Jurnal Teknologi, bil. 25, Disember 1996 hlm. 45–52

PLANT/COCOA: THE CONSTRUCTION OF EXPERT SYSTEM FOR DIAGNOSIS AND ADVISORY OF *THEOBROMA CACAO* PLANTATIONS

MOHAMED OTHMAN
HASHIM HASSAN
Department of Computer Science
Universiti Pertanian Malaysia
43400 Serdang
Selangor Darul Ehsan
Malaysia
E-mail:Mohamed@cs.upm.my

Abstract. This paper discusses the construction of an expert system called PLANT/-cocoa. It provides the consultations on the diagnosis and advisory *Theobroma cacao* plantations. Once the process of diagnosis is completed, this system could determine the existence of any disease on plants and then it will advise the planter to give the suitable method for treatments, preventions and detail descriptions of the diseases. Besides that, it provides the way of planting and replanting with the suitable type of clones. The expert system is implemented using a language called PROgramming in LOGic (PROLOG).

Keywords: Knowledge acquisition, Knowledge-Based, Inference Mechanism.

1.0 INTRODUCTION

During 1980's, few prototypes of an expert system in the field of agriculture was implemented successfully and one of the most popular was developed by Michalski, R. S. et al. [12]. The system is called PLANT/ds which provide the consultancy for the diagnosis of soybean diseases.

Nowadays, there are new technologies and sophisticated tools that have been used in the cocoa industry such as the method of tapping, chocolate production, planting and replanting the *Theobroma cacao* trees and others. As the technologies become more sophisticated, the chances are always more difficult for the planters even engineers to use such tools. Expert system, a branch of artificial intelligence that offers solutions to help diagnose and advise on matters relating to plantations. Computer is one of the new technologies and it uses the concept of artificial intelligence for implementing an expert system in this field.

There are many definitions of expert system. Among them had been defined by artificial intelligence experts such as Barr et al. [1], Buchanan et al. [4] and Bowerman et al. [3]. The author (1988) defines an expert system as a computer program that uses the concept of an artificial intelligence with high potentials of problem solving and behaves like an expert in the professional domain. It should be capable of explaining its decisions and the underlying reasoning.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

2.0 STATEMENT OF PROBLEMS

The major problems that should be known in the domain of plant disease is how to identify the main characteristics of recognizing the existence of any disease. Whenever the plants suffer from a disease it will go through fisiological processes that are changing from the normal state to a new state. It will show symptoms, faults and damages in the plants. It can be caused by the various agents acting singularly or in combination such as bacteria, fungi, viruses, pests, unfavorable environments and others which were mentioned in the text by Wheeler [20].

The main items to be considered are plants, faults and symptoms. Then it can be concluded that:

Plants show symptoms as a result of faults

A fault can be caused by different kind of symptoms in different plants. Usually the symptoms will be similar for most of these plants but occasionally the same fault will produce different symptoms in different plants. For instance, *white root disease* will show yellow leaves in sterculiaceae family, and yellow leaves and brown spots in euphorbiaceae family (Chan et al [5] and Edgar [8]).

Once the plants, faults, symptoms and modes of controls and their relations have been recognised, few questions arise such as:

i. Can we say the existence of a symptom shows the possibility that the plant is suffering from a disease?
ii. If the above question is justifiable, it is true that the existence of a symptom will cause the disease in serious condition or otherwise?
iii. If question (a) is true, is it possible that two or more symptoms will show the plant is suffering from a disease compare with only one symptom, especially if the symptoms have a low contributing factor. In other word, if yellow leaves are the identified symptoms, will all the yellow leaves show the degree of seriousness of the diseases compared with only few number of yellow leaves?
iv. Can a plant suffers from more than one disease at one time?
v. Is it true that at certain time, a symptom can be caused by more than one disease?
vi. Is there any case where the planter is not really sure of the plants having a specific symptom?
vii. Can we identify the existence of any disease with the existence of uncertainty value?
viii. Is it enough based on the uncertainty value for the expert system to belief or refuse of any disease suffer on the plants?

