From Facts to Rules to Decisions: An Overview of the FRD-1 System

Ibrahim F. Imam and Ryszard S. Michalski

Center for Artificial Intelligence
George Mason University
Fairfax, VA. 22030
{iimam , michalski} @aic.gmu.edu

ABSTRACT

One of the central goals of knowledge discovery in databases is to produce knowledge useful for decision making. However, the form in which knowledge can be easily discovered is often different from the form in which its is most readily used for decision making. This paper describes a methodology in which knowledge is discovered in declarative form (decision rules). When it is needed for decision making, it is on-line transferred to a procedural form (a decision structure) optimized for the given decision making situation. The system FRD-1, which implements the methodology, is briefly described, and illustrated by its application to a problem in the area of construction engineering.

1. INTRODUCTION

Research on knowledge mining in databases has so far been primarily concerned with discovering interesting or useful knowledge. The ultimate goal of such a process, however, is not just to store knowledge, but to use it for decision making. Therefore, it is important to develop a system that can dynamically link data and knowledge discovered with decision making. Such a system should be able both to discover knowledge and to transform it to a form that is most suitable for any given decision making situation.

This paper proposes a novel approach to this problem. In this approach, the function of knowledge discovery is separated from the function of using the discovered knowledge for decision making. The first function is performed by an inductive learning program that searches for knowledge relevant to making a given class of decisions, and stores the discovered decision knowledge in the form of decision rules. The second function is performed by a program that transforms discovered knowledge to a decision structure optimized according to the needs of the decision making situation. The second function is performed when there is a need for assigning a decision to new data points in the database (e.g., a classification decision).

By *decision structure* is meant an acyclic graph, whose nodes are assigned binary or multivalued tests that can be single attributes, relations, or multi-attribute equations evaluated to discrete intervals of values. Branches emanating for a node are assigned single values or sets of values from the value set of the test associated with the node. These sets of values (represented by internal disjuction) are mutually disjoint; their union spans all the possible values (outcomes) of the test. Terminal nodes (leaves) are assigned single decisions, or sets of (alternative) decisions with corresponding probabilities. Leaves can also be assigned a special value "?", which stands for "do not know." A decision structure becomes a decision tree when nodes are assigned single attributes, branches are assigned single values (test outcomes), and leaves are assigned single decisions.

The paper briefly describes the proposed methodology, and shows how the system tailors decision structures to different decision making situations. The methodology is illustrated by a problem of determining a decision structure for the wind bracing design in different decision making situations. These situations differ in that they may require 1) a decision structure that avoids some unavailable or costly test; 2) assigning an "unknown" decision in a situation when there is insufficient information, 3) informing the user what new information is needed in order to assign a specific decision with high certainty, and 4) alternative decisions with an estimate of the probability of their correctness, when the needed information can not be provided. A more detailed description of the methodology and results of various comparative studies are presented in (Imam and Michalski, 1994).

2. AN OVERVIEW OF THE FRD-1 METHODOLOGY

The proposed system consists of two subsystems. The first subsystem searches for useful knowledge in the data. Knowledge is represented in the form of decision rules, which are a form of declarative knowledge representation. Such a representation makes no restrictions on the order in which different tests are applied and is relatively easy to modify and improve. To discover knowledge in data, the system employs the inductive learning program AQ15/C (Michalski, et al, 1986). When there is a need to apply the discoverd knowledge to new data for the purpose of decision making, the AQDT-1 system is used to transform the decision rules into a decision stucture that is tailored to the given decision making situation (Imam & Michalski, 1993). Figure 1 shows an architecture of the FRD-1 methodology.

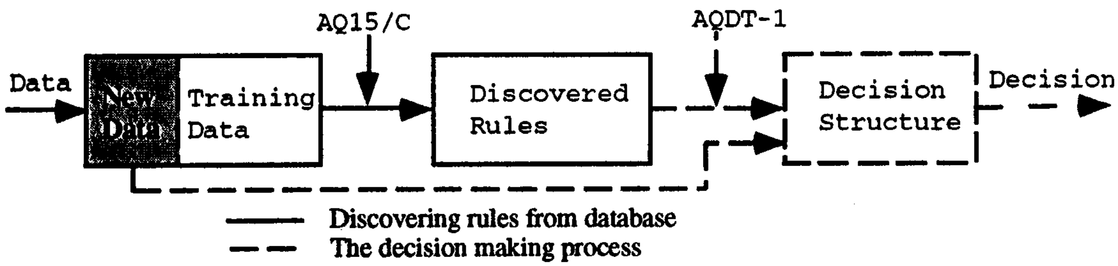


Figure 1: An architecture of the FRD-1 system.

