Bridging Decision Tree Data Mining Model to the Automated Knowledge Base for Rice Plant Agriculture Expert System

Divya Joy
Mtech computer and Information Science,
College Of Engineering Poonjar
India

Sreekumar K.
Prof.Department of Computer Science and Engineering, College of engineering Poonjar,
India

ABSTRACT
Agriculture is one of the most important inventions of human civilization. It is considered to be a primary occupation of the major portion of world population. Agriculture domain is moving towards a production decline for the few decades. Most of the currently available methods and materials in rural agriculture field have been used for decades. The objective of this work is to focus on bringing new techniques in rural area to improve overall agriculture production and management using expert system technology. This paper presents a new methodology for the expert system design and uses a novel approach for the development with some data mining technique and implements a rule based expert system for rice plant. The main idea is that in addition to the predefined rule set already stored in a typical expert system, the proposed system includes the rules that are automatically generated from data set instances of agriculture field. Here the rule generation is done through a process of decision tree based data mining technique. The generated as well as predefined rules together form the automated knowledge base for the rice plant expert system. So this work presents the development of an integrated system for agriculture domain by bridging the data mining aspects with expert system technology. The system is intended to support the decision making and problem solving in respective domain especially for the rural area farmers.

General Terms
Expert system

Keywords
Agriculture, expert system, data mining, decision tree

1. INTRODUCTION
The application of innovative techniques in agriculture field has been on under development for the past few years. This is due to the wider gap between the proportion of the technologies developed by the agricultural research system and technologies adopted by the rural farmers. One of the vital reasons for such wide gap is due to the lack of education and unawareness of emerging technology. In order to bridge the gap between the technologies and the farming community, it is necessary to develop suitable approach for the appropriate extension/technology transfer that can establish better linkage between scientists and farmers in order to facilitate effective production. Information technology opens a door as solution for these problems. Information technology improves the efficiency and productivity of agriculture and facilitates easy agriculture operations. IT can bring qualitatively improved field life by providing timely and quality information inputs for decision making and problem solving by means of a technique-Expert system.

Expert System is one of the important and useful application branches of Artificial Intelligence. An Expert system is an intelligent system or a computer program, which attempt to simulate the specific expertise and reasoning ability of a human expert in a specific domain. They are widely used to solve the complex problem in multiple domains, such as medical field, agriculture, geology, space technology oil exploration, etc. They are mostly useful and helpful in situations where the human expert is not readily available. Expert Systems play a major role in making the useful expert's knowledge available to non expert users, that leads to economic growth and higher standards of living. Expert system technology has been a great support for making decisions in agriculture domain.

This paper deals with a novel approach to make the commonly used agriculture expert system more rich and flexible by using the data mining concepts. Agricultural data increases day by day, hence data mining has achieved a great deal of attention in recent years in agriculture field due to the large availability of wide range of data and the need for extracting useful information and knowledge from that data. This extracted knowledge can be used to aid decision making in various processes in respective domain. Here use the most popular data mining technique decision tree mining to extract the useful knowledge. Decision Trees are widely used classification technique. Its important feature is that it is easy to interpret and explain to users.

2. RELATED WORKS
The application range of expert system is rapidly increasing. Such applications are very useful and effective in situations when the domain expert is not readily available. Many researchers have tried to use data mining as well as expert system technologies in many areas related to agriculture. Expert system can be used in many fields including diseases diagnosis, pest controls, and crop protection, irrigation scheduling, soil protection, market trend analysis etc. Here mention some expert systems developed in agriculture field. exoWHem- "Expert System on Wheat Crop Management" developed by Islam , Division of Computer Applications,IASRI, New Delhi[1]. "AMRAPALIKA: An expert system for the diagnosis of pests, diseases and disorders in Indian mango" is an expert system application in the agriculture domain developed by Rajkishore Prasad, K.R.Ranjan and A.K.Sinha [2]. Another expert system "Dr. Wheat: A Web-based Expert System for Diagnosis of Diseases and Pests in Pakistani Wheat," this expert system developed for the purpose of pest and disease control of Pakistani wheat by F.S.Khan, S.Razzaq, K.Irfan, F.Maqbool, etc. [3]. "Expert System in Detecting Coffee Plant Diseases" developed by Derwin Suhartono eta [4]. " Applying data

mining techniques in the field of agriculture and allied sciences" by Yethiraj N G [5] is paper describing various data mining techniques used in agriculture**. "**Decision Tree Induction Approach for Data Classification Using Peano Count Trees" by B V Chowdary, Annapurna Gummad [6] is a research study on application of data mining. "Study and Analysis of Decision Tree Based Irrigation Methods in Agriculture System" by Ravindra M1 , V. Lokesha2 , Prasanna Kumara3 , Alok Ranjan[7] is a detailed study on the decision tree classification methods of data mining .

