

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Uncertainty Analysis Using Fuzzy Sets for Decision Support System

Mohd Najib Mohd Salleh, Nazri Mohd Nawawi and Rozaida Ghazali
*Universiti Tun Hussein Onn Malaysia, Johor,
Malaysia*

1. Introduction

In agricultural domain application, it is becoming increasingly important to preserve planting material behavior when interact with an environment that is not under its control. However, the uncertainty always inherent such inaccurate decisions when a present of incomplete information in sampling data set (Chao, et al., 2005; Latkowski, 2002; Michelini, et al., 1995). As a result, the proper decision may need to adapt changes in their environment by adjusting its own behavior. Many different methods for dealing uncertainty have been developed. This research work proposes incomplete information with fuzzy representation in objective function for decision modeling. Firstly, we integrate expert knowledge and planting material data to provide meaningful training data sets. Secondly, fuzzy representation is used to partition data by taking full advantages of the observed information to achieve the better performance. Finally, we optimally generalize decision tree algorithms using decision tree technique to provide simpler and more understandable models. The output of this intelligent decision system can be highly beneficial to users in designing effective policies and decision making.

2. Preliminary study

The major problems in decision modeling are missing information of ecological system, such as weather, fertilizer, land degradation, soil erosion and climate variability during planting material selection in physiological analysis. This underlying obstacle will return poor results when the aggregation of all the databases in planting material. If we try to develop a decision modeling to apply knowledge to user, then we have to integrate historical records of planting material behavior, expert knowledge's perception and decision algorithm learning. From an information or knowledge provider's perspective, the challenge is to gain an expert user's attention in order to assure that the incomplete information is valued (Rouse, 2002). One of the more widely applied in representing uncertainty is fuzzy representation (Michelini et al., 1995; Mendonca et al., 2007). This representation required a mathematical support to make further treatment for interpretation of missing values as information, which is compatible with the observed data. We analyze the complexity of ecological system which have missing values caused by the uncertainty to the user, then representing the observed data into plausible values and discuss the outcome of an empirical study for the missing information in induction learning. We have observed the

most commonly used algorithms for decision modeling of missing values (Tokumaru et al., 2009) and resulted in better understanding and comprehensible rules in planting material selection.

2.1 The concept of uncertainty

The uncertainty clearly shows the limitations of the currently adopted approach to the knowledge representation. In particular, it has shown that the result of a measurement represents incomplete knowledge and this knowledge can be usefully employed only if its missing value can be somehow estimated and quantified. However, we have to measure the degree of uncertainty to rank objects for presentation to a user. The approach requires a suitable mathematical theory for handling incomplete knowledge. The practice refers mainly to the probability theory to treat missing values. In the normal situation, the presence of knowledge is recognized, but the exact value is not known, even if it is possible to locate it within a closed interval. By definition, we always take the same value, although unknown, within the estimated interval. The problem of how to estimate the missing value and measure the uncertainty is still open topic in the field, despite much contribution that have been made in the literature over the years. From information or knowledge provider's perspective, the challenge is to gain an expert user's attention, which to assure that the information or knowledge content is valued. The aim of a measurement process is in fact always to take a decision. Moreover, decisions are sometimes required within the measurement process itself.

2.2 Related work

This section reviews some approaches representing missing values in information system when the decision tree is constructed. The development of measurement science has identified the uncertainty concept to quantify the incomplete information during the decision making. The measurement of uncertainty is treated in a purely probabilistic way and has been considered the available mathematical theory capable of handling incomplete information. In some case, the valuable information is obtained in incomplete decision table. The possible decomposition of information table can search possible information within the missing value (Latkowski, 2002). He addressed the degree of similarity between objects on the attributes can be measured by operator product and the imprecise attribute values are expressed in a probability distribution. When the results from possible tables are aggregated, the probabilistic values of decision table can be expressed in a probability distribution and the extended set of possible members of subset is obtained. The elimination of irrelevant features sometimes gives a little or no additional information beyond that subsumed by the remaining features (Koller, et al. 1996). They proposed an efficient algorithm for feature selection which able to compute an approximation to the optimal feature selection criterion. (Chao, et al., 2005) provided comprehensive study dealing with missing values in cost-sensitivity decision trees. He reported that cost-sensitive decision learning algorithms should utilize only known values, it is desirable to have missing values to reduce the total cost of test and misclassification and take advantage of missing values for cost reduction. A conceptual framework of modeling decision making processes with incomplete knowledge in agriculture is presented (Passam, et al., 2003). He compared with existing methodologies and claimed the result to be more realistic, since he took into account the available information without enforcing an artificial accuracy. While Schafer (1997) introduced

the concept of different levels of suitability for learner biases, the fact that no algorithm biases can be suitable for every target concept, the idea that there is no universally better algorithm is fast maturing on the machine learning community. It might do better to map different algorithms to different groups of problems with practical importance.

Often cases with incomplete descriptions are encountered in which one or more of the values are missing. This may occur, for example, because these values could not be measured or because they were believed to be irrelevant during the data collection stage. Considerable attention and significant progress has been devoted in the last decade toward acquiring "classification knowledge", and various methods for automatically inducing classifiers from data are available. Quinlan (1996) had focused several approaches of measurement in high levels of continuous values with small cases on a collection of data sets. A variety of strategies are used in such domain and returned poor result as some imprecise data is ignored.

The construction of optimal decision trees has been proven to be NP-complete, under several aspects of optimality and even for simple concepts (Murthy, et al., 1993). Current inductive learning algorithms use variants of impurity functions like information gain, gain ratio, gini-index, distance measure to guide the search. Fayyad (Fayyad, 1994) discussed several deficiencies of impurity measures. He pointed out that impurity measures are insensitive to inter-class separation and intra-class fragmentation, as well as insensitive to permutations of the class probability distribution. Furthermore, several authors have provided evidence that the presence of irrelevant attributes can mislead the impurity functions towards producing bigger, less comprehensible, more error-prone classifiers.

There is an active debate on whether less greedy heuristics can improve the quality of the produced trees. Others showed that greedy algorithms can be made to perform arbitrarily worse than the optimal. On the other hand, Murthy and Salzberg (1993) found that one-level look-ahead yield larger, less accurate trees on many tasks. (Quinlan, et al., 1995) reported similar findings and hypothesized that look-ahead can fit the training data but have poor predictive accuracy. The problem of searching for the optimal sub-tree can be formulated. The coefficient matrix of the defining constraints satisfies the totally uni-modular property by solving the integer program (Zhang et al., 2005). He provided a new optimality proof of this efficient procedure in building and pruning phase. The cost is imposed on the number of nodes that a tree has. By increasing the unit node cost, a sequence of sub-trees with the minimal total cost of misclassification cost and node cost can be generated. A separate test set of data is then used to select from these candidates the best sub-tree with the minimal misclassification error out of the original overly grown tree. Besides cost-complexity pruning, a number of other pruning algorithms have been invented during the past decade such as reduced error pruning, pessimistic error pruning, minimum error pruning, critical value pruning, etc. It is clear that the aim of the pruning is to find the best sub-tree of the initially grown tree with the minimum error for the test set. However, the number of sub-trees of a tree is exponential in the number of its nodes and it is impractical computationally to search all the sub-trees. The main idea of the cost-complexity pruning is to limit the search space by introducing selection criterion on the number of nodes. While it is true that the tree size decides the variance-bias trade-off of the problem, it is questionable to apply an identical punishment weight on all the nodes. As the importance of different nodes may not be identical, it is foreseeable that the optimal sub-tree may not be included in the cost-complexity sub-tree candidates and hence it will not be selected. An interesting question to ask is whether there is an alternative good algorithm to identify the true optimal sub-tree.

