Proceedings of the 28th Annual Hawaii International Conference on System Sciences — 1995

Rule Induction for Financial Modelling and Model Interpretation

Bob Berry, University of Nottingham, U.K.
Goksan Erdogan, University of Marmara, Turkey
Duarte Trigueiros, ISCTE, Portugal

Abstract

The paper discusses the possibility of applying a specific rule induction algorithm, ID3 to various financial data analysis tasks. These tasks include not only model building but also interpreting the outputs of financial models. The algorithm is shown to have major drawbacks as a modelling tool, some of which carry over to the post processing task. The fact that financial variables are often measured on a ration scale also causes problems. The paper examines solutions to the key problems and provides the basis on which analysts can judge the suitability of the algorithm for their own applications.

1: Introduction

The paper discusses the possibility of applying ID3, Quinlan's rule induction algorithm, to the analysis of financial data [1]. There have been various reported financial applications. For example, Tam and Kiang apply it to the problem of financial failure prediction [2], Braun & Chandler use a development of ID3 known as ACLS for stock market prediction [3], and Race and Thomas [4] use the algorithm to interpret the outputs of a financial simulation model.

Race & Thomas, for example, create a simulation model of an investment decision. Various attributes are assigned values during a run of the model, some according to appropriate probability density functions, and some according to management choice. Each run of the model generates an NPV value for the investment and a set of attribute values. Unfortunately managers had difficulty in determining whether their choices or random events were the key determinants of interesting model outcomes. Race and Thomas therefore ran the model repeatedly to generate a rich set of outputs and used ID3 to extract a set of IF...THEN rules which summarised these outputs.

The algorithm's alternative roles, as a financial modelling tool, and as a post processor of the outputs of financial models, are both commented on in this paper. The algorithm is seen as having major drawbacks as a modelling tool, some of which carry over to the post processing application, though as is described the post processing application also offers significant benefits. An analysis of these drawbacks, together with possible solutions to some of them, are presented here. The trade off between drawbacks and benefits must be evaluated each time it is proposed to try the algorithm in a new context.

2: The ID3 Algorithm in Action

ID3 derives a tree of IF...THEN rules which classifies the observations in a data set. To function, the algorithm requires a set of observations classified by decision outcome and attribute. The outcome variable must be represented by exhaustive, non overlapping categories rather than by a continuous measure. For example, net present value (NPV), originally measured on a ratio scale, would have to be represented by a finite number of NPV categories. NPV of course lends itself to this treatment: $NPV\geq 0$ and $NPV<0$ are categories with decision significance, the former signifying that shareholder value will be increased by the decision under consideration, or at least not worsened. Each attribute variable must also be represented by exhaustive, non overlapping categories. For example the attribute, demand in time period 1, might be represented by the categories, high, medium, and low.

An example of an appropriate data set is given in Table 1. The table shows a situation in which a project's NPV, and hence the choice between projects, is affected by the state of demand in years 1 and 2. There are two projects A and B, NPV is identified as good (g), average, (v) or bad (b), and in each year, demand can be either high (h) or low (l). The situation has occurred repeatedly, and hence it has been possible to record the frequencies with which different combinations of outcome and attribute variable levels have occurred.

In this case NPV is the outcome variable, with its value being determined by three attribute variables, the choice of project, the state of demand in year 1, and the state of demand in year 2.

1060-3425/95 $4.00 © 1995 IEEE

NPV	D1	D2	PROJ	FREQ
g	h	h	A	48
b	h	l	A	12
b	l	h	A	12
b	l	l	A	28
v	h	h	B	48
b	h	l	B	12
b	l	h	B	12
b	l	l	B	28

Table 1
Data Suitable For ID3

Briefly, the ID3 algorithm works as follows:

a) Take each of the attributes in turn and check to see how well each one explains the spread of observations over the outcome variable's categories. This requires the construction of a set of contingency tables of which Table 2, based on the effect of project choice on NPV, is an example. The body of the table includes the frequencies with which particular combinations of attribute and outcome values occur.

	OUTCOME=NPV G	 V	 B
PROJ A	48	0	52
PROJ B	0	48	52

Table 2
Contingency Table For Project Choice

Other similar tables based on state of demand in years 1 and 2 would also be constructed. The single attribute which best explains the pattern of outcomes forms the first branching point in a rule tree. 'Best explains' might be given content by selecting the contingency table with the highest Chi- square value, for example. On this basis, attribute D1 is the most important.

b) New contingency tables, similar to Table 2, but where each table contains only the observations relating to one of the categories of the successful attribute variable chosen at step a) are now constructed. Table 3 shows the relevant contingency table for the attribute, project choice, given that D1=h. Only four of the rows in Table 1, the first, second, fifth and sixth, are relevant. (Another table based on D1=h, and similar tables for state of demand in year 2 would also be constructed.) Once again using Chi-square as a criterion, then given that D1=h has already occurred, demand in year two best explains the data.