3.0 KNOWLEDGE ACQUISITION

Knowledge acquisition is a process of eliciting knowledge about the domain, usually from the expert and transforming it to be executed in a computer program. It is a part of knowledge engineering process and it is the most difficult, time-consuming and expensive task, regarded as the bottleneck. There are many eliciting methods used at this stage; 1) interview 2) protocol analysis and 3) text analysis.

One of the most important strategies of eliciting knowledge from expert is interview and it is divided into six different types of techniques.

i. Forward scenario simulation – an applied situation within a problem domain is selected and investigated under intensive laboratory program. The expert reports on relevant terms and concepts and describes the steps in problem solving and reasoning to achieve a goal.

ii. Goal decomposition – divide the overall problem into subgoals and asks the expert to describe the way for achieving the subgoals.

iii. Problem discussion – explore the kinds of data, knowledge and procedures needed to solve specific problems.

iv. On side observation – watch the expert solving real problem on the job.

v. Pure reclassification – expert statements are differentiated and classified into specific objects and relations between the objects. As a result, object-relations may be reclassified and new relations eventually discovered.

vi. Laddering – the expert is asked to name the important concepts of problem domain in term of questions. These concepts are then used as basis for the interview.

Protocol analysis refers to the automated or semi-automated analysis of thinking aloud protocols, that is, tape-recorded utterances of an expert during a problem solving process. The results can be considered as a path through successive knowledge states representing the sequence of problem solving events. Here, the success of any protocol analysis depends on the quality of the recorded information, detailed and adequate instructions with respect to the attainment of protocol recordings of verbal utterances during the problem solving process.

Text analysis is a phase where the knowledge engineer start to stury and analyse manuals, books, journals, newsletters, pamphlets and documents on the problem domain (Haddon [10], Keane [11], Chan [5]). This can be very time consuming, particularly if the knowledge engineer is supposed to become an expert on that domain before beginning his or her actual work.

Once the incomplete knowledge arise, the knowledge engineer should watch and observe the actual events for the missing components and recovered it. The observations must clear and if it is possible try to use the recording facilities.

The knowledge engineer should discuss along with the expert to solve any problems that may arise, even in the near future. He must avoid any problems concerning the common sense and fuzziness, and as a result, the process of eliciting knowledge is easier and effective.

4.0 DESIGN OF PLANT/COCOA EXPERT SYSTEM

The PLANT/cocoa expert system is implemented by using PROLOG language. This is due to its suitability and manipulating the expert's knowledge. In other word, it has been designed and used with the software engineer's mind rather than computer. It allows the engineer to set up a precise definition of the problem and it enables the computer to apply fundamental mathematical or logical reasoning to solve any problems.

In general, this expert system consists of three major sections, user interface, inference engine and knowledge base as shown in Figure 1. The user interface uses the concepts of menu-driven, pop-up window and question-answering system. It consists of front end and justifier. The front end is representing a language processor which uses the English language as a tool for interaction, and the justifier will answer any questions about reasoning and explanations. For instance, if the user asked on the reasoning such as how, why (figure 2), and what else, the system will analyse the word(s), before making meaningful respond.

The process of analysing includes modelling the question into an understandable form, follow by matching with the best fact and rule in the global memory to find the best answer, and if it failed, the system will pick up the alternative answer. The system also provides