Most of the data mining application reduced to search for the intervals after discretisation phase. All values that lie within this interval are then mapped to the same value. It is necessary to ensure that the rules induced are not too specific. Several experiments have shown that the quality of the decision tree is heavily affected by the maximal number of discretisation intervals chosen. Recent development of estimation uncertainty using fuzzy set theory is introduced. The fuzzy practicable interval is used as the estimation parameter for the uncertainty of measured values. When the distribution of measured values is unknown, results that are very near to the true values can be obtained in this method. In particular, the boundaries among classes are not always clearly defined; there are usually uncertainties in diagnoses based on data. Such uncertainties make the prediction be more difficult than noise-free data. To avoid such problems, the idea of fuzzy classification is proposed. The new model of classification trees which integrates the fuzzy classifiers with decision trees is introduced (Chiang, et.al., 2002). The algorithm can work well in classifying the data with noise. Instead of determining a single class for any given instance, fuzzy classification predicts the degree of possibility for every class. Classes are considered vague classes if there is more than one value for the decision attribute. The vague nature of human perception, which allows the same object to be classified into different classes with different degrees, is utilized for building fuzzy decision trees (Yuan et al., 1995). The extension of earlier work of decision tree C4.5 by Quinlan (1996), each path from root to leaf of a fuzzy decision tree is converted into a rule with a single conclusion. A new method of fuzzy decision trees called soft decision trees (Olaru et.al., 2003) is presented. This method combines tree growing and pruning, to determine the structure of the soft decision tree, with refitting and back-fitting to improve its generalization capabilities. A comparative study shows that the soft decision trees produced by this method are significantly more accurate than standard decision trees. Moreover, a global model variance study shows a much lower variance for soft decision trees than for standard trees as a direct cause of the improved accuracy. Fuzzy reasoning process allows two or more rules or multi branches with various certainty degrees to be simultaneously validated with gradual certainty and the end result will be the outcome of combining several results. Yuan (1995) proposed a novel criterion based on the measurement of cognitive uncertainty and criterion based on fuzzy mutual entropy in possibility domain. In these approaches, the continuous attributes are needed to be partitioned into several fuzzy sets prior to the tree induction, heuristically based on expert experiences and the data characteristics. The effectiveness of the proposed soft discretization method has been verified in an industrial application and results showed that, comparing to the classical decision tree, higher classification accuracy was obtained in testing. The soft discretization based on fuzzy set theory one inherent disadvantage in these methods is that the use of sharp (crisp) cut points makes the induced decision trees sensitive to noise. As opposed to a classical decision tree, the soft discretization based decision tree associates a set of possibilities to several or all classes for an unknown object. As a result, even if uncertainties existed in the object, the decision tree would not give a completely wrong result, but a set of possibility values. Experimental results showed that, by using soft discretization, better classification accuracy has been obtained in both training and testing than classical decision tree, which suggest that the robustness of decision trees could be improved by means of soft discretization.

Basically, the true value is unknown and it could be estimated by an approximation. Sometimes, we found some of the attributes may be form the similar classes with the

presence of missing values. Rough Sets are efficient and useful tools in the field of knowledge discovery to generate discriminant and characteristic rules. This method provides relative reduct that contains enough information to discern objects in one class from all the other classes. From the relative reduct produced by the quick algorithm, rules are formed. If a new object is introduced into the data set with the decision value missing, one could attempt to determine this value by using the previously generated rules. Although rough set have been used for classification and concept learning tasks (De Jong et al, 1991; Janikow, 1993; Congdon, 1995), there is rather little work on their utility as a tool to evolve decision trees. (Jenhani, 2005) proposed an algorithm to generate a decision tree under uncertainty within the belief function framework.

3. Method and material

There are 3300 records of planting material had been collected in physiological analysis. These records are represented as a table of examples which described by a fixed number of features along with a predicted label denoting its class. In oil palm industry, the Elite palms with special characteristic, such as high bunch oil content, slow stem growth and short leaves were identified and selected from backcross and progeny test compact seeds. Effective breeding and selection requires a large genetic variation such as current oil palm breeding populations. Apart from targeting the primary objectives in oil palm breeding, several traits of interest include improvement of physiological traits (such as bunch index and vegetation measurement). Thus, it has become apparent to develop not only high yielding planting materials but also with novel traits. In this research study, we attempt to mine the oil palm germplasm for yield and novel traits for use in breeding and ultimately commercialization.

Total Economic Production (TEP) composites of oil yield and kernel yield per palm per year. Oil yield is derived from a composite of characters and is dependent on a number of components, the most important being fresh fruit bunch (FFB) and fruit composition, such as mesocarp to fruit (M/F), shell to fruit (S/F) and oil to bunch (O/B). The oil palm fruit is a drupe, consisting the mesocarp, shell and kernel. Within the fruit, a major component that determines a high O/B, besides M/F, is S/F. Shell and mesocarp contents of the fruit is negatively correlated, a reduction in shell content subsequently increases the mesocarp content, while the kernel remains unchanged. Among others, a strategy in developing planting materials for high oil yield is through the selection against shell thickness. Fruit size is indicated by the mean fruit weight (MFW) in bunch analysis. Generally, MFW is positively correlated with the oil-related-traits such as mesocarp to fruit (M/F), oil to dry mesocarp (O/DM), oil to wet mesocarp (O/WM), oil to bunch (O/B) and oil yield (OY), and negatively related with kernel yields and F/B. The O/B is the product of oil to bunch is strongly associated with F/B, M/F and O/WM. Beside that, we incorporate vegetation measurements which provide the height, trunk diameter, frond production and leaf area per every palm tree.

In decision modeling analysis decision tree structure refers to all the unique values which some records may contain incomplete information with discrete set of values or continuous attributes. Figure 1 shows the comparison of complete and incomplete information of physiological analysis during the initial study.

