MANAGEMENT CONTROL, A DECISION CRITERION FOR CAPITAL BUDGETING

Robert H. Berry

School of Information Systems

University of East Anglia

Norwich

NR4 7TJ United Kingdom

Phone: +44 (603) 561 61

Fax: +44 (603) 50 77 20

E-Mail: rhb@sys.uea.ac.uk

Duarte Trigueiros

ISCTE

University of Lisbon

Avenida das Forças Armadas

1600 Lisbon, Portugal

Phone: +351 (1) 793 50 00

Fax: +351 (1) 796 47 10

E-Mail: dmt@sara.inesc.pt

Management Control, a Decision Criterion for Capital Budgeting

Abstract

In this study we show that Quinlan's ID3 algorithm (1979) [4], originally intended for extracting rules from sequential data, is very effective in interpreting complex decision trees like those used in Investment Appraisal. As a result of using the ID3, a tree of rules is obtained, in which sequential events are hierarchically placed according to the strength of their causal relationship to outcomes. This tree of rules allows an assessment of the manager's control over the project, by comparing his effectiveness in causing a desired outcome with the one of non-controlled events.

We firstly introduce and discuss some previous research. Then, we explain the algorithm ID3 and the meaning of the rules obtained using it. We point out that the data structure that this algorithm requires is the hierarchical one, and that other characteristics of decision trees used in Investment Appraisal also match its requirements. Next, we show how the ID3 can be used for interpreting decision trees. Finally, we present a well-known example, illustrating this new technique, and we discuss the results in the light of the assessment of management's control over outcomes.

Management Control, a Decision Criterion for Capital Budgeting

This study focus on the interpretation of complex decision trees of the kind used in Investment Appraisal. We expect to show how to assess the degree of control of investors, and the interest of using such a decision criterion.

The analysis of investment decisions describes sequences of events in which the attributes that an investor can control, the decisions, are followed by other uncertain attributes that the investor cannot control. The consequences of a given decision will condition the next decisions to take. The final description of the whole set of possibilities is an hierarchical structure known as a decision tree.

Decision trees for Investment Appraisal are specific in some of their characteristics. Firstly, the outcomes don't need to be continuous-valued. They can be described by a simple nominal variable, having two possible values. These values only need to reflect the fact that the Net Present Value (NPV) of the project turns out to be positive or negative. Under normal conditions, it is unusual to expect high positive outcomes in a project. Such a situation would lead to sharp competition and profits would come down quickly. The real alternatives are between a positive NPV project and a negative one. Thus, information regarding the sign of the outcome is the only one that matters for decisions in Investment Appraisal.

Secondly, in general projects are of a one-off kind or unique. This makes the Expected-Value criterion unrealistic, as it ignores risk-aversion. The solution for this problem would be the use of Utility instead of money. However, investors resisted the use of Utility: They seem to be impractical because of risk atitude instability, difficulty to measure when the decision-makers are a board, and instability through time. Moreover, the assumptions involved in using Utility may easily be violated.

In this study we propose a different approach to the problem of interpreting decision trees. This approach is similar in style to the one of Rosenhead *et al.* (1972) [6] in their search for the robustness of decisions. However, our approach, rather than relating to robustness, takes advantage of the specific characteristics of decision trees used in Investment Appraisal, in order to identify the degree of control that investors have over future outcomes. In this study, we will produce a measure of management power over consequences inherent in an alternative.

1 Management Control and Robustness

This section introduces our concept of Management Control, comparing it with the well-known definition of robustness of decisions.

The intuitive meaning attached to the robustness of a decision is the thickness of the optimum. If an attribute can generate several possible outcomes, the manager will be interested in knowing which of the values of this attribute leads to an optimal outcome. However, the knowledge about this optimum isn't enough: In practice, managers know that attributes suffer deviations during the execution of projects. An outcome is known to be robust when its optimality isn't significantly affected by such deviations. An outcome is fragile when even small deviations from the optimum in the attributes deteriorate or invert its optimality.