3. SYSTEM DESIGN

Expert systems are typically composed of three major components [8]

1. Knowledge Base
2. User and Expert Interface
3. Inference Engine

The knowledge base is considered as the heart of expert system. It is the integration or repository of knowledge collected from the domain experts and facts about the domain. Interfaces are provided to interact with the system; user can input queries and get result through the same interface. The inference engine is known as brain of expert system. It is the main problem solver, which performs the inference process.

3.1 Design framework

Agriculture knowledge mining [9] is an emerging area of computational intelligence applied to accurately analyze the agricultural case studies and to perform diagnosis and protection procedures. Various data mining methods [10] have been proposed to extract useful knowledge from agriculture data, but main techniques adopted by many researchers are rule induction and classification tree generation. The proposed design of system frame work is given in figure1.

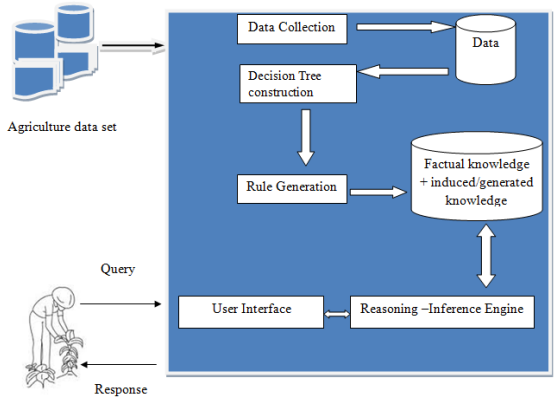

Figure 1.System design framework

The framework composed of four main components: data collection, tree induction, rule generation, and the inference engine. Data integration component is responsible for collecting data from different sources, format transforming, structuring, cleaning etc. These data are to be used by the tree induction component. It results a tree structure that represents the collected valid and useful information. Rule generation component evaluate this tree structure and generates useful rules based on certain criteria or threshold value and form the decision rules. These automatically generated rules as well as the stored rules together used by the inference engine component and performs the reasoning.

4. PHASES OF DEVELOPMENT

To integrate the two computer science application techniques, data mining and expert system technology the overall system should be implemented with three phases. Expert system works based on certain knowledge and facts stored with it. This knowledge is collected using some data mining aspects. In order to build the system required data should be selected from the given data set.

- Decision tree Induction/Construction
- Rule generation
- Reasoning Technique

4.1 Decision tree Construction

Classification is one of the most important tasks regarding data mining. The aim of the classification is to build a model based on some test cases. Then the model is used to predict the target class of new cases. Among various mining algorithms decision tree algorithms is the most commonly used because it provides a modelling technique that is easy for human to understand and that simplifies the classification task. A decision tree is a representation of a decision procedure for determining the class of a given instance.

First phase includes the mining of valid knowledge from the given data set of rice plant and builds the decision tree. In order to build a decision tree [11] from the given data set of case instances, we need to choose the best attribute that contributes the most towards partitioning data to the target or most suited class. For our project the attributes representing the climatic conditions, soil conditions, various symptoms of a particular disease class. The best test attribute is selected by examining their count of occurrences in case instances.

4.1.1 Algorithm

Input: agriculture data set

Output: decision tree

1. Initialization
   Clear knowledge base (KB)

   Scan disease data set to get information about data attributes and instances.
2. Building tree
   Compute count of occurrences of value of each attribute

   Choose the attribute that has highest count value

   Assert edge structure and node information into KB
3. Output
   Output the tree as the decision tree with respective nodes and edges

The attribute with highest value become the top node that means which has highest frequency of occurrence in case instances.

4.2 Rule Generation

In this project for the proposed system framework, here we use a new method for knowledge base creation. Tree induction component gives knowledge mining results as tree structure, which is transformed into a small set of decision rules. The resultant tree is analysed by the rule generation component and traverses each tree branch in order to calculate the similarity of path occurrence. This similarity is defined as the probability or frequency of occurrence of cases. The generated probabilistic rules are then sorted with the probability of case

occurrence. Rules with highest probability value will be used to help the non expert users for making decision about disease diagnosis of rice plant.