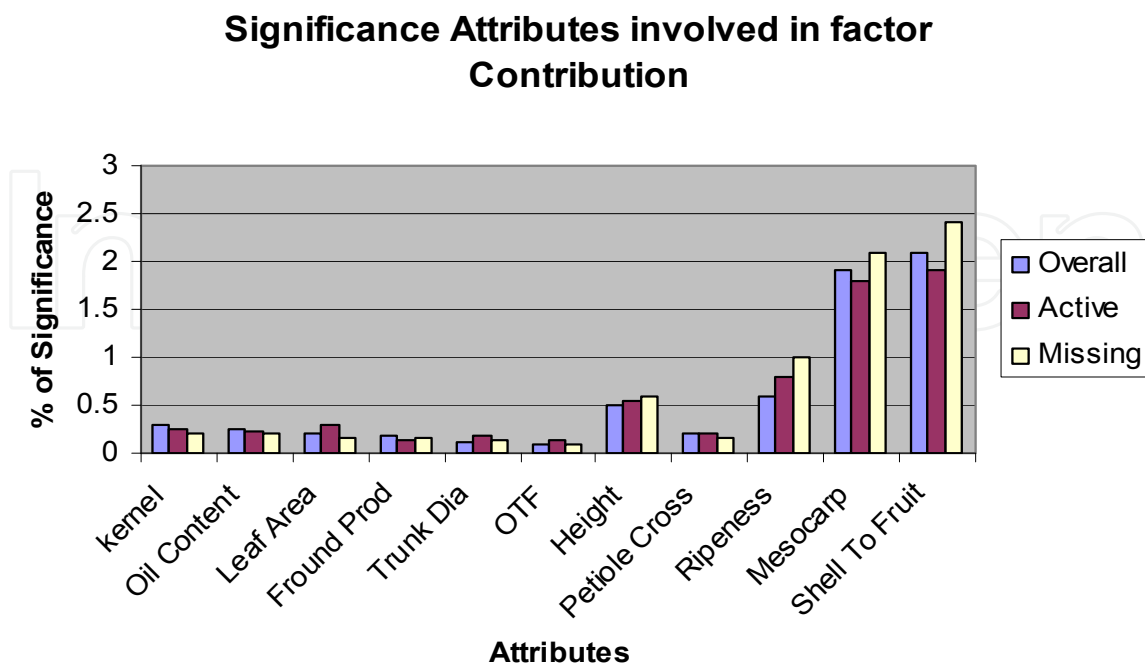


Fig.1. Number of complete and missing attributes in physiological analysis

By integration of expert knowledge and planting material, it seem reasonable to calculate the number of samples in the set T_j which belong to the same predicted class C_j , assigned as (1), then the probability of a random sample of the set T belongs to the class C_j . We estimate an attribute x_i provides data in fuzzy set and gives a membership degree as in (2), where p is the number of specified clusters, k is the number of data points, x_i is the i -th data points, $\mu_j(x_i)$ is a function that returns the membership of x_i in the j -th cluster

$$freq(C_i, T) = \sum_{j=1}^n \mu(C_j). \quad (1)$$

$$\sum_{j=1}^p \mu_j(x_j) = 1 \quad (2)$$

Some data sets can not be separated with clearly defined boundaries due to the data bases which contain incomplete information, but it is possible to sum the weight of data items belong to the nearest predicted class.

3.1 Fuzzy representation

Since fuzzy set theory can be used to describe imprecise information as membership degree in fuzzy clustering, therefore a given item sets are divided into a set of clusters based on similarity. We present fuzzy C-means based on objective function to quantify the goodness of cluster models that comprise prototypes and data partition. Fuzzy cluster analysis assigned membership degrees to deal with data that belong to more than one cluster at the

same time; they determine an optimal classification by minimizing an objective function. A standard minimized objective function simply expressed as

$$l = \sum_{k=1}^p \sum_{i=1}^n (\mu_k(x_i))^m \|x_i - c_k\|^2 \tag{3}$$

It can be seen that the characteristic functions constitute between ambiguity and certainty to the closed subsets x_i of the boundary with respect to the set interval (Heiko et al., 2004). We use more robust method of viewing the compatibility measurement for query x_i takes the weighted average of the predicate truth values. The equation show how the weighted average compatibility index is computed. The average membership approach computes the mean of the entire membership values.

$$x_i = (\sum_{i=1}^n (\mu_i(p_i)) \times w_i) / \sum_{i=1}^n w_i \tag{4}$$

The degrees of membership to which an item sets belongs to the different clusters are computed from the distances of the data point to the cluster centers. With combination expert knowledge and most relevant ecological information, the membership degree can be calculated to determine some plausible data point lies to center of the nearest cluster. An iterative algorithm is used to solve the classification problem in objective function based clustering: since the objective function cannot be minimized directly, the nearest cluster and the membership degrees are alternately optimized. In fuzzy clustering analysis, the calculation of cluster centre value is given as follow, and each of observed data is assigned to the centre using Euclidean distance.

$$c_j = (\sum (\mu_j(x_i))^m x_i) / (\sum (\mu_j(x_i))^m) \tag{5}$$

3.2 Decision tree construction

We develop decision tree with expert knowledge and additional information in Figure. 2(a) and 2(b), the generated nodes form dynamic structure that changes when all elements remain possible to be tested. All available information is applied to derive the fuzzy maximum estimation which describes the imputation of estimates for missing values in cluster centers. In decision tree (Quinlan, 1993) the attribute depends on its entropy computation among the rest of the attributes, it can be simply formulated the uncertain subset x_i from the interval of element x_o are dropped out. The entropy can be measured as the item sets are divided into subset T_j

$$\sum_{j=1}^n \frac{freq(C_i, T_j)}{T_j} \log_2 \frac{freq(C_i, T_j)}{T_j} \tag{6}$$

In the measured value x_i , the possible subset is uniquely determined by the characteristic function. The measurement of the dispersal range of value x_i is relative to the true value x_o .

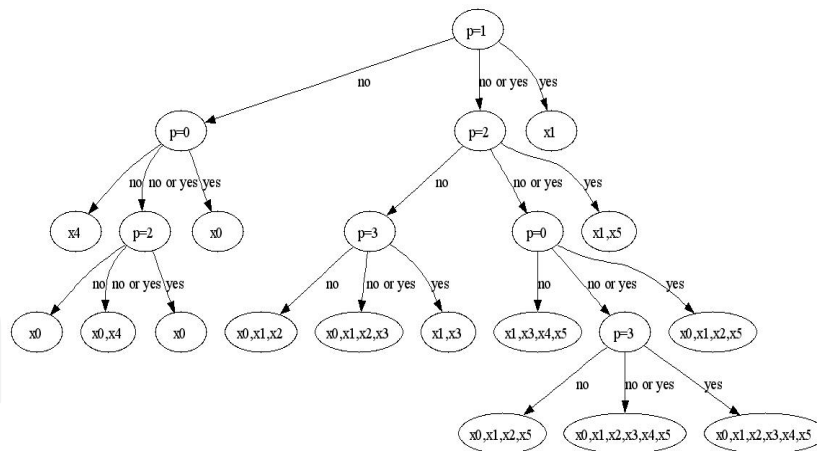


Fig. 2(a). Decision Tree Construction with incomplete information

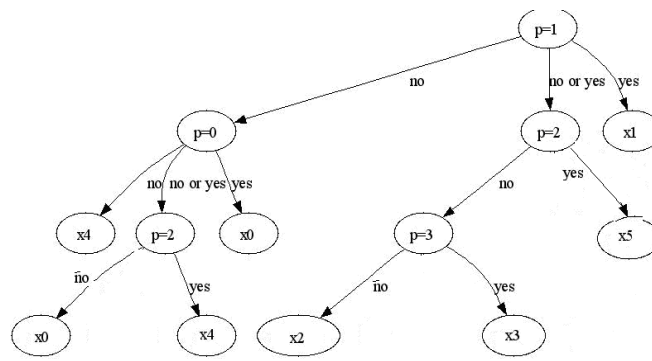


Fig. 2(b). Decision Tree Construction with expert knowledge and additional information