	OUTCOME=NPV G	 V	 B
PROJ A	48	0	12
PROJ B	0	48	12

Table 3
Contingency Table For Project Choice Given D1=h

c) The process continues until either no more attributes are left or all the data has been correctly classified. The outcome of the process is a rule tree, in this case that shown in Figure 1.

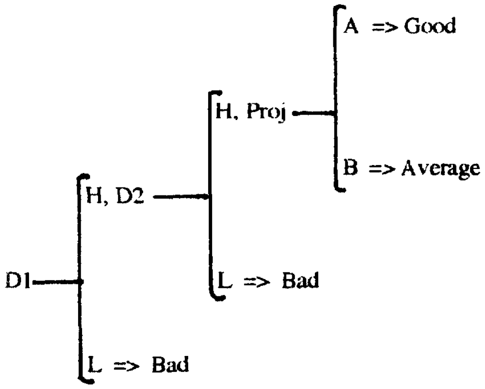

Figure 1
A Simple Rule Tree (Using Chi-square)

Pao [5] gives a description of the ID3 algorithm adequate to support the development of computer code.

The potential attractiveness of the form of output produced by ID3 to managers is obvious: it is a tree of IF...THEN rules growing from a common root, in this case the state of demand in year 1.

3: Branching Criteria

A key question in coding the algorithm is how to decide which attribute best allocates the data into outcome categories. In the description of the algorithm given above Chi-square is used, not unreasonably given that the aim is contingency table analysis. However this is only one of a variety of selection criteria which have been suggested. They link to Shannon's work on information theory [6].

According to Shannon, given a set of possible outcomes of a random event, the amount of information available about which single outcome will occur depends on the probabilities assigned to the outcomes. In general, if all the outcomes have equal probabilities, then there is no information about which outcome will occur. If there are 10 or fewer possible outcomes then a signal resolving the uncertainty need contain only 1 digit, if 100 or fewer outcomes then 2 digits, if 1000 or fewer outcomes then 3 digits. Thus, in general, for N outcomes the number of digits required to signal which outcome will occur is related to the logarithm to base 10 of the number of possible outcomes. $\log_{10}N$ is known as the logical variety of a situation, and measures the missing amount of information required to uniquely identify the eventual outcome of the random event.

Of course if some outcomes are similar and the outcomes can be regarded as a set of categories of outcomes, then the amount of missing information is no longer $\log_{10}N$. If K_i out of N outcomes are similar for each of a set of categories, i=1..I, then the amount of missing information, or entropy, is:

$$H = \log_{10}N - \Sigma(K_i/N)\log_{10}K_i$$

The second part of this equation is Quinlan's branching criterion, the information measure, IM:

$$IM = \Sigma(K_i/N)\log_{10}K_i$$

Quinlan's IM measures the reduction in "difficulty" of assigning observations to outcome category using a single attribute variable. It is calculated for each such variable to allow them to be ranked. Race and Thomas and Mingers [7] use both a simple Chi-square measure and G, a contingency table statistic found in Sokal & Rohlf [8] and which is identical to an information based measure found in Kullback [9], as alternatives to IM . However, IM, G and Chi-square are all closely linked. The Chi-square test statistic and G are both approximations to the Chi-square distribution, but G can be a better approximation in the tails of the distribution [10].

If the outcome variable has only two categories, the choice between the three alternative selection criteria does not appear to have any significant implications. In producing the rule tree in Figure 1, Chi-square was used as much for computational convenience as for any other reason. (As will be discussed later the choice of measure to govern the selection of which attribute variable to branch on is not an insignificant one when the outcome variable has more than two categories.)

Whatever branching criterion is adopted when implementing ID3, experience indicates that a tie breaking mechanism has to be introduced. It is not necessarily the case that a best attribute is unambiguously indicated.

4: ID3 as a Modelling Tool

The algorithm is not difficult to apply if an appropriately structured data set is available. However, there are conceptual problems involved which are not discussed in the literature and which the financial analyst needs to be aware of.

4.1: Modelling v. Complexity Reduction

Quinlan's original work dealt with chess endings. This type of data is complex but deterministic, and the original objective which ID3 was designed to achieve was complexity reduction in deterministic data. In the chess endings example redundancy was eliminated; information was not thrown away. The outcome of the analysis was a representation or description of the data, not a model in which some characteristics had been deemed unimportant and thus ignored.

However, when the aim is to produce a model of a set of stochastic, financial data, the task is to retain regularities which are seen as characteristics of the population, and to throw away details which are seen as sample specific. The resulting model therefore holds less information than the original data sample. ID3 has no automatic control over the complexity of the model it produces. Either all observations will be correctly classified, or the algorithm will simply run out of attributes to use. Without some form of stopping rule, an alternative description of the stochastic data set rather than a model, will be produced. This over fitting problem has implications for the ability of the resulting rule tree to predict or classify when faced with a new data set.

Mingers attempts to cope with this problem of an alternative description being produced, rather than a model being built, by first letting the algorithm generate an entire rule tree, and then pruning it back. Both G and Chi-square, the attribute selection measures he uses, allow a test of statistical significance of a branch in the rule tree. Pruning starts from the tips of the branches and works back until a significant branching point is found. (An alternative, potentially more effective, approach is to evaluate the significance of the entire rule tree before and after deletion of a node. Whatever approach is chosen, some such limitation on the scope of the rule tree produced is necessary for financial modelling purposes.)

4.2: Model Extraction or Model Imposition

There is a second, more significant, problem with the algorithm as a modelling tool. An analogy should serve to make it clear. It is well known that a linear regression produces the best straight line fit to a set of data even if a curvilinear representation is a more appropriate representation of the data set. (Residuals analysis should of course highlight the fact that the linear model is not appropriate before the model is put to use.) A similar issue arises in the application of the ID3 algorithm. A nested hierarchy of IF...THEN rules will be produced even if this hierarchical feature is not actually present in the data. The hierarchical feature was present in Quinlan's chess endings data because a sequence of moves was involved. The investment example data used above potentially have this hierarchical feature because they are generated by the sequence, project choice, demand state in year 1, followed by demand state in year 2.

However, it is easy to envisage data generation mechanisms in finance which are not sequential. The frequent need to model financial statement data by simultaneous equation systems identifies one such situation. Unfortunately the application of ID3 to such data will still generate a model based on the assumption that an hierarchical structure is present. The inappropriateness of the model will only become evident when it is put to use, for predictive purposes say.

4.3: Bayesian Adjustments

A third problem which must be recognised when considering the value of the ID3 algorithm as a modelling tool for stochastic data is that it is necessary to consider the absence of any use of prior probabilities in the approach. The algorithm works with relative frequency data. The available data on relative frequencies stem from the available sample. As with all classification analyses, a key question is whether these sample relative frequencies are adequate reflections of population relative frequencies. Some discussion of Bayesian priors would seem mandatory therefore if a modelling application of ID3 is under consideration.

4.4: Ignoring Available Information

A fourth problem with ID3 as a modelling tool is that it requires that some information which may be present in a data set be thrown away before the analysis can begin. The algorithm assumes that all variables are simply sets of unordered categories. Thus if the original data is measured using a ratio, interval, or even ordinal scale, information must be thrown away before the analysis can begin. This problem is particularly significant in a financial context where the key question is usually not whether alternatives are different, but rather how good they are. Classes, defined by ranges of NPV values replace NPVs measured on a ratio scale, to avoid spurious precision.

4.5: Ordered Categories

This loss of information is unfortunate in itself, but also links to a further problem. The translation from say, NPV measured on a ratio scale, to a pair of NPV classes may not bother an analyst, since difference and ordering are equivalent. In this case ID3, using any of IM, G, or Chi-square will perform acceptably. However if more than two classes are needed for the outcome variable, because a measure of degree of goodness or badness of NPV is required for example, then problems can arise. Experimentation shows that different rule trees can emerge depending on whether IM, G, or Chi-square is used as a branching criterion. If the data given in Table 1 are now reanalysed using IM as the branching criterion then the rule tree shown in Figure 2 emerges. As can be seen this differs from that in Figure 1. Mingers has also observed this type of phenomenon.

The crucial point is that whichever of the three branching criterion is used, ID3 is unable to reliably capture any ordering among the resulting classes, i.e. NPV<0, 0≤NPV<T, NPV≥T, is a problem. The grouping of good and average versus bad in Figure 1 is a coincidence, rather than a consequence of any ability to appreciate rankings which is built into the algorithm. A cursory examination of Figure 2 shows that good is now paired with bad, while knowledge of the ordering involved suggests it should be paired with average, as happens to be the case in Figure 1.

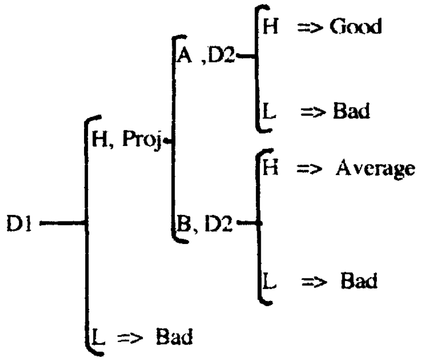

Figure 2
A Simple Rule Tree (Using IM)

5: ID3 as a Post Processor

As has been said earlier. ID3 has been used as a post processor of the output of a financial model. Race and Thomas have attempted to use the algorithm to post process the output of a financial model of an investment decision. This potentially allows a manager to gain more information about the relative importance of the various factors in the model in generating classes of NPV outcomes.

The conceptual problems associated with a model interpretation exercise are less severe than those associated with a model building exercise. Although the data set to be analysed may be generated by a stochastic process, this particular use of ID3 is closer to the work of Quinlan than it is to a modelling activity. The modelling activity has already been done, but the model still produces complex output which it is difficult for a manager to understand. The aim of using ID3 is therefore once again complexity reduction.

Neither the pruning issue nor the issue of Bayesian adjustment are relevant in this context. The pruning issue can be ignored since a variable selection process will have been involved in the building of the original model. The Bayesian adjustment issue will be irrelevant since the model output will by definition reflect the probability patterns present in the model.

However the application of ID3 is still beset with two of the problems previously identified. Firstly the original model may not involve an hierarchical structure, yet the rule tree is always derived on the assumption that it does. If an hierarchical rule tree is relevant for output interpretation, why did it not constitute the original model? The model developed by Race and Thomas is not obviously sequential and there must then be a danger that representing its output using a tree of IF...THEN rules will not enhance a manager's understanding of what is going on.

The second remaining problem is that any order information present in model output continues to be ignored in the rule trees ID3 produces.

6: Non Hierarchical Structures

There is no mechanical solution to this problem of ID3 imposing an inappropriate hierarchical rule structure on a set of data. All that can be done is to be sensitive to the problem. If the system generating the data is a human creation e.g. bank managers classifying loans as at risk, then a rule based model might then be anticipated to be present - though not necessarily an hierarchical rule tree. If the system is a natural one then the appropriateness of a model expressible as rules is more doubtful. All that can be done is to consider the suitability of a tree of rules as an underlying model a priori, and then to test the rule tree on a hold out sample.

The identification of a model to which ID3 can be applied as an output interpreter should be easier than selecting a situation in which to apply ID3 as a model building tool. After all the model structure is already known. The obvious candidate is a decision tree model of an investment decision [11]. There is an outcome variable, and there are decision points and random events to form the set of attributes.

There are two potential advantages to be gained from using ID3 to post process the outputs of a decision tree. Firstly, the relative importance of attributes may be highlighted by examination of their locations in an ID3 generated rule tree. Secondly, the rule tree may be a much simpler construct than the original decision tree. In what follows the reader should take care to distinguish between the *decision tree* being analysed, and the *rule tree* which results from the analysis.

6.1: Identifying Attribute Importance

A simple example shows what is involved in generating a rule tree from a decision tree. The decision involves a choice between project A and project B. Whichever project is chosen, demand will be high or low. The decision tree is shown in Figure 2.

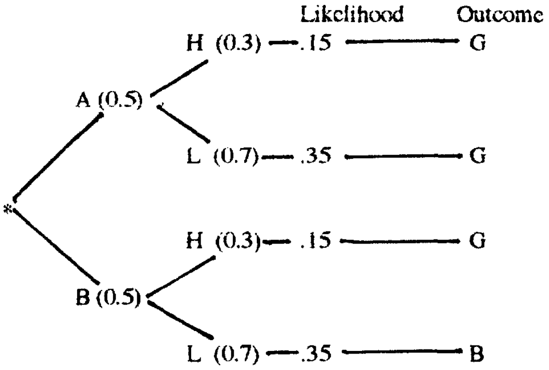

Figure 2
Simple Decision Tree

An unusual element in the tree is the assignment of equal probabilities to the branches of the decision node. The interpretation is that there is no reason prior to the analysis to prefer one project to the other. Multiplying probabilities along a branch gives the "likelihood" of that particular branch's outcome occurring. The data set shown in Table 4 is consistent with the decision tree in Figure 2. The likelihoods shown at the ends of decision tree branches in Figure 2 have been scaled to generate the frequencies in Table 4.

Outcome	Attributes		Freqs.,
	Demand	Project	
G	H	A	15
G	L	A	35
G	H	B	15
B	L	B	35

Table 4
Data Suitable For ID3

In order to decide which attribute is most important two contingency tables are created, Tables 5a and 5b:

		Outcome	
		G	B
Demand	H	30	0
	L	35	35

		Outcome	
		G	B
Proj	A	50	0
	B	15	35

Table 5a **Table 5b**
Contingency Tables For ID3

The value in a cell is the sum of the frequencies associated with the particular course of action/outcome combination. It is worth pointing out at this stage that a scaling factor can be applied to the contingency table elements to ease calculations. There is no absolute size of sample in this context. All the contingency table elements really represent are relative frequencies. Therefore as long as the same scaling factor is applied to each contingency table, relative Chi-square values say, will remain the same. Since there are only two output categories, the choice between Chi-square, IM, and G is irrelevant. The Chi-square figure for the demand attribute is 23.1. For the project attribute the Chi-square value is 53.8. The project decision is then the key attribute in determining the outcome.

The rule tree in this simple case is shown in Figure 3. The implication of this rule tree is that the decision maker is 'in control' of the situation. His decision is at the root of the tree: his decision is the most important determinant of the final outcome. Another way of viewing this is to say that a robust project, A, is available to him. An analysis of this view of robustness, the ability of a decision to produce an acceptable outcome under a wide range of future conditions can be found in Berry [12]. It differs from that popularised by Rosenhead which emphasises the modifiability of a decision sequence [13]. Robustness, as interpreted here, is associated with early appearance of project choice in the rule tree.

The fact that the transformation from decision tree to rule tree can highlight the extent to which a manager is 'in control' rather than subject to the whims of nature is a significant benefit of the technique. However, it should be recognised that the distribution of likelihood across decision tree branches can affect the form of the rule tree. To ensure that information about robustness/degree of control becomes available the rule tree should be carefully examined to see if equivalent trees exist.

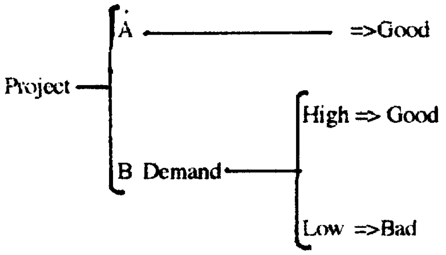

Figure 3
Rule Tree For The Simple Decision Tree

6.2: Complexity Reduction

This simple example demonstrates the mechanics of post processing a decision tree, and highlights the possibility of identifying 'in control' situations. However, in such a simple decision the manager could have seen the control possibility by simply examining the decision tree. Unfortunately, as the decision tree under analysis becomes more complex this task becomes more difficult. Post processing can present a simpler picture if the underlying logic of the decision being investigated allows it.

A more complex example, similar to that presented in Magee [13], will therefore be analysed. There are three alternative decisions, 'A' involves modernising an existing factory, 'B' closure and expansion of facilities located elsewhere, and 'C' modemisation and expansion of the existing factory. The consequences of the decision are modified in the ensuing three time periods by the state of demand. In each time period demand can be high (H), medium (M), or low (L). The probabilities of a given state of demand in periods 2 and 3 are conditional on the state of demand in period 1 and 2 respectively. Three levels of outcome are identified, excellent (3), good (2), and bad (1). The probability values used are as shown in Tables 6, 7, and 8.

Demand	Probability
H	.07
M	.43
L	.5

Table 6
Time Period 1

Demand	D2=H	D2=M	D2=L
D1=H	.6	.4	.0
D1=M	.4	.45	.15
D1=L	.15	.5	.35

Table 7
Time Period 2

Demand	D3=H	D3=M	D3=L
D2=H	.8	.2	.0
D2=M	.6	.35	.05
D2=L	.2	.6	.2

Table 8
Time Period 3

Using this data, and assigning equal probabilities to the three decision alternatives, the likelihood figures associated with each of the 81 decision tree branches can be calculated. They are shown in Table 9. Note that the zeros in the conditional probability tables mean that some branches in the decision tree represented by Table 9 have zero probabilities themselves. They therefore do not feature as possibilities in the reality being modelled, and will not be reflected in any rule tree produced. These branches have therefore been excluded from Table 9.

Demand at Stage			Outcomes		
1	2	3	A	B	C
H	H	H	2	3	3
H	H	M	2	2	2
H	M	H	2	2	2
H	M	M	2	2	2
H	M	L	1	1	1
M	H	H	2	2	2
M	H	M	2	2	2
M	M	H	2	2	2
M	M	M	2	2	2
M	M	L	1	1	1
M	L	H	1	1	1
M	L	M	1	1	1
M	L	L	1	1	1
L	H	H	2	1	1
L	H	M	1	1	1
L	M	H	1	1	1
L	M	M	1	1	1
L	M	L	1	1	1
L	L	H	1	1	1
L	L	M	1	1	1
L	L	L	1	1	1

Table 9
Tabular Representation Of A Decision Tree

The distribution of outcome values 1,2, and 3 is generated by the pattern of causality present in the model. If demand at stage 1 was the sole determining variable of output, the pattern would be different to that which would hold if project choice was of paramount importance. The lack of obvious pattern suggests that the logic represented in Table 9 as it stands is rather more complex than either of the simple causal models just mentioned.

The rule tree that results from the application of ID3 to the data in Tables 5,6, 7, and 8, is shown in Figure 4.

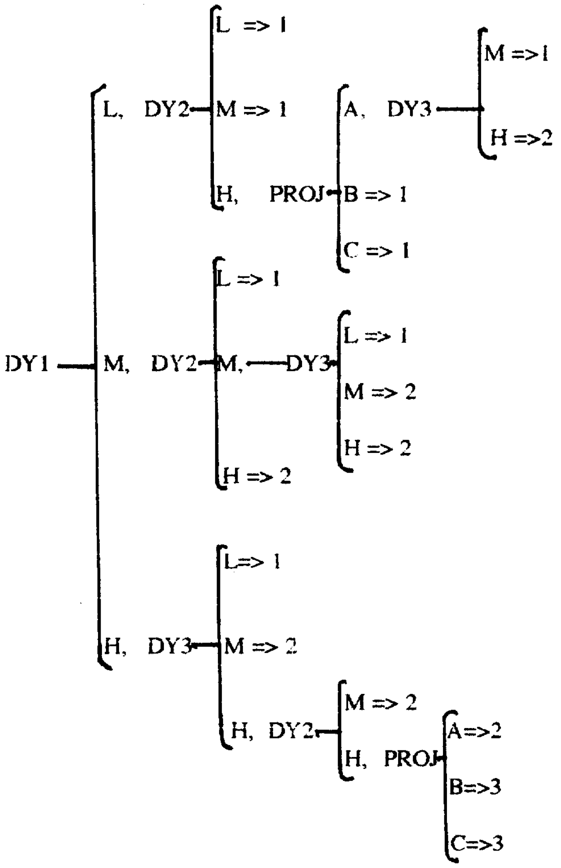

Figure 4
The Rule Tree For Table 9

This is a much simpler construct than the original decision tree. The original 81 outcome branches in the decision tree which were summarised in Table 8 have been reduced to 17 rules. This degree of complexity reduction, 81 decision tree branches to 17 rules, is of course not always possible. Figure 5 shows the rule tree generated by another decision tree generated by one possible set of changes to the distribution of 1s, 2s, and 3s, in Table 9.