Rosenhead *et al.* (1972) [6] elaborated the intuitive idea of robustness presented above. After stressing that robustness only makes sense after establishing the difference between plan and decision, they suggest a definition of robustness near the one of flexibility:

> "Little of what is known about the system being planned for is known with certainty. Much of what is not known cannot be expressed in terms of probabilities. The situation is one of uncertainty. As these uncontrollable and often unpredictable external events unfold, more information becomes available on the desirable future state and how to achieve it. In the light of this information it is natural and appropriate to reconsider and perhaps modify the as yet unimplemented stages of the plan. But if the possibility of making revisions has played no role in the specification of the earlier, implemented decisions, there may no longer be adequate residual flexibility.
>
> All decisions limit the future by commiting the present. A plan whose initial decisions limit the future as little as possible has an evolutionary advantage in an uncertain world.
>
> Consider a planning problem in which one decision must be chosen from a set $D\{\equiv (d_i)\}$ of short-term decisions; and in which one of a set S of alternative plans (or solutions) will be realized in the long run. Any initial decision d_i will restrict the attainable plans to a subset S_i of S.
>
> Suppose that some subset $\hat{S}$ of S is curently considered "good" or acceptable according to some combination of satisaficing criteria. A subset $\hat{S}_i$ of $\hat{S}$ will be attainable after an initial decision d_i. Then the *robustness* of d_i (see also [2] e [1]) is defined as:
>
> $$r_i = \frac{n(\hat{S}_i)}{n(\hat{S})}$$

where $n(S)$ is the number of elements in set S.

Robustness, a measure of the useful flexibility maintained by a decision, has characteristics which make it a suitable criterion for sequential decision making under conditions of uncertainty. It handles the uncertainty of the environment, not by imposing a probabilistic structure, but by stressing the importance of flexibility. (...)"

The line of thought sketched above has been object of further exploring by one of the authors, Rosenhead, in other articles (for example, [5]).

The definition of robustness introduced by Rosenhead *et al.*, despite the useful insights it provides, isn't suited for financial budgetting. The first aspect of such definition is the explicit denying of the use of risk measures such as frequency distributions. When outcomes are uncertain — in the sense used by Rosenhead *et al.* — the information about them is not zero. It is nonexistent. It may be that the likelihood of a given outcome is high, while the one of another is low. This definition ignores such a difference. When Rosenhead *et al.* refer an uncertain world, they mean a world in which managers would make plans whithout any knowkedge about the likelihood of future events.

We think that, despite Rosenhead's remarks, the cases of planning under total uncertainity are rare in Investment Appraisal. By ignoring all the a-priori available information, the suggested criterion throws away data that could be important concerning the decision. As an example, if, in a collection of outcomes, one of them has a likelihood of 99 in 100, this strong expectation would be ignored by the criterion suggested by Rosenhead *et al.*.

A consequence of rennouncing to any a-priori information is the fact that the criterion above can't avoid considering all the set of possible decision as if they had the same effectiveness in obtaining the desired outcome. The only thing that matters for establishing Rosenhead's definition is the total number of paths in each set. In order for such a definition to be adequate, all the paths should exhibit the same effectiveness in causing the desired outcome. But decisions aren't equally effective in causing an outcome. Some decisions are able to bring about an outcome, and some others contribute very little to its happening. We'll show examples of projects in which decisions, apparently important, are almost indifferent regarding the desired outcome.

Also, Rorenhead basis his definition of robustness in the assumption that

All decisions limit the future by commiting the present. A plan whose initial decisions limit the future as little as possible has an evolutionary advantage in an uncertain world.