3.3 Proposed method

During the decision tree construction, the problem of missing values has to be dealt with both when processing the training examples and classifying a case using a learned tree. In some cases, an attribute may have associated missing values which are observed in the training examples and would appear as multiple values or null. When the decision tree is used to classify unseen cases, it is difficult to encounter a test on an attribute whose value is unknown. We present a reactive method for generalizing the decision tree into combination of binary and ternary tree, so called reactive decision tree by removing one or more inconsistency conditions of the attribute values. The method includes setting a preference ordering for conflicting attribute value and forming consistent sub trees or the exact tree leaf, which represents the predicted class at that leaf.

For each attribute, disjoint classes are separated with clearly defined boundary. These boundaries are critical since small changes close to these points will probably cause complete changes in classification. However, the imprecise and missing information often regarded as a difficult problem to define clearly about the class boundaries. This causes several problems for the selection of a test to partition the training set which require comparison of tests based on attributes with different numbers of unknown values. The comparison of the test which had been selected is sometimes in a sensible manner. The existence of imprecise information in real-world problems, the class boundaries may not be defined clearly. To obtain lower and upper approximations, it is necessary to briefly

evaluate the qualitative attribute values, a set of evaluation grades may first be supplied from existing possible values. It provides a complete set of distinct standards for assessing qualitative attributes. In accomplishing this objective, an important aspect to analyze is the level of discrimination among different counting of evaluation grades, in other words, the cardinality of the set used to express the information. The cardinality of the set must be small enough so as not to impose useless precision on the users and must be rich enough in order to allow discrimination of the assessments in a limited number of degrees. We are able to make a decision when we know the exact values of the maximized quantity. However, the method has difficulties for knowledge discovery at the level of a set of possible values, although it is suitable for finding knowledge. This is because the number of possible tables exponentially increases as the number of imprecise attribute value increases. We explore the relation between optimal feature subset selection and relevance. The motivation for compound operators is that the feature subsets can be partitioned into strongly relevant, weakly relevant and irrelevant features (John et al, 1994). The wrapper method (Kohavi et al., 1994) searches for an optimal feature subset tailored to a particular algorithm and a domain. This approach showed the feature subset selection is done using the induction algorithm as a black box, which is no knowledge of the algorithm is needed during the interface. We compare the wrapper approach to induction without feature subset selection and to Relief (Kira et al., 1992), a filter approach to feature subset selection, and FOCUS (Dietterich, et al., 1996). The improvement in accuracy also is achieved for some data sets in Naive-Bayes algorithm. The Maximum Acceptable Error (MAE) provides an improved estimate of the confidence bounds of concentration estimates. This method accommodates even strongly nonlinear curve models to obtain the confidence bounds. The method describes how to define and calculate the minimum and maximum acceptable concentrations of dose-response curves by locating the concentrations where the size of the error, defined in terms of the size of the concentration confidence interval, exceeds the threshold of acceptability determined for the application.

3.4 Decision rule construction

We investigate the possible of rules generated or classifiers during the classification rule development. As we had known that one of the most simple to understand, readable and manipulate is the decision tree. Decision trees represent a series of IF.....THEN type rules which are linked together and can be used to predict properties for our observations based upon the values of various features. We describe approximation boundary of the attribute value process by ambiguity A_x , impossible I_x , necessary values N_x and given information attributes W_x . Generally, we derived our attribute with the values as (A_x, I_x, N_x, W_x) . Initially, we assume that our domain of the attribute x is $D_x = \{ a, b, c \}$ and the values of attribute x is $\{ \emptyset, \emptyset, \emptyset, D_x \}$.

We argue more precise as the rule being uncertainty to the expert as information when there are less information provided which represents decision tree with uncertainty. Decision-tree algorithms suggest a more focused approach to rule extraction. The rule extraction assumes that the user is interested in only one attribute as class as a consequent of every rule. This assumption significantly reduces the search space with the problem of finding association rules between any database attributes. A set of mutually exclusive and exhaustive if-then (production) rules can be easily extracted from a decision tree.

3.5 Improved algorithm and methodology

In this framework, the formal method is an approach towards the extraction of highly similar groups of subsets from a collection of elements described by a set of attributes A . The paradigm occurred within the attributes considered represent binary features with only two possible values true and false.

$$G_A(X) = \begin{cases} 1 & x_i \in A \\ 0 & x_i \notin A \end{cases}$$

We build the functions $G_A(X)$ constitute between ambiguous and certainty to the closed subsets X_i of the boundary with respect to the set inclusion. The function builds operators join and meet to provide the least upper and lower approximation in the set boundary respectively. From set theory viewpoint, the concepts represent complete maximal wide range of subsets which is described by the elements. It may be reduced to the discovery of the closed sets to the test rules. The algorithm used to discover the concept set comes from the set theory. It generates possible subsets in an iterative manner, starting by the most specific test value. At each step, new subsets are generated as members coupled, where the subset is the intersection of the intents of the already existing subsets. We define the border of a partially constructed subset through an element-wise insertion to the root node. The overall decision tree construction is supported by a structure that once the algorithm has finished its work contains the entire three-valued decision tree. Firstly, the set in the root node is a dynamic structure that changes when all elements remain possible to be tested. When the attribute depends on its entropy computation among the rest of the attributes, it can be simply formulated the new border always includes the new element whereas all elements of the old border that are greater than new elements are dropped out.

$$G_A(X) = \left\{ \begin{array}{ll} 1 & x_p \subseteq U_p \cup N_p \cup \{v\} \\ unknown & \forall x \quad U_p \cup \{v\} \neq D_p \\ 0 & v \notin x_p \end{array} \right\}$$

To consider the universal set X and its power set $P(X)$, let K be an arbitrary index set, it can proved that a function Nec is a necessity function if and only if it satisfies the following relationship:

$$Nec\left(\bigcap_{k \in K} A_k\right) = \inf_{k \in K} Nec(A_k)$$

While the function Pos is a possibility function if and only if it satisfies the following relationship:

$$Pos\left(\bigcup_{k \in K} A_k\right) = \sup_{k \in K} Nec(A_k)$$

for any family $\{A_k \mid k \in K\}$ in $P(X)$.

We have analyzed through examples how to aggregate individual elements by considering the overall contributions to the agreement. We now present the considered algorithm in a general and precise way.