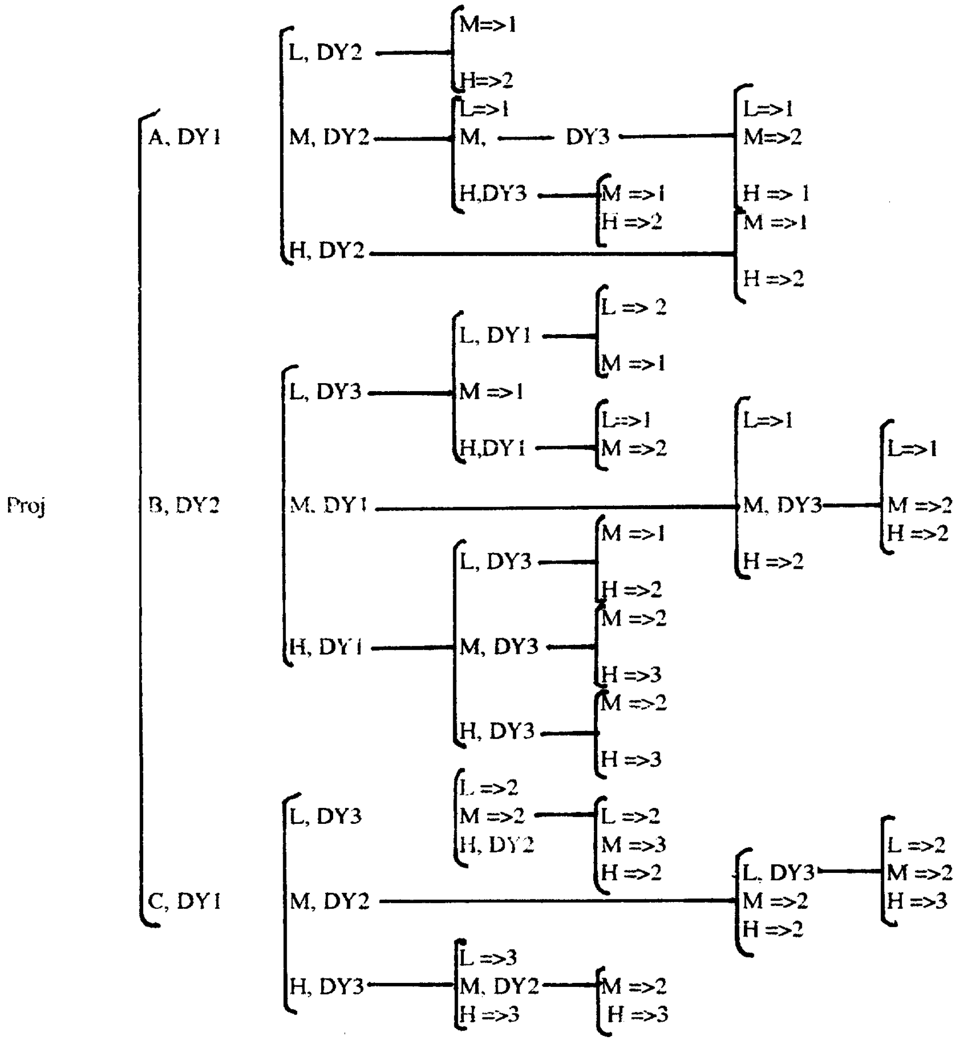

Figure 5
Complex Rule Tree Based On A Variant Of Table 9

The rule tree in this new scenario shows 40 rules. The complexity reduction is therefore much less than was achievable with the original version of the problem. However, project choice is at the root of the rule tree for this new version of the problem. This was not the case for the version of the problem that generated Figure 4. Now management has a degree of control over the outcome not present in the less complex situation originally analysed.

Project C is a robust choice, never leading to a bad outcome. This insight transforms the problem for the manager.

The rule tree of Figure 4 indicates that the manager is not in control. There is therefore an incentive to either delay project start, if that is possible, to allow demand uncertainty to resolve itself. Alternatively the manager may decide to revise the details of the alternative projects to make their performance less sensitive to variations in demand. If a project's cost mix can be changed to reduce the size of the fixed cost component then a positive NPV may be achievable at lower demand levels. However, there is a cost involved in such redesigns, over and above the manpower involved: the lower break even point will have been achieved at the cost of a reduced rate of increase of NPV as demand rises. The rule tree in Figure 5 presents an entirely different message. The manager can select a project which will perform acceptably no matter what demand conditions occur. Should it not be possible to delay the project start date, or if research to clarify future demand conditions is expensive, this is extremely useful knowledge for the manager.

7: Ordered Outcomes

The previous section has illustrated the application of the ID3 algorithm to post processing decision tree models and has emphasised the potential gains in understanding which can stem from such an approach. Any model with a similar sequential structure could similarly benefit. However, the discussion in the last section did not include any discussion of the branching criterion used in applying ID3. The decision analysed involved an outcome variable with three ordered categories, and as has been argued ID3 can have problems with ordered outcome categories. A solution to this problem will now be presented.

7.1: Simplistic Solutions

There are three possible approaches to this problem. The first is to ignore it, and invite the user of the rule tree to engage in further post processing based on any ordering information available outside the algorithm. This is feasible if the rule tree is simple.

A second approach is to use only two categories for the output variable. In this case, as was said earlier, difference and ordering are equivalent. This approach has an added benefit. Choice of measure of attribute importance seems of less significance. IM, Chi-square, and G, perform equivalently when the outcome variable has only two categories.

7.2: Adapting the Algorithm

A final approach to the problem is to change the algorithm. Each time a different branching criterion appears in the algorithm, IM, G, or Chi-square, it seems reasonable to argue that a different algorithm has been created. Mingers when discussing the problem of using ID3 to model noisy data substituted Chi-square for the IM measure and called the resulting algorithm ID5.

There seems no reason to stop at this point. There are many statistical processes which are capable of using the ordering information implied in a categorisation such as Good, Moderate, and Poor. An obvious way to allow ID3 to make use of ordering information therefore is to replace IM by some other statistical test. The appropriate test will depend on the form of the outcome variable. Thus this approach also deals with the problem of providing a reason for preferring one branching criterion to another.

Using as an example the data in Table 1, suppose that it was generated by a set of real outcome variable values, NPVs, as shown in Table 10.

CLASS	NPV
good	608
bad	-177
bad	- 78
bad	-904
avrg	21
bad	-144
bad	-152
bad	-400

Table 10
Original NPV Data

The bad category covers a wide range of experience. Table 11 shows the association between project choice and outcome variable value, with numbers in brackets representing frequencies. (There would be a similar table for each attribute variable.) The choice of which attribute to branch on first here needs to be based on a measure of association which reflects at very least the ordering inherent in the good, average, and bad category labels, and possibly the ranking inherent in the NPV values themselves. The basic question to be asked of the data in Table 11 is whether a difference exists between two batches of data, A and B.

