ISG 1979-1

A PROTOTYPE COMPUTER BASED CONSULTING SYSTEM USING THE VARIABLE VALUED LOGIC SYSTEM VL1: METHODOLOGY AND IMPLEMENTATION

BY

RICHARD LEE CHILAUSKY

B.A., Wabash College, 1967

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Computer Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1979

Urbana, Illinois

ACKNOWLEDGEMENTS

I wish to express my appreciation to Professor R. S. Michalski for his assistance in the areas of variable-valued logic and deductive processing and for actively guiding my work throughout the period of research. Also, I would like to thank Dr. James Larson for his assistance and support during the early stages of my research, the Computing Services Office for an atmosphere conducive to this work, and Debbie Weller for her editorial assistance in the preparation of this thesis.

Finally, thank you Lore, Krista and Amy Beth.

TABLE OF CONTENTS

1. INTRODUCTION

In many areas of expertise, there is a growing interest in computer consulting systems which incorporate knowledge bases on complex subjects. These consulting systems support the decision-making process in a number of ways. Persons who require information or assistance in areas where they do not personally have the necessary background can obtain it from these systems. The systems can be used as an aid to the memory by calling attention to information which might be overlooked. Such systems can be used to shorten decision-making processing by suggesting likely areas of investigation. A number of computer consulting systems have been developed in an effort to provide timely and accurate information and assistance in the medical field:

- DIALOG for general medical diagnosis [Pople 75].
- MYCIN for antimicrobial therapy advice [Shortliffe 74].
- CASNET for disease modelling [Kulikowski 72, 73].

However, consulting systems have not been limited to the field of medicine and interesting work has been done in a number of applications, for example:

- PROSPECTOR for geological predictions [Duda, Hart, Nilson 77].
- Program for assistance with machine repair [Hart 75].
- TELEPLAN includes diagnostic assistance for plant diseases [Harsh 75].

In the development of such a consulting system, it is necessary to use formal structures which adequately represent the knowledge necessary to the consulting area, to acquire the knowledge for the area and translate it into the representation and to provide a processor which can use this information so that advice may be provided concerning individual cases/examples in the area.

In the systems mentioned, the structures provided for knowledge representation take the form of rules The deductive processor, hence DP_1, represents knowlege in the form of logical decision rules expressed in terms of VL_1, a variable-valued propositional logic system [Michalski 72].

Two basic techniques may be used for developing rule based knowledge representation:

- Derivation of rules from expert sources.
- Derivation of rules from examples through an induction process.

When deriving rules based on expert knowledge, information is gathered from specialists or from books and publications in the area in question, and represented in a formal structure. making process of an expert. When deriving rules automatically from examples, a computer program formulates rules inductively from sets of examples. which the information for each set is known a priori. The sets of events have a basis of expert knowledge, i.e. the predefined description space must be adequate to provide for complete and accurate representation of each event and the initial classification of the events must be correct.

The systems mentioned implement knowledge of their subject exclusively from information obtained from specialists or authoritative material. In DP1, knowledge of the subject is implemented using rule sets; developed from information gathered from experts and inductively from examples.

Once the basis and methods for expressing the knowledge base have been selected and the knowledge base has been developed, the next step is to prepare a processor which will deductively apply the knowledge base to specific examples to provide the end product, i.e. information or assistance. To meet this objective, the processor must provide.

- Adequate interface between the user and the processor.
- Suitable deductive processing which includes the necessary evaluation schemes.

The interface between the user and the processor must provide a channel for bi-directional flow of information, i.e. sufficient information to evaluate the data base in terms of a specific event must be obtained for the processor and sufficient information must be presented to the user to allow proper interaction with the processor. Although significant achievements have been made in natural language processing [Winograd 73

Schank 73], the development of such a system for use within a consultant rivals the development of the remainder of the consultant. The processors above use a restricted natural language setting with a core request-response mechanism. The lack of an established methodology for natural language processing prompted the selection of request-response mechanism for DP1. As in MYCIN and PROSPECTOR, additional assistance and information may be obtained from the processor.

Suitable decision processing must be available for the evaluation of the knowledge base which is being considered. As may be observed from the consultants above, the processing schemes are all different, yet each appears to provide a viable, though experimental, alternative to expert consulting. The evaluation methods vary from deductive heuristics to probabilistic and logical based schemes. It is proposed that one reason that these varied schemes are successful is that the areas selected for consulting systems are extremely complex. Therefore, the evaluation methods reflect not only the knowledge gathered from the expert, but also the expert's "artistic feel" for the particular area. The approach taken during the development of the consultant described in this paper was to maintain as much flexibility as possible in the area of rule processing. This approach was used due to the multiplicity of ways which have been proposed for providing semantic evaluations in multiple-valued logic [Reacher 69]. This approach was valid since DP1 makes use of two types of rules; one (inductive) expressed completely in terms of VL_1, and one (expert) expressed, due to its added complexity, in terms of an extended VL_1. Experience shows that the evaluation scheme which is "best" for one rule type is not necessarily the "best" for the other. The desire for flexibility also prompted the use of a menu format for selecting the particular area and topic desired for consulting assistance and a general program structure which would make it possible to incorporate other successful consulting schemes, if desired. This structure provides the user the ability to select the consulting area and the developer to tailor the components of the processor to provide the best results possible.

The initial application of the deductive processor was to a subset of the diseases of soybeans. Specific examples of the consultants will be described in terms of this application. The remainder of this paper is organized as follows:

- Chapter 2 describes the development of the knowledge base.
- Chapter 3 describes the deductive processing mechanisms which are used for DP1.
- Chapter 4 presents a sample session with DP1 which illustrates features which are embodied in the processor.
- Chapter 5 reflects upon the current work and presents areas for expansion.

2. THE KNOWLEDGE BASE

The knowledge base developed for the pilot project of this consultant contains sufficient information for the diagnosis of a subset of the disease of soybeans. The knowledge base contains the following major members:

- Table of descriptors
- Relational rules for descriptors
- Inductively derived decision rules
- Expert derived decision rules

The table of descriptors provides the natural language specification of each variable in the description space and the values which each may take. Each set of rules is expressed via the formal structure supplied by the variable-valued logic system VL_1 although the form of each set is different. Also, all of the rules are expressed using the description space presented below.

2.1 Table of Descriptors

The table of descriptors forms the core of the knowledge base since it specifies the working language of the deductive processor. The variable space for diagnosing the selected soybean diseases was developed by Barry Jacobsen* and Ryszard Michalski**. The variables provided consisted of 35 plant and environmental descriptors and one decision variable. The intent of the particular descriptors selected and their associated values was to provide a description space which was sufficient to describe the diseases of soybeans in terms of macro-symptoms, i.e. those symptoms which could be clearly observed with no sophisticated mechanical assistance. Additionally, the descriptors were prepared with the intention that a layman, e.g. Extension Service Field Agent or farmer, could make reliable observations.

An individual plant can be described by a set of these descriptors. Each descriptor is a function which assigns a specific value from the set called the domain of the descriptor to the plant or its environment. Time of Occurrence (TOC) and Condition of Roots (COR) are examples of a descriptor. The domains of these descriptors for this knowledge base are

D(TOC) = (April, May, June, July, August, September, October)

D(COR) = (Normal, Rotted, Galls or Cysts Present)

---

* Extension Service Pathologist, State of Illinois and Assistant Professor of Plant Pathology, University of Illinois.

** Associate Professor of Computer Science, University of Illinois.

Table 1
Plant Descriptors Used in the Experiment

			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		Cropping History	(4)	(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 Growth	(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	Presence of Fruiting Bodies	(2)	(x_{23})
	3.2.5	External Decay	(3)	(x_{24})
	3.2.6	Presence of Mycelium	(2)	(x_{25})
	3.2.7	Internal Discoloration	(3)	(x_{26})
	3.2.8	Sclerotia - Internal or External	(2)	(x_{27})
3.3		Condition 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})

The values of descriptors may be interrelated, i.e. the values of some variables may not be independent but subject to conditions identified by other descriptors. For example, if the descriptor Condition of Leaves has the value "normal" then those descriptors describing the abnormalities of the leaves of the plant, e.g. Leafspots - Halos, are not applicable. These descriptors take values which reflect the normal condition of the leaves.

The descriptors shown in Table 1 are hierarchically organized. The number in parentheses after each descriptor indicates the number of possible values which the descriptor can take. A plant is described by selecting the appropriate values for the descriptors. Given such a description, it is possible to make decisions concerning the plant. The decision which the consultant was to supply was:

> Can the given description be related to one or more of a set of known diseases?

Fifteen soybean diseases were selected as a representative of the nature and scope of the problems which are faced in the diagnoses of plant diseases. The 15 diseases which were used as the domain of the decision variable follow.

Diaporthe Stem Canker

Charcoal Rot

Rhizoctonia Root Rot

Phytophthora Rot

Brown Stem Rot

Downy Mildew

Brown Spot

Bacterial Blight

Bacterial Pustule

Purple Seed Stain

Anthracnose

Phyllosticta Leaf Spot

Alternaria Leaf Spot

Frog Eye Leaf Spot

The decision variable is included in the table of descriptors so that a total of 36 variables are available for the expression of the rules in the knowledge base. The complete table of descriptors which was used is given in Appendix A.

2.2 Decision Rules

Each of the rules which is used in the knowledge base is a VL_1 rule. A VL_1 rule is expressed as a decision assignment statement [Michalski 78] with premise and consequence:

$$V_P ::> V_C$$

V_P and V_C are extended disjunctive normal VL_1 expressions

($EDVL_1$ expressions). A $EDVL_1$ expression is a union of weighted terms, where a term is a product of selectors. Concatenation is used to express this product. A selector is a statement of the form

$$[L \ rel \ R]_w$$

where:

$w \in [0,1]$ is a constant, assumed 1 if not specified

L is a multiple valued (or discrete) variable

rel stands for one of the relational operators $< \ = \ \neq \ \geq \ \leq \ >$

R is a list of constants which are elements of the domain L.

The degree to which a premise is satisfied by a description indicates its degree of confidence. An example of a decision rule is

$$.9[x_1 = 0{:}3][x_4 \neq 0] \vee .7[X_2 < 5] \vee 0.2$$

$$::> \ [decision = k]$$

is a decision rule which is interpreted: if x_1 is between 0 and 3 and x_4 is not equal to 0, the degree of confidence for decision k is 0.9; if these conditions are not satisfied and x_2 is less than 5, the degree of confidence is 0.7. In every other case, it is 0.2. See section 3.2 page 21 for rule evaluation schemes.

The rules which were developed use the power and flexibility of VL_1 to express the information required.

2.2.1 Rules Describing Restrictions

VL_1 rules derived from the structure of the description space provided ad hoc information about the relations among the descriptors. These rules describe restrictions on the variables of the description space. These restrictions are used to reduce the number of questions it is necessary to ask the user prior to providing a diagnosis and also to provide a mechanism for the inadvertent entry of conflicting information. An example of a relational rule is

[condition of seed = normal] --> [seed mold growth = absent] $\wedge$
[seed discoloration = absent] $\wedge$
[seed size = normal] $\wedge$
[seed shriveling = absent]

If, during the requests for information, it is found that the condition of the seed is normal, then each of the restricted variables is set to the specified value and each rule is evaluated in light of these conditions. The relational rules used in the consultant are given in Appendix B.

2.2.2 Inductively Derived Decision Rules

VL_1 rules generated by automatic induction specify the minimum information required to discriminate a given class from an assumed set of classes. These rules are used as a filter to determine possible class candidates given a minimum amount of information.

The work on inductive processing using the concept of variable-valued logic [Michalski 73, 74a, 74b, 75] and the development of the AQVAL programs at the University of Illinois [Michalski 77] provided the framework for the generation of the inductive decision rules. Initially, Jacobsen provided 20 distinct examples of each of the 15 soybean diseases. These examples for each disease were randomly separated into learning and testing sets of equal size. The program AQVAL/1-V7 [Larson, Michalski 75] was applied to these learning sets as a group to synthesize rules to distinguish the diseases. These rules were tested with the testing sets. The results indicated that additional examples were necessary for five of the diseases. Jacobsen provided the additional examples, AQVAL/1-V7 was then applied to the augmented learning set, and the rules synthesized were tested using the augmented testing set. A detailed description of the methodology, data and results of this work may be found in [Chilausky, Jacobsen, Michalski 76]. Since this effort, two important program developments (ESEL and AQ11) have been made and used in the automatic generation of VL_1 rules. ESEL [Larson, Michalski 77] selects from a set of examples a specified number of examples which are most representative of the set. This distance measuring selection mechanism was used to partition the available examples into learning and testing sets of specified size (Table 2). AQ11 [Larson 76] supports the hierarchical structure of the description space by accepting a priori restrictions of subordinate variables. The restrictions are enforced, when certain conditions on a specified variable are met, to preclude the effects of the subordinate variables. Additionally, AQ11 accepts VL_1 decision rules and sets of examples and modifies these rules, if necessary, to make them consistent with the given examples. This version of AQVAL was used with restrictions for the subordinate variables to generate the inductive rules used in the consultant. Table 2 shows the results of testing the rules generated with the testing sets.

The criteria for accepting a rule as correctly classifying an event were the rule provided the correct classification and either had the highest percentage of satisfied selectors or was within ten percent of satisfying the most selectors.

The form of the inductively derived rules was a disjunction of terms with equal weight for all terms and all selectors. The inductive rules used in the consultant are given in Appendix C.

Table 2
Summary of Testing of the Inductively Derived Rules

Disease	Learning Events	Testing Events	Available Events	Events Correctly Classified	Percentage Correctly Classified
Diaporthe Stem Canker	10	10	20	10	100%
Charcoal Rot	10	10	20	10	100%
Rhizoctonia Root Rot	10	10	20	10	100%
Phytophthora Rot	40	40	88	48	100%
Brown Stem Rot	20	24	44	24	100%
Powdery Mildew	10	10	20	10	100%
Downy Mildew	10	10	20	10	100%
Brown Spot	40	52	92	52	100%
Bacterial Pustule	10	10	20	10	100%
Bacterial Blight	10	10	20	10	100%
Purple Seed Stain	10	10	20	10	100%
Anthracnose	20	24	44	24	100%
Phyllosticta Leaf Spot	10	10	20	10	100%
Alternaria Leaf Spot	40	51	91	49	96%
Frog Eye Leaf Spot	40	51	91	50	98%

2.2.3 Expert Derived Decision Rules

VL_1 rules obtained from expert sources specify the basic characteristics of a class, including those which may not be unique to the class. These rules are used to make the final selection of a class from the candidates which remain after the filtering of the inductive rules.

The expert VL_1 rules developed for the 15 soybean diseases were derived through conferences with an expert in plant pathology. A complete characteristic description was provided for each disease using a slightly modified expression of VL_1 rules. This modification allowed the differentiation of the set of significant characteristics (Q_S) present in a plant when afflicted by a particular disease from the set of characteristics which, although generally present, are merely corroborative (Q_C). When this representation is used, a term is a conjunction of weighted sub-terms, i.e.

$$T = Q_S \wedge Q_C$$

The expert rules used unequal weights for selectors which provide a mechanism for specifying the importance of each selector in relation to the remainder of the selectors. The expert rules used in the consultant are given in Appendix D.

3. USE OF THE KNOWLEDGE BASE

An expermental deductive processor was developed to provide a mechanism to utilize the diagnostic information represented in the knowledge base. The development is general to the extent that the consulting information available is restricted only to that information which can be represented in VL_1 rules. Additionally, the manner in which this information is processed is based upon control information which is part of the knowledge base for the processor. At the moment, only diagnostic information about a subset of soybeans is available; however, consulting in other areas is a matter of developing the rules and control information for the area and adding this information to the knowledge base. DP1 has the capability to select the knowledge base necessary for the indicated area of interest. Also, DP1 will consult on single events or groups of events within an area.

Initially, DP1 requires that a specific area of consulting be specified. This specification controls the initialization of the information base for consulting in the specified area from the knowledge base. Initialization includes the retrieval of the applicable relational, inductive and expert rules, the natural language description of the descriptors and their associated values, and the control information for that area.

In the case of soybean diseases, once initialized, DP1 uses this information in a two-phase process to provide a ranked set of diagnoses. In the first phase, the inductive decision rules are used to rapidly reduce the number of candidate diseases given the specification of the values of a minimum of variables.

The following process is repeated:

1. A variable is selected, using the specified selection mechanism,
2. Each rule is evaluated using the value obtained for this variable,
3. Each rule which has a degree of confidence less than the inductive rule elimination threshold, T_i (0.82 for soybean diseases), is eliminated.

Once the number of rules remaining reaches a specified level (five or less for soybean diseases), the second phase begins. In this phase, the expert decision rules are used to provide a selection of the most likely diseases from the given description. The expert rules are first evaluated using the variables and the values previously given. Then the repetitive process of variable selection and rule evaluation is used for the expert rules until either all variables occurring in the remaining rules are selected or the degree of confidence of all the remaining rules falls below the expert rule elimination threshold, T_e (0.56 for soybean diseases). If the repetitive process ends because values have been obtained for all of the variables, the disease associated with the expert decision rule with the highest degree of confidence is selected. The name of the disease and the degree of confidence is displayed. All diseases associated with expert decision rules which have degrees of confidence within the selection distance, d (0.2 for soybean diseases), of the highest degree of confidence are selected in order of degree of confidence and the name and degree of confidence are displayed as alternate diagnoses. If the values of all the rules drop below T_e, the process indicates that it is not possible to give a diagnosis. Either the plant is healthy or the disease it has contracted is not part of the knowledge base.