1. Decision makers $V = \{v_1, \dots, v_m\}$ sort the alternatives of $X = \{x_1, \dots, x_m\}$ according to the linguistic categories of $\Gamma = \{\ell_1, \dots, \ell_p\}$. Then, we obtain individual weak orders R_1, \dots, R_m which rank the alternatives within the fixed set of linguistic categories.
2. Taking into account the scores s_1, \dots, s_p associated with ℓ_1, \dots, ℓ_p , a score is assigned to each alternative for each alternative for every decision maker: $S_i(x_u), i = 1, \dots, m, u = 1, \dots, n$
3. We aggregate the individual opinions by means of collective scores which are defined as the average of the individual scores:

$$S(x_u) = \frac{1}{m} \sum_{i=1}^m S_i(x_u)$$

and we rank the alternatives through the collective weak order R:

$$x_u R x_v \Leftrightarrow S(x_u) \geq S(x_v)$$

4. We calculate the overall contributions to the agreement for all the decision makers: w_1, \dots, w_m .
 - a. if $w_i \geq 0$ for every $i \in \{1, \dots, m\}$, then we obtain the new collective scores by :

$$S^w(x_u) = \frac{1}{m} \sum_{i=1}^m w_i \cdot S_i(x_u)$$

and rank the alternatives by means of the collective weak order R^w .

$$x_u R^w x_v \Leftrightarrow S^w(x_u) \geq S^w(x_v)$$

- b. Otherwise, we eliminate those decision makers whose overall contributions to the agreement are not positive. We now initiate the decision procedure for the remaining decision makers $V^+ = \{v_i \in V \mid w_i > 0\}$

We apply the method of weighted equivalence classes to information tables containing missing values. We briefly compare the method where uses of indiscernible classes with the method of weighted equivalence classes. The proposed method is based on possible values in the data sets, which consist of precise values, are obtained from an information table. Each possible subset is dealt with by applying rough sets to information tables containing precise information, and then the results from the possible tables are aggregated. In other words, the methods that are already established are applied to each possible table. Therefore, there is no doubt for correctness of the treatment.

The value selection assumes that the algorithm is interested in only one attribute as class as a consequent of the entropy estimation. By reducing the data, we generate different subset from selected condition attribute. This assumption significantly reduces the search space with the problem of analyzing the ambiguity of attribute values. When each attribute has associated a numerical score with respect to decision attribute, each alternative obtain a more focused approach to the selected values of subsets. Then, we obtain a distance between each individual preference and the collective one through the distance among the individual

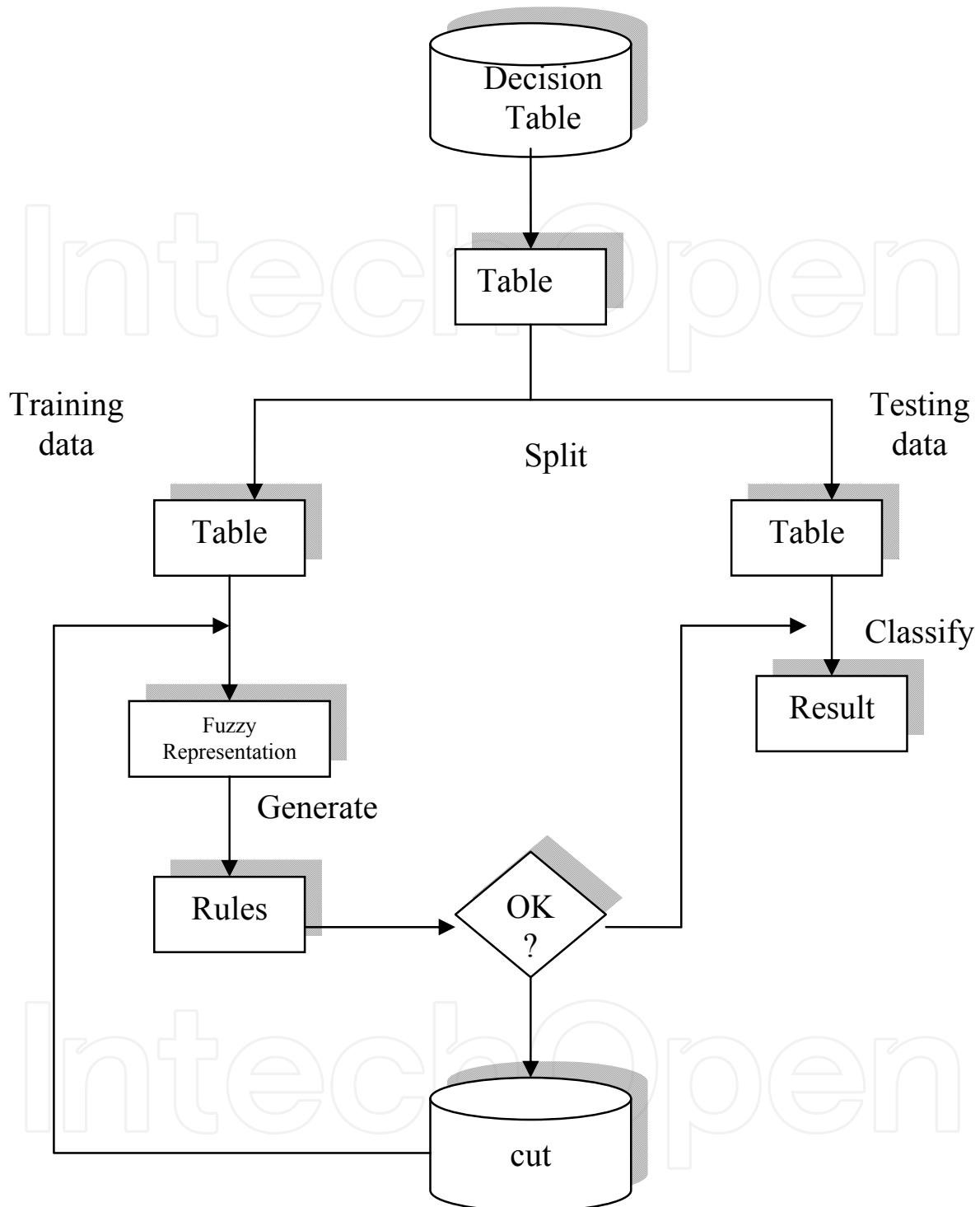


Fig. 3. Framework to determine rough set and induce desirable values to validate a classifier and collective scoring vectors. We measure the agreement in each subset of possible values, and a weight is assigned to each decision makers, the overall contribution to the agreement. Those decision makers whose overall contribution to the agreement is not positive are expelled and we re-initiate the process with the opinions of the decision makers which is positively contribute to the agreement. The sequential process is repeated until it determines a final subset of decision makers where all of them positively contribute to the agreement.

Then, we apply a weighted procedure where the scores each decision makers indirectly assigns to the alternatives are multiplied by the weight of the corresponding decision maker, and we obtain the final ranking of the alternatives. As a result, the mutually exclusive attribute value can be easily extracted from a decision tree in the form of ambiguity reduction. Consequently, we assign the examples with missing values of the test attribute to the 'yes', 'no' and 'yes-or-no' outgoing branches of a tree node.

