [Diagnosis = *Frog eye leaf spot*]

ACKNOWLEDGMENTS

The work presented here would not be possible without the collaboration, support, and interest of Professors J. Sinclair and Barry Jackobsen from the Department of Plant Pathology, University of Illinois, to whom the authors gratefully express their indebtedness. Suggestions, insightful criticisms, and discussions with Bob Stepp and A. B. Baskin were amply helpful in shaping the final form of the paper.

The research was supported in part by grant MCS 79-06614 from the National Science Foundation and in part by grant No. 321512344 from the U.S. Department of Agriculture.

REFERENCES

1. A. B. Baskin and A. Levy, "MEDIKAS—An interactive knowledge acquisition system," *Proceedings of the 2nd Symposium on Computer Applications in Medical Care*, Washington, D.C. (November, 5-7, 1978) (IEEE publication).
2. R. J. Brachman, "On the Epistemological Status of Semantic Networks," Report No. 3807, Bolt Beranek and Newman, Inc., Cambridge, Massachusetts (April 1978).
3. B. G. Buchanan and E. A. Feigenbaum, "Dendral and meta-dendral, their applications dimension," *Artif. Intell.* **11**:5-24 (1978).
4. R. Davis, "Application of Meta Level Knowledge to the Construction, Maintenance and Use of Large Knowledge Bases," STAN-CS-76-552, Department of Computer Science, Stanford University, Stanford, California (July 1976).
5. R. Davis, B. Buchanan and E. Shortliffe, "Production Rules as a Representation for a Knowledge-Based Consultation Program," Memo AIM-266, Stanford Artificial Intelligence Laboratory, Stanford, California (October 1975).
6. T. G. Dietterich and R. S. Michalski, "Learning and generalization of characteristic descriptions: Evaluation criteria and comparative review of selected methods," *Proceedings of the Sixth International Joint Conference on Artificial Intelligence*, Tokyo, Japan (August 1979).
7. R. O. Duda *et. al.*, "A Computer-Based Consultant for Mineral Exploration," Final Report, SRI International, Menlo Park, California (September 1979).
8. R. A. Feigenbaum, "The art of artificial intelligence: I. Themes and case studies of knowledge engineering," *Proc. of the 5th IJCAI* (1977), pp. 1014-1029.
9. F. Hayes-Roth and J. McDermott, "An inference matching technique for inducing abstractions," *Comm. ACM* **21**:401-410 (May 1978).
10. R. S. Michalski, "A Geometrical Model for the Synthesis of Internal Covers," Report No. 461, Department of Computer Science, University of Illinois, Urbana, Illinois (June 1971).
11. R. S. Michalski, "Variable-valued logic: System VL_1." *Proceedings of the Fourth International Symposium on Multiple-Valued Logic*, Morgantown, West Virginia (May 1974).
12. R. S. Michalski, "Synthesis of optimal and quasi-optimal variable-valued logic formulas," *Proceedings of the Fifth International Symposium on Multiple-Valued Logic*, Indiana University, Bloomington, Indiana (1975).
13. R. S. Michalski, "Pattern Recognition as Knowledge-Guided Computer Induction," Report No. 927, Department of Computer Science, University of Illinois, Urbana, Illinois (1978).
14. R. S. Michalski, "An experimental comparison of several many-valued logic inference techniques in the context of computer diagnosis of soybean diseases," *Int. J. Man-Machine Studies* (to appear, 1981).
15. R. S. Michalski and J. B. Larson, "Selection of Most Representative Training Examples and Incremental Generation of VL_1 Hypothesis: The Underlying Methodology and the Descriptions of Programs ESEL and AQ11," Report No. 877, Department of Computer Science, University of Illinois, Urbana, Illinois (May 1978).
16. D. Michie, "Knowledge-Based Systems," Report No. 1001, Department of Computer Science, University of Illinois, Urbana, Illinois (January 1980).

17. T. M. Mitchell, "An analysis of generalization as a search problem," *Proceedings of the 6th IJCAI* (August 20–23, 1979), Vol. 1, pp. 577–582.
18. J. D. Myers and H. E. Pople, "INTERNIST: A consultative diagnostic program in internal medicine," *Proceedings of the 1st Annual Symposium on Computer Applications in Medical Care*, IEEE, New York (1977).
19. E. A. Patrick, *Decision Analysis in Medicine*, (CRC Press, West Palm Beach, Florida, 1979).
20. E. H. Shortliffe, *Computer-Based Medical Consultations: MYCIN* (Elsevier North-Holland, New York, 1976).
21. S. M. Weiss, C. A. Kulikowski, S. Amarel, and A. Safir, "A model-based method for computer-aided medical decision-making," *Artif. Intell.* **11**(12):145–172 (1978).

J. W. Arrowsmith Ltd

Printers
Winterstoke Road
Bristol BS3 2NT

Telephone
Bristol 667545