STATISTICS IN MEDICINE
*Statist. Med.* **18**, 2723–2735 (1999)

# CLASSIFICATION TO ORDINAL CATEGORIES USING A SEARCH PARTITION METHODOLOGY WITH AN APPLICATION IN DIABETES SCREENING

ROGER J. MARSHALL*

*Department of Community Health, University of Auckland, Private Bag 92019, Auckland, New Zealand*

## SUMMARY

A method is proposed for classification to ordinal categories by applying the search partition analysis (SPAN) approach. It is suggested that SPAN be repeatedly applied to binary outcomes formed by collapsing adjacent categories of the ordinal scale. By a simple device, whereby successive binary partitions are constrained to be nested, a partition for classification to the ordinal states is obtained. The approach is applied to ordinal categories of glucose tolerance to discriminate between diabetes, impaired glucose tolerance and normal states. The results are compared with analysis by ordinal logistic regression and by classification trees. Copyright © 1999 John Wiley & Sons, Ltd.

## 1. INTRODUCTION

Classifiers to nominal categorical states can be constructed using various methodologies; the two main competitors are regression methods, for example, binary or polytomous logistic modelling,[1,2] and tree-based partitioning methods. CART,[3] *S* procedures[4] and SPSS's AnswerTree are examples of the latter. When categories are ordinal, regression models can be formulated[5] and tree-based methods can be adapted by using a splitting criterion that accounts for the ordinal structure of the outcome. In CART 'twoing' can be used and AnswerTree has a criterion based on a scalable chi-square measure.[6]

An alternative approach that is suggested here for classification to ordinal states is based on a non-hierarchical partitioning procedure, known as search partition analysis (SPAN). SPAN has been proposed to generate binary classifiers[7] and for identification of subgroups.[8] In this paper I explore its application for the development of a classifier to ordinal categories by combining binary classifiers.

The method is used to develop a screening rule for diabetes. The results of its application are compared to a rule based on ordinal logistic regression and a rule obtained by tree-structured analysis.

* Correspondence to: Roger J. Marshall, Department of Community Health, University of Auckland, Private Bag 92019, Auckland, New Zealand. E-mail: rj.marshall@auckland.ac.nz

CCC 0277–6715/99/202723–13$17.50

*Received March 1998*
*Accepted January 1999*

## 2. EXAMPLE: DIABETES SCREENING

A fasting oral glucose test is usually used to diagnose diabetes and, depending on the results of the test, individuals are classified as on the ordinal scale *normal*, *impaired glucose tolerance* (*IGT*) and *diabetes*. These three categories are determined by internationally accepted cutpoints of plasma glucose: 7·8 mmol/l and 11 mmol/l. Because the test requires 10–16 hour fasting, and a blood test after a two hour glucose load, its routine use is impractical as a screening instrument.

Using data from a work force survey of about 5500 people in New Zealand, it was hoped that a simpler test could be developed. All participants in the survey were asked to take a fasting glucose tolerance test and in addition other risk factors for diabetes were measured. These included a non-fasting fructosamine test,[9] which had been suggested as an alternative test for diabetes, as well as body mass index, blood lipids, hypertension and race, among others. The objective of the analysis here is to use these measures to provide a simple rule for classification to each of the three ordinal states *normal*, *impaired glucose tolerance* and diabetes.

Without, for the moment, going into details, a SPAN derived rule that was developed for discrimination between *diabetes* and the combined *normal* and *impaired* states was:

assign to *diabetes* if: positive fructosamine and albumin tests

or if: positive fructosamine and triglyceride tests and Polynesian ethnicity.

If $F$, $T$ and $U$ denote positive fructosamine, triglyceride and urinary albumin tests, respectively, and $E$ is Polynesian (Maori or Pacific Island) ethnic group, then the rule can be written in symbols as

$$(F \cap U) \cup (F \cap E \cap T) \tag{1}$$

where $\cap$ stands for 'and' and $\cup$ stands for 'or'.

A SPAN derived rule was also obtained to discriminate between *normal* and combined *impaired glucose tolerance* and *diabetes* categories. In symbols the rule was assign to *diabetes* or *impaired* if

$$(F \cap U) \cup (F \cap T) \cup (F \cap H) \tag{2}$$

where $H$ denotes hypertension. These two rules are sufficient for three-way classification; anyone who does *not* satisfy rule (2) would be assigned to *normal* and someone who satisfies (2) but not (1), would be classified to *impaired*. Comparing rules (1) and (2) it is clear that anyone satisfying rule (1) necessarily also satisfies rule (2), that is, rule (1) is a subset of rule (2). Were this not the case, the rules would be inconsistent, as discussed below.

In the following two sections the methods used to derive rules such as these are described and I return to the example, with more details of the analysis, in Section 5.