4.2.1 Algorithm

Input: decision tree and a probability threshold

Output: set of decision rules ranking with the highest probabilities

1. Scan and traverse tree branches from a root node to each leaf node
2. Count number of data instances and collect edge information.
3. Compute probability as a proportion

   (Number of instances at leaf node) / (total data instances in a data set)
4. Assert a rule with the format (attribute-value pair, target class, probability value) into KB
5. Sort rules in the KB in descending order according to the rules probability
6. Select rules that have probability greater than the specified threshold value.
7. Store the selected rules into the KB and return KB as an output

4.3 Reasoning Technique

The rule generation component results a set of decision rules. These generated rules and some set of predefined rule together form the knowledge base .Hence for the development of agriculture expert system for rice plant the automated knowledge base is formed form the given set of rice plant data instances. The respective rules stored in knowledge base will provide direction for reasoning process.

4.3.1 Algorithm

Input: KB containing mined knowledge and stored facts, and a new case

Output: most suited decision for new case

1. Read all observation conditions (attribute-value pairs) given in the current case
2. Compare the conditions with relevant rules in the KB
3. Compute the decision possibility as

   (Number of matched attribute-value pair) * (probability of the decision rule)
4. Output a final decision based on the percentage of possibility

The real forte of expert systems is their capacity to make inferences or the drawing of conclusions from premises. This is precisely what makes an expert system intelligent. The proposed system is intended to support decision making and problem solving in agriculture domain. In this expert system the basic reasoning process take place based on the generated decision rules. Reasoning process mainly aims to complete the various processes provided in user interface such as disease diagnosis, plant protection, crop selection, and market trend analysis.

5. SYSTEM IMPLEMENTATION

The project titled Bridging data mining model to the automated knowledge base of agriculture expert system aims to develop an efficient and effective expert system for rural farmers in order to make the rice plant production and management better and act as a virtual agriculture expert for providing right advice and recommendations at right time. The proposed system for rice plant cultivation is sometimes called a consultation system. The user can ask questions and reach conclusions. This consultation system has two components: a questioning strategy and a decision making strategy. The questioning strategy interviews a person with a problem, and then examines structure of the decision rules and its satisfied components to determine the most possible next question. The decision making mechanism is a process of using right rule at right time from a set of decision rules. In order to make this process more accurate the knowledge base should contain valid and accurate rules. This can be done with decision tree data mining technique.

To implement the designed system frame work of rice plant agriculture expert system it is necessary to implement all the process done by various system components, data collection component, tree induction component, rule generation component and inference engine component

5.1 Decision-tree construction

Based on the facts and data set about disease on rice plant and the symptoms, the construction of the decision tree is shown in Figure 2.

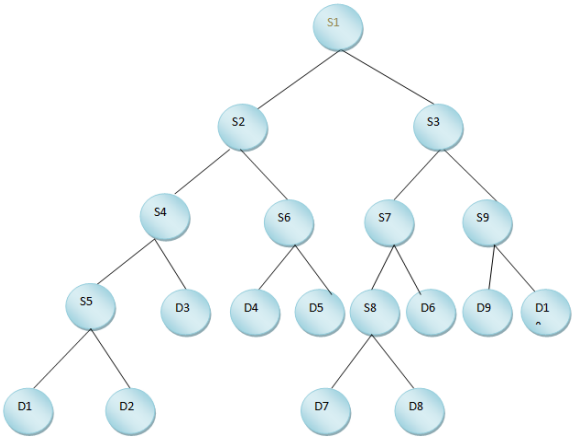

Figure 2. Decision tree structure

This is the corresponding tree structure constructed from the given rice plant disease data set. The respective disease and symptom descriptions are given in table below. Table 1 describes various observation conditions includes the symptoms, climatic and soil conditions of the rice plant diseases which is represented by the symbols start from S1 to S9. Table 2 represents the various diseases found named with D1 to D10. The system starts with various observation conditions and is moving forward to the appropriate decision. Here the decision represents the rice plant diseases (such as Sheath Blight, Brown Spot, rice blast. grain discolouration etc).