This is indeed an interesting statement in negotiation processes but not in Finance. When negotiating, the goal is to achieve an agreement between the parts involved, no matter which. In Capital Budgeting the goal is a positive NPV. In most of the negotiation processes, the particular shape of the solution found is not important when compared with the fact that the parts were brought to agree. This strongly contrasts with the importance of the outcome in financial projects. Moreover, the second statemant of the above paragraph is eventually against the common experience of managers: A plan designed so that its decisions will commit the future as little as possible, will not have any special advantage, even in an uncertain world. The idea of marching to the desired outcome through non-commiting decisions seems somehow alien to Management. To walk through decisions carefully chosen so as to avoid any commitment, is generally the best way of not attaining any goals. This is because the extent to which a decision commits the future isn't, in general, independent of its causal weight. A strong commitment often means the perception by the manager that the decision is important for causing an outcome. In the limit, if a decision doesn't commit the future at all, its capacity for causing the outcome is probably zero.

Finally, the flexibility referred to by Rosenhead *et al.* depends on the complexity of the model. It is achieved by means of a growing sophistication. Only after introducing many alternatives is it possible to select the flexible ones. Therefore, only sophisticated, very complete models can be checked according to this criterion. Simplified models will not allow such an assessment. It is well known that simplicity and generality are linked. Analysts often try to prune out the non-significant branches or attributes of a model so that the resulting one reflects general trends and avoids very particular ones. In this sense, simplicity means also a sort of robustness: With a large number of unimportant attributes, it becomes more likely that the model will fail to recognize the desired goal based on its attributes.

We suggest that, instead of measuring the robustness of attributes, managers should assess their causal effectiveness or degree of control over future events. These concepts are, to some extent, similar because an effective decision is also robust and flexible. However, our concept is more specific than the above ones. According to it, good decisions should be those which, in a project, would be more able to cause the desired outcome than the attributes the manager can't command. A project would be reliable when the most effective attributes were decisions. It would be an unreliable one when the most causal attributes were out of reach of the manager. A decision with causal weight means that the manager has in his hands a real capacity of control. Before its execution he can shape it so that the desired outcome is the most likely.

Our definition is not dependent on particular structures of decision or on the degree

of sophistication of the model. The same decision tree can have all sorts of causal power associated with each attribute, depending on the probabilities present in the model. It is also a measure able to use all the information available about future events. How to assess the Management Control? We'll see that the algorithm ID3 can do it in the particular case of sequential decision structures having nominal outcomes.

2   An Introduction to ID3

This section introduces the algorithm ID3. The application of this algorithm is discussed. Finally, a simple example is given illustrating its use in sequential decision problems having nominal outcomes.

The ID3 algorithm (Quinlan, 1979) [4] is a hierarchical selection of the most informative attributes for explaining outcomes in seqential processes. The criterion for selecting attributes is the gain in information about the outcomes they apportion. Originally, algorithms for rule induction were intended to convert complex deterministic experience into logical structures. For example, the ID3 algorithm was presented as a tool able to extract rules from Chess ends. We now introduce the concept of Gain in information.

Probabilities are one amongst several ways of assessing expectations. In order for interpreting probabilities we must know the number of possible outcomes. For example, a probability of 1/2, in a game having two outcomes (the toss of a coin) means something different from the same probability of 1/2 in a six-outcomes one. In the first game it expresses no expectations or the absense of prior information about the outcome. In the second one it expresses a trend towards a given outcome and few chances associated with the other ones. In general, if N is the number of outcomes in a game, a probability of $1/N$ means the absense of prior knowledge or expectation about the outcome.

It seems useful to gather in a unique measure of uncertainty both the collection of probabilities, and the number of outcomes. This would allow the comparing of the uncertainty involved in games. Given a game having N outcomes, the number of digits needed to signal one of them is proportional to the logarithm of N. Thus, $logN$ is known as the logical variety of a game, and it measures the missing amount of information required to uniquely identify an outcome.

When outcomes show some regularity, the missing amount of information about them is no longer $logN$. There is a gain in information by knowing that, say, K_i of the N outcomes share the same attribute and can be lumped into one class known in advance. The knowledge of this attribute brings about some amount of information, which must be subtracted to the

logical variety, $logN$ in order to obtain the real uncertainty about outcomes.