## 3. USING SPAN FOR ORDINAL OUTCOMES

In SPAN a search is carried out among different Boolean combinations of predictive factors to find a binary partition, $P = \{\bar{A}, A\}$, of the sample space. Each combination is tested against the data to establish how well the partition induces a split of the data into two groups, such that each are relatively homogeneous with respect to an outcome variable $y$. If $y$ is nominal, homogeneity can be measured by, for example, an entropy measure or Gini index of diversity. These measures can also be applied if $y$ is ordinal, although the ordinal nature of $y$ would be not be taken into account.

 

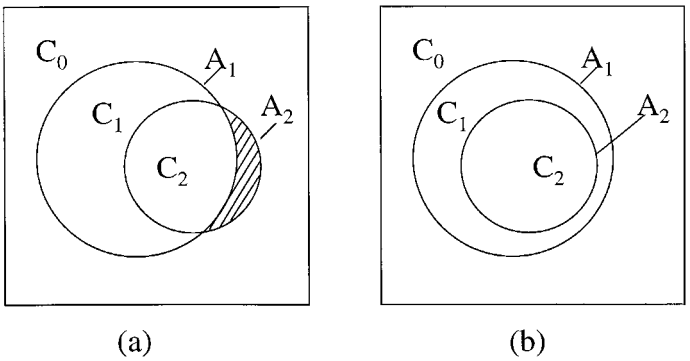


Figure 1. Binary partitions $\{\bar{A}_1, A_1\}$ and $\{\bar{A}_2, A_2\}$. (*a*) in which $A_2$ is not a subset of $A_1$; (*b*) in which $A_2$ is a subset of $A_1$ enabling partition into $\{C_0, C_1, C_2\}$

Further, whether $y$ is nominal or ordinal, if it has more than two states, the binary partition will obviously not constitute a classifier for $y$. However, categories of $y$ can be collapsed to create a binary outcome and a binary partition found to classify to $y$. If there are $t+1$ categories of $y$ there are $2^t - 1$ possible ways to combine categories, but if $y$ is ordinal it is natural to only consider combining adjacent categories in the ordinal pathway so that the binary outcome becomes an indicator of 'above' and 'below' a certain value of $y$. Then obviously there are only $t$ possible ways of combining $t+1$ categories.

Consider, for example, the case where $y$ has three ordinal states 0, 1 and 2. Either states 0 and 1 can be combined, or states 1 and 2. Suppose states 1 and 2 are combined and a binary partition, say $P_1 = \{\bar{A}_1, A_1\}$ is found, where $A_1$ is predictive of 'either 1 or 2' and $\bar{A}_1$ predicts category 0. On the other hand, states 0 and 1 could be collapsed and a partition, $P_2 = \{\bar{A}_2, A_2\}$, found where $A_2$ is predictive of state 2.

To form a three-way partition $\{C_0, C_1, C_2\}$ as a classifier to states 0, 1 and 2, it would be natural to make $C_0 = \bar{A}_1$ and $C_2 = A_2$, and, for the central state, $C_1 = \bar{A}_2 \cap A_1$. However, $C_0$ and $C_2$ will not necessarily be mutually exclusive and an individual could simultaneously be predicted to be in category 2 and category 0 (Figure 1(*a*), shaded area). However, if $A_2$ is a subset of $A_1$ (Figure 1(*b*)), the classifier $C_2 = A_2$, $C_0 = \bar{A}_1$ and $C_1 = \bar{A}_2 \cap A_1$ *is* a valid partition. Notice that $C_1$ is, in Figure 1(*b*), the 'space' between $A_2$ and $A_1$.

For illustration, suppose there are three ordinal outcome categories of atherosclerosis in symptomatic heart disease patients, *mild*, *moderate* and *severe*. Also suppose predictive attributes are: elderly, that is, age over 60 ($E$); male gender ($M$); raised serum cholesterol ($S$). If the presence of any one of the attributes is a reasonable predictor of either *moderate* or *severe*, while presence of all three is a predictor of *severe*, we would have

$$A_1 = E \cup M \cup S$$

and

$$A_2 = E \cap M \cap S.$$

Clearly in this case $A_2$ is a subset of $A_1$, so that classification to the moderate category is accomplished with $C_1 = A_1 \cap \bar{A}_2$, which, in words, is any one, or two, of the three attributes, but not all three.

