# AUTOMATED ACQUISITION OF DECISION RULES: THE PROBLEMS OF ATTRIBUTE CONSTRUCTION AND SELECTION

BY

PAUL WILLIAM BAIM

B.S., University of Florida, 1981

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Computer Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1984

Urbana, Illinois

# Dedication

**To LJH now LHB for still thinking I was worth the wait...**

## Acknowledgements

Many people were part of this effort. First, I would like to thank Professor Ryszard Michalski for his inspiration, guidance, and editorial aid, as well as providing technical guidance for much of this work. I would like to thank Bob Stepp for his tireless aid when the chips were down. Thanks also to those others who contributed time and effort such as Kent Spackman, MD for his help with the mysteries of Craniostenosis, and Bill Ruesink, PhD for his help with Black Cutworm habits. Carl Uhrik and Bob Reinke deserve thanks for reviewing versions of this work and offering insight and criticisms when necessary. June Wingler helped when last minute typing was needed on too many occasions to list. Gayanne Carpenter was ready even at the eleventh hour to check the format. To all these people, and others not mentioned, thanks. I would also like to thank my parents for their constant moral support. Finally, I wish to acknowledge the Office of Naval Research for their support under Grant No N00014-82-K-0136, and the National Science Foundation for their support under Grant MCS-82-05166.

## TABLE OF CONTENTS

# 1 INTRODUCTION

Expert systems are computer programs which contain and are capable of applying expert knowledge of a particular problem domain to solving problems within the domain. Examples of such systems are Mycin [Shortliffe,1982], used for medical diagnosis, and Plant/ds [Michalski,1982a], used for diagnosing soybean plant diseases. Each system identifies a given input example as belonging to a particular class within the system's domain of expertise. The Plant/ds system, for example, arrives at the most plausible diagnosis for an afflicted soybean plant based on the symptoms the plant exhibits. In this case, the domain of soybean diseases has been divided into a number of categories (classes) each representing a different disease. The task of the expert system is to identify which class (disease category) a given plant's symptomology represents.

Traditional expert system development techniques require an expert in the domain of interest to specify the important factors he uses to arrive at expert determinations. In the Mycin system, the experts were required to formulate a set of rules to apply to test cases to determine a plausible course of treatment for each case. The experts found that while they were capable of reaching determinations easily, they were less able to articulate the process they used to arrive at them. This approach requires the expert to analyze his own methodology from a perspective to which he is unaccustomed. The systems which result are often inefficient because they use unnecessary data in the process of analyzing a test case, and unreliable because contingencies not previously encountered may result in faulty or indeterminate results [Clancey,1981] [Hayes-Roth,1980]. To repair these shortcomings, the system designer resorts to iterative development methods in which the expert is presented with the faulty results and asked to augment the current system so that correct results are produced. When further testing uncovers new flaws, the development loop is repeated.

This paper describes a system which seeks to relieve the domain expert of the burden of analyzing his own methods. He uses his knowledge to reach expert determinations (classifications) for examples from his domain of expertise. Once the expert has classified each sample case, the system uses techniques patterned on a general model of the expertise development process to attempt to discover the relationships prevalent in each class used by the expert, and formulates general rules for classifying future examples. This approach, termed *learning from examples* minimizes the need for the expert to codify his technical knowledge. The system consists of three programs which are described in detail later in the paper. The first, CONVART, computes attributes of the time-varying characteristics of examples. The second, VARSEL, selects the "most relevant" attributes for formulating classification rules from those produced by CONVART and those initially available. The third program, GEM, uses the attributes identified by VARSEL to formulate classification rules for each of the classes represented by the set of examples. VARSEL was devised and written by the author, while CONVART and GEM were written by others [John Davis, and Robert Stepp, et. al., respectively].

## 1.1 Current Expert System Development Techniques

### 1.1.1 Knowledge Encoding

One traditional approach to the creation of expert systems is shown in Figure 1. A *knowledge engineer* works in concert with a domain expert to characterize the process used by the expert to reach decisions [Hayes-Roth,1980]. Unfortunately, domain experts are often unable to describe their own mental processes clearly enough to enable the production of efficient, complete and reliable systems. This problem makes many iterations through the development loop necessary. The knowledge engineer encodes the expert's knowledge, applies the resulting system to sample cases, uncovers insufficiencies or inconsistencies, and returns to the expert to overcome these difficulties. This iterative process strains the knowledge representation scheme and the process control structure until the system becomes unwieldy and difficult to use and maintain.

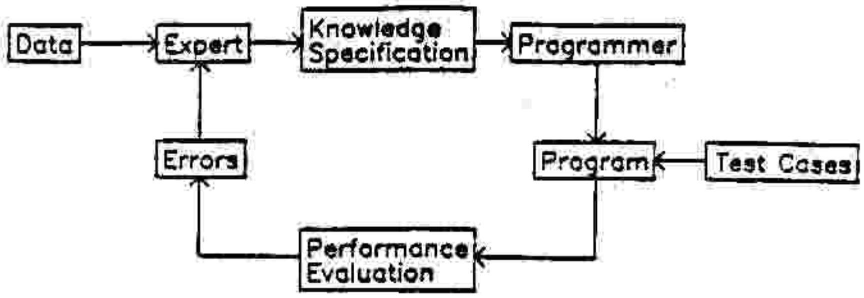


Figure 1. The traditional expert system development loop

A second failing of the knowledge encoding approach arises from the necessity of having a domain expert in the development process. This requirement precludes the possibility of expert-system development in domains in which no experts have developed because the area is too new or obscure. The development of domain expertise is a slow process because of the tremendous difficulties inherent in discovering relevant avenues of inquiry into the underlying processes and structures within the domain.

### 1.1.2 Automated Inductive Inference

A relatively new approach toward expert system development involves *inductive learning* systems which use empirical evidence, in the form of examples, to derive rules by which determinations may be made for data presented to the system. Such systems use *automated inductive inference*, a method whereby rules for classifying events are derived using computer programs which systematically process the data to discover patterns and interelationships which may be useful for distinguishing each class from the others [Michalski, 1972, 1980, Angluin,1982]. The failing of these systems arises from the limitations inherent in dealing with statistically valid populations of sample data. For some automated inductive inference methods, an exponential relationship exists between data-set size and processing time required [Rabin,1974]. This causes a problem termed *combinatorial explosion* where the processing time for a large data set is often far too lengthy to be of practical value. Modern inductive learning systems, e.g., AQ11 or ID3 are able to avoid this problem by applying appropriate heuristic methods. However, the task of looking for patterns of interaction and interdependency within data is so complex, and since the amount of data should be large if the results are to be reliable, the modern systems must use some means of further *data-reduction* if the processing time for the learning programs is to be acceptable. Traditionally, system designers rely on domain experts to tell them which data are relevant and which should be ignored. Consequently, expert system development using automated methods is also prone to iterative development because the domain experts frequently misjudge which data are important or relevant. This produces systems with the same basic flaws encountered in knowledge engineered systems: inefficiency and unreliability.

## 1.2 Requirements for Better Expert System Development

The major shortcoming of current expert system development techniques is the continued dependency on domain experts to provide information they are normally unprepared to provide. System development techniques requiring the aid of domain experts for initial development are doomed to inadequate performance unless new insights can be gained into the ways in which domain expertise arises and is implemented. This paper describes a data-driven system which is largely domain independent and can operate without the explicitly stated *a priori* knowledge of the problem domain usually needed by system developers [Buchanan,1979]. Assuming that the necessary knowledge is contained in the pairing of sample events with expert determinations, sufficiently powerful methods for extracting the knowledge are needed [Tunstall,1974] [Shapiro,1981]. The ability to do this requires a method which can assess the relevance of attributes reliably while remaining so computationally efficient that extensive pre-selection of attributes by a domain expert is not required [Chen,1974]. Hence, the role played by the domain expert in ordinary automated system development paradigms is now also largely automated. The system can deal with large amounts of data and indicate promising avenues of exploration for researchers in the problem domain and expert system implementors alike.

## 2 A MODEL OF THE EXPERTISE DEVELOPMENT PROCESS

The following definitions will be used throughout the discussion which follows:

- An *exemplar* or *example* is an object which exemplifies a given decision class.
- An *attribute* is a measurable characteristic of an exemplar within the problem domain. Examples are color, height, and annual rainfall. Attributes fall into two categories: static attributes measure characteristics which have a constant value while dynamic attributes measure characteristics whose values vary relative to some other variable such as time, for a given exemplar.
- A *selector* is a relation between a given attribute and the value of that attribute. Examples include color≠red, height=1.75m, and weight≤95kg.
- An *event* is a vector of attribute values characterizing a given exemplar. An event may be represented by a conjunction of selectors.
- A *decision class* or *class* represents the membership of one or more exemplars in a category characterized by some common denotation.
- An *event set* is a set of events which are members of known decision classes.

An example of an event set is given in Figure 2. The event set is comprised of two classes of two events each and one class of one event. Each event is in the form of a conjunction of four selectors. A Dodge-Dart, for example, weighs 6000lb, is twelve feet long, and can carry six passengers.

A model of the expertise development process, patterned after one in [Dietterich,1981], is presented in Figure 3. The sequential model of expertise development presented here has four major processing steps and feedback from any step to any previous step [Kanal,1978]. Each of the four processing steps— attribute construction, attribute selection, rule formation, and rule implementation— are described in detail below.

**Class : Car**
**Examples :** [model=Dodge-Dart] ∧[weight=6000lb] ∧ [length=12ft] ∧ [passengers=6]
[model=Toyota-Corolla] ∧ [weight=5000lb] ∧ [length=11ft] ∧ [passengers=5]

**Class : Truck**
**Examples :** [model=Mack] ∧ [weight=36000lb] ∧ [length=35ft] ∧ [passengers=3]
[model=Peterbilt] ∧ [weight=340000lb] ∧ [length=35ft] ∧ [passengers=3]

**Class : Bus**
**Example:** [model=Greyhound] ∧ [weight=21000lb] ∧ [length=32ft] ∧ [passengers=30]

**Figure 2. A sample event set.**

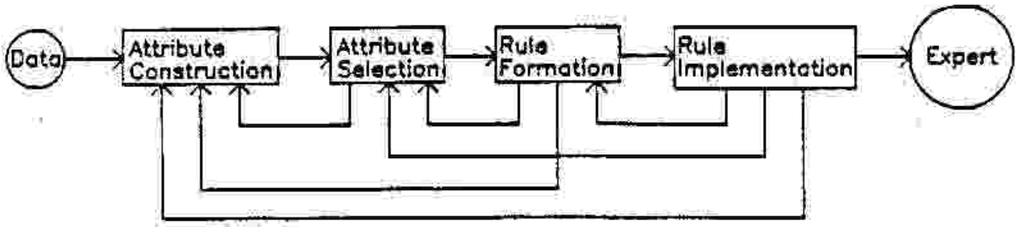


**Figure 3. A model of the expertise development process**

## 2.1 Attribute Construction

Once an initial set of attributes have been selected by the expert as potentially relevant, the first processing step involves the construction of derived attributes. Derived attributes are produced by applying a variety of logical, statistical, or mathematical operations to the initial attributes. For example, a derived attribute may be a mathematical product of some numerical attributes or a logical expression of propositional attributes (properties that are true or false).