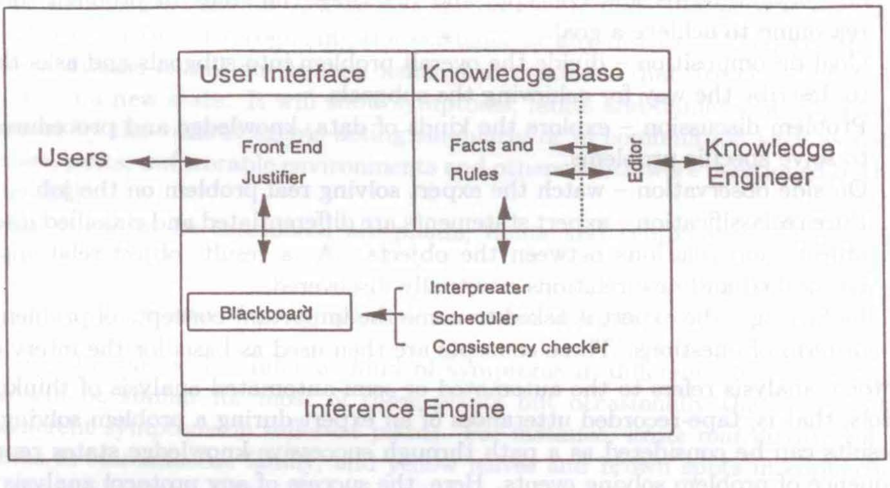

Fig. 1 Architecture of PLANT/cocoa Expert System

additional facilities such as help, ignore unnecessary questions and reasoning for particular question. Levels of steps have been provided for ignore unnecessary questions; 1) skip the rest of the questions, 2) jump to the next group of questions and 3) going to the next question. With these facilities, the users can choose any questions which are relevant to his or her observations. These facilities are very useful and as a result, it will make the user more confident and satisfied with the answer given by the system.

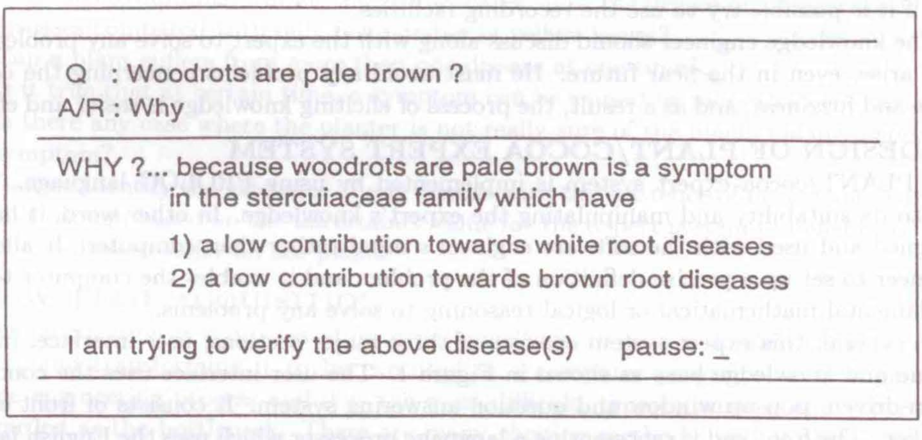

Fig. 2 Reasoning for particular question

The knowledge base is a storage of facts and rules on the problem domain, both in general and modular form. Knowledge of preventions, treatments, and detail of the diseases are modeled in the form of several facts. While the knowledge of the diseases and symptoms are represented in a collection of production rules in the form of IF-THEN. For example in Prolog form, it can be shown as in Figure 3(a). Rule 2 in Figure 3(a) can be converted into the English sentence as shown in Figure 3(b). Whenever the IF statements is fulfilled by the facts about symptoms that was selected by the user then the THEN statement is executed. This execution can be said that the production rule is fired.

The construction and debugging of the knowledge base is the most difficult and time consuming and it is very expensive task. Several approaches have been explored for eliminating the knowledge acquisition bottleneck. The editor can be used to aid in extracting knowledge from the experts that does some clever consistency checkers against the existing knowledge base. It also can be use to add new rules or facts to the knowledge base, beside maintenance the existing knowledge. The knowledge engineer and the expert are fully responsible to make the knowledge base up-to-date.

The inference engine is a mechanism to show conclusions based on data entered, rules and facts in the knowledge base and other information. It can be said as a driver program for the system and it is problem independent so that it does no vary from one expert system to the other. The main consideration regarding the information flow in the system, is how to ask the planters concerning to the questions relevant to his or her plants. The most important factor which influence this, is the knowledge that the plant may suffer from more than one disease at any time.

A global memory contains the emerging partial solutions. The knowledge in the memory is divided into several separate knowledge sources that can be observed and modified. They cannot communicate among themselves. The interpreter will interpret the rules and facts in order to infer new knowledge and a scheduler that decide the order in which the rules should be applied. The consistency checker will checks the semantics of the rules and data to see if there is any conflict with the knowledge in the global memory.

The concept of heuristic is used for manipulating the rules, facts, and data in order to make conclusions and their reasoning strategy in term of exact solution, and several alternative solutions. It also used the combination of backward chaining and breadth-first search in and AND/OR graph. This sort of technique can give direct conclusion and very systematic approach during the consultations.

```
rule(2, 'red root diseases',
     [sterculiaceae],
     [leaves('are yellow'): 2,
     leaves('are falling'): 2,
     roots('are covered with red brown mycellium'): 8,
     fungus('are growing margin usually creamy white'): 5,
     brackets('are dark brown on the upper surface'): 4
     ]).
```

Fig. 3(a) A rule 2 in Prolog form

```
Rule 2: IF plants are belong to sterculiaceae family and
   they are show the following symptoms:
   leave are yellow : 2 AND
   leaves are falling : 2 AND
   roots are covered with red brown mycelium : 8 AND
   fungus are growing margin usually creamy white : 5 AND
   brackets are dark brown on the upper surface : 4
 THEN plants are suffering from red root diseases.
```

Fig. 3(b) A rule 2 in English form

5.0 COMPUTATION OF UNCERTAINTY VALUES

Many concepts have been used for computing the uncertainty values such as Bayesian inference, the concept of probabilities and others. For example, PLANT/ds is an agriculture expert system which used its own reasoning mechanism as mentioned by Michalski [12].

The concept of true, false and unknown is used to compute their uncertainty values in this system. The idea is to promote relationship with the concept of connectionist model in the future as mentioned by Wielinga [21], Gallant [9], and Spiegelhalder [17]. It can model the expert's knowledge in the form of networks.

This system used the above concept to indicate its belief, the answer is to be true which was discovered from the expert's belief, while the severity value represents the planter's belief which was neglected by this system. Thus, it is still being used to compare between the expert's and planter's belief.

The computation of uncertainty and severity values can be put into mathematical expressions as follows

$$UV = \sum_{i=1}^{n} A_i \times AV_i$$

and

$$UV = \sum_{i=1}^{n} A_i \times CV_i$$

where i = selected symptom
UV = uncertainty value
SV = severity value
A_i = time
CV_i = user certainty value for symptom i
AV_i = associative value for symptom i

The degree of beliefs in a proposition are based on three different values; 1) a constant, 2) an associative value and 3) a user certainty value. These balues show certain level of seriousness of symptoms present in the plants. The first two values were identified by the expert through his or her experience, observation and knowledge of expertise. While the third value was identified by the planters through his experience and observation.

Whenever a symptom, i is selected, the system will asked "How sever is the symptom?" and the planter has to enter a value to show certain level of seriousness of the symptom present at that time. This value ranged from one to nine to indicate that there may be a belief and it is called a user certainty value, CV_i. Indirectly, a constant A_i is selected and the value is equal to ten. This indicates the totally belief otherwise the value is zero, indicates the disbelief, as describe below:

$$A_i = \begin{cases} 10 & \text{if symptom } i \text{ is selected} \\ 0 & \text{if symptom } i \text{ is not selected.} \end{cases}$$

A constant A_i is proportional to the value of a user certainty CV_i for every symptom i. The different is that the value of A_i is used to compute the uncertainty value whereas the value of CV_i is used to compute the severity value. Every symptoms has their own associative value AV_i to show the seriousness of the symptom present and it was assigned by the expertise. The manipulation of these three values were not based on any rigorous mathematical theory although it is intuitively reasonable. Actually, there are many factors influenced in the computation of the above values:

(i) Each symptom has its own value showing the degree of seriousness of the symptom present.

(ii) The combination of any value from several symptoms will determine the present of the disease with a stronger belief compared with a symptom.

On the above computation, it can be concluded that if the graph has one branch, another word only one symptom then the degree of belief of the evidence that the plants are having any disease is less confidence. On the contrary, if the graph have few branches, then the degree of belief of the evidence that the plants are having any diseases is close to the answers given by the experts.

6.0 CONCLUSION AND FUTURE WORKS

Research efforts on application of expert system programming techniques in the domain of cocoa industry are still in progress and it is hoped that more system will be developed in the near future. With reference to the existing expert system developed so far, perhaps it can be conclude that the emergence of truly expert system is far ahead. Although the existing expert systems are able to shows some levels of expertise, however their capability are still limited, proposition of the processed by software engineer.

Progress in this field to date has been very interesting and new research avenues in the use of expert system in cocoa industry are found which includes the knowledge of soils, minerals, weathers and natural disasters. Many enhancement has been made which include interfacing with other packages together with help facilities. Currently, the neural networks package is been used for analysing data on weather, see Neuralware Inc., [15].

7.0 ACKNOWLEDGEMENT

The author is grateful to all staffs from Department of Plant Protection, Universiti Pertanian Malaysia for their co-operations and contributions during the implementation of this project.

REFERENCES

[1] A. Barr & E. Feigenbaum, *The Handbook of Artificial Intelligence 1*, Addison-Wesley Pub. Co., 1981a.

[2] A. Barr & E. Feigenbaum, *The Handbook of Artificial Intelligence 2*, Addison-Wesley Pub. Co. 1, 1981a.

[3] R. G. Bowerman & D. E. Glover, *Putting Expert System into Practice*, Van Reinhold Co., New York, 1988.

[4] B. G. Buchanan & E. H. Shortliffe, *Rule-based Expert System: The MYCIN Experiment of the Stanford Heuristics Programming Project*, Addison-Wesley Pub. Co., New York, 1984.

[5] C. L. Chan & B. S. Lee, *A Preliminary Survey of Cocoa Diseases in West Malaysia*, MARDI Res. Bull. 1(2), 1973.

[6] W. F. Clocksin & C. S. Mellish, *Programming in Prolog*, Springer-Verlag, Berlin, Heidelberg, Germany, 1984.

[7] R. J. Duda, J. Gasching & Hart, *Model Design in the Propector Consultant System for Mineral Exploration in: Expert System for Microelectronic Age*, Michie D, Edinburgh Press, 1979.

[8] A. T Edgar, *Manual of Rubber Planting (Malaya)*, Percetakan Season Sdn. Bhd. Kuala Lumpur, 1958.

[9] S. I. Gallant, *Connectionist Expert Systems*, Communications of the ACM, 31(2), February 1988.

[10] A. V. Haddon, *Variety Trails of Seeding Cocoa in Malaya*, Malaysian Agriculture Journal, 43, 1961.

[11] P J. Keane & P. D. Turner, *Vascular Streak Dieback of Cocoa in West Malaysia*, Proc. Int. Soc. Planters Conference.

[12] R. S. Michalski, J. H. Davis, V. S. Bicht & J. R. Sinclair, *PLANT/ds: An Expert System Consulting for the Diagnosis of Soybean Diseases*, Proceeding 1982 European Conference on Artificial Intelligence, 1972.

[13] Mohamed Othman, *A Prototype of an Expert System for Diagnosing the Rubber Plant Diseases in Malaysia*, Proceeding of the Third National Symposium on Mathematical Science, UPM, 1988.

[14] Mohamed Othman, *Logic and Prolog Database System*, Journal of Technology, December 14, 1989.

[15] Neuralware Inc, *NeuralWorks Professional II/Plus*, USA, 1989.

[16] Nicholas D. Stone & Bernard A. Engel, *Knowledge-based System in Agriculture and Natural Resource Management*, AI Magazine, 11(3), 1990.

[17] D. J. Spiegelhalder, *A Statistical View of Uncertainty in Expert System in Artificial Intelligence and Statistics*, Gale A. W., Addison-Wesley Pub. Co., 1989.

[18] P. H. B. Talbot & P. J. Keane, *Oncobasidium a New Genus of Tulasnelloid Fungi*, Aust. J. Bot, 19, 1971.

[19] J. F. Traub, G. W. Wesilkowski & H. Wozniakowski, *Information, Uncertainty, Complexity*, Addison-Wesley Pub. Co., New York, 1983.

[20] B. E. J. Wheeler, *An Introduction to Plant Disease*, Wiley Co., London, 1978.

[21] B. J. Wielinga & J. A. Brouker, *Model of Expertise in: Advance in Artificial Intelligence II, B. D. Boulay and J. Steels*, Elsevier Science Pub. Co., 1987.

[22] 80386 Prolog-2 Language Reference Manual, issue 2 (1986).

[23] Expert System Ltd, *80386 Prolog-2 Option Reference Manual, issue 2*, 1986.