In the case of a multiple classification, the remaining missing information needed to correctly identify each case is the difference between the missing information before the classification, $logN$, and the average amount of information that the classification carries with it:

$$H = logN - \sum_i \frac{K_i}{N} logK_i$$

This difference is known as "Entropy". It measures the missing information in a game: When $H = 0$ there is no missing information. Outcomes can be predicted by knowing its attributes. When $H = logN$ there is total ignorance about outcomes. Between those extreme cases, any expectation can be assessed.

The use of the Entropy of games simplifies the process of decision making. It allows the direct comparing the uncertainty of games. Instead of a collection of probabilities and N, only a unique measure is required. Also, Entropy is less misleading than the direct observation of probabilities. For example, in a two-outcomes game, the information gained by passing from a set of probabilities $\{p = 1/2, 1 - p = 1/2\}$, to $\{p = 1/3, 1 - p = 2/3\}$, is very small despite the apparently important difference. In the first case $H = 0.3$ and in the second one $H = 0.28$. Thus, the gain in information about the future resulting from a biased probability of $1/3$ is just 0.02. Notice that the concept of gain in information and the one of causal power are similar. In the above example, we could also say that a bias implies some power or control over the future outcome, while a maximum in Entropy implies no control at all.

The gain in information caused by some prior knowledge about outcomes,

$$G = \sum_i \frac{K_i}{N} logK_i, \tag{1}$$

is what the ID3 maximizes through a hierarchical selection. It begins by discovering the attribute which best predicts outcomes when G is used as criterion. Next, the sample is devided into as many sub-samples as the classes of the selected attribute, and a similar selection is performed in each of them. This selection obtains, for each sub-sample, the remaining attribute able to best explain the outcomes according to G. By repeating this process, a hierarchical structure is obtained, having the most informative attribute in its root, and the least informative ones in its leaves. Thus, the ID3 gathers probabilities, number of classes and structure, into one unique information measure. Sequential structures involving probabilities are converted into an hierarchical display of the relative importance of each attribute to causing the outcome.
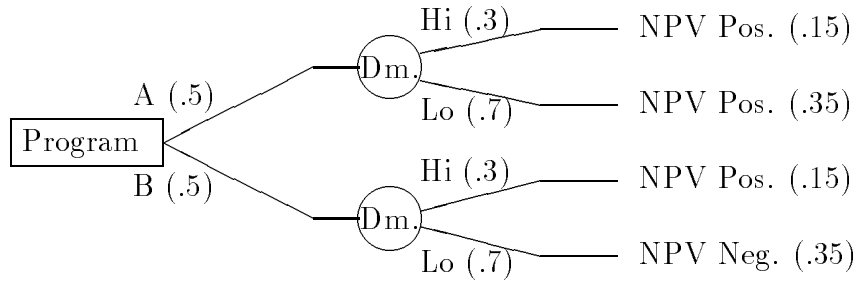
Figure 1: A simple decision tree. *Dm.* is the attribute "Demand". The likelihood of each class is also shown.

Net Present Value:	Demand	Program	Frequence
Positive (+)	High	A	15
Positive (+)	Low	A	35
Positive (+)	High	B	15
Negative (−)	Low	B	35

Table 1: The set of Outcomes, their attributes and expected frequencies.

How to implement the ID3 Algorithm: The ID3 is easy to omplement. A simplified example will show the entire process. Figure 1 is a decision tree. The attribute "Project" is a decision. The attribute "Demand" is an uncertain event. The outcomes (NPV) are the ones displayed. In order to implement the ID3 we follow these steps:

Consider only nominal outcomes: NPV are "positive" or "negative".

Consider any prior trend associated with the decision attributes: If no prior trend exists, the decision attributes are equally alike. So, in this case, a probability of 1/2 is assigned to each of them.

Build the set of observations: We assess the likelihood of each outcome by multiplying all the probabilities along every path of a decision tree. Outcomes, with their expected frequencies and attributes, form the collection of observations that the ID3 transforms. This collection is displayed in table 1.