Table 1. Symptom Description

Symptom	Symptom description
S1	Circular or oval spot on leaf under normal or humid condition
S2	Greenish / grayish lesions under dry land or wet land

S3	Brown lesions under purely rain fed
S4	Purple color margin
S5	White centers with poorly filled grain or dead of leaf blade
S6	Ash color centre with wilting or rolling up
S7	Reddish brown margin on leaf under terraced land
S8	Yellowish color or brown color spot on grains
S9	Under humid condition orange-pinkish color grains or black discoloration of glumes

Table 2.Disease Description

Disease	Disease Description
D1	Sheath Blight
D2	Rice Blast
D3	False smut
D4	Bacterial Blight
D5	Leaf Blast
D6	Stem Rot
D7	Rice Turngo disease
D8	Brown Spot
D9	Grain discoloration
D10	Bacterial leaf streak

From this decision tree structure decision rules are generated based on threshold value measurement. One of rules generated by the proposed algorithm is as follows

#sheath blight
{
Circular or oval spot on leaf under normal (yes)
Greenish / greyish lesions under dry land (yes)
Purple colour margin (yes)
White centres with poorly filled grain or dead of leaf blade (yes)
}

The generated rules compare with the observation conditions reported in the case studies of given data set. Then count of occurrences are calculated, the rules having the similar case instances above the given threshold value transformed as decision rules.

When the user approaches the system with query then the coming observation conditions are examined and compared with available decision rules in knowledge base. Then system draws the most appropriate decision for that specific case.

6. EVALUATION RESULTS

The correctness and performance of developed expert system can be evaluated with real time application. The best appropriate evaluation is comparing the results obtained with proposed system and existing system with decision taken by a human agriculture expert. For this evaluation the developed agriculture system is presented to rice plant farmers and studied problems of rice plant production and collected some observation conditions from paddy field. These conditions include symptoms, soil and climatic conditions. Since this current work aims to transfer the technology and techniques used in research work to rural area field it should be tested and evaluated with real time cases in rice plant field. For this purpose we visited rice plant farmers from different areas with different soil, climatic condition, plantation type. And collected commonly occurred as well as rarely occurred disease conditions and symptoms randomly.

To measure the performance we collected a test set containing 20 different real time cases of observation conditions from rice plant growers obtained from different rice field. These observation conditions of each case instance are given as user query for the developed rice plant expert system. And obtained most possible decision, here decision represents the most suited disease name. Then the same set of case instances again tested with the existing system of general agriculture expert system. The decisions made by current work of developed agriculture system and the general existing expert system for rice plant with the observation conditions are evaluated by an agriculture officer, specialised on disease diagnosis. Rice plant disease diagnosis expert examines the observation conditions of each case instance and give the correct decision or disease name using his domain knowledge. Corresponding results are given below in Table 3.

In the result table the blue coloured decision is marked as undetected cases. And red coloured decision represents incorrect decision. All the results made by both the systems are evaluated with agriculture expert. Of the 20 samples of cases 18 cases are detected and of the 18 detected cases it could make the 17 correct decisions. This result is compared with performance of existing agriculture system. It can detect only 13 samples of cases and of the detected cases only 9 cases are correctly measured.

Table 3. Evaluation Result

Test Cases	General ES	Proposed ES	Human Agriculture Expert decision
1	Sheath Blight	Sheath Blight	Sheath Blight
2	Rice Blast	Rice Blast	Rice Blast
3	Undetected	False smut	False smut
4	Sheath Blight	Bacterial Blight	Bacterial Blight
5	Undetected	Leaf Blast	Leaf Blast
6	Rice Blast	Rice Blast	Rice Blast
7	Undetected	Undetected	Rice Turngo
8	Sheath Blight	Sheath Blight	Sheath Blight
9	Undetected	Grain discoloration	Grain discoloration
10	Rice Blast	Rice Blast	Rice Blast
11	Undetected	Undetected	False smut
12	Undetected	Leaf Blast	Leaf Blast
13	Brown Spot	Brown Spot	Brown Spot
14	False smut	Brown spot	Stem Rot
15	Rice Blast	Rice Blast	Rice Blast
16	Sheath Blight	Sheath Blight	Sheath Blight
17	Brown Spot	Brown Spot	Brown Spot
18	Undetected	Grain discoloration	Grain discoloration
19	Rice Blast	Rice Blast	Rice Blast
20	Sheath Blight	Sheath Blight	Sheath Blight

From the above experiments the performance evaluation of developed system and general system with human expert is shown using a graphical representation. The measures used for the comparison are detection rate and error rate.

- Detection rate

It is measurement of no of cases detected by a method of system when test set is applied. For the existing system 13 cases are detected and the proposed expert system detected 18 test cases. Result is shown in Figure (3).