A	B
(48) 608	(48) 21
(12) -177	(12) -144
(12) - 78	(12) -152
(28) -904	(28) -400

Table 11
Effect Of Project Choice On NPV

This looks like a question which could be answered by applying a simple t test. However, in this context the normality assumption of the t test may not be met. Therefore a non parametric approach is needed. Further it is unreasonable to assume that an attribute variable can have only two classes. Therefore a test which can hold when more than two is required. These considerations imply that a non parametric equivalent to ANOVA is wanted. The Kruskal-Wallis test is generally recommended in these circumstances. A description of this test can be found in Miller, Freund and Johnson [14]. There remains a further problem which must be considered; there are a large number of ties (equal values) present in the data set. The test must therefore be amended to take this into account as described in Siegel [15].

Applying the ID3 algorithm, based on Kruskal-Wallis, to the combination of Table 1 data as enriched in Table 10, generates again the rule tree in Figure 1. This reflects the ordering information available. The amended version of ID3 performs as hoped. Similar results have been obtained with other decision trees. This approach suggests the nature of the measures of attribute and outcome variables used determines which branching criterion is required. There is of course a limit on the form of variables which can be considered. Every attribute or outcome variable value is capable of generating a branch in a rule tree. There is a strong case for parsimony. As few categories as is possible in the context of the decision being made, should be the aim.

8: Summary and Conclusions

ID3, (and by implication similar algorithms), has serious defects as a model building tool. It also has defects as a post processing tool for model output. These are overcome when the algorithm is applied in a limited domain. Decision tree structures, for example, are within that domain. This structure is sequential and consistent with a nested rule set of the kind that ID3 seeks to identify.

The application of ID3 is capable of producing a rule tree which carries information about the key sources of risk. The factors near the root of the rule tree have greatest significance in determining outcomes. If in an investment example project choice is at the root of the tree, the manager is in command since project choice determines outcome to a significant extent. If however, state of demand is at the root of the rule tree, then fate is in command.

Decision trees can become a bushy messes. In these circumstances the visual nature of the model, so often a boon in communicating the analysis to management, ceases to perform. Numerical output then dominates management discussion. Even sensitivity analysis is of limited use because the key factors determining the optimal output value are not clear. In these circumstances ID3 can perform a useful function, identifying key factors, identifying robust alternatives, and often achieving substantial complexity reduction.

Even in this limited application area however, there remain problems with the ID3 algorithm, most notably the choice of branching criterion, and the loss of information about the ordering relation between outcome variable categories. The approach developed in this paper, which deals with both these issues, is to introduce a new statistical method into the algorithm. For the types of data typical of decision tree examples Kruskal-Wallis has proved a useful approach.

References

[1] J. R. Quinlan, Discovering rules by induction from large collections of examples. in *Expert Systems in the Micro Electronic Age*, (D. Michie, Editor), Edinburgh, 1979.

[2] K. Y. Tam & M. Y. Kiang, Managerial applications of neural networks: the case of bank failure predictions, *Management Science*, 38, 1992.

[3] H. Braun & J. S. Chandler, Predicting stock market behaviour through rule induction: an application of the learning-from-example approach, *Decision Sciences*, 18, 1987.

[4] P. Race & R. Thomas, Rule induction in investment appraisal. *Journal of the Operational Research Society*, 39, 1988.

[5] Y-H. Pao, *Adaptive Pattern Recognition and Neural Networks*, Addison Wesley, 1989.

[6] C. E. Shannon, A mathematical theory of communication. *Bell System Technical Journal*, 370-432, 623-659, 1948.

[7] J. Mingers, Expert systems - rule induction with statistical data. *Journal of the Operational Research Society*, 38, 1987.

[8] R. Sokal & F. Rohlf, *Biometry*, Freeman, 1981.

[9] S. Kullback, *Information Theory and Statistics*, Dover, 1967.

[10] H. Theil, *Economics and Information Theory*, North Holland, 1967.

[11] G. Erdogan, *Interpreting the outputs of financial models using ID3*, Unpublished PhD dissertation, University of East Anglia, 1992.

[12] R. H. Berry, *An examination of the analysis process underlying the decision to invest in reclamation and disposal facilities*, Unpublished PhD dissertation, University of Warwick, 1984.

[13] J. Rosenhead, Robustness analysis: keeping your options open. In *Rational Analysis for a Problematic World* (J. Rosenhead, Editor), Wiley, 1989.

[14] J. Magee, How to use decision trees in capital investment. *Harvard Business Review*, 79-96, 1964.

[15] I. Miller, J. E. Freund & R. A. Johnson *Probability and Statistics For Engineers*, Prentice Hall, 1990.

[16] S. Siegel, *Non parametric Statistics*, McGraw Hill, 1989.