The control information which is obtained during initialization includes T_i, T_e, d, the variable selection mechanism to be used during each phase and the evaluation functions to be used during each phase.

3.1 Variable Selection

Two variable selection schemes may be selected for use in DP1:

- Linear Selection
- Frequency Selection

When the rules for the consulting area are loaded, two tables of frequencies of descriptors are constructed - one for the inductively derived decision rules and one for the expert derived decision rules. Each table specifies the number of occurrences of each descriptor in the respective rule set. These tables are used with both selection schemes. Once a variable is selected, its frequency is set to 0. If a rule is eliminated during the evaluation process, the frequencies in the table of all the unevaluated variables in the rule are decremented by the number of occurrences of each.

During the linear selection process, each descriptor is selected in the order that it is stored in the table of descriptors if it remains to be evaluated. Therefore, for n descriptors, this process selects $descriptor_1, \ldots, descriptor_n$ in order providing the frequency of $descriptor_i$ is greater than 0.

During the frequency selection process, each descriptor is chosen dependent upon its present frequency in the appropriate table of frequencies and a bias which may be given each descriptor. Bias is considered first; frequency, second. A descriptor is biased when it provides especially significant diagnostic information, or is a hierarchical variable. In the case of soybeans, the following descriptors were biased equally:

- Time of Occurrence
- Condition of the Leaves
- Condition of the Stem
- Condition of the Fruit Pods
- Condition of the Seed
- Condition of the Roots

The descriptors, Time of Occurrence and Condition of the Roots, were biased for their diagnostic significance. The remainder were biased since they are hierarchical descriptors which establish the applicability of a set of descriptors.

Since all of the biased descriptors are weighted equally, they are selected in the order of occurrence in the table of descriptors; Time of Occurrence is the first descriptor chosen.

Once a descriptor is selected, it and its associated values are presented with a request to select a specific value. Once a value is selected, if the descriptor is hierarchical and the value is normal, all dependent descriptors are set as described by the relational rules. The rules are evaluated and any decrementing of frequencies necessary due to the elimination of rules is done. Subsequently, another descriptor is chosen. The biased descriptors are exhausted first followed by the remaining descriptors in order of dynamically changing frequencies.

3.2 Rule Evaluation

The degree of confidence (DC) in a specific VL_1 rule is determined by the DC of the $EDVL_1$ expression, which is the premise of the rule. DC is in the real interval [0,1]. If DC = 1 for a premise, it indicates that the consequence of the rule is extremeley likely. If DC = 0 for a premise, it indicates that the consequence of the rule is extremely unlikely.

When a VL_1 rule is evaluated for a given event, evaluation schemes are necessary for each of the well-formed parts of the premise. Due to the difficulty of precisely determining the interdependence of the elements of the premise, a number of schemes were developed for the evaluation of each well-formed part of an $EDVL_1$ expression. The schemes for evaluating the component parts of an $EDVL_1$ expression cover a spectrum with the minimum-maximum multiple-valued logic evaluation (which assumes strong dependence) proposed by Lukasiewicz [Reacher 69] at one end. On the other end of the spectrum is the multiplication-summation probabilistic evaluation (which assumes strong independence) suggested by Suppes [Huntikka, Suppes 66].

Thus, to determine the DC for a VL_1 rule, the following evaluation schemes are necessary:

DC(S,e)	Degree to which a selector S is satisfied by an event e.
DC(T,e)	Degree to which a term T (a product of selectors) is satisfied by an event e.
DC(V,e)	Degree to which an $EDVL_1$ expression V (a union of terms) is satisfied by an event e.

When a premise $EDVL_1$ expression is evaluated for a given description, each selector is interpreted as a condition. A selector is satisfied if the value of the descriptor in the description satisfies the condition; otherwise, it is not satisfied. The following evaluation schemes used in DP1 assume $DC \in [0,1]$ and $w \in [0,1]$. The evaluation scheme for selectors

$$DC(S,e) = \begin{cases} 1, \text{ if satisfied} \\ 1-w, \text{ if not satisfied} \\ *, \text{ if the value is not specified} \end{cases} \tag{1}$$

is used for the following evaluation schemes for terms:

Minimum

$$DC(T,e)^* = \min_{i \in I} \{DC(S_i)\} \tag{2}$$

*The special case where w=1 for all selectors and the minimum-maximum schemes are used is treated in detail in [Chilausky, Jacobsen, Michalski 76]

Probabilistic

$$DC(T,e) = \prod_{i \in I} DC(S_i) \tag{3}$$

Average

$$DC(T,e) = \sum_{i \in I} DC(S_i)/ns \tag{4}$$

where ns is the number of selectors in the term.

In the evaluation schemes (2) and (3), the effect of an unsatisfied selector upon DC(T,e) is reflected by its weight. The greater the weight of a selector, the greater the reduction of confidence in the term; the smaller the weight, the smaller the reduction in confidence. In (3), the effect of unsatisfied selectors increases quickly even when the selectors are weakly weighted when two or more selectors are not satisfied.

The evaluation scheme (4) provides a method for specifying the effect an unsatisfied selector has on the ratio of satisfied selectors over the total number of selectors in the term. The greatest effect that any selector may have is 1/ns. This scheme provides a definition of DC(T,e) which does not rely upon the complete dependence of the descriptors involved. It gives a measure of the closeness to full satisfaction of a term, i.e. EF(T,e) = 1.

In the following evaluation schemes for selectors and terms suggested by Michalski, DC $\in$ [-1,1].

$$DC(S,e) = \begin{cases} w, \text{ if satisfied} \\ -w_i, \text{ if not satisfied} \\ *, \text{ if the value is not satisfied} \end{cases} \tag{5}$$

$$DC(T,e) = \sum_{i \in I} DC(S_i) / \sum_{i \in I} w_i \tag{6}$$

The scheme (5) provides a method for specifying the positive and negative effects of the evaluation of a selector. When (5) and (6) are used, if DC(T,e) is near 1, it indicates that the evidence given strongly supports the term. If DC(T,e) is near -1, it indicates that the evidence given strongly contradicts the term. If DC(T,e) is near 0, no indication of support or contradiction for the term may be assumed.

When weighted sub-terms are used specifying significant and corroborative consulting information as in some expert rules for soybeans (Section 2.2.3), the evaluation of a term is

$$DC(T,e) = sDC(T_s,e) + cDC(T_c,e) \tag{7}$$

where s and c are coefficients expressing the relative weights of the significant and corroborative information and $s + c = 1$. In the pilot project, $s = 0.9$ and $c = 0.1$.

Each sub-term is evaluated independently using the same evaluation functions, i.e. (2), (3), (4) or (6). The combination of the sub-term evaluations is done at the time of evaluation for the term so that a current value for that term is available.

Once the evaluations of the terms of an $EDVL_1$ expression have been completed, the expression is evaluated using the Lukasiewiez maximum or the Suppes summation. The maximum is

$$DC(V,e) = \max_{i=1,n} \{EF(T_i)\} \tag{8}$$

where n is the number of terms in V.

The summation is expressed for two terms as

$$DC(V,e) = DC(V,e) + EF(T_i) - EF(V,e)EF(T_i) \tag{9}$$

where DC(V,e) = 0 initially and is evaluated for each term in the expression, V.

In (8), each term is assumed to independently cover sets of similar cases of a class, here symptoms of a soybean disease. Therefore, the "best" or most representative term for the class is the one with the maximum value. In (9), each term is assumed to cover the entire class and the value of each term, no matter how small, provides evidence that the particular class is present. This function assumes the dependence of terms in an expression to adequately describe the class.

Finally, the DC of a rule is set equal to DC(V,e) for the premise expression.

The various evaluation schemes provide a necessary flexibility in the environment in which the prototype deductive processor functions. This environment includes a knowledge base with two sets of rules for soybean diseases - inductive and expert. The inductive rules are unweighted and each selector present is assumed equally important. Also, the method used to generate the rules prevents intersection between them. The expert rules have weighted selectors which reflects the relative importance of a condition to disease. These weights, as well as the specification of which set of selectors were significant and which were merely corroborative, were provided by a plant pathologist.

Each of the evaluation schemes has been tested to determine their effectiveness, both in the filtering process and the selection process. Three hundred and forty cases of the 15 soybean diseases were used for these tests. The inductive and expert rules are evaluated using different schemes in the most successful test to date. In this test, the inductive rules for filtering were evaluated using the Lukasiewiez minimum-maximum scheme; the expert rules for selecting, the weighted average - maximum scheme. The results of this test are shown in Appendix E.