3.6 New decision tree construction

New extended ID3 algorithm based on T2, T3 calculates optimal decision trees up to depth 3 by using two kinds of decision splits on nodes. Discrete splits are used on discrete attributes, where the nodes have as many branches as there are possible attribute values. Interval splits are used on continuous attributes where a node has as many branches as there are intervals. The number of intervals is restricted to be either at most as many as the number of existing classes plus one, if all the branches of the decision node lead to leaves, or to be most as 2 otherwise. The attribute value "unknown" is treated as a special attribute value. Each decision node has an additional branch, which takes care of unknown attribute values. In fact, this way of treating unknown attributes is reported to perform better than that of C4.5.

In the tree-building phase, at each node, all attributes are examined, in order to select one on which a split will be performed for the node. When the attribute is discrete, the relative frequencies of all of its possible values are calculated. For continuous attributes, the same approaches would be inefficient because of the number of possible values and the resulting low frequencies of them. For that reason, local discretisation is used. Finally, a split is performed on an attribute if it results in maximum accuracy. Consequently T3 produces a tree hierarchy, which determines how important is an attribute in the classification process, in contrast to C4.5 which uses the gain ration. To carry out this local discretisation of a continuous attribute, its values have to be partitioned into multiple intervals. The set of intervals that minimizes the classification error is found by a thorough exploration instead of heuristically applying recursive binary splitting. The search for these intervals is computationally expensive, so T3 restricts decision trees to three levels tests, where only the third level employs binary splits of continuous attributes. T3 does not use a pruning technique. Instead it uses a parameter called Maximum Acceptable error (MAE). MAE is a positive real number less than 1, used as a stopping criterion during building the tree. T2 was observed to use a greedy approach when building the tree, thus further splitting at a node would stop only if the records already classified in this node, belonged to a single class. However, this greedy approach is not optimal, because minimizing the error in the leaf nodes does not necessarily result in minimizing the overall error in the whole tree. In fact, it was proved that a strategy choosing locally optimal splits necessarily produces sub-optimal trees. It should be noted here that classification error indicates how many instances of a training set have incorrectly classified, while generalization error indicates how many instances of a testing set have been incorrectly classified. Furthermore, even minimizing classification error does not always cause minimization of the generalization error due to over-fitting.

By introducing MAE, we allow the user to specify the level of purity in the leaves and stop to further building of the tree at a potential node split. We set MAE to have 4 distinct values, namely 0.0 to 0.3, meaning that splitting at a node stops even if the error in that node is equal to or below a threshold of 0 to 30 percent respectively. More precisely, building the tree would stop at a node in two cases. In the first case, building stops when the maximum

depth is reached, i.e. 3 when T3 is used or 2 when T2 is used. In the second case, building stops at that node only all the records remaining there to be classified to the same class in a minimum proportion of 70 to 100 percent. We used eight different version of T3 in our experiments, with different depth and length with MAE set to 0.1 - 0.3.

We associate a score to each value and aggregate each individual values by means of the average of the individual scores, providing a collective weak order on the set of alternatives. Then we assign an index to each value which measures their overall contribution to the alternatives. Taking into account these indices, we weight individual scores and obtain a new collective ranking of alternatives. Once the exact value is chosen, a branch relative to each value of the selected attribute will be created. The data are allocated to a node according to the value of the selected attribute. This node is declared as a leaf when the gain ratio values of the remaining attributes do after excluding the opinions of those decision makers whose overall contributions to the agreement are not positive. The new collective ranking of alternatives provides the final decision. Since overall contribution to the agreement indices usually is irrational numbers, it is unlikely that the weighted procedure provides ties among alternatives. The proposed decision procedure penalizes those individuals that are far from consensus positions, this fact incentives decision maker to moderate their opinions. Otherwise, they can be excluded or their opinions can be underestimated. However, it is worth emphasizing that our proposal only requires a single judgment to each individual about the alternatives. We can generalize our group decision procedure by considering different aggregation operators for obtaining the collective scores. Another generalization consists in measuring distances among individual and collective scoring vectors by means of different metrics.

4. Experimental result

The presence of the incomplete information in ecological system impacts on classification evaluation of planting material behavior in physiological analysis, since the semantics are no longer obvious and uncertainty is introduced. In figure 3, the graph shows the observed data are more likely to satisfy with additional information. The experiment combines planting material with expert knowledge and ecological information to generate an ROC curve. It can be seen that, in most of the data sets, the number of data belonging to the various categories do not exactly match the results of physiological analysis. This is because some of the planting material that should be classified in same type A has been misclassified into type B and vice versa.

In this experiment, we examine gradually the selected records of complete, ranked features and missing records of real physiological traits. The result, in Table 1 shows that the features in missing values compare to the others complete records. By adding additional features of ecological information to physiological trait, it shows less correlation between each feature in missing values and the others. The possible clusters rearrange their structure and rules are generated mostly come from combination of planting material and ecological information. As a result, the selected proper fuzzy values in decision tree construction provides less and simple rules. It removes some irrelevant features during the selection of subset of training data.

Figure 4 shows some of the rule production which obtained from the experiment. From it, the program obtained two decision tables for each clinician, one consisting of those with an 'induce' response and the other with a 'Don't induce' response (these tables are not shown). Rather than attempt to create a reduct from indistinguishable rules at this point we

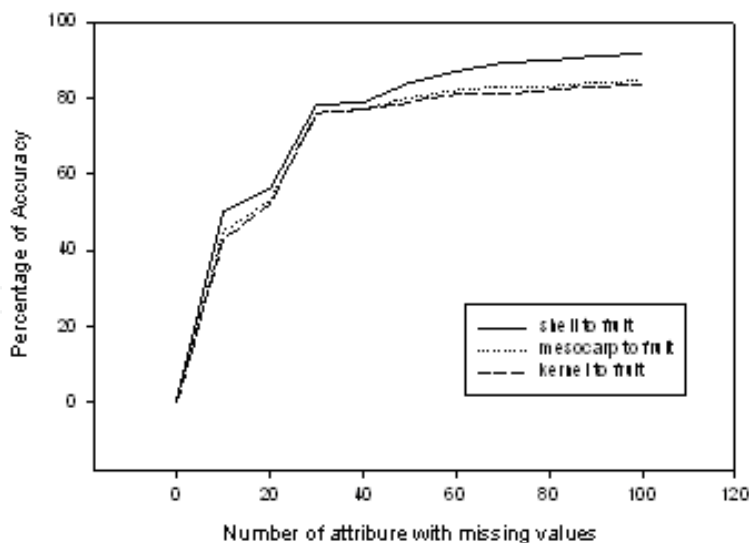


Fig. 3. Comparison of most observed attribute in missing values

	Training		Testing	
	Agree	Disagree	Agree	Disagree
Modeling Decision with Fuzzy Set	83.37%	16.63%	81.09%	18.91%
Modeling Decision with incomplete information	78.82%	21.18%	76.68%	23.32%

Table 1. Summary of comparisons of agrees and disagree classification based on decision tree algorithms

generated all the possible rules from the decision table and then removed those that contradicted each other from the 'induce' and 'don't induce' set. By this means we hoped to prevent the production of spurious rules, while still producing a reasonable number of rules from a small decision table.