Additional attributes may be derived by first noting that some attributes change over time. Rainfall would be one example of such an attribute. The rainfall on a given field could be given as a series of values representing "rainfall" on each day for a given time interval. New attributes can be constructed from this type of attribute using statistical operations such as averaging (e.g. "the average rainfall per week"). Arithmetic operators can be applied to numeric attributes. For example, if we have the attributes length, width, and height, we might apply the multiplication operator to construct the attribute "volume." Neither mathematical nor statistical operators can be applied to non-numeric (i.e., *symbolic*) attributes such as "blood type" or "eye color."

To perform attribute construction, operators are applied to the attribute values in the hope of discovering new attributes which have greater class-differentiating abilities than the original attributes. This process is termed *constructive induction* [Michalski,1982b] [Davis,1981].

## 2.2 Attribute Selection

After attribute construction is completed, the set of constructed attributes is joined to the set of initial attributes and each attribute is evaluated in terms of its potential *relevance* (i.e., its potential utility for differentiating classes based on its value for a given exemplar) [Kodratoff,1982]. This is the process of attribute selection in which unusable or less informative attributes are discarded and only those attributes which best distinguish the classes are retained.

## 2.3 Rule Formation

The third processing step is rule formation. Rule formation entails characterizing the classes by logic expressions involving selectors. The most interesting expressions are those which most compactly capture the knowledge necessary to classify a new event reliably. The two basic conditions that a set of such expressions (rules) must meet are *consistency* and *completeness*. Completeness requires that each exemplar must satisfy some rule from the rule set, and consistency requires that each exemplar satisfy *at most* one class from the set of classes.

A set of simple discriminant rules for the event set of Figure 2 is shown in Figure 4. The first rule reads: "If the length is less than 20 feet, the vehicle is a car." The second rule reads: "If the length is greater than 20 feet and the vehicle can carry less than four passengers, then it is a truck."

| | |
|---|---|
| Rule 1: | $[\text{length} < 20\text{ft}]$ ::> car |
| Rule 2: | $[\text{length} \geq 20\text{ft}] \wedge [\text{passengers} \leq 3]$ ::> truck |
| Rule 3: | $[\text{length} \geq 20\text{ft}] \wedge [\text{passengers} > 3]$ ::> bus |

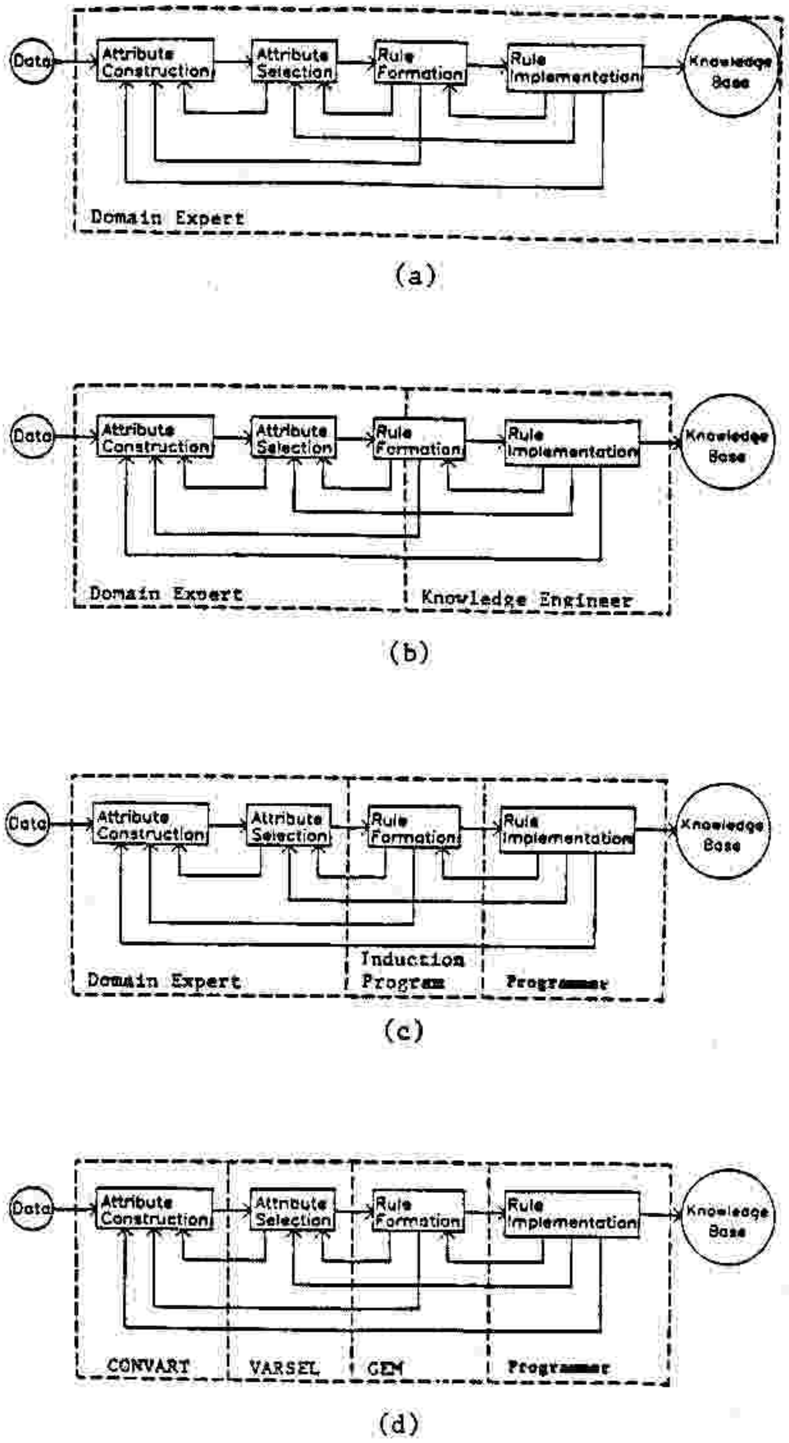
Figure 4. A set of discriminant rules derived from the event set in Figure 2.

## 2.4 Rule Implementation

The final processing step is rule implementation. This is the encoding of the rules generated in the previous step in such a fashion that they may be applied to new events supplied by a user. Although implementing the rules is a simple process, designing the user interface is not. The most important characteristics of an expert system which will be used by people who are not intimately familiar with the software are: easy data entry, comprehensive presentation of results including measures of how confident the system is in its analysis, and the ability to explain the reasoning processes which led to the result presented.

## 2.5 Conventional Methods in Terms of the Model

Figures 5a-c show the model of Figure 2 configured to correspond to its application to human expertise development, the knowledge engineering approach to expert system development, and the automated induction method respectively. In Figure 5a, human expertise development, we see that the human expert is responsible for all processing, implementation, and application of the resulting expertise. In Figure 5b, the knowledge engineering approach, the expert is still responsible for all attribute processing but now the knowledge engineer must share the task of rule generation and becomes responsible for implementing the rules on a computer. Figure 5c shows the automated inductive inference method in terms of the general model. The human expert is now responsible only for deciding which attributes are relevant. The data is then presented to an inference program in terms of the specified attributes, and the program develops rules which characterize the data. The software engineer must then implement the rule set on the computer. Figure 5d shows the software described below as it relates to the model.



(a)

(b)

(c)

(d)

Figure 5. A progression toward increasing automation of the expert system development process:

(a) Pre-expert-systems method
(b) Classical expert systems development
(c) Automated rule formulation
(d) Automated attribute construction and selection and automated rule formulation using the software described in this paper

# 3 AN IMPLEMENTATION OF THE MODEL

## 3.1 Attribute Construction Using the Program CONVART

### 3.1.1 A Rationale for Attribute Construction

Time-dependent attributes are those that capture the value of some characteristic of an event as it changes with time. Such attributes are, therefore, useful for describing procedures or processes. If such attributes could be used by an inference engine, a broader range of problem domains would be approachable by expert systems. Current inference engines, however, are only capable of processing so-called static attributes which have only a single value for each event. If inductive inference techniques are to be applied to these problem domains, some method for representing dynamic (time-dependent) attributes by static attributes is needed. Two ways of describing dynamic attributes statically may be employed. The first involves characterizing the progression of values for the attribute over time as a polynomial or other functional entity. This is the method used in the series of Bacon programs [Langley,1982].

The second method is the computation of fundamental descriptive quantities for the set of measured values. Examples of such quantities for numeric attributes (i.e., attributes whose possible values have a natural partial-ordering) are maximum, minimum, average, etc. Examples of quantities for nominal attributes (i.e., attributes whose possible values have no inherent partial-ordering such as color or shape) are most-common-value and least-common-value. This is the method used by CONVART [Davis,1981]. CONVART is composed of a number of routines which compute such measures for dynamic attributes. It is implemented in Pascal on a Vax 11/780.

Another problem found in both dynamic and static attributes is the use of continuous variables which may take on any real value within some range of values. Interpreting the meaning of a given value for such an attribute is often difficult, and applying inductive inference techniques to them often produces unwieldy results. The solution CONVART applies to continuous attributes is *discretization*. Discretization is the process of dividing the range of possible values for an attribute into a manageable set of discrete values based on the distribution of the observed values throughout the range of possible values. CONVART uses a nearest-neighbor algorithm to arrive at plausible discrete ranges.

### 3.1.2 Using CONVART in An Expertise Development System

Given that CONVART produces many derived attributes from each dynamic attribute and adds them to the original static attributes, the burden falls on the inference engine to weed out the bad attributes and use only the relevant ones. Although a rudimentary relevancy evaluator was incorporated in CONVART, the developer admits that it is far from sufficient for the task [Davis,1981]. Therefore, the rule constructor is still faced with the task of dealing with many attributes measured (or computed) for many events. Because of the complexity of the algorithms on which inductive inference programs are based, the processing required for event sets containing many attributes is unacceptable. Attribute selection is needed to make the process of expert system development practical for "real" event sets.

## 3.2 The Attribute Selection Program VARSEL

The program VARSEL performs the task of attribute selection based on a measure of attribute relevancy for class discrimination calculated using the method described below. It is implemented in Pascal on a Vax 11/780. The program evaluates each attribute individually and then compiles a subset of attributes which completely differentiates each class from the others. The user may select one of two ways in which the compilation is to proceed. The first method is an adaptation of a procedure proposed by [Lbov,1965] in which attributes are chosen randomly using a weighted selection scheme in which more relevant attributes are more likely to be chosen. The method described here uses this principle of *random adaptive search*, modified so that small subsets of attributes are evaluated and the individual relevancy measure of each constituent attribute is improved or degraded based on the performance of the selected subset as a whole

[Smith,1980] [Bethke,1981]. If the process converges, the most relevant attributes are then indicated by high relevancy measure.