 

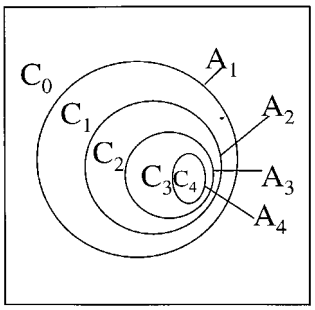


Figure 2. Nested binary partitions enabling construction of classifier $\{C_0, C_1, \ldots, C_4\}$

On the other hand suppose we had

$$A_1 = E \cup M$$

that is, either male or elderly predicts either *moderate* or *severe* atherosclerosis, and

$$A_2 = (E \cap M) \cup S$$

so that *severe* is predicted if you are an elderly male or have raised serum cholesterol. In this case $A_2$ is not contained in $A_1$ since someone who is neither male nor elderly, that is, a younger woman, yet has raised serum cholesterol is in $A_2$ but not $A_1$. Therefore $C_0 = \bar{A}_1$, $C_1 = A_1 \cap \bar{A}_1$, $C_2 = A_2$ is not a partition. We would have to decide how to classify a younger female with raised serum cholesterol.

## 4. MORE THAN THREE CATEGORIES

Consider now extending these ideas to $t + 1$ ordinal categories, $\{0, 1, \ldots, t\}$. We require a partition $C_0, C_1, \ldots, C_t$ from which individuals in $C_k$ will be classified to ordinal category $k$. The ordinal categories can be split into two groups at a point $k$ on the ordinal scale. Let $R_k = \{k, \ldots, t\}$ be the set of ordinal states in the range $k$ to $t$ and $\bar{R}_k = \{0, 1, \ldots, k-1\}$ is its complement. Suppose a SPAN partition $P_k = \{\bar{A}_k, A_k\}$ can be obtained for discrimination between $\bar{R}_k$ and $R_k$. This can be done for each $k = 1, 2, \ldots, t$ to give the sequence $A_1, A_2, \ldots, A_t$.

Now if the $A_k$ are nested in the sense illustrated in Figure 2, that is, if $A_t$ is a subset of $A_{t-1}$, which is itself a subset of $A_{t-2}$, and so on, then the sets $C_k = A_k \cap \bar{A}_{k+1}$ for $0 < k < t$, with $C_0 = \bar{A}_1$ and $C_t = A_t$, provide a valid classifier. Looking at Figure 2, the $C_k$ sets are essentially the spaces between concentric 'shells'.

### 4.1. Generating nested partitions

Unfortunately, if SPAN, or any other method, is used to independently establish the partitions $P_1, P_2, \ldots, P_t$, there is no guarantee that the partitions will be nested as required. However, partitions that *are* nested can be generated by a series of searches that are not independent. Instead, each search is constrained by the results of the previous search.

Specifically, we could begin by finding $P_t = \{\bar{A}_t, A_t\}$ in an unconstrained SPAN search for discrimination between $\bar{R}_t$ and $R_t$. Next $P_{t-1} = \{\bar{A}_{t-1}, A_{t-1}\}$ is found by another SPAN search,

 

to discriminate between $\bar{R}_{t-1}$ and $R_{t-1}$, with $A_{t-1}$ constrained so that $A_t$ is a subset of it. The constraint can be achieved by conducting the search so that, for each partition generated, $A_t$ is 'forced' into the Boolean expression as follows: if $P = \{\bar{A}, A\}$ is one of the generated partitions of the search, it is augmented by constructing the 'forced partition' $P^* = \{\bar{A}^*, A^*\}$ where $A^* = A \cup A_t$. The best $P^*$ on the search is used for $P_{t-1}$. Similarly, given $P_{t-2}$ is found by forcing $A_{t-1}$ into another search to discriminate between $\bar{R}_{t-2}$ and $R_{t-2}$, and the process is repeated.

This procedure is 'top down', in the sense that it starts with $P_t$ followed by $P_{t-1}$ and so on. It is not the only way to generate a nested sequence. It is possible, for example, to work 'bottom up' beginning with an unconstrained partition $P_1 = \{\bar{A}_1, A_1\}$ to discriminate between $\bar{R}_1$ and $R_1$. The next search to form $P_2$ is carried out to discriminate $\bar{R}_2$ from $R_2$ with $A_1$ forced into a generated partition, by creating $A^* = A \cap A_1$ and the best such partition becomes $P_2$. Yet another possibility, if there are more than three ordinal categories, is to begin in the middle and work out. For example, start with unconstrained $P_l$ for some $l$ in $1 < l < t$. Then, working 'upwards' for $k > l$ force $A^* = A \cap A_{k-1}$, and, working 'downwards', for $k < l$ force $A^* = A \cup A_{k+1}$.

### 4.2. Positive attributes

One of the features of SPAN is that binary partitions are formed by Boolean combinations of predictive factors that are positive with respect to the binary outcome. For example, 'raised serum cholesterol' is a positive attribute for a binary indicator of presence of atherosclerosis; its complement '*not* raised serum cholesterol' is a corresponding negative attribute. On the other hand *low* HDL cholesterol would be considered a positive attribute for that disease. An attribute can be considered positive if it is more common for $y = 1$ than it is for $y = 0$. Usually specifying an attribute in advance to be positive is not restrictive.