Apply Rule Induction to the set of observations. Simple experiments can be carried out using the selective "query" facilities of a data-base software associated with contingency table facilities and popular statistics like the Chi-Square.

In this simple case, the first step of ID3 would generate two contingency tables: Outcome by Program, and Outcome by Demand:

7

Demand	Outcome		Program	Outcome	
	$NPV > 0$	$NPV < 0$		$NPV > 0$	$NPV < 0$
High	30	0	A	50	0
Low	35	35	B	15	35
Qui-Square: 23.1			Qui-Square: 53.8		

The Chi-Square can be used instead of the gain in information, as the former is a second-order approximation of the later. In our example, knowledge about the Program attribute discriminates between outcomes better than the corresponding knowledge about the Demand to pursue. So, the first attribute to be selected for building the tree of rules is the Program.

Next, we consider separately Program "A" and "B" observations, and we repeat the process for each of them. The result would be a logical structure like this:

If	Program	is A then:	the outcome is
			NPV Positive.

If	Program	is B then:	
	If	Demand is High then:	the outcome is
			NPV Positive.
	If	Demand is Low then:	the outcome is
			NPV Negative.

This structure is a tree of rules. It is important to consider the result as a tree, not only as a set of rules. The above rule tree expresses the effectiveness of each attribute in causing the outcome. In the given, simple example, the Program attribute turns out to be significantly more powerful than the Demand attribute for causing the outcome. Since investors can control the most causal attribute, the conclusion is that the projected investment is reliable.

ID3 and Capital Investment: Decision trees having nominal outcomes, and the corresponding rule trees generated by ID3, are re-presentations of the same information. Both structures are equivalent. The ID3 only re-arranges complex data so as to eliminate redundant information. This is because decision trees are hierarchical, and such a structure is precisely the one the ID3 supposes to exist in the data. If we apply the ID3 to data that isn't hierarchical, the resulting tree of rules is no longer a transformation: It will be an attempt to build a hierarchical model of non-hierarchical data.

Hierarchical data is frequent in Investment Appraisal. Sequential processes involving decisions and uncertain events, lead to hierarchical data. Quinlan original work on extraction of rules from Chess ends is also a typical example of the same kind of data: The earlier

Figure 2: Part of the decision tree for Prism Paints Inc.

PRISM PAINTS INC.		Year 2			Year 3		
		If Demand in year 1 is:			If Demand in year 2 is:		
Level of Demand	Year 1	Low	Medium	High	Low	Medium	High
Low	.50	.35	.15	0	.20	.05	0
Medium	.43	.50	.45	.40	.60	.35	.20
High	.07	.15	.40	.60	.20	.60	.80

Table 2: Prism Paints Inc: Probabilities that demand will be low, medium and high.

movements of the player condition all the next ones. The conclusion is that the characteristics of ID3 and the ones of problems in Investment Appraisal seem to match.

Moreover, the ID3 can only handle the poorest type of measurements, the simple nominal categories. But this isn't a problem in Investment Appraisal, as only two situations are really important (corresponding to the sign of the final NPV of the project)

3   An Experiment: Prism Paints Inc

In this section we use the well-known decision tree described by Magee (1964) [3] to illustrate the application of Management Control to a real-world problem. Firstly, we shall quote some paragraphs of Magee's introduction.

> "Prism Paints Inc. must decide what to do with one of its manufacturing plants which is rather small and unable to supply the quality of products required in the current market. There is considerable managerial controversy over the proper course of action — whether to modernize the operation by construction

of better facilities at the location, or to scrap the existing plant and supply the area involved from the company's facilities elsewhere. (...) There are three basic patterns of of operation offering promise:

Program A: To modernize the plant in question and also expand elsewhere. This program is less expensive when annual demand is less then a known threshold.

Program B: To close the plant in question and expand elsewhere. This program is less expensive when annual demand is between two known values.