The second method involves what has been termed a *greedy* search scheme in which attributes are added to a subset of attributes until a sufficiently discriminatory attribute set has been found. Both methods are described in detail below.

### 3.2.1 A Rationale for Attribute Selection

Attributes may be of three types. Numeric attributes have real or integer values which represent measured quantities directly. Examples of numeric attributes are "temperature" and "number of legs." Ordinal attributes have integer values which capture a true partial ordering present in the attribute values. An example of such an attribute is "quality" which may have the values "very bad", "bad", "acceptable", "good", and "excellent." The third type of attribute is a symbolic attribute. The values of symbolic attributes have no inherent partial ordering. One example of such an attribute is "shape." Another term often applied to symbolic and ordinal attributes is *nominal*.

The problem of selecting the most relevant set of attributes to describe objects for the purpose of inductive learning is traditionally approached using various methods of attribute selection such as factor analysis, multidimensional scaling, and linear transformation, developed in the fields of pattern recognition and statistical decision theory. These methods involve statistical or information-theoretic measures which identify the principal attributes in the event set [Andrews,1972] [Harmon,1960] [Lawley,1963]. These methods are most effective for numeric attributes when the size of the event set is statistically significant. When attributes are symbolic (categorical or propositional) and the event set is small, these methods are not adequate [Stearns,1976] [Zadeh,1981]. In fact, using the methods mentioned above, "it is almost impossible to formulate general guidelines regarding the selection of physical [attributes] and structural [attributes]" [Tou,1974]. To analyze such *nominal* attributes mathematically, numeric values must be assigned to each possible value of the attribute (e.g., color might be encoded as: red=1, yellow=2, blue=3, etc.). When statistical methods are applied to such attributes, the results from the statistical methods may vary if the ordering of the values changes (e.g., red=2 and yellow=1). Thus, although the actual values of the attribute have not changed, the measures derived from them have. This is clearly erroneous. Therefore, there is a need for new methods applicable to multivalued or propositional symbolic attributes and for small numbers of events.

### 3.2.2 A Relevancy Measure

The difficulty of performing satisfactory attribute selection using statistically-based methods may be overcome by analyzing the observed values of the attributes using a method based on heuristic knowledge of the properties of attribute values typically found in event sets of interest. Such a system could deal effectively with symbolic and ordinal attributes since these attributes do not behave statistically in the same way that numeric attributes do.

The computation time necessary to process a set of events is large when the event set contains many attributes. A computationally inexpensive means of rapidly eliminating irrelevant attributes would reduce the computation time required. The method described below accomplishes all of these goals.

Classical information theory has been used to attack the problem of attribute relevancy measurement by modeling decision trees as information sources and attribute values as "messages." The information contained in a message "M" depends on the probability of the message and is expressed by:

$$I = \log_x \left( \frac{1}{P(M)} \right) \qquad (1)$$

where: P(M) is the probability of message "M."

In event sets of interest, a given value (message) may occur in events representing more than one class. If we can assume that the message is *correct* for only one of the classes in which it occurs, then the information provided by the message depends not only on the probability of the message, but on the probability that the message is correct P(c∩M):

$$I = \log_x\left(\frac{1}{P(c \cap M)}\right) \tag{2}$$

Others have postulated that the information in such a "questionable" message is [Quinlan,1982]:

$$I = \log_x\left(\frac{1}{p^+}\right) + \log_x\left(\frac{1}{p^-}\right) \tag{3}$$

where: $p^+$ is the probability of the correct occurence of the message.

$p^-$ is the probability of the incorrect occurence of the message.

The derivation of this expression relies on one fundamental assumption. The correct and incorrect messages are probabilistically independent. This, however, is incorrect. The two messages are, in fact, mutually exclusive by their very nature. This invalidates the expression. We might then hypothesize that, given the information measure in equation 2, that for "v" possible values (messages) for a given attribute:

$$I_{tot} = \log_x\left(\frac{1}{P(c \cap M_1)}\right) + \log_x\left(\frac{1}{P(c \cap M_2)}\right) + \cdots + \log_x\left(\frac{1}{P(c \cap M_v)}\right) \tag{4}$$

holds if the messages are independent. Unfortunately, the most frequent case is that the different messages are not necessarily independent and such an assumption has an associated risk. In addition, these measures effectively assume that we are interested in viewing the messages as "correct" or "incorrect." While such a binary view is sometimes useful, the multi-class nature of event sets of interest calls for a more robust measure. Therefore, we require a measure which retains the multi-class view of the world and is immune to the vagaries of data collection and random occurrence of events which often produce widely disparate class sizes within an event set.

We approach this problem by first making a simple assumption: since we wish to measure a characteristic of the data (which, hopefully, reflects the nature of the domain of interest) we may take advantage of the fact that for a given attribute, for a given event, only one value may be observed in the data. This assumption allows us to further assume that the possible messages (values) for a given event are not independent, but mutually exclusive. Unlike the case of equation 3 above, we realize that dealing with such messages by combining their information contents can be misleading. Instead, we will manipulate their probabilities directly. The total probability of all possible messages given mutual exclusivity is:

$$P_{tot} = P(M_1)+P(M_2)+ \cdots +P(M_v) \tag{5}$$

If each of the terms is further divided into probability of correctness $P(c \cap M_i)$ and error $P(e \cap M_i)$, the total probability is:

$$P_{tot} = P(c \cap M_1)+P(e \cap M_1)+P(c \cap M_2)+P(e \cap M_2)+ \\ \cdots +P(c \cap M_v)+P(e \cap M_v) \tag{6}$$

This equation may be separated into two:

$$P_c = P(c \cap M_1)+P(c \cap M_2)+ \cdots +P(c \cap M_v) \tag{7a}$$

$$P_e = P(e \cap M_1)+P(e \cap M_2)+ \cdots +P(e \cap M_v) \tag{7b}$$

capturing the probability of correctness or incorrectness for all messages (values). Each term in Equation 7b may be expanded into:

$$P(e \cap M) = \frac{n_{e_1} + n_{e_2} + \cdots + n_{e_{(m-1)}}}{N_{tot}} \tag{8}$$

where: m is the number of classes in the event set.

$n_{e_i}$ is the number of occurrences of the message in error class "i" of which there are (m-1) since one of the classes is "correct," not in error.

$N_{tot}$ is the total number of events in the event set.

We see that this value is sensitive only to the *relative sizes* of the class for which the value is correct and the remaining classes in the event set. Since this is not desirable as mentioned above, we introduce normalization factors into each term to remove this size bias. Since this will change our result into something other than a pure probability, we will designate our new result the *likelyhood of error for a given message (value)* $l_{e,M}$:

$$l_{e,M} = \frac{1}{m-1}\left(\frac{n_{e_1}}{N_{e_1}} + \frac{n_{e_2}}{N_{e_2}} + \cdots + \frac{n_{e_{(m-1)}}}{N_{e_{(m-1)}}}\right) \tag{9}$$

where: $N_{e_i}$ is the total number of events in error class "i."

Given the likelyhood that each possible value will be interpreted incorrectly, the total likelyhood of error is:

$$L_{e,a} = \sum_{M=1}^{v} l_{e,M} \tag{10}$$

since the number of possible messages equals the number of possible values "v" of the attribute. The value of this measure is between 0 and 1 but the range is inverted in meaning (i.e., the best attribute has likelyhood of error of 0). The final step is to invert the range to arrive at a measure we will call the *relevance* of an attribute denoted by ρ:

$$\rho = 1 - L_{s,a} \tag{11}$$

Which has value 0 for worst-case attributes and value 1 for perfectly discriminant attributes. The final consideration is the determination of which class is the "correct" one for a value which appears in more than one class. We have assumed, for the purposes of this study, that the class for which the value is most likely is the correct class. In practice, this is the class for which the value of the term $n_{m_{class}} / N_{class}$ is greatest for a given attribute value "m."

An example will show the application of equation 11. In Figure 6a, a event set of three attributes measured for three classes is given. Figure 6b shows that, for attribute $x_2$, the normalized probabilities of occurrence in each class of the value "1" is 1.0, 0.0, and 0.0 respectively. Since there are 3 out of 4 events with value 1 in class 1 (0.75), 0 out of 2 events with value 1 in class 2 (0.0), and 0 out of 1 events with value 1 in class 3 (0.0). The same analysis for the other messages (values) completes the table. Figure 6c shows the calculation of likelyhood of error for each of the values. For example, assuming the "correct" class for value 2 is class 2 since the term in Figure 6b for value 2 is greatest for class 2, The calculation is the sum of the remaining 2 "error" classes times 1/(m-1) which is 1/2. Finally, Figure 6d gives the remaining calculation of relevance.

Three experiments were performed to examine the performance of equation 11 using the program PROMISE which calculates the value of $\rho$ for attributes in a given event set. The inductive inference engine was the program AQ11 [Michalski,1978]. Both PROMISE and AQ11 were implemented in Pascal on a Cyber 175 computer.

| | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|
| $c_1$ | 1 | 1 | 1 |
| | 1 | 1 | 1 |
| | 1 | 1 | 4 |
| | 1 | 2 | 2 |
| $c_2$ | 2 | 2 | 2 |
| | 2 | 4 | 2 |
| $c_3$ | 3 | 3 | 2 |

(a)

| | $\frac{n_{M_{class}}}{N_{class}}$ | Class 1 | 2 | 3 |
|---|---|---|---|---|
| Message | 1 | 0.75 | 0.00 | 0.00 |
| | 2 | 0.25 | 0.50 | 0.00 |
| | 3 | 0.00 | 0.00 | 1.00 |
| | 4 | 0.00 | 0.50 | 0.00 |

(b)

$$l_{e,1} = \frac{1}{2}\left(\frac{0}{2} + \frac{0}{1}\right) = 0.000$$

$$l_{e,2} = \frac{1}{2}\left(\frac{1}{4} + \frac{0}{1}\right) = 0.125$$

$$l_{e,3} = \frac{1}{2}\left(\frac{0}{4} + \frac{0}{2}\right) = 0.000$$

$$l_{e,4} = \frac{1}{2}\left(\frac{0}{4} + \frac{0}{1}\right) = 0.000$$

(c)

$$L_{e,x_2} = 0 + 0.125 + 0 + 0 = 0.125$$

$$\rho = 1 - 0.125 = 0.875$$

(d)

**Figure 6. Calculating relevance:**

**(a) A sample event set**

**(b) A chart showing the number of events with a given value in each class divided by the total number of events in the class for each of the possible values of $x_2$**

**(c) The calculations of likelyhood of error for each the possible values of $x_2$**

**(d) The final steps in the calculation of the relevance of $x_2$**

### 3.2.2.1 Experiment I

The relevance $\rho$ for more than one attribute at a time may be calculated by considering the values of several attributes in an event as a single *compound* attribute. To test the behavior of $\rho$ for this type of processing, the "Animals" event set, described in [Michalski,1975], was processed by PROMISE. The Animals event set is shown in Figure 7. The data and a set of rules for Animals are given in Appendix D. The protozoan creatures in each of the fourteen classes may be described by thirteen attributes:

- $x_1$ is the number of black circles on the body.
- $x_2$ is the number of tails.
- $x_3$ is the number of crossmarks on tails.
- $x_4$ is the number of easily distinguished extremities.
- $x_5$ is the body texture.
- $x_6$ is the number of empty circles on the body.
- $x_7$ is the number of empty squares on the body.
- $x_8$ is the number of empty triangles on the body.
- $x_9$ is the type of tail.
- $x_{10}$ is the shape of the body.
- $x_{11}$ is the number of sharp or straight angles.
- $x_{12}$ is the number of "eyes" (half-black circles).
- $x_{13}$ is the number of black squares on the body.