Proposed agriculture expert system:
*(18/20) * 100% = 90%*
Existing agriculture expert system:
*(13/20) * 100% = 65%*

- Error rate

It is measurement of no of cases accurately diagnosed by system when test set is applied. For the existing system 9 cases are accurately diagnosed of 13 detected cases and the proposed expert system detected accurately diagnosed 17 test cases of 18 detected cases. Corresponding result is given in Figure (4)

Proposed agriculture expert system
*(1/18) * 100% = 5.55%*

Existing agriculture expert system
*(2/13) * 100% = 15.38%*

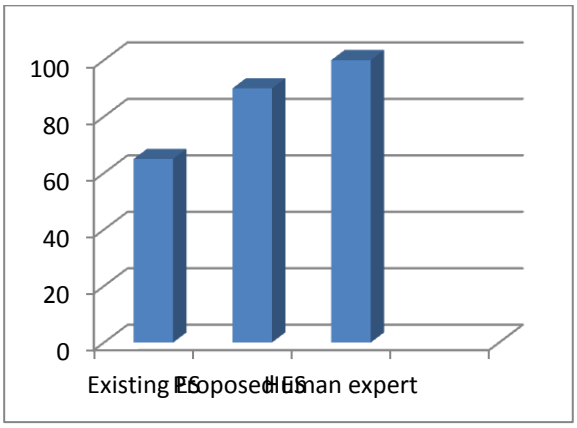

Figure 3. Detection rate

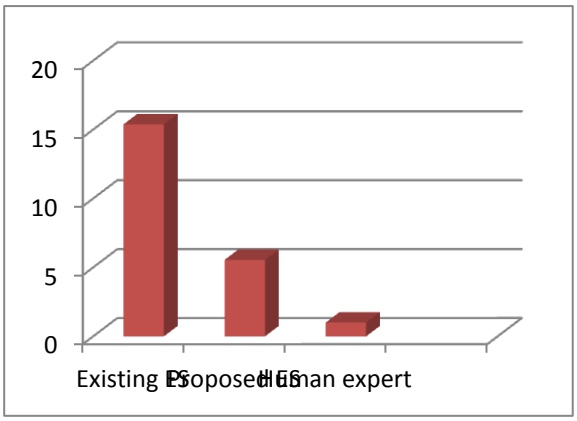

Figure 4. Error rate

7. CONCLUSION

This work attempted to play a role of a virtual agricultural expert who can give timely expert advice to a large community of rice plant farmers, specific to their need in the form of agriculture expert system. Agricultural knowledge mining is a research area that employing decision tree data mining in the current work to generate the most accurate and useful rules and ease the agriculture problem solving and decision making process. Hence this work achieved the goal by bridging data mining aspects to an automatically generated knowledge base for a rice plant agriculture expert system. So by this way the project work will help in transfer of emerging technology to rural area for achieving better agriculture production and management will contribute to enhancement of the marketing level of agriculture product.

This expert system is developed purely based on the rice plant production and management. This work can be extended by making agriculture expert system for multiple crops and plants and include novel methods for rule generation and reasoning and decision making steps.

8. REFERENCES

[1] Islam et.al. IASRI, New Delhi. 2010, Expert System on Wheat Crop Management.

[2] Rajkishore Prasad, K.R.Ranjan and A.K.Sinha 2006. AMRAPALIKA: An expert system for the diagnosis of pests, diseases and disorders in Indian mango.

[3] Fahad Shahbaz Khan eta Proceedings of the World Congress on engineering 2008 Vol I. Dr. Wheat: A Web-based Expert System for Diagnosis of Diseases and Pests in Pakistani Wheat.

[4] Derwin Suhartono et al 2013 Expert System in Detecting Coffee Plant Diseases.

[5] Yethiraj N G 2012 Applying data mining techniques in the field of agriculture and allied sciences.

[6] B V Chowdary, Annapurna Gummad 2012 Decision Tree Induction Approach for Data Classification Using Peano Count Trees.

[7] Ravindra M1, V. Lokesha2, Prasanna Kumara3, Alok Ranjan 2012 Study and Analysis of Decision Tree Based Irrigation Methods in Agriculture System.

[8] Ohio Journal of Science: Volume 90, Issue 5 (December, 1990) Research Review: Application of Expert Systems in the Sciences.

[9] Raorane A.A 2013Review- Role of Data Mining in Agriculture.

[10] Hetal Patel 2014 A Brief survey of Data Mining Techniques Applied to Agricultural Data".

[11] Kittisak Kerdprasop and Nittaya Kerdprasop 2012 Bridging Data Mining Model to the Automated Knowledge Base of Biomedical Informatics.