It is assumed that the database is not static, but is regularly updated. A decision making problem arises when there is a case or a set of cases to which the system has to assign a decision based on the knowledge discovered. Each decision making case is defined by a set of attribute values. Some attribute values may be missing or unknown. A new decision structure is obtained such that it suits the given decision making problem. The learned decision structure associates the new set of cases with the proper decisions.

3. SUBSYSTEM I: DERIVING KNOWLEDGES FROM DATA

The first subsystem applies the AQ15/C inductive learning program to discover knowledges in the data. The program creates decision rules from examples of decisions, using the STAR methodology (Michalski, 1986). The simplest algorithm based on this methodology, called AQ, starts with a "seed" example of a given decision class, and generates a set of the most general conjunctive descriptions of the seed (alternative decision rules). Such a set is called the "star" of the seed example. The algorithm selects from the star a description that optimizes a user-defined criterion reflecting the needs of the problem domain. If the criterion is not defined, the program uses a default criterion that selects the description that covers the largest number of positive examples (to minimize the total number of rules needed) and, if there is a tie, it selects the description that involves the smallest number of attributes.

If the selected description does not cover all examples of a given decision class, a new seed is selected from uncovered examples, and the process continues until a complete class description is generated. The algorithm can work with few examples or with many examples. It can optimize the description according to a variety of easily-modifiable hypothesis quality criteria.

AQ15/C can generate decision rules that represent either *characteristic* or *discriminant* concept descriptions, depending on the settings of its parameters. A characteristic description states properties that are true for all objects in the concept. The simplest characteristic concept description is in the form of a single conjunctive rule (in general, it can be a set of such rules). The most desirable is the maximal characteristic description, that is, a rule with the *longest* condition part which states as many common properties of objects of the given class as can be determined. A discriminant description states properties that discriminate a given concept from a fixed set of other ("context") concepts. The most desirable is the minimal discriminant descriptions, that is a rule with the *shortest* condition part.

To illustrate the AQ15/C rule learning, let us consider a problem of determining a decision structure for wind bracing design (Arciszewski, et al, 1992). The wind bracing data contains four decision classes representing different categories of the buildings: high (C1), medium (C2), low (C3) and infeasible (C4). Examples of design have been expressed in terms of seven attributes that are shown in Table 1. The data consists of 335 examples, 220 randomly selected examples were used as training examples for learning a concept description and the remaining 115 examples for testing.

In the experiment, AQ15/C was applied to the training examples to obtain rules from the data. Figure 2 shows the discovered rules and their strengths. The *t*-weight (*total-weight*) of a rule for some class is the number of examples of that class covered by the rule. The u-weight (*unique-weight*) of a rule for some class is the number of examples of that class covered only by this rule.

Attribute	Attribute value
x1	number of stories
x2	bay length
x3	wind intensity
x4	number of joints
x5	number of bays
x6	number of vertical trusses
x7	number of horizontal trusses

Table 1: The symbolic attributes of the wind bracing data and their full names.

C1: High Buildings

1 [x1 = 1] [x5 = 1 v 2] [x6 = 1] [x7 = 1 v 2 v 3] (t:18, u:09)
2 [x1 = 1] [x2 = 2] [x5 = 1 v 2] [x6 = 1] (t:10, u:02)

⋮

C2: Medium Buildings

1 [x1 = 2 v 3 v 4] [x4 = 3] [x5 = 2 v 3] [x7 = 2 v 3] (t:28, u:02)
2 [x1 = 2 v 3 v 4 v 5] [x3 = 1] [x5 = 2 v 3] [x6 = 1] [x7 = 1 v 2 v 3] (t:27, u:10)

⋮

Figure 2: A sample of rules discovered by AQ15 for the wind bracing data.

4. SUBSYSTEM II: DRIVING DECISION STRUCTURE FROM DECISION RULES

AQDT-1 generates decision structures from decision rules by determining the "best" test (here, attribute) at each step of the decision process. The attribute is selected by analyzing the rules. The method aims at producing decision structures with the minimum number of nodes and/or the minimum cost (where the "cost" was defined as the total cost of classifying unknown examples, given the cost of measuring individual attributes and the expected probability distribution of examples of different decision classes). For more detailed explanation of the algorithm and its comparison with other systems, see Imam & Michalski (1993).

AQDT-1 chooses attributes on the basis of an *attribute utility* criterion, which combines three elementary criteria: 1) *disjointness*—that captures the attribute effectiveness in discriminating among decision rules of different decision classes, 2) *dominance*—that measures the attribute relevance by counting the number of rules that contain the attribute, and 3) *extent* —that measures the number of different attribute values present in the rules.