First, the relevance of the individual attributes were evaluated by PROMISE, and the attributes were arranged in order of increasing value of relevance. Next, all projections on pairs of attributes were evaluated and the results were analyzed as follows: The combinations were ranked by increasing value of $\rho$ and the range of observed values of $\rho$ was divided into 10 equal-sized sub-ranges. The number of occurrences of a particular attribute in the set of pairs with relevance values in a sub-range is expressed as a percentage of the total number of occurrences of all

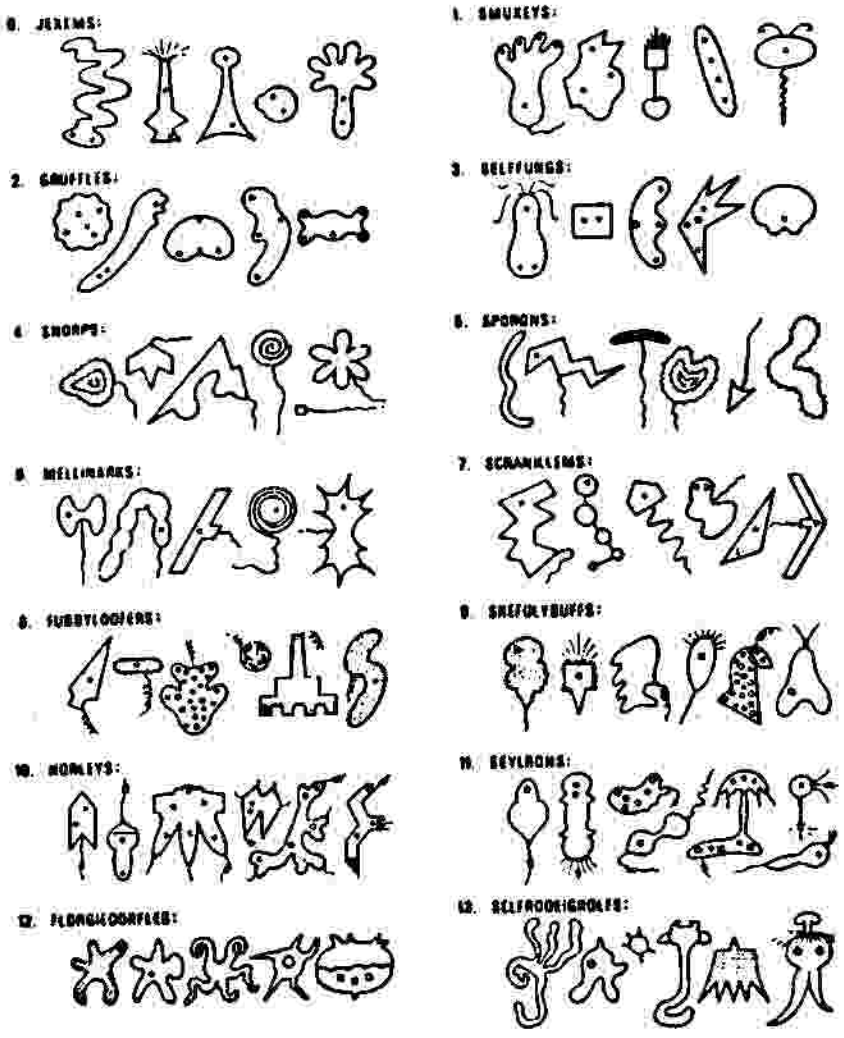


Figure 7. An event set showing fourteen species of "Animals."

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| $c_1$ | 1 | 1 | 1 | 1 |
| | 1 | 1 | 1 | 1 |
| | 1 | 1 | 1 | 1 |
| | 1 | 2 | 3 | 1 |
| $c_2$ | 2 | 2 | 2 | 2 |
| | 2 | 2 | 2 | 2 |
| | 2 | 2 | 2 | 2 |
| $c_3$ | 3 | 2 | 3 | 2 |
| | 3 | 2 | 3 | 2 |
| | 3 | 3 | 3 | 3 |

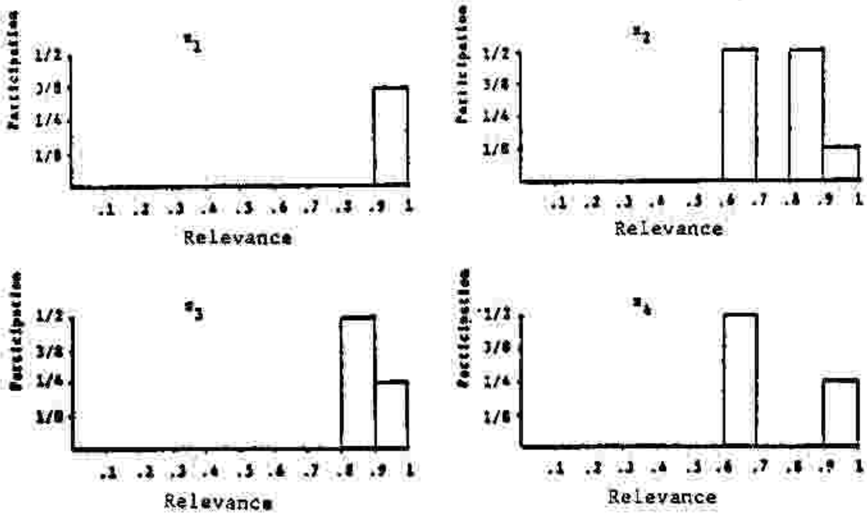
(a)

| Pair | $\rho$ |
|---|---|
| $x_1x_2$ | 1.00 |
| $x_1x_3$ | 1.00 |
| $x_1x_4$ | 1.00 |
| $x_3x_4$ | 1.00 |
| $x_2x_3$ | 0.87 |
| $x_2x_4$ | 0.67 |

(b)

| Attribute | Frequency of appearence of attributes from table above by relevance value. | | | | |
|---|---|---|---|---|---|
| | 0- <.6 | .6- <.7 | .7- <.8 | .8- <.9 | .9-1.0 |
| $x_1$ | 0 | 0 | 0 | 0 | 3/8 |
| $x_2$ | 0 | 1/2 | 0 | 1/2 | 1/8 |
| $x_3$ | 0 | 0 | 0 | 1/2 | 2/8 |
| $x_4$ | 0 | 1/2 | 0 | 0 | 2/8 |

(c)



(d)

Figure 8. Sample analysis of attribute groups for determining individual attribute relevance:

(a) Sample event set
(b) Value of $\rho$ for all distinct pairs of attributes
(c) Table of counts of the number of times each attribute appears in each subrange of promise, divided by the total number of attributes appearing in the range
(d) Histogram plots of the values from the table in (c)

attributes in the subrange and plotted as a histogram. An example of such an analysis is presented in Figure 8. A relevant attribute should exhibit greater participation in the high-valued sub-ranges and irrelevant attributes should exhibit greater participation in low-valued sub-ranges. When linear regression analysis is applied to these plots, relevant attribute "profiles" should have more positive slopes and lesser y-intercepts and irrelevant attribute profiles should have more negative slopes and greater y-intercepts. Such an analysis was performed for all combinations of three and four attributes as well. The histogram plots for all of the analyses are given in Appendix D. The resulting attribute rankings are shown in Figure 9. Actual values from the analyses are not given since values computed for different size groupings of attributes are not directly comparable. The table shows that the relevance value measured for each attribute is independent of the interactions between attributes in this event set because the rankings are fundamentally the same independent of the size of the groupings used. Therefore, the values indicated by PROMISE for individual attributes are independent of the interactions between attributes in this experiment, and can be used to order the attributes.

| Combinations | Best | | | | | | | | | | | | Worst |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Single Attributes | $x_1$ | $x_9$ | $x_6$ | $x_{10}$ | $x_3$ | $x_2$ | $x_4$ | $x_8$ | $x_{13}$ | $x_5$ | $x_{11}$ | $x_{12}$ | $x_7$ |
| Pairs | $x_1$ | $x_9$ | $x_6$ | $x_{10}$ | $x_4$ | $x_{13}$ | $x_3$ | $x_2$ | $x_8$ | $x_{11}$ | $x_5$ | $x_{12}$ | $x_7$ |
| Triples | $x_1$ | $x_9$ | $x_6$ | $x_{10}$ | $x_2$ | $x_3$ | $x_4$ | $x_8$ | $x_{13}$ | $x_{11}$ | $x_5$ | $x_{12}$ | $x_7$ |
| Quadruples | $x_1$ | $x_9$ | $x_6$ | $x_{10}$ | $x_4$ | $x_{13}$ | $x_2$ | $x_3$ | $x_8$ | $x_{11}$ | $x_5$ | $x_7$ | $x_{12}$ |

Figure 9. Attributes ranked by their relevance when evaluated interdependently in pairs, triples, and quadruples. A sample of such an analysis is given in Figure 7 and all of the analyses are given in Appendix D

### 3.2.2.2 Experiment II

The event set used for experiment II is given in Figure 10. It includes attributes with widely varying degrees of relevance. The values of one attribute were arranged so that the classes could be distinguished by the value of that attribute alone. Another attribute had the same value in all events. Two more attributes were pseudo-random, and the rest differentiated the classes to varying degrees. The data comprised five classes (3 with 8 events, 2 with 5 events). The result of the analysis of the attributes individually is shown in Figure 11. The ordering of the attributes by $\rho$-value (Figure 11) matches the order of attribute relevancy designed into the event set. The results show that PROMISE evaluates attributes based on their ability to distinguish classes in the event set. This indicates that a set of attributes which uniquely characterizes each class in an event set and contains few extraneous variables can be obtained by examining attributes beginning with the most relevant attributes and adding more attributes in order of decreasing relevance until the projection of the event set on the attributes has $\rho=1$. For example, when the most relevant attribute is excluded, the next three attributes can distinguish the classes uniquely (two of them comprise the minimum number that do so in this event set when the best attribute is excluded).

### 3.2.2.3 Experiment III

Experiment three was undertaken to determine the magnitude of the computational advantage, if any, realized by removing as many extraneous attributes as possible using PROMISE before the data is processed by AQ11.