Program C: To modernize and expand the plant in question. This program is less expensive when demand is above a known threshold."

"Evidence on market demand and the data underlying the demand forecasts lead to an estimation of how likely it is that the demand will fall in the low, medium, or high range in each of the three stages of the project. These estimations must be made in relation to the demand in the preceding stage. That is, demands are conditional on previous stage's demand. Table 2 contains these estimations."

This problem is outlined by the decision tree of figure 2 (page 9). The three choices under discussion, Program "A", "B" or "C", are followed by three years, P 1, P 2 and P 3, each of them having an uncertain demand level (Low, Medium, or High). Thus, the resulting decision tree has $3^4 = 81$ final leaves. Some refinements of the original problem, like the allowances for program shifts, weren't considered here.

Knowing the initial investment that each alternative involves, and establishing a required return on investment, it is possible to express outcomes in terms of Net Present Values (NPV). In our experiment, simulation was used to create several different NPV scenarios. In this study we apply ID3 to one of those scenarios. We allowed the simulated NPV to have three possible outcomes ("Negative", "Positive" and "Extra Profits"), in order to make models more complex.

The tree of rules obtained is shown in figure 3 (page 11). The original decision tree of 81 leaves was reduced to a new tree of only 17 rules. The redundant information discarded by the ID3 was significant. The ID3 hierarchically scaled attributes according to their causal weight. In our example there is only one decision attribute, the choice of the program. Since the question is whether to select program "A", "B" or "C", the branches of the rule tree in which this attribute appears far away from the root, show that a manager's control over the project would be small. This is clearly the case in figure 3. A low demand in year 1 leads to important differences between programs, while for a medium or high demand the

```
                          ┌── = L, NPV = 1
                          │
          ┌── = L,   D 2  │── = M, NPV = 1