The analysis uses the trait set to produce good rules and is examined to see if any, and the contradicted rules are removed from the final rule - set. This is done by removing all unknown or missing values set with the same parameters that have equal or worse values than the original rule. We extracted the relevant parameters for each of these trait under their physiological analyse. Those traits induced for reasons other than our indications were removed from the database. We then calculated the 'actual' rate of induction for the relevant indications. The rules obtained above are then applied to the relevant database. Experimental results have shown that our proposed method produces relatively small sized and comprehensible trees with high accuracy in generalisation and classification. It improves the performance of existing classifier in terms of both generalisation accuracy and particularly classification accuracy. Our proposed also outperforms C4.5 in terms of tree size and classification accuracy. However, this method's generalisation accuracy remains lower than that of C4.5

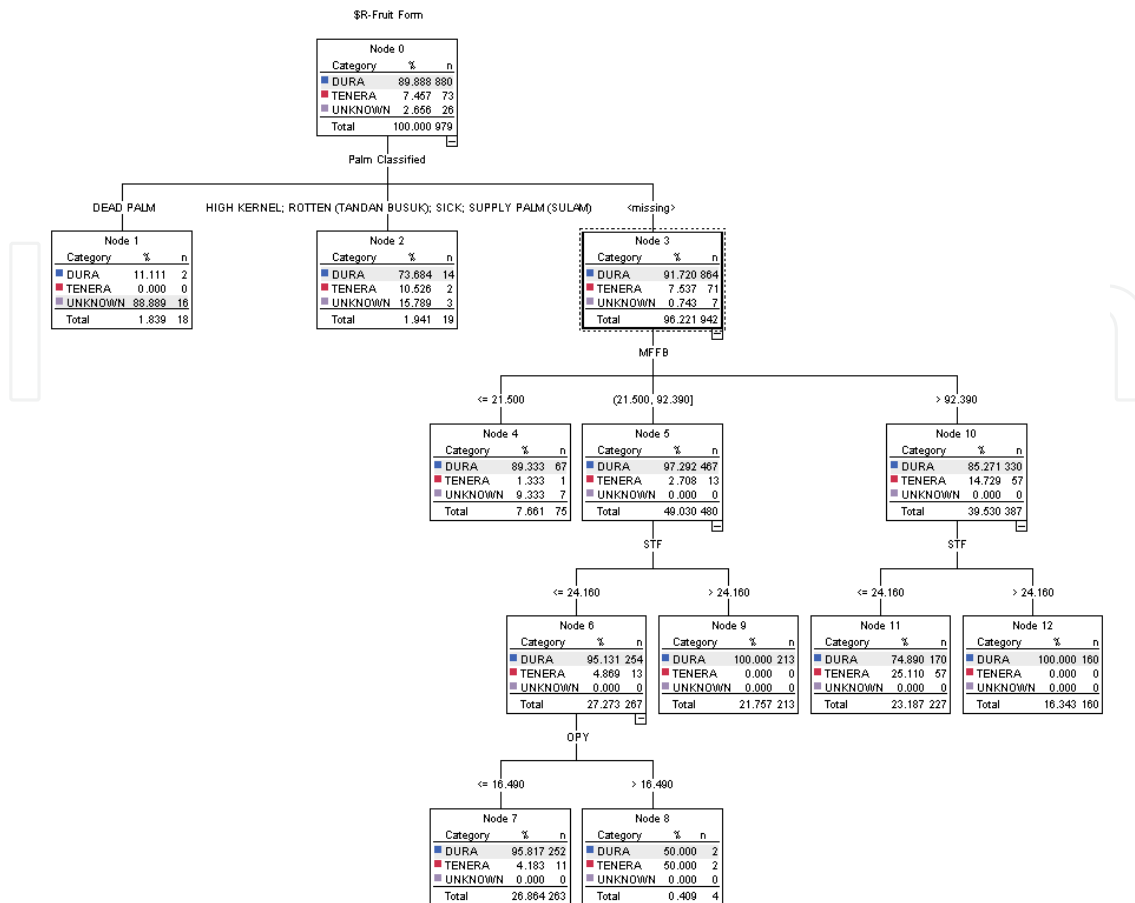


Fig. 4. The production rules by proposed method

5. Discussion and summary

This study has focused on domains with relatively high levels of unknown values and small testing set. It can be seen that the proposed method can be used for small samples and produces better result with expert knowledge and additional information. We have measured the degree of uncertainty using membership function and the proposed method is near to the true values of the uncertainty and the relative errors of the estimation are very small. However, this method still produces fairly large error after pruning with more missing values. This shows that the proposed method can only be applied to several conditions. The uncertainty estimation of measured values can be directly obtained using fuzzy representation and it is no longer necessary to estimate the standard deviation by mean value. For systems with small samples and unknown distributions, the proposed method is more suitable. When the training set is partitioned, ignoring cases with unknown values of the tested attribute leads to very inferior performance. During classification, attempting to determine the most likely outcome of a test works well in some domains, but poorly in others. Combining all possible outcomes is more resilient, giving better overall classification accuracy in these domains.

6. Conclusion

A missing value is often described as an uncertainty in the development of a decision system. It is especially difficult when there is a genuine disagreement between experts in the

field and also complex and unstated relationships between the variables that are used in the decision. The field of planting materials is particularly difficult to study because of the wide variation in environmental factors, and the differing population groups of genotypes serve. This decision support system allows the breeders or experts to make decisions in a way that is similar to their normal practice, rather than having to declare their knowledge in a knowledge engineering sense. Feedback from the experts in their domain should also be collected to refine the system especially in the evaluation of the decision trees themselves.

In this research work, we points out the extension of the splitting criterion in the missing values together. In addition to this, the splitting criterion also discards the irrelevant predictor attributes for each interior node. From the experiments reported, the classification error using our method was still low as compared to the original complete data sets by including additional attributes. It showed that discovered rules have a strong predictive power in missing values; this rule was not captured by C4.5 algorithm. On the other hand, the proposed method with his splitting criteria allows normally undeclared rules to be discovered. An added advantage here is that every expert has the opportunity to study the rules and make decisions on planting materials selection. On the other hand, the rough sets technique produce coarse splitting criterion and are often used for knowledge discovery from databases. It is also useful in any situation where a decision table can be constructed. It has an advantage in producing a set of comprehensible rules. The fact that this technique produces a lower and upper approximation of the true value, it allows a degree of uncertainty to be represented. The rules that are generated by the decision tree could be applied to a database of 'standard' procedures, or one that reflects their planting materials to obtain standard rules or drawing up guidelines for the development of an expert system in the oil palm industry.