The data for the third experiment were the same data used in experiment II but with the most-relevant attribute removed and a number of pseudo-random attributes added (42 in this experiment) to simulate the common occurrence of event sets which include no perfect attributes and many irrelevant ones. The attributes were processed by PROMISE and ordered by increasing relevance value. Because the three most relevant attributes are not distinguished by different $\rho$-values, they were all accepted initially and the projection of the data onto those three attributes was

|  | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $c_1$ | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 3 |
|  | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 3 | 4 |
|  | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 3 | 0 |
|  | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 1 |
|  | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 | 4 | 1 |
|  | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 4 | 2 |
|  | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 2 |
|  | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 | 3 | 0 |
| $c_2$ | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 5 | 3 |
|  | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 4 | 5 | 3 |
|  | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 4 | 5 | 2 |
|  | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 3 | 5 | 1 |
|  | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 3 | 5 | 3 |
|  | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 5 | 2 |
|  | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 4 | 5 | 4 |
|  | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 3 | 5 | 3 |
| $c_3$ | 2 | 1 | 0 | 0 | 0 | 3 | 0 | 2 | 0 | 2 |
|  | 2 | 1 | 0 | 0 | 0 | 3 | 0 | 3 | 2 | 0 |
|  | 2 | 1 | 0 | 0 | 0 | 3 | 0 | 3 | 0 | 0 |
|  | 2 | 1 | 0 | 0 | 0 | 3 | 0 | 1 | 0 | 0 |
|  | 2 | 1 | 0 | 0 | 0 | 3 | 0 | 1 | 1 | 3 |
|  | 2 | 1 | 0 | 0 | 0 | 3 | 0 | 1 | 1 | 1 |
|  | 2 | 1 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 4 |
|  | 2 | 1 | 0 | 0 | 0 | 3 | 0 | 2 | 2 | 3 |
| $c_4$ | 2 | 2 | 1 | 1 | 0 | 4 | 0 | 4 | 6 | 3 |
|  | 2 | 2 | 1 | 1 | 0 | 4 | 0 | 0 | 6 | 3 |
|  | 2 | 2 | 1 | 1 | 0 | 4 | 0 | 2 | 6 | 1 |
|  | 2 | 2 | 1 | 1 | 0 | 4 | 0 | 1 | 6 | 0 |
|  | 2 | 2 | 1 | 1 | 0 | 4 | 0 | 4 | 6 | 3 |
| $c_5$ | 2 | 2 | 2 | 0 | 1 | 5 | 0 | 3 | 3 | 4 |
|  | 2 | 2 | 2 | 0 | 1 | 5 | 0 | 2 | 1 | 0 |
|  | 2 | 2 | 2 | 0 | 1 | 5 | 0 | 0 | 4 | 3 |
|  | 2 | 2 | 2 | 0 | 1 | 5 | 0 | 2 | 1 | 1 |
|  | 2 | 2 | 2 | 0 | 1 | 5 | 0 | 1 | 0 | 3 |

**Figure 10. An event set containing attributes with a wide range of relevance used to test the performance of equation 11**

Best ——————— Worst

$x_6$ $x_1$ $x_2$ $x_3$ $x_4$ $x_5$ $x_8$ $x_9$ $x_7$

Figure 11. Attributes from the test event set ranked by individual relevance

evaluated by PROMISE and found to have $\rho=1$. Next, the projected event set was processed by the inductive-learning program AQ11 which derived rules to discriminate the classes. The rules derived from the projected event set were identical to those derived from the entire event set. Figure 12 shows a comparison of the time needed to derive the same set of rules using AQ11 on the entire event set versus running the program using only the three best attributes determined by PROMISE. In this instance, the computation time required to filter the data using PROMISE and derive rules using AQ11 was approximately one tenth the time required using AQ11 alone.

### 3.2.3 The Random Adaptive Search Algorithm

The relevance measure is intended for the analysis of single attributes independently of other attributes to find those that are most relevant. If such evaluation is not sufficient due to a high degree of interdependency between attributes, the Randomized Adaptive Search (RAS) algorithm may be used.

| | Using PROMISE | Not Using PROMISE |
|---|---|---|
| PROMISE | 0.319 CPU sec. | - |
| AQ11 | 0.227 CPU sec. | 4.576 CPU sec. |
| Total | 0.546 CPU sec. | 4.576 CPU sec. |

**Figure 12. A comparison of CPU time required to formulate identical rule sets using and not using PROMISE to process the data before processing by inductive learning program AQ11**

The algorithm iteratively evaluates groups of attributes and continuously updates an indicator of the relevancy of the individual attributes based on the performance of the groups. First, the relevance of each of the attributes is evaluated and stored as the initial *relevance index* (so-called because the values will no longer be based on $\rho$ as the algorithm progresses) for that attribute. A user-specified number of attributes is then chosen by weighted random selection based on these values. Attributes with better relevancy values are more likely to be chosen. The relevance of the group of attributes is then computed as a whole. Based on a comparison between the value for this group and a reference value (0.5 initially), the relevance index for each of the constituent attributes in the group is increased by a small, fixed amount (0.05 for this study) if the group performed better than the reference, or decreased by the same amount if the the group score was worse than the reference. The individual values are kept within the 0 to 1 range of $\rho$. The reference value is then set to the new group's relevance. The attributes are re-ordered by the new values of their relevance indices and a new group is chosen by weighted random selection. The process iterates until the reference value converges to a constant value or until a specified number of iterations have transpired.

The object of RAS is to find interdependent groupings of attributes and promote their selection as a group, even if one or more of them has a low value of $\rho$ when evaluated independently. As the algorithm progresses, groups of attributes which perform well are promoted and groups which perform poorly are suppressed. When the algorithm converges (if it does), a stable, high-relevance group of attributes has risen to the top of the list of indices and the poor attributes have fallen to the bottom.

### 3.2.4 The Greedy Attribute Selection Algorithm

A second method for applying the relevance measure to attributes is Greedy Attribute Selection (GAS). When attribute interdependency is not a problem, GAS may be an effective way to arrive at a quasi-minimal set of attributes in a less computationally intensive way than RAS.

The GAS algorithm begins with the independent evaluation of the relevance of each of the attributes. The list of attributes is rank-ordered by decreasing relevance. The first two (those two with the highest relevance) are chosen and evaluated as a group. If the relevance of the pair is greater than the relevance of the most relevant attribute alone, the second most-relevant attribute is deemed to have added useful information (as measured through the increased relevance) and is kept. If no improvement is noted, the second attribute is discarded since it added no information to that from the first attribute. The third most relevant attribute is then added and the group evaluated. The third attribute is also kept or discarded based on whether it contributes to an improvement in relevance score for the group. The scheme continues until a group of attributes is found which is perfectly discriminatory (*i.e.*, $\rho=1$) or the list of attributes is exhausted.

The Greedy Attribute Selection scheme is based on the assumption that the relevance measure is a good indicator of the value an attribute has for discrimination and its ability to evaluate groups of attributes in a way that yields results which can be compared meaningfully.

## 3.3 The Rule Generation Program GEM

### 3.3.1 The $A^q$ Algorithm for Rule Generation

The $A^q$ algorithm is a method for generating generalized descriptions which *cover* all of the positive events (i.e., those within the class to be described) and none of the negative events (i.e., those within the other classes) [Michalski,1978]. The process of developing a cover involves partially computing the complement of the set of negative events and selecting logical conjunctions of selectors, called *complexes*, which cover positive events. The final cover may be a single complex or a disjunction of complexes. The algorithm proceeds depth-first using the method of *disjoint stars*. A positive event, $e_1$, is chosen and an approximation of the set of all prime implicants which cover $e_1$ and are in the compliment of the set of all negative events is developed. This set is called a *star*. The best complex in the star, *lq*, is chosen using a lexicographic evaluation functional (see [Diettrich,1980]). The events covered by *lq* are removed from further consideration. The process is then repeated but each new event, $e_1$ is chosen so that it has not been covered by any element of any previous star. This ensures that disjoint, well-separated stars are built. The process is repeated until all events have been covered by at least one star.

### 3.3.2 Using GEM in an Expertise Development System

The $A^q$ algorithm is implemented in a program called GEM written in Pascal on a Vax 11/780. The $A^q$ algorithm (and, consequently, GEM) is an attractive choice for an inductive inference engine because of its flexibility. GEM is domain dependent only within the confines of the data structures used. Any problem domain which can be described by events that can be characterized using acceptable data structures can be processed using GEM. The data structures used required for input to GEM are patterned after relational tables. This structure has proven to be very flexible.

# 4 EXPERIMENTATION

## 4.1 Experiment 1: Black-Cutworm Damage Prediction

### 4.1.1 Problem Domain

Black Cutworms are insect larvae which damage between two and ten percent of the corn acreage in the Midwest annually. The name derives from the effect of cutworm action on corn plants: the severing of the stalk just above the soil line. In mid-April, Black Cutworm (BCW) moths are carried into Illinois by southerly winds and they land in the fields they find most attractive and lay their eggs. The growth cycle of the cutworm is short enough to allow three generations of worms to mature each growing season. Because more mature plants are more resistant to the ravages of the larvae, only the first generation of worms typically causes damage to field corn. Two major factors have been identified by corn entomologists in explaining damage mechanisms. The first is the attractiveness of a given field for the moths. A more attractive field will be the target of more moths' egg-laying. One of the most commonly postulated factors in field attractiveness is weediness at the time of moth flight. The second factor has been termed *synchrony* or the correspondence in time between corn maturation and cutworm maturation. Both corn and cutworm larvae mature at rates proportional to temperature. When the corn is young and the larvae are large, damage will be severe. When the corn is mature before the larvae mature, damage is slight. Also, if the larvae mature into pupae before the corn emerges from the soil, the damage will be slight [Boulanger, 1983]. Many factors may effect the rate of cutworm development and the size of cutworm populations. The difficulty of identifying the most important factors lies in the lack of sufficient quantities of high quality data due to both the rarity of cutworm damage and the lack of sufficient manpower for data collection. The system described in this paper was applied to a selected subset of the data to attempt to uncover some of the important factors.

### 4.1.2 Data

The data for this experiment consists of seventeen static attributes and two dynamic attributes as well as time and cutworm damage percentage figures for each of 210 events (given in Appendix D) recorded for the 1978 growing season. A sample event is shown in Figure 13. The breakdown of the data by damage percentage is shown in Figure 14. The zero number of fields damaged between 45 and 50% was thought to be a logical class division. CONVART used the original two dynamic attributes to construct 10 new attributes which characterize the behavior of the dynamic attributes over time. The attributes (both original and constructed) are given in Appendix A, with the range of possible values each may have. The original values for some of the attributes in some of the events were missing so values deemed plausible by an expert corn entomologist were inserted

```
Static Portion

   [fieldowner=smith][year=1978][previous crop=clover]&
   [bcw history=yes][adjacent water=no][surface slope=north]&
   [surface character=level][fall tillage=plow][spring till=disc]&
   [manure used=no][fertilizer regimen=none][insecticide=none]&
   [planting date=june 1]&


Dynamic Portion

   [date=mar 24]=> [weed species=horseweed][weed density=heavy]
   [date=mar 24]=> [weed species=smartweed][weed density=light]
   [date=apr 15]=> [weed species=smartweed][weed density=heavy]
   [date=may 1 ]=> [weed species=weedkill ][weed density=none ]

                                        ::> [damage= 75%]
```

Figure 13. A sample event for the black cutworm damage data showing static and dynamic attributes

where appropriate. If no plausible value was apparent, the value was left as "unknown". The final event set is missing approximately 20% of the data values and many of the events have limited time-dependent data. These factors combined to complicate the experiment considerably.

