

PLANT/DS REVISITED : NON-HOMOGENEOUS EVALUATION SCHEMA IN EXPERT SYSTEMS

Carl T. Uhrik

Department of Computer Science
University of Illinois at Urbana-Champaign

ABSTRACT

This paper describes several deficiencies of PLANT/ds, an expert system for the diagnosis of soybean diseases. This production system employs both human and machine derived rules. The rule application mechanism for both rule groups suffer from improper treatment of incomplete data. The problem is illustrated by several examples, and a solution is proposed. Implementation of a new version of PLANT on the empty expert system ADVISE is seen as the crux of future research.

INTRODUCTION

Within the expert system PLANT/ds [2] for diagnosis of soybean diseases in Illinois, two rule groups operate simultaneously. One rule set is derived through the painstaking effort of plant pathologists to encode their knowledge in the form of machine executable rules. The other rule set is the product of the automated induction program Aq11 run on exemplary cases of the diseases [3]. PLANT has enjoyed relatively high popularity among the agronomists exposed to the system, yet the current rules suffer from a number of maladies.

From a pragmatic basis, two problems are of concern :

- 1) A rule may fail to recognize a true instance of a disease (false negative), or
- 2) A rule all-too-freely declares that some disease other than the correct one is present (false positive).

Only the latter of these problems is noticeably present. A typical situation is: "System reports most probable occurrence of disease is A with evidence that B, C, D, and E are present, B being most likely alternative and E being least likely. However, the expert making the inquiry states that disease C is the correct diagnosis and, moreover, that B is completely erroneous."

The "confusion" matrix, Fig. 1, summarizing some 340 test cases, shows this problem rather dramatically. The fact that the diagonal entries of the matrix are almost solidly 100% for both human and machine rules attests to the skill of the knowledge engineer and prowess of Aq11. Conversely, relatively high numbers in the off-diagonal entries of the table evince the severity of problem (2). This syndrome, loosely labelled

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

Correct diagnosis	Test cases	Assigned decision														
		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
<i>Diaporthe stem canker</i> (D1)	10	100										100			70	
<i>Charcoal rot</i> (D2)	10		100													
<i>Rhizoctonia root rot</i> (D3)	10			100	100											
<i>Phytophthora root rot</i> (D4)	48				100											
<i>Brown stem rot</i> (D5)	24			8	100			4						8	8	
<i>Powdery mildew</i> (D6)	10					100										
<i>Downy mildew</i> (D7)	10						100	90				30	90	100		
<i>Septoria brown spot</i> (D8)	52				10			100				88	100	100		
<i>Bacterial blight</i> (D9)	10							100	50			10	80	80		
<i>Bacterial pustule</i> (D10)	10							20	100	10		30				
<i>Purple seed stain</i> (D11)	10							40	10		100	10	60	60		
<i>Anthraxnose</i> (D12)	24				8							100			4	
<i>Phylloticta leaf spot</i> (D13)	10								100			100	100	90		
<i>Alternaria leaf spot</i> (D14)	51									100		22	100	100		
<i>Frog eye leaf spot</i> (D15)	51				2						100	82	4	100	100	

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

Correct diagnosis	Test cases	Assigned decision														
		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
<i>Diaporthe stem canker</i> (D1)	10	100									40				40	
<i>Charcoal rot</i> (D2)	10		100													
<i>Rhizoctonia root rot</i> (D3)	10			100							90					
<i>Phytophthora root rot</i> (D4)	48				27	8	100	2								
<i>Brown stem rot</i> (D5)	24								87			4			4	
<i>Powdery mildew</i> (D6)	10									100						
<i>Downy mildew</i> (D7)	10										80	100	30		30	
<i>Septoria brown spot</i> (D8)	52	37									40	100	38	37	90	
<i>Bacterial blight</i> (D9)	10										50	100	90		30	
<i>Bacterial pustule</i> (D10)	10										10	70	50	100	30	
<i>Purple seed stain</i> (D11)	10										20	10	10	80	60	
<i>Anthraxnose</i> (D12)	24	50									4	4		54	96	
<i>Phylloticta leaf spot</i> (D13)	10												20	100	50	
<i>Alternaria leaf spot</i> (D14)	51												39	100	20	
<i>Frog eye leaf spot</i> (D15)	51												4	39	63	

Fig. 1 - Confusion Matrix for Two Rule Sets [2]

* This work is supported in part by U.S. Dept. of Agriculture Grant No. 321512344 and in part by the Office of Naval Research Grant No. N00014-82-K-0186.

as "confusion," can be further conceptualized as:

- a) assigning diagnoses to diseases on the basis of evidence which is too weak, or
- b) assigning diagnoses to too many diseases at once.

Although the two cases overlap to a large degree, there is some distinction. Case a) amounts to "jumping the gun" due to excessive sensitivity in evaluation; case b) involves a lack of proper discrimination/filtering between the diseases either by the intrinsic nature of the disease or lack of proper granularity in the rules. Assuming a solution, in the first case we simply adjust the evaluation parameters, but case b) requires revision of the descriptors and/or rules.

PARTIAL SATISFACTION OF RULES

A rule is said to be *satisfied* when some of the user input data matches all of the *selectors* in the rule. The basic unit of a rule is a *selector* which tests whether a selector's variable takes a value from a reference set defined by the selector. Further, a weight or *confidence* can be assigned as the value of the selector, contingent upon the value of the selector variable. The concept of a selector is fairly intuitive, but for a complete discussion refer to [4]. Selectors combine via an AND function to produce *terms* which further combine under an OR function to produce the left hand sides of rules. Thus, a rule $[x_2=3][x_3=1,3] \vee [x_4 < 4] ::> [decision=A]$ can be interpreted: Decision A is taken if x_2 is 3 and x_3 is 1 or 3, or if x_4 is smaller than 4.

In a case of perfect knowledge of symptom descriptors, there is little doubt that the rules are more than adequate to discriminate various cases of the diseases. However, sparse information on the state of diseased plants requires efficient use of the little data available. Hence, an *evaluation scheme* is defined to handle the numerous cases of *partial satisfaction* of a group of selectors in a rule. Such a scheme includes the semantics of the AND and OR functions (eg., MIN/MAX) and the subsequent assignment of confidences to right-hand sides of rules. Fig. 2 describes the particular differences in the

structure and evaluation schema between the PLANT expert rules and induction rules.

The greatest confusion occurs either among the various leaf spot diseases or between the 3 leaf spot diseases and another disease. Since the leaf spot diseases have the longest rules (number of terms), one might expect a high potential for confusion in the case of partial satisfaction. Consider the following rules for downy mildew and brown spot:

```
[leafspots halos=p][seed mold growth=p][stem=n]
[leaf mildew growth=on lower leaf surface]
::> [diagnosis=downy mildew]
```

and

```
[precipitation>n][# yrs crop repeated>1]
[damaged area = not whole field][roots=n]
[leaves=abn][leafspots halos=no yellow halos]
[leafspots water soaked margin=abs]
[leafspot size>1/8"][leaf malformation=abs]
V [precipitation>n][leafspot size>1/8"]
[leaves=abn][leafspots halos=no yellow halos]
[leafspots watersoaked margin = abs][roots=n]
V [time=apr..jun][damaged area=not whole field]
[leaves=abn][leafspot halos=no yellow halos]
[leafspots water soaked margin=abs][roots=n]
::> [diagnosis=brown spot]
```

Both of these rules adequately cover the standard textbook cases of the two diseases. However, consider the situation where leafspot halos exist with no yellow halos, roots are normal, leaf mildew growth is observed, water soaked margins are absent, and leaf spot size is greater than 1/8". Using this data in the rules causes 3 of 4 selectors to be satisfied in the first rule. Let us assume the evaluation scheme for the INDUCTION RULES of PLANT (Fig.2), giving .5 confidence. Since many of the same selectors occur in the second rule, its confidence level must also be evaluated. The first term satisfies

```
[leaves=abn][leafspots halos=no yellow halos]
[roots=n][water margins=p]
```

for confidence of .33. The second term satisfies

```
[leaves=abn][leafspots halos=no yellow halos]
[roots=n][leaf spots>1/8"]
```

 for a confidence of .33. The third satisfies this again for .33, which brings the cumulative confidence to .95. Thus, there is good evidence for downy mildew, but brown spot is indicated as the most probable choice on the basis of this inflated confidence.

Evaluation :

Selector	S : $[x_1 \# R : w]$	1) A selector is assigned a confidence value of + if satisfied and - if unsatisfied or unknown.
Term	T : $S_1 S_2 S_3 \dots$	2) Confidence of a term is the AND of its selector confidence values.
Disjunctive Term	DT : $T_1 \vee T_2 \vee T_3 \dots$	3) Confidence of terms is the OR of their constituent term confidence values.
Linear Expression	$q_s \cdot DT_s + q_c \cdot DT_c$	4) Linear expressions arithmetically combine confidence of disjunctive terms.

Note : w, q_s, q_c are optional weights found only in EXPERT RULES.
 $\#$ denotes a relational operator.
R is a subset of x_1 's value set.