4. SAMPLE SESSION

This chapter contains an example of an interactive consulting session with DP1 in which advice concerning the diagnosis of soybean diseases is computed. The specific disease which is described during the session is Brown Spot.

The user's input always follows a question mark and is underlined here for clarity. Comments (in parentheses) were added when an explanation seemed appropriate.

```
(Once the login and access procedures are completed, the session
 begins with the following.)

DO YOU WISH TO SEE THE INTRODUCTION TO THE DEDUCTIVE PROCESSOR (YES OR NO)?
? YES

     THE EXPERIMENTAL DEDUCTIVE PROCESSOR WAS DEVELOPED TO PROVIDE
CONSULTING ASSISTANCE IN A NUMBER OF FIELDS.  AT THE MOMENT, ONLY
ONE AREA IS AVAILABLE - DIAGNOSIS OF SOYBEAN DISEASES.
     WHEN USING THE PROGRAM, YOU WILL BE ASKED TO PROVIDE INFOR-
MATION CONCERNING SPECIFIC CONDITIONS PERTINENT TO THE AREA.  EACH
CONDITION IS ASSOCIATED WITH A LIST OF DESCRIPTIONS.  TO SELECT
A SPECIFC DESCRIPTION, ENTER THE CORRESPONDING LETTER AT THE QUES-
TION MARK(?)  PROMPT.

     AT ANY PROMPT, YOU MAY USE ANY OF THE FOLLOWING COMMANDS:

         HELP - TO OBTAIN MORE DETAILED INFORMATION ABOUT THE
                CONDITION AND THE DESCRIPTIONS PROVIDED.
          WHY - TO SEE WHAT DECISIONS THE CONDITION AND
                SPECIFIC DESCRIPTIONS SUPPORT.
      SUMMARY - TO OBTAIN A SUMMARY OF THE SEQUENCE WHICH
                PROVIDED THE SPECIFIC CONSULTATION.
         QUIT - TO TERMINATE THE EXECUTION OF THE PROGRAM.
```

THE FOLLOWING AREAS ARE AVAILBLE FOR ASSISTANCE AND DEDUCTIVE PROCESSING AT THE PRESENT TIME:

A. PLANT DISEASE DIAGNOSIS

SELECT ONE OF THE ABOVE.
? A
PLANT DISEASE DIAGNOSIS IS AVAILABLE FOR THE FOLLOWING PLANTS:

A. SOYBEANS

SELECT ONE OF THE ABOVE.
? A

EACH QUESTION WHICH FOLLOWS WILL HAVE A LIST OF ANSWERS. CHOOSE THE BEST ANSWER, TYPE THE CORRESPONDING LETTER, THEN PRESS RETURN.

TIME OF OCCURRENCE?
A. APRIL
B. MAY
C. JUNE
D. JULY
E. AUGUST
F. SEPTEMBER
G. OCTOBER
H. DO NOT KNOW
? B

CONDITION OF LEAVES?
A. NORMAL
B. ABNORMAL
C. DO NOT KNOW
? B

CONDITION OF STEM?
A. NORMAL
B. ABNORMAL
C. DO NOT KNOW
? WHY

(The user would like to see what decisions specific evidence supports.)

IF [CONDITION OF STEM = NORMAL] THEN THE FOLLOWING ARE INDICATED:
DOWNY MILDEW
BACTERIAL PUSTULE
PHYLLOSTICTA LEAF SPOT
ALTERNARIA LEAF SPOT

```
IF [CONDITION OF STEM = ABNORMAL] THEN THE FOLLOWING ARE INDICATED:
          CHARCOAL ROT
          RHIZOCTONIA ROOT ROT
          PHYTOPHTHORA ROT
          BROWN STEM ROT
          ANTHRACNOSE
          FROG EYE LEAF SPOT
```

(The user is queried again.)

```
-----------------------------

CONDITION OF STEM?
A.  NORMAL
B.  ABNORMAL
C.  DO NOT KNOW
? A
-----------------------------

CONDITION OF FRUIT - PODS?
A.  NORMAL
B.  DISEASED
C.  FEW OR NONE PRESENT
D.  DOES NOT APPLY
E.  DO NOT KNOW
? A
-----------------------------

CONDITION OF SEED?
A.  NORMAL
B.  ABNORMAL
C.  DO NOT KNOW
? A
-----------------------------

CONDITION OF ROOTS?
A.  NORMAL
B.  ROTTED
C.  GALLS OR CYSTS PRESENT
D.  DO NOT KNOW
? A
-----------------------------

LEAF SPOTS - HALOS?
A.  ABSENT
B.  WITH YELLOW HALOS
C.  WITHOUT YELLOW HALOS
D.  DO NOT KNOW
? C
-----------------------------
```

LEAF SPOTS - MARGIN?
A. WITH WATERSOAKED MARGIN
B. WITHOUT WATERSOAKED MARGIN
C. DOES NOT APPLY
D. DO NOT KNOW
? B

PRECIPITATION?
A. BELOW NORMAL
B. NORMAL
C. ABOVE NORMAL
D. DO NOT KNOW
? C

DAMAGED AREA?
A. SCATTERED PLANTS
B. GROUPS OF PLANTS IN LOW AREAS
C. GROUPS OF PLANTS IN UPLAND AREAS
D. WHOLE FIELDS
E. DO NOT KNOW
? B

LEAF SPOT SIZE?
A. LESS THAN 1/8"
B. GREATER THAN 1/8"
C. DOES NOT APPLY
D. DO NOT KNOW
? B

SHOT-HOLING/SHREDDING?
A. ABSENT
B. PRESENT
C. DO NOT KNOW
? A

LEAF MILDEW GROWTH?
A. ABSENT
B. ON UPPER LEAF SURFACE
C. ON LOWER LEAF SURFACE
D. DO NOT KNOW
? HELP

(The user would like a more details specification of the condition and descrfption.)

IF THE LEAVES OF THE PLANT SHOW NO EVIDENCE OF MILDEW GROWTH ON EITHER SURFACE OF THE LEAVES, SELECT ABSENT.

IF THE LEAVES OF THE PLANT HAVE A MILDEW GROWTH WHICH IS WHITE IN COLOR AND IS PREDOMINATELY ON THE UPPER SURFACE OF THE LEAVES, SELECT ON UPPER LEAF SURFACE.

IF THE LEAVES OF THE PLANT HAVE A MILDEW GROWTH WHICH IS GRAY IN COLOR AND IS PREDOMINATELY ON THE LOWER SURFACE OF THE LEAVES, SELECT ON LOWER LEAF SURFACE.

(The user is queried again and replies with A.)

LEAF MILDEW GROWTH?
A. ABSENT
B. ON UPPER LEAF SURFACE
C. ON LOWER LEAF SURFACE
D. DO NOT KNOW
? <u>A</u>

TEMPERATURE?
A. BELOW NORMAL
B. NORMAL
C. ABOVE NORMAL
D. DO NOT KNOW
? <u>B</u>

CROPPING HISTORY?
A. CROP DIFFERENT THAN LAST YEAR
B. CROP SAME AS LAST YEAR
C. CROP SAME FOR LAST THREE YEARS
D. CROP SAME FOR FOUR OR MORE YEARS
E. DO NOT KNOW
? <u>A</u>

THE EVIDENCE PRESENTED SUGGESTS

BROWN SPOT WITH A DEGREE OF CONFIDENCE 1.00

ALTERNATIVELY,

PHYLLOSTICTA LEAF SPOT WITH A DEGREE OF CONFIDENCE 0.82

DO YOU WISH TO CONTINUE IN THIS AREA: YES OR NO?
? <u>SUMMARY</u>

(The user would like to see the processing sequence which reached this result.)