A SPAN search involves defining a set of $m$ positive attributes, say $T_m$, with respect to the outcome variable and the search is conducted from among the class of 'regular' Boolean combinations of the attributes in $T_m$. Here regular means that any partition $P = \{\bar{A}, A\}$ that is derived from $T_m$ is such that $A$ is a Boolean combination, using only the $\cup$ and $\cap$ operators, of the positive attributes.[7,8]

To establish an ordinal classifier by the forcing process just described, it is necessary to define a $T_m$ for each search, that is, a set of attributes that are positive with respect to the collapsed categories. Matters are simplified if the same attribute set is assumed for each search. This simplification can be justified if, as often occurs, the frequency of an attribute $X$ increases with steps up the ordinal scale, that is

$$P(X|0) \leqslant P(X|1) \leqslant \cdots \leqslant P(X|t). \tag{3}$$

In this case, with regard to the collapsed dichotomy, it can be shown that

$$P(X|\bar{R}_k) \leqslant P(X|R_k)$$

for each $k = 1, \ldots, t$. Therefore the attribute $X$ would be considered positive for each $k$ and, if all attributes each have the property (3), it is justifiable to assign a *global* or identical set of positive attributes for each search.

Assuming a global attribute set recognizes the ordinal nature of the outcome and adds additional structure on the partitions when, as is often the case, (3) holds. It guards against generating illogical and contradictory Boolean expressions. However, sometimes (3) may not

 

hold, for example, if an attribute is more common in intermediate states than at either extreme. In this case, consideration would have to be given to a separate attribute set for each search.

If a global positive attribute set is used, then it is clear that the 'end' classifiers $C_0$ and $C_t$ are, respectively, Boolean combinations of negative and positive attributes. The partition is then 'regular' in the sense that a combination of negative attributes predicts the lower end of the ordinal scale while a combination of positive attributes predicts the upper end. Intermediate points are predicted by combinations of positive and negative attributes. This seems to be a reasonable structure for an ordinal classifier.

## 5. ANALYSIS OF DIABETES SCREENING DATA

The background to this study was outlined in Section 2. Here I present three approaches to developing prediction rules, beginning with the proposed SPAN method.

### 5.1. SPAN method

The attributes that were used to develop the classification rules were: raised non-fasting fructosamine test; high body mass index; raised serum cholesterol raised serum triglyceride; raised urinary albumin; presence of hypertension, and ethnicity Maori, Pacific Islander or Asian – all factors known to be associated with impaired glucose tolerance or diabetes. As these factors are each more frequent with worsening glucose tolerance, it was justifiable to use a global attribute set for an ordinal SPAN analysis.

Adopting a top down procedure, the lower two ordinal categories, *normal* and *impaired*, were first collpsed into one and an unconstrained SPAN analysis was done using a reduction in diversity index

$$G = i(\pi) - p_A i(\pi_A) - p_{\bar{A}} i(\pi_{\bar{A}}) \tag{4}$$

where $i(x) = x(1 - x)$ and $\pi$, $\pi_A$ and $\pi_{\bar{A}}$ are sample proportions of the binary outcome means in the whole sample, in $A$ and in $\bar{A}$, respectively, and $p_A = 1 - p_{\bar{A}}$ is the proportion of the sample in $A$. An iterative complexity penalized procedure[8] was used for the analysis. In this procedure a plot of $G$ against a measure of complexity, $c$, of a partition is generated and the upper envelope of the plot is the 'complexity hull'. An optimal complexity penalized partition, that is, one that offsets increasing $G$ against complexity, must be one of the points on the hull.[8]

In this case the points on the hull were at $(c, G)$ co-ordinates: (1, 0·00200), (4, 0·00270) and (6, 0·00277). There is substantial improvement in $G$ between the $c = 1$ and $c = 4$ partitions, but little is gained by the complexity $c = 6$ partition. The complexity $c = 4$ partition is

$$A_2 = (F \cap U) \cup (F \cap E \cap T)$$

that is the rule in expression (1).

Next the upper two categories, *impaired* and *diabetes*, were collapsed into one and a SPAN analysis done with $A_2$ forced into the search using $A^* = A \cup A_2$, as described above. Points on the complexity hull were at $(c, G)$ co-ordinates: (1, 0·0050), (4, 0·0072) and (6, 0·0075). Again the optimal partition is at complexity $c = 4$.

$$A_1 = (F \cap U) \cup (F \cap T) \cup (F \cap H)$$

which is the rule in equation (2). It is easy to confirm that $A_2$ is a subset of $A_1$.