For expert rules :
(+ = 1, - = 0, AND = mean, OR = MAX,
 $q_s=0.8, q_c=0.2$)

For induction rules :
(+ = 1, - = -1, AND = mean, OR(x,y)= $x+y-xy$)

Fig. 2 - Structure & Evaluation of 2 Rule Sets

Granted, if we select the threshold low enough, we can empirically guarantee for large test sets that the correct diagnosis of the disease will always be indicated, but we are only increasing the confusion between similar diseases when we do this. The problem of soybean diagnosis may only be trivially solved for easiest cases (unless the cases the machine readily discriminates are different than those previously noted by soy pathologists).

This leads to a more general question of how commonalities between rules should be treated in the evaluation scheme. Groups of selectors can occur in clusters in more than one term in the same rule. The simultaneous satisfaction of such a group, as

```
[roots=n][leaves=abn][leafspot=no yellow halos]
in the above example, allows this one condition to
artificially inflate the significance of the rule
by raising the confidence value out of proportion.
It would seem that these terms should be factored
out into a separate rule. Often, this indicates a
significant generalization on its own and can be
assigned a meaningful name, as with the term
common to Brown Spot, Frog Eye Leaf Spot,
Alternaria Leaf Spot, and Phyllosticta Leaf Spot :
```

```
[leaves=abn][leafspot halos=no yellowhalos]
[leafspots water soaked margin =abs]
[leafspot size>1/8"]
::> leafspot disease ,
where leafspot disease can serve within a selector
with the proper weight in the contexts of the
previous occurrence of the cluster. Several of
these factorings of terms have been noted in Fig.
3 for both expert and induction rules.
```

Another problem exists in the homogeneity of the evaluation scheme. The leafspot diseases are simply too similar to assume that the same scheme for distinguishing the most diverse diseases will also distinguish the fine shades of the leaf diseases. There is a need for different evaluation schemes to evaluate different rules (perhaps topology dependent, eg. f(# of terms)). The average function may serve well as the default

since it has been empirically proven the best homogenous scheme, but the minimum function, which yields a stricter separation of confidences, suits the leafspot disease rules better. The switching between various levels of rules might be effected through the use of separate rule groups in the ADVISE system. The intersection (AND) of the exclusion sets of the selectors in a rule group may serve to define the left-hand side of a rule for switching rule groups. As soon as the confidence of that rule surpasses all the current active rule group, the new rule group would be activated (and optionally deactivate the other). See Fig. 4 for an illustration of this rule group switching.

The probabilistic sum used in the induction evaluation is certainly justified in rules where there are a small number of distinct terms, but the leafspot diseases indicate the flaw with trying to apply the technique too loosely. If there is only marginal evidence in a great many terms, we have suddenly inflated the significance of these cumulatively (in fact they probably have no significance when considered en masse). Since there is clearly a variant need, dependent on the rule, as to the choice of evaluation scheme, the scheme should be flexible in this respect.

CONTEXTS FOR SELECTOR AND SELECTOR GROUPS

As stated earlier, one shortcoming seems to be a sparse use of weight functions or weight values within selectors to establish the proper "context" for a selector or group of selectors. This is in fact a tiresome and tedious task accomplished through trial and error. We accomplish this somewhat heuristically in the present system. However, we might imagine an increasing wealth of knowledge being accumulated in an online learning environment for the proper weighting of terms based on an expert's interaction with the system. Now, the particular partial evaluation function for a term might likewise be learned in this manner.

Induction Rules	Expert Rules
[stem=abn][leaf malformed=abs] ::> stem disease	[leaves=abn][stem=abn] ::> extensive damage
[stem=abn][leaf=abn] ::> stem & leaf disease	[plant stand<norm][extensive damage]
[leaves=abn][leafspot=pres][leaf spot size>1/8"]	[roots=rotted][stem cankers=present]
[leafspot water soaked margin=abs] ::> leafcond1	[canker lesion color=brown,black] ::> root rot
[roots=norm][prec=>norm]	[leaves=abn][leafspot halos=no yellow halos]
[leafcond1]	[prec=>norm] ::> halo syndrome
[leafspot water soaked margin=pres]	[leaves=abn][leafspot halos=no yellow halos]
[leafspot halos=no yellow halos] ::> leafcond2	[leafspot water soaked margin=abs]
[leaves=abn][leafspot size>1/8"] ::> leafcond3	[leafspot size>1/8"] ::> leafspot disease
[leafcondition3][stem=a] ::> leafcond4	

Fig. 3 - Partial Term Factors Occurring in Multiple Terms/Rules

The following is a rule group
to distinguish between mildew diseases:

```
[leaves=abn][leaf malformation=abs]
[leaf mildew growth=on upper surface][roots=n]
    ::> [Diagnosis=Powdery mildew]

[leafspots halos=p][stem=n][seed mold growth=p]
[leaf mildew growth=on lower leaf surface]
    ::> [Diagnosis=Downy mildew]
```