SUMMARY

DISEASES ELIMINATED DURING THE EVALUATION OF THE INDUCTIVELY DERIVED RULES

DIAPORTHE STEM CANKER
BECAUSE OF THE FAILURE TO SATISFY THE FOLLOWING CONDITIONS:
[TIME OF OCCURRENCE = MAY]

CHARCOAL ROT
BECAUSE OF THE FAILURE TO SATISFY THE FOLLOWING CONDITIONS:
[CONDITION OF STEM = NORMAL]

RHIZOCTONIA ROOT ROT
BECAUSE OF THE FAILURE TO SATISFY THE FOLLOWING CONDITIONS:
[CONDITION OF LEAVES = ABNORMAL]
[CONDITION OF STEM = NORMAL]

PHYTOPHTHORA ROT
BECAUSE OF THE FAILURE TO SATISFY THE FOLLOWING CONDITIONS:
[CONDITION OF STEM = NORMAL]

BROWN STEM ROT
BECAUSE OF THE FAILURE TO SATISFY THE FOLLOWING CONDITIONS:
[CONDITION OF LEAVES = ABNORMAL]
[CONDITION OF STEM = NORMAL]

BACTERIAL BLIGHT
BECAUSE OF THE FAILURE TO SATISFY THE FOLLOWING CONDITIONS:
[TIME OF OCCURRENCE = MAY]

PURPLE SEED STAIN
BECAUSE OF THE FAILURE TO SATISFY THE FOLLOWING CONDITIONS:
[CONDITION OF LEAVES = ABNORMAL]
[CONDITION OF SEED = NORMAL]

ANTHRACNOSE
BECAUSE OF THE FAILURE TO SATISFY THE FOLLOWING CONDITIONS:
[TIME OF OCCURRENCE = MAY]
[CONDITION OF STEM = NORMAL]

ALTERNARIA LEAF SPOT
BECAUSE OF THE FAILURE TO SATISFY THE FOLLOWING CONDITIONS:
[TIME OF OCCURRENCE = MAY]
[LEAF SPOTS - MARGIN = WITHOUT WATERSOAKED MARGIN]

FROG EYE LEAF SPOT
BECAUSE OF THE FAILURE TO SATISFY THE FOLLOWING CONDITIONS:
[TIME OF OCCURRENCE = MAY]
[LEAF SPOTS - MARGIN = WITHOUT WATERSOAKED MARGIN]
[CONDITION OF STEM = NORMAL]

DISEASES ELIMINATED DURING THE EVALUATION OF THE
EXPERT DERIVED RULES

POWDERY MILDEW
BECAUSE OF THE FAILURE TO SATISFY THE FOLLOWING CONDITIONS:
[TIME OF OCCURRENCE = MAY]
[LEAF MILDEW GROWTH = ABSENT]

BACTERIAL PUSTULE
BECAUSE OF THE FAILURE TO SATISFY THE FOLLOWING CONDITIONS:
[TIME OF OCCURRENCE = MAY]
[CROPPING HISTORY = CROP DIFFERENT THAN LAST YEAR]
[LEAF SPOT SIZE = GREATER THAN 1/8"]
[SHOT-HOLING/SHREDDING = ABSENT]

ELIMINATED BECAUSE VALUES NOT WITHIN 0.20 OF THE "BEST" VALUE

DOWNY MILDEW
BECAUSE OF THE FAILURE TO SATISFY THE FOLLOWING CONDITIONS:
[TIME OF OCCURRENCE = MAY]
[DAMAGED AREA = GROUPS OF PLANTS IN LOW AREAS]
[LEAF MILDEW GROWTH = ABSENT]
[CONDITION OF SEED = NORMAL]

DIAGNOSIS

BROWN SPOT HAD A DEGREE OF CONFIDENCE OF 1.00
THE FOLLOWING CONDITIONS WERE NOT SATISFIED FOR THIS DIAGNOSIS
ALL CONDITIONS OF THE RULE SATISFIED

PHYLLOSTICTA LEAF SPOT HAD A DEGREE OF CONFIDENCE OF 0.[illegible]
THE FOLLOWING CONDITIONS WERE NOT SATISFIED FOR THIS DIAGNOSIS
[TIME OF OCCURRENCE = MAY]
[TEMPERATURE = NORMAL]
[DAMAGED AREA = GROUPS OF PLANTS IN LOW AREAS]
[SHOT-HOLING/SHREDDING = ABSENT]

DO YOU WISH TO CONTINUE IN THIS AREA: YES OR NO?
? QUIT

(The user exits the program.)

The sample session with the deductive processor was run on the CDC CYBER 175 at the University of Illinois. The approximate cost of this session was one dollar.

5. CONCLUSIONS

Computer consulting systems for a number of areas are being investigated as a method of providing assistance which might otherwise be difficult to obtain. A description of such a system which has been applied to the diagnosis of soybean diseases has been presented. This system uses a knowledge base which was derived both from examples and expert sources. The knowledge base is represented as sets of rules - inductive and expert. The inductive rules are used as a filter to quickly reduce the number of possible diagnoses. The expert rules are then used on this reduced set to provide the final selection. More than one diagnosis is possible in the environment of soybean diseases since a plant may have more than one disease. The experimental results indicate that the VL_1 representation and its application to this area was very successful. It appears that it will be possible to provide adequate diagnostic tools to the user to diagnose soybean diseases at a minimal cost.

5.1 Possible Extensions

The deductive processor is in the experimental stage and many extensions in the representation and methodology are possible. Among these extensions should be

- The development of a knowledge base for all soybean diseases in this geographical area.

- The investigation of other schemes of knowledge representation, in particular VL_2 [Michalski 74c] - a variable-valued predicate logic, as a means of representing knowledge.

- The development of a variable selection processor based on extended entry decision tables [Michalski 78] and conditional probabilities of variables and values.

- The addition of empirical weights in the evaluation schemes as data about the success of rule evaluation is compiled and experience is gained with the system.

LIST OF REFERENCES

Chilausky, R. L., B. J. Jacobsen, and R. S. Michalski. An application of variable-valued logic to inductive learning of plant disease diagnostic rules. Proc. Sixth Ann. Intern. Symposium on Multiple-Valued Logic. Utah State Univ., Logan, Utah, 1976.

Harsh, S. B. A progress report on TELEPLAN activities, Dept. of Agr. Econ., Michigan State Univ., East Lansing, Michigan, 1975.

Hart, P. E. Progress on a computer-based consultant. Artificial Intelligence Tech. Note 99, Stanford Res. Inst. Menlo Park, California, 1975.

Hart, P. E., and R. O. Duda. PROSPECTOR -- a computer-based consultation system for mineral exploration. Artificial Intelligence Tech. Note 155, Stanfard Res. Inst. Menlo Park, California, 1977.

Huntikka, J., and P. Suppes. Aspects of Inductive Logic. North-Holland Publ., Amsterdam, 1966.

Kulikowski, C., and S. Weiss. The medical consultant program - glaucoma (MCP-G1). Dept. of Computer Sci.Rpt. CBM-TR-5, Rutgers Univ. New Brunswick, New Jersey, 1972.

Kulikowski, C., and S. Weiss. An interactive facility for the inferential modeling of disease. Proc. Seventh Ann. Princeton Conf. on Information Sciences and Systems. Princeton, New Jersey, 1973.

Larson J. B., and R. S. Michalski. AQVAL/1 (AQ7) user's guide and program description. Dept. of Computer Sci. Rpt. No. 731. Univ. of Illinois, Urbana, Illinois, 1975.

Larson, J. B. AQ11: a program for multi-step formation of VL_1 hypothesis and evaluation VL_1 formulas. Univ. of Illinois, Urbana, Illinois, 1976.

Larson, J. B., and R. S. Michalski. A user's guide for ESEL: a VL_1 event selection program. Dept. of Computer Sci. Univ. of Illinois, Urbana, Illinois, 1977.

Michalski, R. S. A variable-valued logic system as applied to picture description and recognition. Graphic Languages, Proc. of IFIP Working Conf. on Graphic Languages. Vancouver, Canada, 1972.

Michalski, R. S. AQVAL/1 -- computer implementation of a variable-valued logic system and the application to pattern recognition. Proc. of First Intern. Joint Conf. on Pattern Recognition. Washington, D.C., 1973.

Michalski, R. S. Variable-valued logic: system VL_1. Proc. of Fourth Intern. Symposium on Multiple-Valued Logic. West Virginia Univ., Morgantown, West Virginia, 1974 (a).

Michalski, R. S. A variable-decision space approach for implementing a classification system. Second Intern. Joint Conf. on Pattern Recognition. Lyngby-Copenhagen, Denmark, 1974 (b).

Michalski, R. S. Learning by inductive inference. J. C. Simon (ed.), NATO Advanced Study Institute on Computer Oriented Learning Processes, Series E Apple. Sci. 14: 321-337. Boras, France, 1974 (c).

Michalski, R. S. Synthesis of optimal and quasi-optimal variable-valued logic formulas. Proc. of Fifth Intern. Symposium on Multiple-Valued Logic. Indiana Univ., Bloomington, Indiana, 1975.

Michalski, R. S. Designing extended entry decision tables and optimal decision trees using decision diagrams. Dept. of Computer Sci. Rpt. No. 898. Univ. of Illinois, Urbana, Illinois, 1978.

Michalski, R. S. Pattern recognition as knowledge-guided computer induction. Dept. of Computer Sci. Rpt. No. 927. Univ. of Illinois, Urbana, Illinois, 1978.