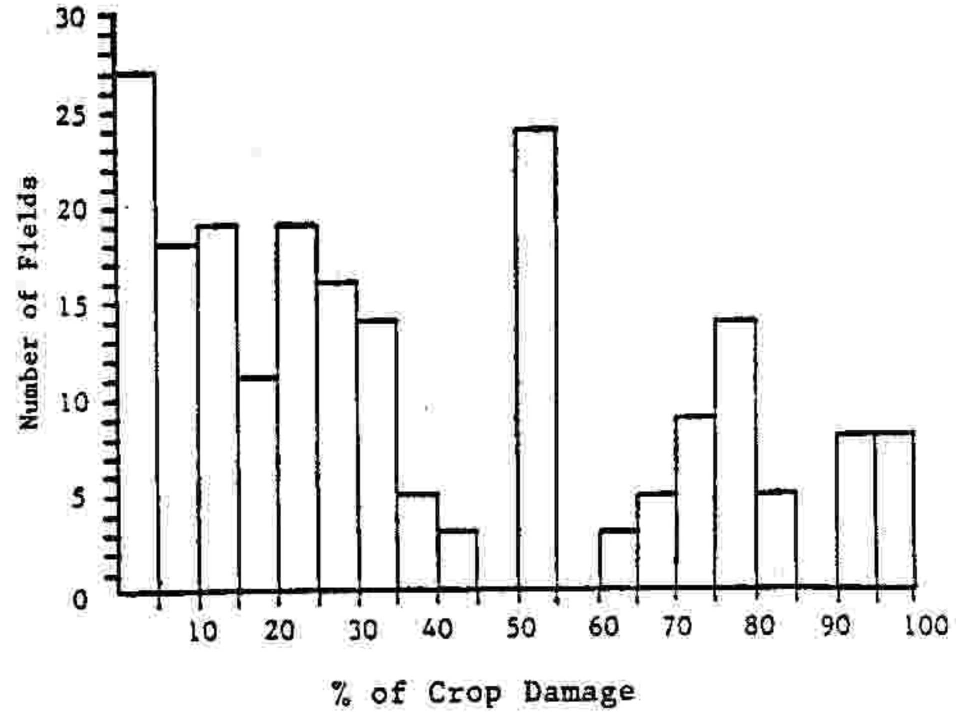


Figure 14. Histogram of the relative frequency of occurrence of different ranges of damage percentage in the 210 events used for this study

### 4.1.3 Results

Rules generated by applying GEM to this data are given in Appendix B. As the sparsity of the data could inject a significant amount of noise into the data, a corn entomologist was asked to identify those complexes which might be extraneous or unnecessarily convoluted. A comparison of the performance of the random adaptive search method and the greedy method is given in Figure 15. The output of the test runs is given in Appendix D. The lists of attributes were comparable with only minor differences among the least relevant attributes.

| Search Scheme | No. Attributes Per Sample | CPU Time | No. Iterations | No. Attributes Finally Chosen |
|---|---|---|---|---|
| Greedy | - | 34 sec | 1 | 11 |
| Random | 5 | 46 | 18 | 10 |
| " | 10 | 48 | 48 | 12 |
| " | 15 | 45 | 7 | 13 |

Figure 15. A comparison between different search schemes and varying search parameters for selecting a quasi-minimal set of attributes

### 4.1.4 Analysis

A comparison between the rules generated and the model outlined in 4.1.1 shows substantial agreement between the factors considered most important by corn entomologists and those identified by the processing. For example, factors such as planting date, several weediness measures, and tillage regimens reflect considerations of field attractiveness and synchrony presented earlier. Insecticide usage effects the survival of larvae during periods of low availability of food and during other periods of adverse conditions. The value of CONVART for this processing was considerable since it was required for the production of the attributes which capture the weed population trends that indicate both synchrony and field attractiveness.

Comparison between the performance of random adaptive search and greedy search show that, for this event set, the differences in results in terms of attribute set size is small and the penalty for processing using the random adaptive search scheme is large. The results of the random search shows that no highly synergistic combinations of attributes are contained within the event set since RAS (designed to find such interaction) provided the same results as greedy search.

## 4.2 Experiment 2: Craniostenosis Syndrome Identification

### 4.2.1 Problem Domain

The study and classification of medical syndromes is a well established field of medicine and medical diagnosis. The task of diagnosing these syndromes is increasingly difficult due to the growing number of rare genetic disorders and poorly understood patterns of malformation and malfunction. In addition, the problem is compounded by disease symptomes and hereditary traits which resemble the indications of many of the disorders of interest. A genetics clinic must be capable of reliably diagnosing approximately 8000 disorders on a day to day basis.

One of the fundamental problems in this field has been the rarity of most of the syndromes. With only a few (sometimes only one or two) examples of a given disorder reported, it is difficult to differentiate true indications of the syndrome from individual peculiarities of the patients. This problem has been approached in the past by the merging of the findings of several geographically separate observers of the same findings to recognize patterns in the data.

The class of syndromes studied in Experiment 2 is named *Craniostenosis*. The cranium of an infant is composed of bony plates joined at their edges by flexible joints or sutures. When these joints ossify (become bony), the skull assumes its final shape and size. The now-hard joints are known as synostes and occur normally during the course of growth and development. If a suture should harden before or after the others, the skull will develop abnormally, leading to asymmetry of the skull and face or abnormalities such as unusually limited cranial size. Due to the large number of observable anomalies of the face and skull and the sparseness of the available data, the methods described in this paper were applied in an attempt to pinpoint the most definitive anomalies for differentiating the syndromes in this class of disorders.

### 4.2.2 Data

The data consists of 231 observed cranio-facial anomalies for 80 patients diagnosed as having craniostenosis (the data may be found in Appendix D). Craniostenosis may be divided into four major syndromes with many patients undiagnosable due to the large overlap between the indications of the syndromes. Despite this overlap, doctors have characterized each syndrome according to certain individual tendencies [Spackman,1983]:

- *Crouzon's Syndrome* patients exhibit premature craniosynostosis, shallow orbits and frontal bossing, and maxillary hypoplasia with or without a parrot-like nose. About a quarter of the reported cases appear to be fresh mutations.
- *Saethre-Chotzen Syndrome* is highly variable in almost all of its features. Among the more common abnormalities are: synostes of the coronal sutures, low-set hairline, facial asymmetries, shallow orbits, ptosis (drooping eyelid), small ears, and partial webbing of two or more fingers or toes.
- *Apert's Syndrome* includes both craniosynostosis and severe syndactly (fusion of two or more fingers).
- *Pfeiffer's Syndrome* also involves craniosynostosis and syndactly, although it is milder in almost every similar feature.

Among the 81 patients, many were not diagnosed as to individual syndrome. When a single "normal" patient is added as a sixth control class, the data may be broken down as shown in Figure 16. A sample event is shown in Figure 17. A list of the 231 anomalies is given in Appendix C.

| Syndrome | # Patients |
|---|---|
| Apert | 16 |
| Crouzon | 24 |
| Saethre-Chotzen | 8 |
| Pfeiffer | 1 |
| Undiagnosed | 31 |

Figure 16. Breakdown of data for craniostenosis patients by syndrome

### 4.2.3 Results

The attributes were all binary (i.e., syndrome present=1, not present=0), VARSEL identified a set of 20 attributes, among the original 231 in the event set, which successfully characterized each of the syndromes without characterizing any of the patients of indeterminate diagnosis or the normal patient. The chosen attributes are given in Figure 18. GEM then analyzed the new event set and used 16 of the twenty attributes to produce the rules shown in Figure 19. A physician who had been attempting to perform the same task manually, had invested approximately 4 man-months in the job. He also discovered twenty attributes but had only succeeded in characterizing about 75% of the patients successfully. These attributes are also given in Figure 18. The rules he derived using GEM and these attributes are given in Appendix D. The total real-time required for the computer processing was less than an hour.

If we compare the generated rules in Figure 19, with the clinical profiles presented earlier, we may note close correspondence between the rules for each class and the clinical findings. An important consideration is that the clinical profiles presented have a great deal of overlap so they are not a perfect "benchmark" for deciding rule validity. In addition, the data collected by different clinicians often varies as to quality and scope.

```
[patient number=000][date=1/1/81][flat forehead=present][syndactly=present]&
                     [craniosynostosis=absent] ... 228 other anomalies

                                        ::> [syndrome=Pfeiffer]
```

Figure 17. A sample event from the craniostenosis event set

| Chosen by VARSEL | Chosen by Expert |
|---|---|
| Craniosynostosis | Craniosynostosis-General |
| Craniosynostosis-Coronal | Ear Malformations |
| Craniosynostosis-Saggital | Impaired Hearing |
| Facial Assymetry | Facial Assymetry |
| Flat Forehead | Flat Forehead |
| Assymetric Forehead | Beaked Nose |
| Ptosis-Eye | Ptosis |
| Shallow Orbit | Hypertelorism |
| Exophthalmos/Proptosis-L Eye | Proptosis |
| Exophthalmos/Proptosis-R Eye | Hallux Valgus |
| Byzantine Palate | Byzantine Palate |
| Midface Hypoplasia | Maxillary Hypoplasia |
| Cutaneous Syndactly-Hand | Syndactly of Fingers |
| Syndactly-Foot | Syndactly of Toes |
| Partial Syndactly-Foot | Webbing of Toes |
| Hypertonia | Webbing of Fingers |
| Plagiocephaly | Strabismus |
| Pyloric Stenosis | Cleft Palate |
| Position Anomolies-Digits | Tear Duct Stenosis |
| Undescended Testes | Proptosis |

Figure 18. A comparative listing of the attributes chosen as most relevant by VARSEL and by a human expert for the discrimination of craniostenosis syndromes from the data used in this study

```
                                                                          (NEW,IND,COV)
R1: [syndactly_foot=absent][craniosynostosis=present]&                      (18,16,18)
    [exophthalmos_eye=present][flat_forehead=absent]
                  v                                                         (1,1,3)
    [syndactly_foot=absent][byzantine_palate=present][ptosis_eye=absent]
                  v                                                         (1,1,1)
    [syndactly_foot=absent][exophthalmos_eye=present][plagiocephaly=present]
                  v
    [craniosynostosis=absent][craniosynostosis_coronal=absent]&             (2,2,2)
    [midface_hypoplasia=present][cutaneous_syndactly_hand=absent]
                  v
    [shallow_orbit=absent][craniosynostosis_coronal=absent]&                (1,1,1)
    [midface_hypoplasia=absent][flat_forehead=present][plagiocephaly=absent]

              ::> [syndrome=Crouzon]

                                                                            (5,3,5)
R2: [cutaneous_syndactly_hand=present]
                  v
    [midface_hypoplasia=present][flat_forehead=present]&                    (3,3,5)
    [byzantine_palate=absent]

              ::> [syndrome=Sathre-Chotzen]

                                                                            (16,16,16)
R3: [syndactly_foot=present]

              ::> [syndrome=Apert]

                                                                            (1,1,1)
R4: [pyloric_stenosis=present]

              ::> [syndrome=Pfeiffer]
```

Figure 19. A set of discriminant rules for the craniostenosis event set (the numbers in parentheses are the number of events first covered by this complex, the number of events only covered by this complex, and the total number of events covered by this complex)

Random adaptive search failed to converge for this event set, probably due to the low likelyhood of synergistic interaction since the minimum number of binary attributes required to differentiate six classes would be three and 16 were actually required. The system has clearly operated cost-effectively on this event set, producing plausible rules within the limitations of the event set in a computationally attractive amount of time. The resulting data reduction from the use of VARSEL as a preprocessor was over 91% since only 20 of the original 231 attributes were needed.