Table 2 shows the disjointness and dominance values for each attribute. The disjointness criterion is selected as the first criterion for selecting tests to be nodes in the decision tree. Figure 3 shows a decision structure learned in the default setting of FRD-1 parameters from AQ15/C rules. It has 5 nodes and 9 leaves. Testing this decision structure against 115 testing examples results in 102 examples matched correctly and 13 examples mismatched. For comparison, when the C4.5 system (Quinlan, 90) for learning decision trees from examples was applied to the same set of examples. It produced a decision structure with 17 nodes and 43 leaves. The decision tree matched correctly 97 and mismatched 18 examples.

Criterion's name	Attributes						
	x1	x2	x3	x4	x5	x6	x7
Attribute Disjointness	5	3	0	11	3	18	3
Attribute Dominance	45	34	42	33	40	30	54

Table 2: Initial values of selection criteria for each attribute of the wind bracing problem.

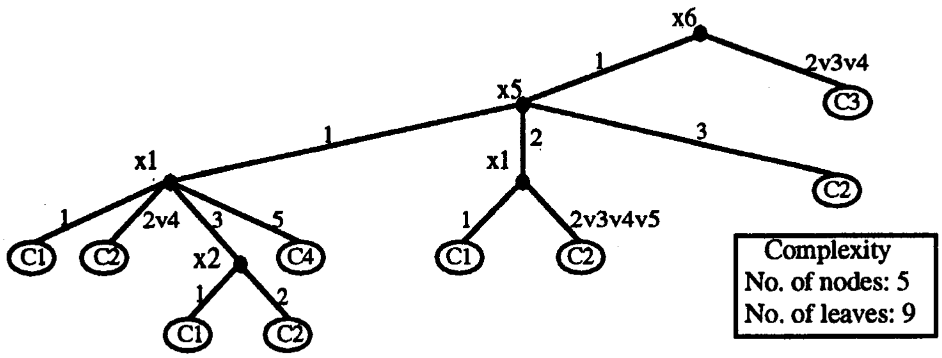


Figure 3: A decision structure learned by AQDT-1 from AQ15/C wind bracing rules.

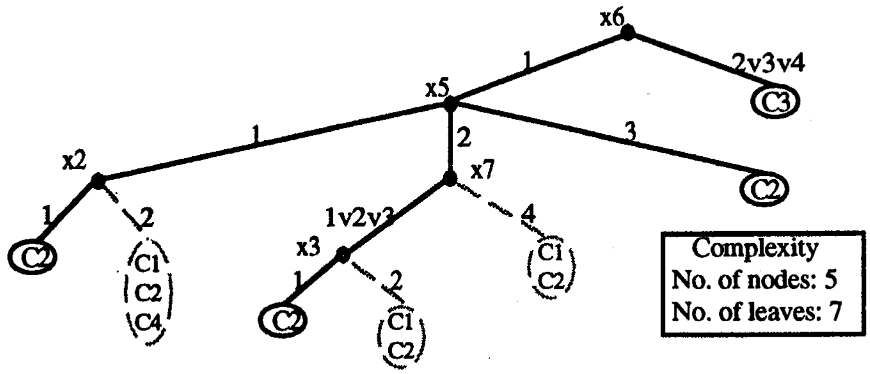
5. SPECIAL CASES IN DECISION MAKING

Decision making situations can vary in several respects. In some situations, a complete information about a data item is available (i.e., values of all attributes are specified), in others the information may be incomplete. To reflect such differences, the user specifies a set of parameters that control the process of creating a decision structure from decision rules. FRD-1 provides several features for handling different decision making problems: 1) generates a decision structure that tries to avoid unavailable or costly attributes (tests); 2) generates "unknown" leaves in situations where there is insufficient information for generating a complete decision structure; 3) provides the user with the most likely decision when measuring a required test is impossible; 4) provides alternative decisions with an estimate of the likelihood of their correctness, when the needed information can not be provided.

Avoiding unavailable attributes. The cost of an attribute refers to the cost of measuring that attribute during the process of decision making. Sometimes an attribute cannot be measured at all, in which case cost is assumed to be "infinite". Quinlan (1990) proposed a method to handle the "unknown attribute value" problem, in which the relative probability of an example belonging to different classes is used to determine class membership. FRD-1 provides a new way of solving this problem. Specifically it attempts to generate a decision structure that either does not contain the unavailable attribute or shifts it to the bottom of the decision structure.

For example, assume that the value of x1 is unavailable for a given data item. FRD-1 generates the decision structure shown in Figure 4. Assume now that x6 can be measured but has higher cost than other attributes. In this case, AQDT-1 generates the decision structure shown in Figure 5.