Pople, H. E., J. D. Myers, and R. A. Miller DIALOG: a model of diagnostic logic for internal medicine. Proc. Fourth Int. Joint Conf. on Artificial Intelligence 2: 848-855. Tbilisi, Georgia, USSR, 1975.

Reacher, N. Many-Valued Logic. McGraw-Hill. New York, New York, 1969.

Schank, R. C. Identification of conceptualizations underlying natural language. R. C. Schank and K. M. Colby (eds.), Computer Models of Thought and Language. W. H. Freeman. San Francisco, California, 1973.

Shortliffe, E. H. MYCIN: a rule based computer program for advising physicians regarding antimicrobial therapy selection. Computer Sci. Dept. Rpt. No. CS-74-465, Stanford Univ., Stanford, California, 1974.

Winograd, T. A procedural model of language understanding. R. C. Schank and K. M. Colby (eds.), Computer Models of Thought and Language. W. H. Freeman. San Francisco, California, 1973.

APPENDIX A

The 36 variables which were used in the experiment are given below. In the format for each variable, the first line gives the formal name of the variable, i.e. $x_1, x_2, \ldots, x_{35}$, and its natural language description, i.e. the attribute of the plant or environment which the variable describes. Following this is a list of the numerical values which the variable may take followed by the corresponding descriptions of these values.

x_1 := TIME OF OCCURRENCE

0 April
1 May
2 June
3 July
4 August
5 September
6 October

x_2 := PLANT STAND

0 Normal
1 Less Than Normal

x_3 := PRECIPITATION

0 Below Normal
1 Normal
2 Above Normal

x_4 := TEMPERATURE

0 Below Normal
1 Normal
2 Above Normal

x_5 := OCCURRENCE OF HAIL

0 No
1 Yes

x_6 := CROPPING HISTORY

0 Crop Different Than Last Year
1 Crop Same As Last Year
2 Cròp Same For Last Three Years
3 Crop Same For Four Or More Years

x_7 := DAMAGED AREA

0 Scattered Plants
1 Groups Of Plants In Low Areas
2 Groups Of Plants In Upland Areas
3 Whole Fields

x_8 := SEVERITY

0 Minor
1 Potentially Severe
2 Severe

x_9 := SEED TREATMENT

0 None
1 Fungicide
2 Other

x_{10} := SEED GERMINATION

0 90-100%
1 80-89%
2 Less Than 80%

x_{11} := PLANT HEIGHT

0 Normal
1 Abnormal

x_{12} := CONDITION OF LEAVES

0 Normal
1 Abnormal

x_{13} := LEAF SPOTS HALOS

0 Absent
1 With Yellow Halos
2 Without Yellow Halos

x_{14} := LEAF SPOTS MARGIN

0 With Watersoaked Margin
1 Without Watersoaked Margin
2 Does Not Apply

x_{15} := LEAF SPOT SIZE

0 Less Than 1/8"
1 Greater Than 1/8"
2 Does Not Apply

x_{16} := SHOTHOLING/SHREADING

0 Absent
1 Present

x_{17} := LEAF MALFORMATION

0 Absent
1 Present

x_{18} := LEAF MILDEW GROWTH

0 Absent
1 On Upper Leaf Surface
2 On Lower Leaf Surface

x_{19} := CONDITION OF STEM

0 Normal
1 Abnormal

x_{20} := PRESENCE OF LODGING

0 No
1 Yes

x_{21} := STEM CANKERS

0 Absent
1 Below Soil Line
2 At Or Slightly Above Soil Line
3 Above Second Node

x_{22} := CANKER LESION COLOR

0 Does Not Apply
1 Brown
2 Dark Brown Or Black
3 Tan

x_{23} := FRUITING BODIES ON STEM

0 Absent
1 Present

x_{24} := EXTERNAL DECAY OF STEM

0 Absent
1 Firm And Dry
2 Watery And Soft

x_{25} := MYCELIUM ON STEM

0 Absent
1 Present

x_{26} := INTERNAL DISCOLORATION OF STEM

0 None
1 Brown
2 Black

x_{27} := SCLEROTIA INTERNAL OR EXTERNAL

0 Absent
1 Present

x_{28} := CONDITION OF FRUIT PODS

0 Normal
1 Diseased
2 Few Or None Present
3 Does Not Apply

x_{29} := FRUIT SPOTS

0 Absent
1 Colored Spots
2 Brown Spots With Black Specks
3 Distorted Pods
4 Does Not Apply

x_{30} := CONDITION OF SEED

0 Normal
1 Abnormal

x_{31} := SEED MOLD GROWTH

0 Absent
1 Present

x_{32} := SEED DISCOLORATION

0 Absent
1 Present

x_{33} := SEED SIZE

0 Normal
1 Smaller Than Normal

x_{34} := SEED SHRIVELING

0 Absent
1 Present

x_{35} := CONDITION OF ROOTS

0 Normal
1 Rotted
2 Galls Or Cysts Present

x_{36} := SOYBEAN DISEASE

0 Diaporthe Stem Canker
1 Charcoal Rot
2 Rhizoctonia Root Rot
3 Phytophthora Rot
4 Brown Stem Rot
5 Powdery Mildew
6 Downy Mildew
7 Brown Spot
8 Bacterial Blight
9 Bacterial Pustule
10 Purple Seed Stain
11 Anthracnose
12 Phyllosticta Leaf Spot
13 Alternaria Leaf Spot
14 Frog Eye Leaf Spot

APPENDIX B

Rules Describing Restrictions Used in the Consultant

Condition of Leaves

[condition of leaves = normal] -->

[leafspots halos = absent]
[leafspots margin = does not apply]
[leafspot size = does not apply]
[leaf shredding = absent]
[leaf malformation = absent]
[leaf mildew growth = absent]

Leafspots - Halos

[leafspots halos = absent] -->

[leafspots margin = does not apply]
[leafspot size = does not apply]

Condition of Stem

[condition of stem = normal] -->

[presence of lodging = absent]
[stem cankers = absent]
[canker lesion color = does not apply]
[fruiting bodies on stem = absent]
[external decay of stem = absent]
[mycelium on stem = absent]
[internal discoloration = none]
[sclerotia internal or external = absent]

Condition of Fruit - Pods

[condition of fruit pods = normal] -->

[fruit spots = absent]

Condition of Seed

[condition of seed = normal] -->

[seed mold growth = absent]
[seed discoloration = absent]
[seed size = normal]
[seed shriveling = absent]

APPENDIX C

Inductively Derived Discriminant Diagnostic VL_1 Rules for 15 Soybean Diseases

Diaporthe Stem Canker

D1: [time = Jul, Aug, Sep, Oct] [precipitation = above normal]
[leaf malformation = absent] [condition of stem = abnormal]
[stem cankers = above second node] [external decay = firm and dry]
[condition of fruit pods = normal]

::> [Diagnosis = Diaporthe Stem Canker]

Charcoal Rot

D2: [leaf malformation = absent] [condition of stem = abnormal]
[internal discoloration = black]

::> [Diagnosis = Charcoal Rot]

Rhizoctonia Root Rot

D3: [condition of leaves = normal] [condition of stem = abnormal]
[stem cankers = below soil line] [canker lesion color = brown] V
[leaf malformation = absent] [condition of stem = abnormal]
[stem cankers = below soil line] [canker lesion color = brown]

::> [Diagnosis = Rhizoctonia Root Rot]

Phytophthora Root Rot

D4: [plant stand = less than normal]
[precipitation = normal, above normal]
[temperature = below normal, normal]
[plant height = abnormal]
[condition of leaves = abnormal] [leaf malformation = absent]
[condition of stem = abnormal] V
[time = Apr, May, Jun, Jul, Aug] [plant stand = abnormal]
[damaged area = groups of plants in low areas]
[plant height = abnormal] [condition of leaves = abnormal]
[condition of stem = abnormal]
[external decay = absent, watery and soft]

::> [Diagnosis = Phytophthora Root Rot]

Brown Stem Rot

D5: [leaf malformation = absent] [condition of stem = abnormal]
[internal discoloration = brown] V
[condition of leaves = normal] [condition of stem = abnormal]
[internal discoloration = brown]

::> [Diagnosis = Brown Stem Rot]

Powdery Mildew

D6: [condition of leaves = abnormal] [leaf malformation = absent]
[leaf mildew growth = on upper leaf surface]
[condition of roots = normal]

::> [Diagnosis = Powdery Mildew]

Downy Mildew

D7: [leafspots halos = with yellow halos, without yellow halos]
[leaf mildew growth = on lower leaf surface]
[condition of stem = normal] [seed mold growth = present]

::> [Diagnosis = Downy Mildew]