 

Table I. Re-classification table of SPAN derived three-way partition for ordinal categories 0, 1, 2 (*normal*, *impaired*, *diabetes*) of glucose tolerance using SPAN method with a diversity index. Equal false positive and false negative costs.

| Actual | $C_0$ | $C_1$ | $C_2$ | |
|---|---|---|---|---|
| 0 | 5114 | 115 | 49 | 5278 |
| 1 | 124 | 18 | 16 | 158 |
| 2 | 36 | 17 | 27 | 80 |
| Total | 5274 | 150 | 92 | 5516 |

With $C_0 = \bar{A}_1$, $C_1 = A_1 \cap \bar{A}_2$ and $C_2 = A_2$, the re-substitution classification of actual and predicted counts is shown in Table I. 'Pre-test' probabilities are $P(0) = 95{\cdot}6$ per cent, $P(1) = 2{\cdot}9$ per cent and $P(2) = 1{\cdot}5$ per cent. The estimated 'post-test' probabilities, or predictive values, are $P(0|C_0) = 5114/5274 = 97{\cdot}0$ per cent, $P(1|C_1) = 18/150 = 12$ per cent and $P(2|C_2) = 27/92 = 29{\cdot}4$ per cent. The overall misclassification rate, $Q = 6{\cdot}5$ per cent, is the sum of the off-diagonal counts divided by the total 5540. As the states are ordinal, it is appropriate to also calculate a weighted misclassification rate, say $Q'$, in which twice the weight is assigned to misclassifications from 0 to 2, and from 2 to 0. For Table I $Q' = 8{\cdot}0$ per cent.

A bottom up approach was also adopted in which the top two categories are collapsed and an analysis done to first find $A_1$ and then $A_2$. In this instance, the resulting $A_1$ and $A_2$ were identical to those by the top down procedure. Generally, however, this may not always be the case.

It is evident from Table I that applying these rules may lead to rather many false negatives. This may be unacceptable for such a screening programme. One way to generate rules that decreases the number of false negatives is to introduce misclassification costs. In the above analysis the cost of a false negative is implicitly the same as the cost of a false positive. A cost analysis can be achieved by 'altering' the prior probabilities in a similar way to that described in Breiman *et al.*[3] Details are given in the Appendix. I chose to make the cost of a false negative ten times that of a false positive ($\alpha = 10$) and then repeated the analysis as above.

Working top down to find $A_2$, the points on the hull were at $(c, G)$ co-ordinates (1, 0·1018), (4, 0·1110) and (5, 0·1126). Evidently little is gained by the more complex partitions; the $c = 1$ partition was simply raised fructosamine

$$A_2 = F. \tag{5}$$

Next the upper two categories were collapsed into one and a SPAN analysis done with $A_2$ forced into the search using $A^* = A \cup A_2$, as described above. The SPAN search resulted in points on the complexity hull at $(c, G)$ co-ordinates (1, 0·0379), (3, 0·0445), (4, 0.0450) and (6, 0·0462). Little seems to be gained by a partition with complexity greater than $c = 3$. The $c = 3$ partition is

$$A_1 = (F) \cup (H \cap B). \tag{6}$$

 

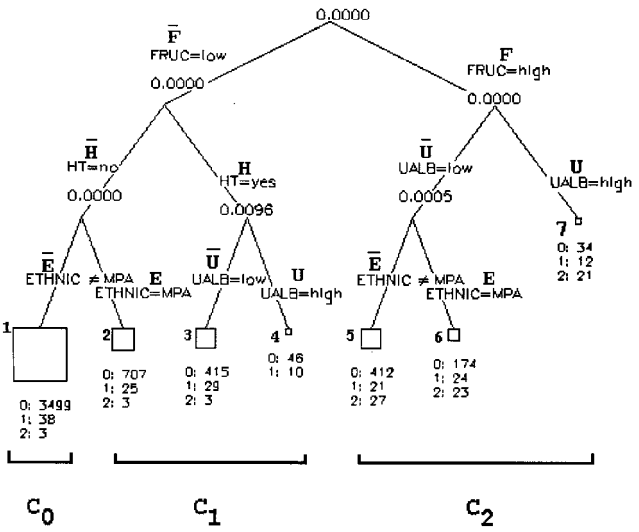


Figure 3. Tree diagram for classification in the diabetes screenig example. Outcome categories are 0, 1, 2 which corresponding node assignments $C_0, C_1, C_2$ and frequencies as indicated (for example, 0:707 indicates 707 individuals with category 0). Node squares are proportional to node sample size. Figures before each node split are $p$-values associated with a chi-square test. The tree partition (see text) is based on this tree but with branches 'FRUC = high' and 'HT = yes' pruned.