The intersection of the exclusion sets of
the selectors in the above rules forms a left
hand side of a rule for switching rule groups:

```
[leaves=n][roots=abn][seed mold growth=abs]
[stem=abn][leafspots halos=abs]
    ::> [Rule Group#Mildew Group]
```

Fig. 4 - Formulation of Rule Group Switching Rules

DISCUSSION

Confusion arises at several levels, but the root of the cause seems to be the assignment of confidence values to the partial evaluations of rules which has been misguidedly forced to be homogenous between rules. Furthermore, expert rules were derived under somewhat ill-defined pretexts. When the human comes up with the rule, he is not aware of the consequences of the machine's partial evaluation of his rule. He is only considering the relevant variables in totality, not the exception driven cases necessary for accurate partial evaluation. There is a violation of knowledge granularity between man & machine since the man assumes (upon the recommendation of the intrepid knowledge engineer) that if he works at the highest level of his experience (i.e., complete knowledge), he can trust the machine to accomplish the task of exception handling (which he has built into his own knowledge base) in a manner similar to that which he himself would have used. In such a manner, results in the machine evaluation scheme appear that are quite different than anticipated. Moreover, since this inherent assumption is relatively easy to make and somewhat obscure, another expert may readily verify the rectitude of the first expert's rules but will fail just the same to see the reason for the rules not stacking up to expectations.

Surprisingly, humans possess an uncanny proclivity for exception handling. Children of age 2-7 encounter no difficulty in breaking down their internalized models of language and mores and physical reality to rebuild them in the light of conflicting environments until a model consistent with their surroundings is achieved. We continue this process throughout our adult life, but with a different emphasis and motivation. Exceptions trigger demon processes so subtly that often we are not aware of any exceptional condition having occurred. This is the true power of human thought. It no longer becomes necessary to have complete models when the exception handling is well integrated.

Another problem, which is more a fault of the expert than the machine, is that there does not seem to be a well balanced use of descriptors in the expert rules. Since the expert himself devises the descriptors, it is unlikely that the unused ones are irrelevant; they are simply irrelevant to this particular expert. Asking several other experts to augment the rules to try and add supporting cases and exceptions to our existing expert rules is one possible solution. In effect, this integrates multiple sources of knowledge [5] and avoids "inbreeding". Regardless of the approach, diversity in choice of experts will build a stronger system.

CONCLUSION

The prime deficiency of PLANT is currently its inability to utilize more than one evaluation scheme within a rule group. This does not allow simultaneous assessment of the diagnosis plausibility of several classes of diseases (e.g., the required evaluation scheme for stem diseases may differ from the scheme for leaf spot diseases but within a PLANT rule group only one evaluation scheme is allowed). Given sufficient real world data, this presents no problem to the system, but sparse or incomplete field data mandates optimal use of the user's input. Consequently, non-homogeneous evaluation is proposed to alleviate the problem. Non-homogeneity can be effected by imposing a supervisory control mechanism above the existing evaluation schema, which will ultimately dispatch the proper subgroup of rules once sufficient evidence precludes satisfaction of any other subgroup. Alternatively, a special rule group to accomplish rule group switching might be defined to direct the system to the proper subgroup with its associated evaluation scheme. The current ADVISE [Baskin & Michalski] implementation of PLANT is particularly conducive to this approach. In addition, the facilities in ADVISE, an "empty expert system", allow a wealth of other questions regarding evaluation schema to be quickly and painlessly determined.

REFERENCES

- [1] Chilausky, R. "A Prototype Computer Based Consulting System Using the Variable Valued Logic System VL1 : Methodology and Implementation," Thesis, Dept. of Comp. Sci., Univ. of Illinois, Urbana, January, 1979.
- [2] Chilausky and R. Michalski, " Knowledge Acquisition by Encoding Rules versus Computer Induction from Examples: Case Study Involving Soybean Pathology," International Journal for Man-Machine Studies, No.12, 1980, pp.63-87.
- [3] Larson, J.B. and Michalski, "Selection of Most Representative Training Examples and Incremental Generation of VL Hypotheses: The Programs ESEL and AQL1," Report No.877, Dept. of Comp.Sci.,Univ. of Ill., Urbana, May, 1978.
- [4] Michalski, "A Variable-Valued Logic : System VL1," Proc. of the 4th Intl. Symposium on Multi-Valued Logic, West Virginia Univ., Morgantown, May, 1974, pp.323-346.
- [5] D. Michie, "Expert Systems", The Computer Journal, Vol.23, No.4, pg.369, January, 1980.