Brown Spot

D8: [precipitation = normal, above normal]
[cropping history = crop same as last year, crop same for last three years, crop same for four or more years]
[damaged area = scattered plants, groups of plants in low areas, groups of plants in upland areas]
[condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = without watersoaked margin]
[leafspot size = greater than 1/8 inch]
[leaf malformation = absent] [condition of roots = normal] V
[precipitation = above normal] [condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = without watersoaked margin]
[leafspot size = greater than 1/8 inch]
[condition of roots = normal] V
[time = Apr, May, Jun]
[damaged area = scattered plants, groups of plants in low areas, groups of plants in upland areas]
[condition of leaves = abnormal]

[leafspots halos = without yellow halos]
[leafspots margin = without watersoaked margin]
[leafspot size = greater than 1/8 inch] [leaf shredding = absent]
[leaf malformation = absent] [condition of roots = normal]

::> [Diagnosis = Brown Spot]

Bacterial Blight

D9: [time = Jun, Jul, Aug, Sep] [temperature = normal, above normal]
[condition of leaves = abnormal]
[leafspots halos = with yellow halos, without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = less than 1/8 inch]
[condition of fruit pods = normal] [condition of roots = normal]

::> [Diagnosis = Bacterial Blight]

Bacterial Pustule

D10: [condition of leaves = abnormal]
[leafspots halos = with yellow halos]
[leafspots margin = without watersoaked margin]
[leafspot size = less than 1/8 inch]
[condition of stem = normal] [condition of fruit pods = normal] V
[leafspots halos = with yellow halos, without yellow halos]
[leafspot size = less than 1/8 inch]
[condition of stem = normal] [condition of roots = rotted] V
[time = May] [precipitation = normal]
[condition of leaves = abnormal]
[leafspots halos = with yellow halos]

::> [Diagnosis = Bacterial Pustule]

Purple Seed Stain

D11: [plant stand = normal] [precipitation = above normal]
[severity = minor] [plant height = normal]
[leafspots halos = without yellow halos]
[condition of seed = abnormal]
[seed discoloration = present] [seed size = normal] V
[condition of leaves = normal] [condition of seed = abnormal]
[seed size = normal]

::> [Diagnosis = Purple Seed Stain]

Anthracnose

D12: [precipitation = above normal] [leaf malformation = absent]
[condition of stem = abnormal]
[stem cankers = at or slightly above soil line, above second node]
[condition of seed = abnormal]
[condition of roots = normal] V
[time = Aug, Sep, Oct] [precipitation = above normal]
[condition of leaves = normal] [stem cankers = above second node]
[condition of fruit pods = diseased]
[fruit spots = brown spots with black specks] V
[temperature = above normal] [leafspots halos = absent]
[leaf malformation = absent] [condition of stem = abnormal]
[external decay = firm and dry]

::> [Diagnosis = Anthracnose]

Phyllosticta Leaf Spot

D13: [time = Jun, Jul] [precipitation = below normal, normal]
[severity = minor] [leafspots halos = without yellow halos]
[leafspots margin = without watersoaked margin]
[condition of stem = normal] [condition of roots = normal] V
[precipitation = below normal] [condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = without watersoaked margin]
[condition of roots = normal] V
[plant stand = less than normal] [precipitation = normal]
[occurrence of hail = no] [leafspots halos = without yellow halos]
[leafspots margin = without watersoaked margin]
[condition of stem = normal] [condition of roots = normal]

::> [Diagnosis = Phyllosticta Leaf Spot]

Alternaria Leaf Spot

D14: [time = Aug] [precipitation = above normal]
[seed treatment = none] [condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = greater than 1/8 inch]
[leaf mildew growth = absent] [condition of stem = normal]
[condition of fruit pods = normal] V
[time = Sep, Oct] [precipitation = above normal]
[damaged area = scattered plants, groups of plants in low areas, whole fields]
[seed germination = 90-100%, 80-89%]
[condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = greater than 1/8 inch]
[condition of stem = normal] V

[time = Aug, Sep, Oct]
[damaged area = scattered plants, groups of plants in low areas]
[seed germination = less than 80%] [plant height = normal]
[condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = greater than 1/8 inch]
[leaf mildew growth = absent] [condition of stem = normal] V
[time = Oct] [seed germination = 80-89%, less than 80%]
[condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = greater than 1/8 inch]
[leaf mildew growth = absent] [condition of stem = normal] V
[time = Aug, Sep, Oct]
[damaged area = groups of plants in upland areas, whole fields]
[seed treatment = none, other]
[seed germination = 90-100%, 80-89%] [condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margins]
[leafspot size = greater than 1/8 inch] [leaf mildew growth = absent]
[condition of stem = normal] [condition of fruit pods = normal] V
[occurrence of hail = no] [damaged area = scattered plants]
[severity = potentially severe] [seed germination = 90-100%, 80-89%]
[condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = greater than 1/8 inch] [leaf mildew growth = absent]
[condition of stem = normal] V
[time = Aug, Sep, Oct] [temperature = normal]
[seed treatment = fungicide] [seed germination = 80-89%]
[condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = greater than 1/8 inch] [leaf mildew growth = absent]
[condition of stem = normal] [condition of fruit pods = normal] V
[time = Sep, Oct] [condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = greater than 1/8 inch] [leaf shredding = present]

::> [Diagnosis = Alternaria Leaf Spot]

Frogeye Leaf Spot

D15: [precipitation = normal, above normal]
[cropping history = groups of plants in low areas, groups of plants in upland areas, whole fields]
[condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = greater than 1/8 inch] [leaf shredding = absent]
[leaf mildew growth = absent] [condition of stem = abnormal]

[condition of roots = normal] V
[time = Jul, Aug, Sep] [precipitation = normal, above normal]
[temperature = normal] [occurrence of hail = no]
[damaged area = groups of plants in low areas, whole fields]
[seed treatment = fungicide] [condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = greater than 1/8 inch]
[leaf shredding = absent] [leaf malformation = absent]
[condition of roots = normal] V
[time = Aug, Sep] [precipitation = normal, above normal]
[damaged area = groups of plants in low areas, groups of plants in upland areas] [severity = minor] [condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = greater than 1/8 inch]
[leaf shredding = absent] [leaf mildew growth = absent]
[condition of seed = normal] [condition of roots = normal] V
[time = Jul, Aug] [precipitation = above normal]
[cropping history = crop same as last year, crop same for last three years, crop same for four or more years]
[damaged area = scattered plants]
[seed treatment = none, other] [condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = greater than 1/8 inch]
[leaf shredding = absent] [leaf mildew growth = absent]
[condition of roots = normal] V
[precipitation = above normal]
[cropping history = crop different than last year, crop same as last year, crop same for last three years]
[damaged area = scattered plants, groups of plants in upland areas]
[severity = potentially severe] [seed germination = less than 80%]
[condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = greater than 1/8 inch] [leaf mildew growth = absent]
[condition of roots = normal] V
[time = Jul] [occurrence of hail = yes]
[condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = greater than 1/8 inch] [leaf mildew growth = absent]
[condition of stem = normal] V
[plant stand = normal] [precipitation = normal, above normal]
[cropping history = crop same for last three years]
[condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = greater than 1/8 inch] [leaf shredding = absent]
[leaf mildew growth = absent] [condition of seed = normal]
[condition of roots = normal]

::> [Diagnosis = Frogeye Leaf Spot]

APPENDIX D

Expert Derived Descriptive Diagnostic VL_1 Rules for 15 Soybean Diseases

Q_s indicates significant conditions.

Q_c indicates corroborative conditions.

Diaporthe Stem Canker

D1: Q_s([time = Aug, Sep] (.9[precipitation = above normal] V
.7[precipitation = normal]) [stem cankers = above second node]
[fruiting bodies = present] [condition of fruit pods = normal])
$\wedge$
Q_c([temperature = normal, above normal]
(.9[cropping history = crop same for four or more years] V
.8[cropping history = crop same for last three years] V
.7[cropping history = crop same as last year V .2)
[canker lesion color = brown])

::> [Diagnosis = Diaporthe Stem Canker]

Charcoal Rot

D2: Q_s([time = Jul, Aug] [precipitation = below normal, normal]
[temperature = normal, above normal] [plant growth = abnormal]
[condition of leaves = abnormal] [condition of stem = abnormal]
[sclerotia internal or external = present]
[condition of roots = rotted] [internal discoloration = black])
$\wedge$
Q_c([damaged area = groups of plants in upland areas]
[severity = severe] [seed size = less than normal]
(.9[cropping history = crop same for last three years, crop same for four or more years] V .6[cropping history = crop same as last year V .2))

::> [Diagnosis = Charcoal Rot]

Rhizoctonia Root Rot

D3: Q_s([time = May, Jun] [plant stand = less than normal, normal]
[temperature = less than normal, normal]
[precipitation = less than normal, normal]
[condition of leaves = abnormal] [condition of stem = abnormal]
[canker lesion color = brown] [condition of roots = rotted]