Obviously $A_2$ is a subset of $A_1$ and the classification rule, $C_1 = A_1 \cap \bar{A}_2$, to the intermediate state is simply

$$C_1 = \{F \cup (H \cap B)\} \cap \bar{F} = (\bar{F} \cap H \cap B)$$

that is, low fructosamine, hypertension and raised body mass. With $C_0 = \bar{A}_1$ and $C_2 = F$ the re-substitution classification of actual and predicted counts is shown in Table II. The estimated 'post-test' probabilities are $P(0|C_0) = 4289/4359 = 98{\cdot}4$ per cent, $P(1|C_1) = 34/418 = 8{\cdot}1$ per cent and $P(2|C_2) = 71/749 = 9{\cdot}5$ per cent. The $Q$ and $Q'$ misclassification indices are now 20·5% and 31·8% respectively.

### 5.2. Tree analysis

For a tree-based analysis the three ordinal outcome states were treated as if they were nominal and the Gini index of diversity was used to determine split effectiveness. The tree was initially grown to four tiers and subsequently pruned. Pruning a node was done 'manually' by lopping a branch if a chi-square test of a split did not achieve statistical significance ( $P < 0{\cdot}01$), or if one of its child nodes was small ($< 30$). The resulting tree is shown in Figure 3. The rectangles drawn at each node in Figure 3 are, in area, proportional to the number of observations at each node.

If a Bayes (maximum empirical probability) rule was used to assign each node to one of the three categories then, on the basis of the node frequencies, each node would be assigned to class 0, that is, *normal*. This is clearly not useful. An alternative way to assign nodes is by assigning the category with the largest likelihood ratio, $\pi(j|s)/\pi(j)$, that is, the extent to which the category probability $\pi(j|s)$, at node $s$, is altered from its prior $\pi(j)$. By this rule node 1 is assigned to category 0, nodes 2, 3 and 4 to category 1, and nodes 5, 6 and 7 are assigned to category 2, as

 

Table II. Re-classification table of SPAN derived three-way partition for ordinal categories 0, 1, 2 (*normal*, *impaired*, *diabetes*) of glucose tolerance using SPAN derived partition taking misclassification costs into account. Note: differences in row totals in Tables I to IV arise because of missing data

| Actual | $C_0$ | $C_1$ | $C_2$ | |
|---|---|---|---|---|
| 0 | 4289 | 376 | 621 | 5286 |
| 1 | 64 | 34 | 57 | 155 |
| 2 | 6 | 8 | 71 | 85 |
| Total | 4359 | 418 | 749 | 5526 |

Table III. Re-classification table of three-way partition for ordinal categories 0, 1, 2 (*normal*, *impaired*, *diabetes*) of glucose tolerance. Note: differences in row totals in Tables I and IV arise because of missing data

| Actual | $C_0$ | $C_1$ | $C_2$ | |
|---|---|---|---|---|
| 0 | 3499 | 1169 | 670 | 5338 |
| 1 | 38 | 64 | 57 | 159 |
| 2 | 3 | 6 | 71 | 80 |
| Total | 3540 | 1239 | 798 | 5577 |

shown in Figure 3. These rules imply that branches from nodes 'HT = yes' and 'FRUC = high' can be pruned. The rules for assignment to the three categories are

$$C_0 = \bar{F} \cap \bar{H}$$

$$C_1 = (\bar{F} \cap H) \cup (\bar{F} \cap \bar{H} \cap E)$$

$$C_2 = F.$$

Another way to assign the nodes is by changing prior probabilities. For example, by specifying equal prior probabilities to the three categories, the same assignment rule as that just given is obtained. However, as the data are representative of the population in which a screening rule is likely to be applied, adjusting the priors in this way seems unjustfiable.

Note that $C_1$ in the above simplifies to $C_1 = (\bar{F} \cap H) \cup (\bar{F} \cap E)$ showing the redundancy of $\bar{H}$ in the rule, a situation which can often arise with tree rules and which can be misleading.[6] Note also that the partition is regular with respect to the designated positive attributes, in the sense discussed in Section 4.2, because of the simplicity of the pruned tree.

Table III gives the re-substitution classification table using the above rule. The classification rates are $P(0|C_0) = 3499/3540 = 98{\cdot}8$ per cent, $P(1|C_1) = 64/1239 = 5{\cdot}2$ per cent and $P(2|C_2) = 71/798 = 8{\cdot}9$ per cent. The misclassification rates are $Q = 34{\cdot}8$ per cent and $Q' = 46{\cdot}9$ per cent.