# 5 CONCLUSIONS

A model has been presented which describes the process of domain expertise development in terms of the sequential application of four sub-processes with multiple feedback paths. The subprocesses include attribute construction (implemented in CONVART), attribute selection (implemented in SELECT), rule formation (implemented in GEM), and rule implementation (left for the system builder).

A system has been constructed which is capable of applying the model for the purpose of deriving expert decision rules from data using a minimum of explicitly stated domain knowledge and minimal iterative processing. The system contains modules for: the construction of attributes which describe time-dependent behavior of other attributes so that time-varying processes can be analyzed, the selection of most relevant attributes for class discrimination using a recently developed measure of attribute relevancy, and inductive rule inference based on established methods of automated inductive inference. The rule implementation process in which a set of rules is embedded in a program which can apply the rules to new events, remains as a task for the programmer.

## 5.1 System Performance

The constructive induction program CONVART was vitally important in the processing of Black-Cutworm damage data because cutworm damage is closely tied to time-dependent processes. It was of not used for Craniostenosis data since the attributes present were all static. The variable selection program VARSEL was useful in both cases but different search strategies were of different utility. Random adaptive search showed explicitely the lack of synergistic interaction between attributes in the BCW data since the attribute sets chosen by RAS were the same as those chosen by greedy search (these results are given in Appendix D). RAS failed to produce useful results of any sort for the craniostenosis data because of the poor discriminatory value of binary attributes for multi-class event sets. Greedy search was effective in both cases since significant data reduction was achieved in both cases and appropriate experts favorably evaluated the resulting attribute sets. The inductive inference engine GEM produced good rules from the pre-processed event sets which were thought to be reasonable and consistent with the input data when examined by experts from within the problem domains.

## 5.2 Indicated Future Directions of Inquiry

The potential for significant future inquiry exists throughout the processing chain used in this system. Specific avenues of study for the concepts embodied in CONVART and GEM are described in detail elsewhere [Michalski,1982b] [Davis,1981]. Future directions for applying and studying the measure of relevancy presented earlier include exploring its use as an indicator of potentially relevant groups of attributes for constructive induction purposes, and further refinement of the measure to enhance its resolution. In addition, new implementations of the greedy and random adaptive search schemes may overcome the limitations of the current implementations, and new search schemes may prove more effective than either.

# APPENDICES

## Appendix A: Attributes in BCW Damage Event-Set

The following is a list of the attributes, both original and constructed, used for the analysis of black cutworm damage in the state of Illinois.

* means this attribute was chosen by RAS.

** means this attribute was chosen by GAS.

*** means this attribute was chosen by both RAS and GAS.

| | | Attribute | Possible Values (domain) |
|---|---|---|---|
| Initial Attributes | Static Attributes | | |
| | 1. | Damage | =<50%, >50% |
| | 2. | Field Owner | too numerous to list |
| | 3. | Growing Year | 1978 |
| | ***4. | Previous Crop | Corn, Beans, Weeds, Sorghum |
| | 5. | BCW History | Yes, No |
| | ***6. | Permanent Border Vegetation | Yes, No |
| | ***7. | Permanent Border Water | Yes, No |
| | ***8. | Surface Direction | North, South, East, West, None |
| | 9. | Surface Character | Level, Rolling, Bottomland |
| | ***10. | Fall Tillage | None, Plow, Chisel, Disc |
| | ***11. | Spring Tillage | None, Plow, Chisel, Disc |
| | 12. | Manure Usage | Yes, No |
| | 13. | Fertilizer Usage | Yes, No |
| | ***14. | Insecticide | None, Yes (nonspecific), Preventative, Rescue Treatment |
| | *15. | Planting Date | Jan1-Apr10, Apr11-Apr20, Apr21-May15, May16-Jun9, After Jun9 |
| | 16. | Planting Rate | <100%, 100% |
| | Dynamic Attributes | | |
| | 17. | Weed Species | None, Weeds (nonspecific), Very Few Weeds, Onion, Grass, Legume, Winter Annual, Other Broadleaf, Weedkill |
| | 18. | Weed Density | None, Light, Heavy, Heavy Patches |
| Constructed Attributes | Static Attributes | | |
| | *19. | Most Common Weed Species | None, Weeds (nonspecific), Very Few Weeds, Onion, Grass, Legume, Winter Annual, Other Broadleaf, Weedkill |
| | 20. | Least Common Weed Species | None, Weeds (nonspecific), Very Few Weeds, Onion, Grass, Legume, Winter Annual, Other Broadleaf, Weedkill |
| | 21. | Number Of Observations Of Same Weed Species | 1, 2 |
| | ***22. | Number Of Different Weed Species | 0, 1, 2, 3, >3 |
| | *23. | Average Density | None, Light, Heavy, Heavy Patches |
| | *24. | Intercept of Density vs Time | |
| | *25. | Slope Of Density vs Time | -.057, -.047, -.044, -.040, -.036, -.034, -.032, -.030, -.029, -.026, -.024, -.020, -.018, -.015, >0 |
| | ***26. | Maximum Weed Density | None, Light, Heavy, Heavy Patches |
| | 27. | 1st Time Of Max. | Before Mar24, Mar24-Apr13 |
| | *28. | Minimum Weed Density | None, Light, Heavy, Heavy Patches |
| | ***29. | 1st Time of Min. | Apr5-Apr18, Apr19-22, Apr23-Apr27, Apr28-Apr30, May1-May2, May3-May7, May8-May12, May13-May19, May20-May21, May22-May26, May27-Jun3 |

## Appendix B: BCW Damage Estimation Rules

The following two pages give rules for predicting black cutworm damage severity derived by GEM from 1978 BCW data after new attributes were constructed by CONVART and selection was performed by VARSEL. Each rule is quite complicated in that each has many complexes. The reason for this becomes clear when one examines the parameters on the right-hand side of the page:

- NEW - The first number is the number of events covered by this complex which were not covered by previous complexes in the list.
- IND - The second is the number of events covered by this complex alone of all complexes.
- COV - The third is the total number of events covered by this complex.