          │               │                       ┌── = M, NPV = 1
          │               │    = "A",   D 3  ┌──── │
          │               └── = H,   Prog. ── │    └── = H, NPV = 2
          │                                   │── = "B", NPV = 1
          │                                   └── = "C", NPV = 1
          │
          │               ┌── = L, NPV = 1
          │               │                   ┌── = L, NPV = 1
  D 1 ────│── = M,   D 2  │── = M,    D 3 ──── │── = M, NPV = 2
          │               │                   └── = H, NPV = 2
          │               └── = H, NPV = 2
          │
          │               ┌── = L, NPV = 1
          │               │
          └── = H,   D 3  │── = M, NPV = 2
                          │                   ┌── = M, NPV = 2           ┌── = "A", NPV = 2
                          └── = H,    D 2 ──  │                          │── = "B", NPV = 3
                                              └── = H,   Proj. ── ────── │── = "C", NPV = 3
```
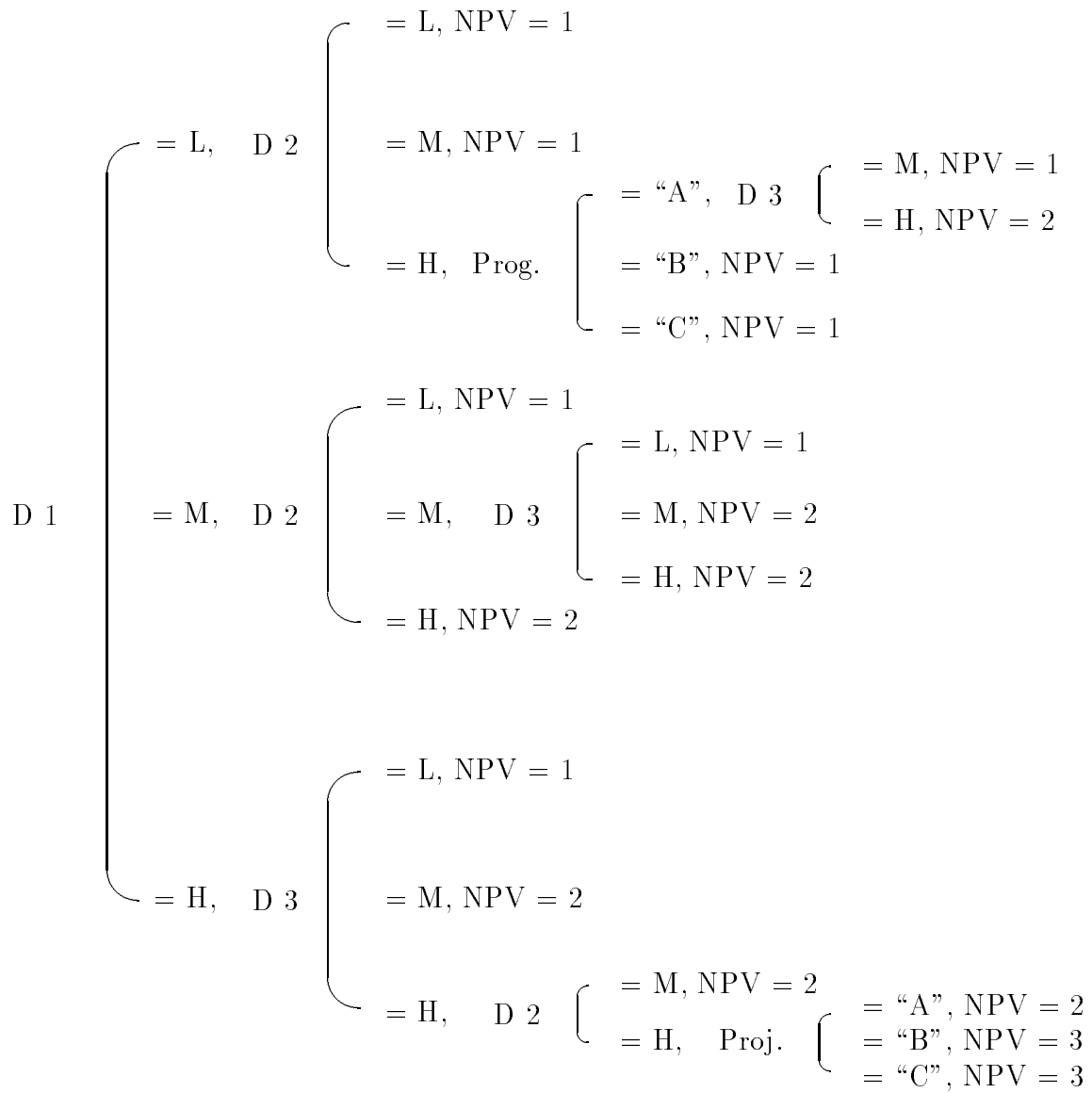
Figure 3: Prism: Rule Tree, first simulation.

program doesn't influence the outcome. The decision attribute doesn't appear in first place and this means that whatever managers decide, the year 1 demand is what really condition the results.

This simple experiment seems to show that Rule Induction can be used to interpret sequential decision models. The way ID3 scales attributes can provide managers a measure of a project's dynamic behaviour.

References

[1] J. Friend and W. Jessop. *Local Government and Stratergic Choice*. Tavistock, London, 1969.

[2] S. Gupta and J. Rosenhead. Robustness in sequential investment decisions. *Management Sciences*, 15:B–18, 1968.

[3] J. Magee. How to use decision trees in capital investment. *Harvard Business Review*, pages 79–96, September 1964.

[4] J. Quinlan. Discovering rules from large collections of samples — a case study. In D. Michie, editor, *Expert Systems in the Micro Electronic Age*. Edimburgh University Press, 1979.

[5] J. Rosenhead. Planning under uncertainty. *Journal of the Operational Research Society*, 31:209–216, 1980.

[6] J. Rosenhead, M. Elton, and S. Gupta. Robustness and optimality as criteria for stratergic decisions. *Operational Research Quarterly*, 23(4):413–431, 1972.