 

Table IV. Ordinal logistic regression re-classification table of three-way partition for ordinal categories 0, 1, 2 (*normal*, *impaired*, *diabetes*) of glucose tolerance. Note: differences in row totals in Tables I–IV arise because of missing data

| Actual | $C_0$ | $C_1$ | $C_2$ | |
|---|---|---|---|---|
| 0 | 4776 | 448 | 53 | 5277 |
| 1 | 94 | 51 | 13 | 158 |
| 2 | 15 | 49 | 16 | 80 |
| Total | 4885 | 548 | 82 | 5515 |

Table V. Comparison of re-classification rates. Each measure expressed as a percentage

| Method* | $P(0\|C_1)$ | $P(1\|C_1)$ | $P(2\|C_2)$ | $Q$ | $Q'$ | Predictors used‡ |
|---|---|---|---|---|---|---|
| SPAN* | 97·0 | 12·0 | 29·4 | 6·5 | 8·0 | *F*, *H*, *E*, *T*, *U* |
| SPAN† | 98·4 | 8·1 | 9·5 | 20·5 | 31·8 | *F*, *H*, *B* |
| Tree | 98·8 | 5·2 | 8·9 | 34·8 | 46·9 | *F*, *H*, *E* |
| Logistic | 97·7 | 9·3 | 19·5 | 12·2 | 13·4 | *F*, *H*, *E*, *T*, *U*, *B* |

* Using SPAN rules (1) and (2)
† Using SPAN rules (5) and (6)
‡ *F*, *T* and *U* are positive fructosamine, triglyceride and urinary albumin tests, respectively. *H* is hypertensive, *B* body mass index $> 25$, *E* is Maori or Pacific Island ethnic group

### 5.3. Ordinal logistic regression analysis

An analysis was also done by logistic regression using a model in which the ordinal structure is explicitly accounted for by the form[5]

$$\text{logit}\, P(Y \leqslant j) = \alpha_j + \beta X.$$

for $j = 0, 1$. The predictors were all binary indicators and stepwise model reduction was done. Six predictors were highly statistically significant in the model (see Table V) and the direction of their effects the same as their designation as positive attributes in the SPAN analysis.

To be consistent with the tree analysis above, I based classification on a likelihood ratio rule, that is, $\pi(j|i)/\pi(j)$ where $\pi(j|i)$ is the predicted probability of category $j$ for individual $i$, and the classifications in Table IV were obtained. The predictive values of the rule are $P(0|C_0) = 4776/4885 = 97{\cdot}7$ per cent, $P(1|C_1) = 51/548 = 9{\cdot}3$ per cent and $P(2|C_2) = 16/82 = 19{\cdot}5$ per cent. The overall misclassification rate is $Q = 12{\cdot}2$ per cent and $Q' = 13{\cdot}4$ per cent.

Table V present a summary of the estimated predictive values, as are reported above, by the three methods. Overall, the first SPAN rule looks the best, but it will give more false negatives than the others. The second SPAN rule, which is very simple, gives fewer false negatives at the expense of more false positives and a greater overall misclassification rate. The tree-based rule

 

gives the largest overall misclassification rate. Logistic regression, on the other hand, fares comparatively well.

## 6. DISCUSSION

A method has been proposed for ordinal classification by applying the search partition analysis (SPAN) approach.[7,8] SPAN is based on a class of logical Boolean expressions, which is determined by specification of a set of positive attributes. The class can be thought of as analogous to the role that a model plays in regression methods in the sense that it formalizes a sensible framework within which to search for the 'best fit'. This framework is expanded in the proposal made here to apply SPAN to the problem of ordinal classification.

For this purpose a series of SPAN analyses for collapsed categories of the ordinal outcome has been proposed. The analyses together provide a partition of the predictor space for ordinal classification. Recognition of the ordinal nature of the outcome is made explicit in this procedure by two features. First, by the very reasonable way that ordinal categories are collapsed into above/below a point on the scale. Secondly, by an assumption, which is also reasonable for an ordinal outcome, of a global set of positive attributes for each search. The SPAN partition is then regular, in the sense discussed in Section 4.2. Regularity seems to be a sensible requirement for an ordinal classifier. Tree derived classifiers are unlikely to be regular since combinations of positive and negative attributes are forced together as a tree is grown. However, in the diabetes example, the partition *is* regular, but only because the pruned tree is very simple.

Although the suggested method is ostensibly for ordinal categories, it may be possible to also apply it to categories on a nominal scale. For example, nominal categories can always be labelled to correspond to increasing prevalence of some attribute $X$, to satisfy (3). If, with this labelling, property (3) holds for every other attribute then it would be justifiable to apply the method. It is, anyway, sometimes sensible to 'ordinalize' a nominal scale.[10]

Ordinal structure of the outcome can be accounted for in tree growing by using an ordinal splitting criterion, as in SPSS's AnswerTree (which invokes a scalable chi-square criterion[6]), or by 'twoing' in CART. Although the tree analysis described above was done using software (see below) which uses neither of these devices, I also ran an analysis with SPSS AnswerTree, designating the outcome as ordinal and the default 'C&RT' options. The resulting partition, after pruning and combining nodes as above, was identical to the tree analysis reported above.