([occurrence of hail = no] --> [stem cankers = below soil line, at or slightly above soil line])
([occurrence of hail = yes] --> [stem cankers = above second node]))

∧

Q_c([fruiting bodies = absent] [external decay = firm and dry]
[mycelium = absent])

::> [Diagnosis = Rhizoctonia Root Rot]

Phytophthora Root Rot

D4: Q_s((.7[time = Apr, Aug] V .9[time = May, Jun, Jul])
[plant stand = less than normal]
[precipitation = (normal <--> [time = Apr, May, Jun],
above normal <--> [time = Jul, Aug])]
[temperature = (above normal <--> [time = Apr],
normal <--> [time = May, Jun, Jul, Aug])]
[damaged areas = groups of plants in low areas]
[plant growth = abnormal] [condition of leaves = abnormal]
[condition of stem = abnormal]
[stem cankers = at or slightly above soil line]
[canker lesion color = (dark brown or black <-->
[time = May, Jun, Jul, Aug])]
[condition of roots = rotted])

∧

Q_c(.9[cropping history = crop same for last three years,
crop same for four or more years])

::> [Diagnosis = Phytophthora Root Rot]

Brown Stem Rot

D5: Q_s([time = Jul, Aug, Sep] [precipitation = above normal]
[temperature = below normal, normal]
[condition of leaves = abnormal] [condition of stem = abnormal]
[internal discoloration = brown] [presence of lodging = yes])

∧

Q_c([seed size = less than normal]
(.7[cropping history = crop same for last three years, crop same for four or more years V .5[cropping history = crop same as last year] V .1[cropping history = crop different than last year]))

::> [Diagnosis = Brown Stem Rot]

Powdery Mildew

D6: Q_s([condition of leaves = abnormal]
[leaf mildew growth = upper leaf surface])

$\wedge$

Q_c [time = Aug, Sep]

::> [Diagnosis = Powdery Mildew]

Downy Mildew

D7: Q_s([time = Jun, Jul, Aug] [precipitation = normal, above normal]
[damaged areas = whole fields] [condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leaf mildew growth = lower leaf surface]
[condition of seed = (abnormal <--> [time = Sep, Oct]), normal]
[mold growth on seed = present])

::> [Diagnosis = Downy Mildew]

Brown Spot

D8: Q_s([condition of leaves = abnormal]
[leafspots halos = with yellow halos, without yellow halos]
[leafspots margin = without watersoaked margin]
[leafspot size = greater than 1/8 inch])

$\wedge$

Q_c([time = May, Aug, Sep] [precipitation = normal, above normal])

::> [Diagnosis = Brown Spot]

Bacterial Blight

D9: Q_s([time = Apr, May, Jun, Aug, Sep]
[precipitation = (normal, above normal <--> [time = Apr, May, Jun]), (above normal <--> [Aug, Sep])]
[temperature = (normal <--> [time ≠ Aug]), below normal]
[condition of leaves = abnormal]
[leafspots halos = with yellow halos]
[leafspots margin = with watersoaked margin]
[leafspot size = less than 1/8 inch]
[leaf shredding = present])

::> [Diagnosis = Bacterial Blight]

Bacterial Pustule

D10: Q_s([time = Jun, Jul, Aug] [precipitation = normal, above normal]
[condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspot margins = without watersoaked margin]
[leafspot size = less than 1/8 inch]
[leaf shredding = present])

$\wedge$

Q_c [cropping history = crop same as last year, crop same for last three years, crop same for four or more years]

::> [Diagnosis = Bacterial Pustule]

Purple Seed Stain

D11: Q_s([time = Sep, Oct] [condition of seed = abnormal]
[seed discoloration = present] [seed size = smaller than normal])

$\wedge$

Q_c([time = Aug, Sep] [precipitation = normal, above normal]
[condition of leaves = abnormal])

::> [Diagnosis = Purple Seed Stain]

Anthracnose

D12: Q_s([time = Aug, Sep, Oct] [precipitation = normal, above normal]
[condition of stem = abnormal] [canker lesion color = brown]
[fruiting bodies = present]
[condition of seeds = (abnormal <--> [time = Sep, Oct]),
(normal, abnormal <--> [time = Aug])]
[fruit spots = absent, brown spots with black specks])

$\wedge$

Q_c [damaged area = whole fields]

::> [Diagnosis = Anthracnose]

Phyllosticta Leaf Spot

D13: Q_s([time = Apr, May, Jun, Jul]
[precipitation = normal, above normal]
[condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = without watersoaked margin]
[leafspot size = greater than 1/8 inch]
[leaf shredding = present])

$\wedge$

Q_c([damaged area = whole fields]
[temperature = (below normal <--> time = Jun]), normal])

::> [Diagnosis = Phyllosticta Leaf Spot]

Alternaria Leaf Spot

D14: Q_s([time = Jul, Aug, Sep, Oct] [condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = without watersoaked margin]
[leafspot size = greater than 1/8 inch] [leaf shredding = absent])

$\wedge$

Q_c([condition of fruit pods = normal, (diseased <-->
[time = Sep, Oct])] [fruit spots = (colored spots <-->
[condition of fruit pods = diseased])
[condition of seed = normal, (abnormal -->
[seed discoloration = present])])

::> [Diagnosis = Alternaria Leaf Spot]

Frogeye Leaf Spot

D15: Q_s([time = Jul, Aug, Sep] [precipitation = normal, above normal]
[condition of leaves = abnormal]
[leafspots halos = without yellow halos]
[leafspots margin = without watersoaked margin]
[leafspot size = greater than 1/8 inch])

$\wedge$

Q_c([fruit spots = absent, (colored spots <--> [time = Sep])]
[stem canker = above second node]
[canker lesion color = tan] [fruiting bodies = absent])

::> [Diagnosis = Frogeye Leaf Spot]

APPENDIX E

Figure 1 shows the results of testing the 15 sets of diagnostic rules (inductive and expert) to determine the accuracy with which they classified distinct descriptions of soybean plants with known diseases. If two or more rules satisfied the description, the specific description was multiply classified and the decision was not unique. The labels for the figure are defined as follows:

CORRECT ASSIGN — The correct decision for a testing set. The decision which should be assigned by the rules generated. Di implies that the disease description belongs to Class i.

INDECISION RATIO — The number of rules which were selected as best satisfying the descriptions of a class over the number of testing descriptions in the class. The measure is presented as a ratio. 1.0 means that only one rule was best satisfied for each testing description. Greater than 1.0 implies that more than one rule best satisified one or more testing descriptions, i.e. the higher the ratio the less unique the decisions for the testing of a class. Uniqueness does not imply correct assignment of decisions.

TIES — The number of testing descriptions which had two or more rules describe them equally well.

MAXIMUM RULES PER TIE — The maximum number of rules which were best satisified when ties occurred for a testing set.

TEST EVENTS — The number of events used to test the rule for the class associated with the CORRECT ASSIGN decision.

ASSIGNED DECISION — The percentage of decisions made for each decision class where CORRECT ASSIGN is the proper decision.

For clarity, the percent of correctly assigned diagnoses are shown between the diagonal lines in the figure.

Figure 1
Confusion Matrix Summarizing the Diagnosis of 340
Testing Events Using Discriminant and Descriptive VL Rules

CORRECT ASSIGN	INDECISION RATIO	TIES	MAXIMUM RULES PER TIE	TEST EVENT	ASSIGNED DECISION D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
Diaporthe Stem Canker(D1)	1.3	3	2	10	100							30							
Charcoal Rot(D2)	1.3	3	2	10		100				30									
Rhizoctonia Root Rot (D3)	2.0	10	10	10			100									100			
Phytophthora Root Rot(D4)	1.7	19	3	48		29	35	100	2										
Brown Stem Rot(D5)	2.4	19	2	24			71		100	66									
Powdery Mildew(D6)	1.0	0	0	10						100									
Downy Mildew(D7)	2.1	8	3	10						20	100		30				30		30
Septoria Brown Spot(D8)	2.4	48	4	52	27						25	100		23			65		
Bacterial Blight(D9)	2.2	9	4	10						10	70		100	40					
Bacterial Pustule(D10)	2.2	8	5	10							20		30	100			40	10	20
Purple Seed Stain(D11)	1.8	6	3	10					30		10		10		100	10		10	
Anthracnose (D12)	1.8	17	3	24					29						54	100			
Phyllosticta Leaf Spot(D13)	2.7	10	4	10							20	100		50			100		
Alternaria Leaf Spot(D14)	3.1	51	4	51							100						8	100	100
Frog Eye Leaf Spot(D15)	1.8	31	3	51			2	6	43		14							14	100