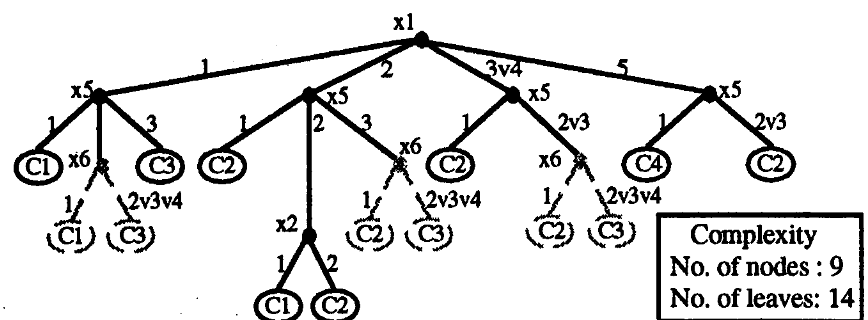
Completeness of Decision Structure: Another decision making problems that can be solved by FRD-1 is when the testing example does not match any branch in the structure. To explain this problem, let us assume that the value of x1 is unavailable for a given decision making problem. The generated decision structure in this case is not complete (in Figure 4, the dark lines only). Assume that the available information is (x1 = 1; x2 = 2; x3 = 2; x5 = 2; x6 = 1; x7 = 2). In this case, decision structure indicates likely decisions of which an estimate of their likelihood. The estimate reflects the frequency of classes when the data points satisfy the conditions on the path from the root to the leaf. Details on how to calculate these frequencies are in (Imam & Michalski, 1994).



The ratio present the likelyhood of each decision at the given leave.

Figure 4: A decision structures learned where x1 is unavailable.

The *leaf weight* for a class C shown in figure 4 is calculated as the sum of the t-weights of the rules satisfied by conditions in the decision structure from the root to the given leaf divided over the sum of the t-weights of all rules of the class C. This is an estimate of the weight of this leaf in assigning the decision class C to an event satisfying the condition of the path from the root to the leaf. The weight of the leaf for class C is viewed as a likelihood of an event satisfying the conditions on the path to the leaf.



The dotted lines indicate situation in which x6 can be evaluated

Figure 5: A decision structures learned with high cost of x6.

6. CONCLUSION

The paper outlined a methodology for discovering decision rules from data and transfering them to decisions structures most suitable for any particular decision making situation. The preliminary experiments show that it is easier to generate a decision structure tailored to any given decision making problem from rules than to modify a decision structure once created from examples. FRD-1 has several novel features including: provides a method for creating decision structure from rules; to generate nodes with approximate decisions in situations when there is insufficient information to generate a complete decision structure. These features are were illustrated in an experiment concerned with determining a decision structure for wind bracing design.

ACKNOWLEDGMENTS

The authors thank Eric Bloedorn and Ken Kaufman for their useful comments on the earlier draft of the paper.

This research was conducted in the Center for Artificial Intelligence at George Mason University. The Center's research is supported in part by the Advanced Research Projects Agency under the grant No. N00014-91-J-1854, administered by the Office of Naval Research, and under the grant No. F49620-92-J-0549, administered by the Air Force Office of Scientific Research, in part by the Office of Naval Research under grant No. N00014-91-J-1351, and in part by the National Science Foundation under grant No. IRI-9020266.

REFERENCES

Arciszewski, T, Bloedorn, E., Michalski, R., Mustafa, M., and Wnek, J., (1992). "Constructive Induction in Structural Design", Report of Machine Learning and Inference Labratory, MLI-92-7, Center for Artificial Intelligence, George Mason University.

Imam, I.F. and Michalski, R.S., (1993). "Learning Decision Structures from Decision Rules: A method and initial results from a comparative study", in *Journal of Intelligent Information Systems JIIS,* Vol. 2, No. 3, pp. 279-304, Kerschberg, L., Ras, Z., & Zemankova, M. (Eds.), Kluwer Academic Pub., MA.

Imam, I.F. and Michalski, R.S., (1994), "From Data to Knowledge to Optimal Decision Structures: A Methodology and Expriments", Report of Machine Learning and Inference Labratory, MLI-94-xx, Center for AI, George Mason Un.. (to appear)

Michalski, R.S., Mozetic, I., Hong, J. & Lavrac, N., (1986). "The Multi-Purpose Incremental Learning System AQ15 and Its Testing Application to Three Medical Domains", *Proceedings of AAAI-86*, (pp. 1041-1045), Philadelphia, PA.

Quinlan, J. R. (1990). "Probabilistic decision structures," in Y. Kodratoff and R.S. Michalski (Eds.), *Machine Learning: An Artificial Intelligence Approach*, *Vol. III*, San Mateo, CA, Morgan Kaufmann Publishers, (pp. 63-111), June.