The classification rates that are presented are all based on the re-substitution method and may be optimistic. Nevertheless, sample size is large and classification rates may not be unduly affected. Further, depending on the criteria and methods that are used, there is usually a number of contender partitions that are nearly 'optimal' in both SPAN and tree methods. Equally, of course, there is no 'correct' linear model. The rules presented here are the result of different judgements made in the course of the analysis and it is acknowledged that other equally valid partitions may exist. Deciding which partition to report is often difficult. Tree analysis, methods to pick from a 'forest' of trees[11] may be adaptable.

The search methods of analysis that are presented here (including the tree analysis) were done using Windows 95 SPAN statistical software, developed by the author and available from URL http:/www.auckland.ac.nz/mch/span.htm. SAS procedure LOGISTIC was used for the ordinal logistic regression analysis.

 

## APPENDIX: ALTERED PRIORS TO ACCOUNT FOR COSTS

Suppose $c_{A0}$ is the cost of misclassification of a false positive and $c_{\bar{A}1}$ that of a false negative. Let $\alpha = c_{\bar{A}1}/c_{A0}$ be the relative cost. The expected costs are

$$c_{A0}\pi P(A|0) + c_{\bar{A}1}(1-\pi)P(\bar{A}|1).$$

Putting $\pi' = \pi c_{A0}$ and $1-\pi' = (1-\pi)c_{1\bar{A}}$ as 'altered priors' gives the expected costs as

$$\pi' P(A|0) + (1-\pi')P(\bar{A}|1)$$

which suggests using altered priors that are user specified in the SPAN analysis. The effect that this has on the $G$ criterion in equation (4) is to adjust the weight given to the diversities of $A$ and $\bar{A}$ as follows.

Without loss of generality we can make

$$c_{A0} = \frac{1}{\pi + \alpha(1-\pi)}$$

and

$$c_{\bar{A}1} = \frac{\alpha}{\pi + \alpha(1-\pi)}$$

ensuring the altered prior $\pi'$ is a proper prior probability.

Replacing these in the reduction in diversity index, $G$, in equation (4), leads to

$$G' \propto i(\pi) - Rp_A i(\pi_A) - \bar{R}p_{\bar{A}} i(\pi_{\bar{A}}) \tag{7}$$

where the constant of proportionality is $c_{A0}c_{\bar{A}1}$ and $R = \pi_A/\pi'_A$ and $\bar{R} = (1-\pi_A)/(1-\pi'_A)$. Here $\pi'_A$ is the expected sample proportion in $A$ under the altered priors, specifically $\pi'_A = P(A|0)\pi' + P(A|1)(1-\pi')$. Hence, comparing (4) and (7) shows that the altered criterion weights the diversities of $A$ and $\bar{A}$ by factors $R$ and $\bar{R}$, respectively.

## REFERENCES

1. Wijesinha, A., Begg, C. B., Funkenstein, H. H. and McNeil, B. J. 'Methodology for the differential diagnosis of a complex data set: a case study using data from routine CT scan examinations', *Medical Decision Making*, **3**, 133–154 (1983).
2. Marshall, R. J. 'Hypothesis testing in the polychotomous logistic model with an application to detecting gastrointestinal concer', *Statistics in Medicine*, **4**, 337–344 (1985).
3. Breiman, L., Freidman, J. H., Olshen, R. A. and Stone, C. J. *Classification and Regression Trees*, Wadsworth International, California, 1984.
4. Clark, L. A. and Pregibon, D. 'Tree-based models', in Chambers, J. M. and Hastie, T. J (eds), *Statistical Models in S*, Wadsworth and Brooks/Cole, Pacific-Grove CA, 1992, pp. 377–420.
5. McCullagh, P. and Nelder, J. A., *Generalized Linear Models*, 2nd edn, Chapman & Hall, New York, 1989.
6. Magidson, J. 'Chi-squared analysis of a scalable dependent variable', in Proceedings of the 1992 Annual meeting of the American Statistical Association, Education Statistics Section, 1992.

 

7. Marshall, R. J. 'Partitioning methods for classification and decision making in medicine', *Statistics in Medicine*, **5**, 517–526 (1986).
8. Marshall, R. J., 'A program to implement a search method for identification of clinical subgroups', *Statistics in Medicine*, **14**, 2645–2659 (1995).
9. Baker, J., Metcalf, P., Scragg, R. and Johnson R. 'Fructosamine test-plus, a modified fructosamine assay', *Clinical Chemistry*, **37**, 552–555 (1991).
10. Feinstein, A.R. *Clinimetrics*, Yale University Press, 1987.
11. Chipman, H., George, E. I. and McCulloch, R. 'Making sense of a forest of trees', Proceedings of the 29th Symposium on the Interface: Computing Science and Statistics, 29, 84–92 (1998).

 