Examination of the numbers shows that the rules reveal a subtle and complex interaction between three factors: weediness, synchrony, and pesticide use. The interaction is suggested by the coverage numbers. Note that each complex covers several events but covers very few uniquely. Therefore, the complexes have significant overlap in coverage but each one accounts for a slight variation on one or more of these themes.

```
                                                                                    (NEW,IND,COV)

21:  [previous crop=corn,weeds,sorghum][permanent vegetation=no]&
     [adjacent water=yes][terrain=level,rolling]&
     [fall tillage=none,chisel,disc][planting date=after may15]                      (5,3,5)
                                    v
     [previous crop=soybeans,weeds,sorghum][fall tillage=plow,disc]&
     [spring tillage=none,plow,disc]&
     [number of diff. weed species<=3][maximum weedensity=heavy..heavy patches]      (7,2,7)
                                    v
     [spring tillage=none,disc][planting date=after may15]&
     [number of diff. weed species=3][maximum weedensity not=heavy patches]&
     [1st minimum weedensity=after may2]                                             (9,2,10)
                                    v
     [permanent vegetation=yes][terrain=level][fall tillage=none,disc]&
     [planting date=jan1..may15][maximum weedensity=heavy patches]                   (2,1,2)
                                    v
     [terrain=level,bottomland][fall tillage=plow][spring tillage=none,plow,disc]&
     [planting date=after may15][number of diff. weed species=3]                     (5,3,9)
                                    v
     [adjacent water=no][fall tillage=none,chisel][spring tillage=none,chisel]&
     [number of diff. weed species=3..>3][maximum weedensity=heavy]                  (3,1,3)
                                    v
     [previous crop=corn,weeds,sorghum][permanent vegetation=yes]&
     [fall tillage=chisel][planting date= after may15]                               (3,3,3)
                                    v
     [permanent vegetation=no][1st minimum weedensity=aftjun3]                       (1,1,4)
                                    v
     [adjacent water=yes][maximum weedensity=heavy patches]&
     [1st minimum weedensity=may19..may21]                                           (1,0,1)
                                    v
     [previous crop=corn,weeds,sorghum][terrain=level][fall tillage=disc]&
     [number of diff. weed species=<3]                                               (1,1,2)
                                    v
     [permanent vegetation=yes][fall tillage=chisel,disc][insecticide=no]&
     [number of diff. weed species>2][maximum weedensity not= heavy patches]         (2,1,6)
                                    v
     [permanent vegetation=yes][terrain=level,bottomland]&
     [fall tillage=chisel,disc][insecticide=no][maximum weedensity=heavy]            (2,2,4)
                                    v
     [permanent vegetation=yes][adjacent water=no][fall tillage=chisel]&
     [spring tillage=none,plow,disc][planting date=after apr30]&
     [maximum weedensity=no..lt][1st minimum weedensity=before may2]                 (1,1,3)
                                    v
     [permanent vegetation=yes][terrain=level][fall tillage=none,chisel,disc]&
     [spring tillage=none,disc][insecticide=preventive][maximum weedensity=heavy]    (1,1,1)
                                    v
     [permanent vegetation=no][terrain=level,bottomland]&
     [fall tillage=none,plow,disc]&
     [spring tillage=none,disc][planting date=after may15]&
     [number of diff. weed species=3]                                                (3,1,8)
                                    v
     [previous crop=corn,weeds,sorghum][terrain=level,rolling]&
     [fall tillage=none,chisel]&
     [spring tillage=none,chisel,disc][planting date=after may15]&
     [maximum weedensity=no..lt][1st minimum weedensity=before apr30]                (2,2,3)
                                    v

     [permanent vegetation=no][fall tillage=chisel]&
     [spring tillage=none,plow,chisel][maximum weedensity=heavy..heavy patches]      (1,1,3)
                                    v
     [permanent vegetation=yes][fall tillage=none,disc]&
     [insecticide=no,rescue treatment][number of diff. weed species=3]               (1,1,5)
                                    v
     [permanent vegetation=no][adjacent water=no][terrain=level,bottomland]&
     [spring tillage=none,disc][planting date=after may15]&
     [number of diff. weed species=3]                                                (1,1,7)

                              ::> (damage >= 50%)
```

```
                                                                                  (NEW,IND,COV)
R2:  [previous crop=soybeans,weeds,sorghum][fall tillage=none,plow]&
     [spring tillage=none,chisel,disc][insecticide=preventive,rescue treatment]&
     [number of diff. weed species=<3]                                              (12,2,12)
                                v
     [terrain=rolling][fall tillage=plow,chisel]&
     [maximum weedensity not= heavy patches][1st minimum weedensity=before apr30]  (10,3,11)
                                v
     [previous crop=soybeans,sorghum][adjacent water=yes]&
     [terrain=level,rolling][insecticide=preventive,rescue treatment]              (13,4,16)
                                v
     [permanent vegetation=yes][fall tillage=none,plow,disc]&
     [spring tillage=none,disc]&
     [number of diff. weed species=>3][1st minimum weedensity=before may21]        (4,1,6)
                                v
     [permanent vegetation=yes][fall tillage=plow,chisel,disc]&
     [spring tillage=none,chisel][insecticide=preventive,rescue treatment]         (6,2,7)
                                v
     [adjacent water=yes][terrain=level,bottomland][fall tillage=none,plow]&
     [spring tillage=none,plow,disc][number of diff. weed species<3]               (7,3,7)
                                v
     [terrain=level,bottomland][planting date=before apr30]                        (9,4,18)
                                v
     [previous crop=corn,soybeans,sorghum][permanent vegetation=yes]&
     [insecticide=preventive,rescue treatment][maximum weedensity=heavy patches]   (4,1,10)
                                v
     [permanent vegetation=yes][terrain=level,rolling][fall tillage=none,chisel]&
     [insecticide=no,rescue treatment][number of diff. weed species<3]&
     [maximum weedensity=no..lt]                                                   (4,1,4)
                                v
     [previous crop=soybeans,sorghum][fall tillage=plow,disc]&
     [insecticide=preventive,rescue treatment][planting date=after may15]          (4,3,8)
                                v
     [permanent vegetation=yes][fall tillage=none,chisel,disc]&
     [insecticide=no,rescue treatment][number of diff. weed species<3]&
     [maximum weedensity=heavy]                                                    (1,1,1)
                                v
     [previous crop=soybeans,sorghum][permanent vegetation=no]&
     [terrain=level,rolling][fall tillage=chisel,disc]&
     [insecticide=preventive,rescue treatment][number of diff. weed species>2]&
     [1st minimum weedensity=before may26]                                         (2,1,6)
                                v
     [permanent vegetation=yes][fall tillage=none,chisel]&
     [insecticide=preventive,rescue treatment][number of diff. weed species<4]&
     [maximum weedensity=heavy]                                                    (1,1,6)
                                v
     [permanent vegetation=no][terrain=level,bottomland]&
     [spring tillage=none,plow,chisel][maximum weedensity=no..lt]                  (5,0,15)
                                v
     [permanent vegetation=yes][terrain=level,bottomland][fall tillage=none]&
     [maximum weedensity=no..heavy]                                                (2,2,12)
                                v
     [permanent vegetation=no][adjacent water=no][terrain=rolling]                 (2,1,11)
                                v
     [adjacent water=no][fall tillage=disc]&
     [insecticide=preventive,rescue treatment][number of diff. weed species>2]     (1,1,7)
                                v
     [permanent vegetation=no][adjacent water=yes][terrain=bottomland]&
     [fall tillage=plow,chisel]                                                    (2,2,4)
                                v
     [permanent vegetation=yes][number of diff. weed species=<3]&
     [1st minimum weedensity=may2..may21]                                          (2,2,4)
                                v
     [previous crop=soybeans,sorghum][permanent vegetation=yes]&
     [terrain=level,bottomland][insecticide=preventive,rescue treatment]&
     [number of diff. weed species>2]                                              (1,1,5)
                                v
     [previous crop=soybeans,weeds,sorghum]&
     [insecticide=preventive,rescue treatment]&
     [number of diff. weed species=>3][maximum weedensity=heavy patches][1,3]      (1,1,3)
                                v
     [previous crop=corn,weeds,sorghum][planting data=apr30]                       (1,1,12)
                                v
     [terrain=level][spring tillage=none,chisel][maximum weedensity=no..lt]        (1,1,16)
                                v
     [permanent vegetation=no][adjacent water=no][fall tillage=disc]               (1,1,10)

                            ::> [damage < 50%]
```

## Appendix C: Attributes for Craniostenosis Event Set

The following list gives the anomalies present in the original craniostenosis data-set grouped by body system effected. See Figure 18 for lists of which attributes were chosen by VARSEL and which were chosen by a human expert.

```
Palate
   Incomplete Median Cleft
   Median Cleft
   Submucous Cleft (3 varieties)

Skull
   General
   Assymetric
   Macrocephaly
   Craniosynostosis
   Craniosynostosis-Coronal
   Craniosynostosis-Sagittal
   Faulty Sutures
   Shape Anomalies
   Brachycephaly
   Cloverleaf Skull
   Plagiocephaly
   Trigonocephaly
   Prominent Coronal Suture

Forehead
   General
   Assymetric
   Large
   Typical Aperts
   Bossing
   Prominent/Bulging
   Elongated
   Flat
   Midline Defect
   Typical Crouzons
   Supraorbital Ridge Anomalies
```

Midface

- General
- Facial Assymetry
- Midface Hypoplasia
- Hypertelorism
- Hypotelorism
- Assymetric Orbital Placement (Right Higher)
- Hypoplasia (Left Side)

Jaw

- General
- Assymetric
- Micrognathia/Hypoplasia
- Antegonial Notching
- Prognathia
- Wide Gonial Angle

Left Eye

- General
- Anophthalmia
- Microphthalmia
- Exophthalmos/Proptosis
- Prominent/Protruding
- Setting Sun Sign
- Devil's Eye
- Shallow Orbit
- Small Orbit
- Malpositioned Orbit
- Blepharitis
- Ptosis
- Antimongoloid Slant
- Epicanthal Fold
- Synoche of the Lids
- Esotropia
- Strabismus
- Nystagmus

Right Eye

- General
- Anophthalmia
- Microphthalmia
- Exophthalmos/Proptosis
- Prominent/Protruding
- Orbit Anomalies
- Devil's Eye
- Shallow Orbit
- Small Orbit
- Blepharitis
- Ptosis
- Lids Fail To Close
- Antimongoloid Slant
- Epicanthal Fold
- Syneche of the Lids
- Esotropia
- Strabismus
- Nystagmus

Left Ear
- Small
- Preauricular Pit/Sinus
- EAC Atresia
- Low Set
- Posteriorly Set
- Ossicular Anomalies
- Cupped
- Lopped/Protruding

Right Ear
- Small
- Preauricular Tag
- Rotated
- Low Set
- Cupped
- Lopped/Protruding

Nose
- General
- Assymetric
- Bifid
- Narrow
- Broad/Bulbous
- Alae Anomalies
- Cleft Nostrils/Alae
- Pinched Nares
- Beaked
- Saddle Shaped/No Bridge/Flat Bridge
- Choanal Atresia/Stenosis
- Deviated Bridge/Nose
- Deviated Septum

Oral Cavity
- High Arched Palate
- Byzantine Palate
- Torus Paltinus
- Narrow Palate
- Maxillary Assymetry
- Commissural Lip Pits
- Macrostomia

Tongue
- Short Frenulum/Tongue Tie
- Dry

Dentition
- Open Bite
- Crossbite
- Dental Crowding
- Missing Teeth
- Typical Aperts

Mimetic Musculature
- Motor Problem
- Paresis

Muscles of Mastication
   Motor Problem
   Paresis

Neck
   General
   Torticollis
   Short

Abdominal Wall
   Hernia, Unspecified
   Umbilical Hernia
   Inguinal Hernia

Chest Wall
   General
   Asymmetry
   Prominent
   Pectus Excavatum

Back
   General
   Scoliosis
   Kyphosis/Kyphoscoliosis

Respiratory System
   Chronic URI or Other Respiratory Disease

Cardiovascular System
   Cardiac Anomalies
   Valve Anomalies
   Rythm Anomalies
   Aortic Arch Anomalies

GI System
   Liver,Spleen
   Pyloric Stenosis

Genital
   General
   Hypospadias
   Undescended Testes
   Scrotal Anomalies
   Uterine Anomalies
   Kidney Anomalies

Skin and Adnexia
   Alopecia/Bald
   Thin/Sparse Hair
   Echzema

Left Arm
   General

Upper Arm Anomalies
Joint
Dislocation/Subluxation
Contractures

Right Arm
General
Upper Arm Anomalies
Joint
Contractures

Left Hand
General
Arachnodactly
Contractures
Syndactly
Complete Syndactly
Cutaneous Syndactly
Dermatoglyphics
Simian Crease
Phalangeal Anomalies
Position Anomalies/Digits

Right Hand
General
Arachnodactly
Contractures
Syndactly
Complete Syndactly
Cutaneous Syndactly
Simian Crease
Position Anomalies

Left Leg
Contractures
Joint
Knee

Right Leg
Contractures
Joint
Knee
Overlapping Toes

Left Foot
Overlapping Toes
Bifid Toe
Broad/Large Digits
Syndactly
Partial Syndactly

- Cutaneous Syndactly
- Clubbing
- Abnormal Position
- Cleft/Seperation, Toes
- Abnormal Postion/Foot

Right Foot

- Broad/Large Toes
- Syndactly
- Partial Syndactly
- Cutaneous Syndactly
- Clubbing
- Abnormal Position
- Overlapping Toes
- Cleft/Seperation, Toes
- Abnormal Postition/Foot

Nervous System

- General
- Facial Nerves
- Retardation
- Developmental Retardation
- Epilepsy/Seizure Disorder
- EEG Abnormal
- Hyperactive
- Reflexes Hyperactive
- Cerebral Anomalies
- Corpus Collosum Anomalies
- Aplusia/Speech Problems

Skeletal (Primary Axial)

- General
- Cervical
- Cervical Fusions
- C-1 Anomalies
- Occipitalization, C-1
- Basilar Invagination
- Scoliosis
- Hip Anomalies
- Hip Dislocations
- Joint Anomalies, Generalized
- Spina Abifida

Muscular System

- General
- Hypotomia
- Hypertonia
- Contractures