We have focused on measurement of uncertainty in decision modeling. In this study, a special treatment of uncertainty is presented using fuzzy representation and clustering analysis approach in constructing the decision model. The uncertainty is not only due to the lack of precision in measured features, but is often present in the model itself since the available features may not sufficient to provide a complete model to the system. The result of the study shows that uncertainty is reduced and several plausible attributes should be considered during classification process. This formalization allows us to the better understanding and flexibility for selecting planting material in acceptance of the classification process.

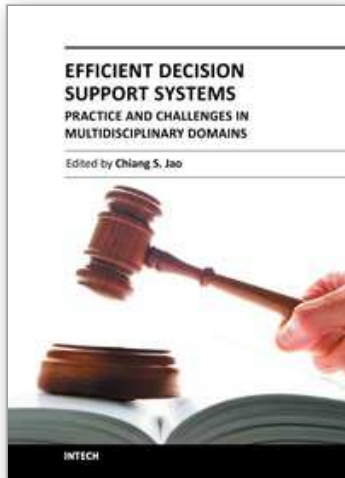
7. Acknowledgement

The author would like to thank Universiti Tun Hussein Onn Malaysia for funding the research grant.

8. References

- Chao, S., Xing Z., Ling X. & Li, S., (2005). Missing is Useful: Missing Values in Cost-Sensitive Decision Trees. *IEEE transactions on knowledge and data engineering*, vol. 17, no. 12.
- Chiang I-Jen, Jane Hsu Y.J., (2002). Fuzzy classification trees for data analysis. *Fuzzy Sets and Systems, International Journal in Information Science and Engineering, Fuzzy Sets and Systems*, Volume 130, Issue 1, Pages 87-99
- Congdon, C.B., (1995). A comparison of genetic algorithms and other machine learning systems on a complex classification task from common disease research, The University of Michigan.

- De Jong, & Spears. M., (1991). " Learning Concept Classification Rules using Genetic Algorithms". Proceedings of the Twelfth International Joint Conference on Artificial Intelligence .
- Dietterich, T., Kearns, M., & Mansour, Y. (1996). Applying the weak learning framework to understand and improve C4.5. *13th International Conference on Machine Learning*, pp. 96-104.
- Fayyad, U. M. & Irani, K. B. (1993). Multi-interval discretization of continuous valued attributes for classification learning. *13th International Joint Conference on Artificial Intelligence*, pp. 1022-1027.
- Fayyad, U. M. (1994). Branching on attribute values in decision tree generation. *12th National Conference on Artificial Intelligence*, pp. 601-606.
- Heiko, T., Christian, D. & Rudolf, K. (2004). Different Approaches to Fuzzy Clustering of Incomplete Datasets. *International Journal of Approximate Reasoning* 35 239-249.
- I., Jenhani, Z., Elouedi, N. B., Amor, K., Mellouli. Qualitative Inference in Possibilistic Option Decision Trees. In Proceedings of ECSQARU'2005. pp.944~955
- Janikow., Z. (1993). Fuzzy Processing in Decision Tree. In *Proceedings of International Symposium on Artificial Intelligence*.
- Janikow C.Z, (1996). Exemplar learning in fuzzy decision trees, in *Proceedings of FUZZ-IEEE*.
- Koller, D. and Sahami, M. (1996). Toward Optimal Feature Selection, *13th International Conference on Machine Learning*, pages 284-292.
- Latkowski, R. (2002). Incomplete Data Decomposition for Classification. *3rd International Conference*, Malvern, PA, USA.
- Mendonca, F., Vieira, M., & Sousa, M. (2007). Decision Tree search Methods in Fuzzy Modeling and Classification, *International Journal of Approximate Reasoning*.
- Micheleni, R. & Rossi, G. (1995). Measurement uncertainty: a probabilistic theory for intensive entities Measurement, 15 143-57.
- Murthy, S., Kasif, S., Salzberg, S., and Beigel, R. (1993). OC1: Randomized induction of oblique decision trees. *11th National Conference on Artificial Intelligence*, pp. 322-327.
- Olaru, C. & Wehenkel, L. (2003). A complete fuzzy decision tree technique. *Fuzzy sets and systems* 138, 221 - 254, Elsevier.
- Passam., C., Tocatlidou. A., Mahaman, B. D. & Sideridis., A. B. (2003). Methods for decision making with insufficient knowledge in agriculture. *EFITA Conference*, 5-9 July. Debrecen, Hungary.
- Quinlan, J. (1993). Introduction to Decision Tree. Morgan Kaufman Publisher.
- Quinlan, J. (1996). Improved use of continuous attributes in c4.5. *Journal of Artificial Intelligence Research*, 4:77-90.
- Rouse, B. (2002). Need to know - Information, Knowledge and Decision Maker. *IEEE Transaction On Systems, Man and Cybernetics-Part C: Applications and Reviews*, Vol 32, No 4.
- Schafer, J., (1997). Analysis of Incomplete Multivariate Data. *Chapman and Hall*, London.UK.
- Schafer, J., and Graham, J., (2002). Missing data: Our view of the state of the art. *Psychological Methods* 7, 147-177.
- Tokumar, M. & Muranaka, N. (2009). Product-Impression Analysis Using Fuzzy C4.5 Decision Tree. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.13 No 6.
- Wang, X. & Borgelt. C. (2004). Information Measures in Fuzzy Decision Tree. *Fuzzy Systems, Proceedings and IEEE International Conference*.
- Yuan, Y. & Shaw., J. (1995). Induction of fuzzy decision trees. *Fuzzy Sets System*. 69:125-139.
- Zhang, Y. & Huang, C. (2005). Decision Tree Pruning via Integer Programming.



Efficient Decision Support Systems - Practice and Challenges in Multidisciplinary Domains

Edited by Prof. Chiang Jao

ISBN 978-953-307-441-2

Hard cover, 478 pages

Publisher InTech

Published online 06, September, 2011

Published in print edition September, 2011

This series is directed to diverse managerial professionals who are leading the transformation of individual domains by using expert information and domain knowledge to drive decision support systems (DSSs). The series offers a broad range of subjects addressed in specific areas such as health care, business management, banking, agriculture, environmental improvement, natural resource and spatial management, aviation administration, and hybrid applications of information technology aimed to interdisciplinary issues. This book series is composed of three volumes: Volume 1 consists of general concepts and methodology of DSSs; Volume 2 consists of applications of DSSs in the biomedical domain; Volume 3 consists of hybrid applications of DSSs in multidisciplinary domains. The book is shaped decision support strategies in the new infrastructure that assists the readers in full use of the creative technology to manipulate input data and to transform information into useful decisions for decision makers.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Mohd Najib Mohd Salleh, Nazri Mohd Nawi and Rozaida Ghazali (2011). Uncertainty Analysis Using Fuzzy Sets for Decision Support System, Efficient Decision Support Systems - Practice and Challenges in Multidisciplinary Domains, Prof. Chiang Jao (Ed.), ISBN: 978-953-307-441-2, InTech, Available from: <http://www.intechopen.com/books/efficient-decision-support-systems-practice-and-challenges-in-multidisciplinary-domains/uncertainty-analysis-using-fuzzy-sets-for-decision-support-system>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen