<table>
<thead>
<tr>
<th><strong>Statistica Sinica Preprint No:</strong> SS-14-164R2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Manuscript ID</strong></td>
</tr>
<tr>
<td><strong>URL</strong></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
</tr>
<tr>
<td><strong>Complete List of Authors</strong></td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
</tr>
<tr>
<td><strong>E-mail</strong></td>
</tr>
</tbody>
</table>

Notice: Accepted version subject to English editing.
SCORE-Scale Decision Tree for Paired Comparison Data

Kuang-Hsun Liu and Yu-Shan Shih

National Chung Cheng University

Abstract: A new decision tree method for analyzing paired comparison data is proposed. It finds the preference patterns of the subjects based on some covariates. A scoring system is implemented first and the total scores associated with each object for each subject are counted. The GUIDE regression tree for multi-responses is then applied to the score outcomes and the average scores of the objects are used to give the preference scale of the subjects in each terminal node. This way of preference ranking is identical to that given by the Berry-Terry model when the 2-1-0 scoring system is employed. Our tree method itself is free of selection bias. Simulation and real data analysis are given to demonstrate its usefulness.

Key words and phrases: Bradley-Terry model; GUIDE; regression tree; scoring system; selection bias.

1. Introduction

Paired comparison data are collected by comparing objects in couples. The ultimate goal is to find the preference patterns (ranks) of the subjects. However, it may be easier for people to compare pairs of objects than to rank a list of items (Cattelan; 2012). The Bradley-Terry model (Bradley and Terry; 1952; Davidson; 1970) which gives a latent preference scale to the objects is commonly used to analyze such data. Other paired comparison models include Thurston’s model, Mallows’ model and the Babington Smith model (Marden; 1995). Moreover, the preference scaling of a group of people may not only depend on characteristics of the objects but also on some covariates related to the people themselves. Cattelan (2012) reviews various methods on modeling paired comparison data. Among them, Strobl et al. (2011) proposes a model-based tree method to incorporate the covariates.

Classical tree methods including classification and regression trees are commonly used in data mining, machine learning and statistics. Starting at the root node, the methods recursively partition the data into two or more subnodes. A
split is used to partition the sample at each node. One way to decide the split is by using node impurity criteria. A different approach to determine the split is by conducting statistical tests. The final tree model is determined by either a direct stopping rule or a pruning method (Loh; 2011). The tree methods have been applied to analyze various data collected from diverse fields (Breiman et al.; 1984; Hothorn et al.; 2006; Loh; 2011). Easy interpretation is the key feature of these methods. In order to achieve this goal, the associated split selection method should be free of bias. That is, the probability of each covariate being selected is equal when the response variable is independent of the covariates. Several tree methods are proposed to eliminate such bias. Among them are QUEST (Loh and Shih; 1997), CTREE (Hothorn et al.; 2006), GUIDE (Loh; 2002, 2009; Loh and Zheng; 2013) and MOB (Zeileis et al.; 2008). For paired comparison data, Strobl et al. (2011) takes the subject related information into account first by fitting a Bradley-Terry model at each node. After such fitting, the split which consists of a covariate and a set is chosen by conducting parameter instability tests on the fitted model (Zeileis and Hornik; 2007). A direct stopping procedure based on \( p \)-values is performed to obtain the final tree.

In this article, we propose an alternative method to construct decision trees for paired comparison data. It relies on a scoring system which gives two points for a win, one point for a tie, and zero point for a loss for each paired comparison and counts the total scores associated with each object for each subject. We then treat the scores as multi-response outcomes and use the GUIDE regression tree method to help us identify possible preference patterns within each terminal node. In each terminal node, the average scores of the objects are used to give the preference scale to the objects. A nice property of using the 2-1-0 scoring system and its average scores to give preference ranking is given in Proposition 1.

An example of such tree is illustrated with a training-delivery-mode data set where there are 198 trainees with 3 subject-specific covariates and 5 objects. The goal of the study is to compare training delivery modes (objects) among trainees (subjects). These modes include computer-based (CO), TV-based (TV), paper-based (PA), audio-based (AU) and classroom-based (CL) training. The three covariates are age, learning personality type (1 accommodating, 2 diverging, 3 converging, 4 assimilating) and sex (1 male, 2 female). Complete data descrip-
Table 1.1: Mean scores in the terminal nodes of the GUIDE tree for the TRDEL data (Figure 1.1).

<table>
<thead>
<tr>
<th></th>
<th>CO</th>
<th>TV</th>
<th>PA</th>
<th>AU</th>
<th>CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>node 4</td>
<td>5.53</td>
<td>4.00</td>
<td>3.16</td>
<td>3.21</td>
<td>4.11</td>
</tr>
<tr>
<td>node 5</td>
<td>5.12</td>
<td>3.27</td>
<td>4.38</td>
<td>1.69</td>
<td>5.54</td>
</tr>
<tr>
<td>node 3</td>
<td>5.11</td>
<td>2.70</td>
<td>4.44</td>
<td>2.07</td>
<td>5.67</td>
</tr>
</tbody>
</table>

Partition is given in Section 4. Starting from the top node, the data are recursively partitioned by a split which is determined by the values of the covariates. This tree growing process continues until some terminal nodes are formed. For those terminal nodes, the plot of the average scores of the objects is shown and their values are given in Table 1.1. For this data set, there are three terminal nodes and they are further presented in Figure 1.1 (nodes 3, 4 and 5). From Figure 1.1, we can quickly summarize that the male (sex=1) trainees whose learning type is accommodating or diverging (node 4) rate their training modes in the following order: CO>CL>TV>AU>PA. For the female trainees of the same learning types (node 5), their preference is in the order of CL>CO>PA>TV>AU. The differences between these two groups show on their first two priorities (CO vs. CL) as well as the last three ones. For the last group of trainees whose learning type is either converging or assimilating (node 3), their preference is CL>CO>PA>TV>AU which is the same as that of the subjects in node 5. Furthermore, Table 1.1 shows that the largest scaling difference between CO and CL is 1.42 (5.53 – 4.11) in node 4. The amounts of the scaling difference between CL and CO are .56 and .42 in node 3 and node 5, respectively. The resulting tree helps us understand how the covariates are related to the preference pattern of the subjects in each terminal node. Therefore, the preference pattern of future trainees can be predicted.

The rest of the paper is organized as follows. The Bradley-Terry model is introduced in Section 2. Our proposed method and the tree method of Strobl et al. (2011) are described and contrasted in Section 3 and 4. In Section 5, simulation experiments are conducted to compare these two methods. Conclusion is given in Section 6.

2. The Bradley-Terry model

We consider $n$ subjects who compare all unordered pairs of $J$ objects. As a
Figure 1.1: GUIDE decision tree for the TRDEL data. At each intermediate node, an observation goes to the left branch if and only if the condition is satisfied. Sample sizes are below nodes. The plot also shows the average scores of the objects in each terminal node (CO: computer-based, TV: TV-based, PA: paper-based, AU: audio-based and CL: classroom-based).
result, each subject performs \( k^* = J \cdot (J - 1)/2 \) comparisons. Each comparison, say \((j, k)\), yields a choice for an answer \( c \in \{1, 2, 3\} \) where “1” means a win, “2” means a loss and “3” means a tie for object \( j \). Bradley and Terry (1952) introduces a probability model to fit such data with no ties. Davidson (1970) extends the model to allow tie (no preference) between two objects. The probabilities of three possible outcomes when comparing object \( j \) and \( k \) are defined as

\[
\begin{align*}
p_{jk1} &= \frac{\pi_j}{\pi_j + \pi_k + \nu \sqrt{\pi_j \pi_k}}, \\
p_{jk2} &= \frac{\pi_k}{\pi_j + \pi_k + \nu \sqrt{\pi_j \pi_k}}, \\
p_{jk3} &= \frac{\nu \sqrt{\pi_j \pi_k}}{\pi_j + \pi_k + \nu \sqrt{\pi_j \pi_k}},
\end{align*}
\]

where the \( \pi_j > 0, j = 1, \ldots, J \) are the locations of objects on the preference scale known as the worth parameters and \( \nu \geq 0 \) is a discriminant constant controlling the probability of ties. The worth parameters are scaled to sum to unity and their values give the preference scaling of the subjects. The MLE of the parameters, \( \hat{\Pi} = (\hat{\pi}_1, \ldots, \hat{\pi}_J) \) and \( \hat{\nu} \), can be obtained through an iterative procedure (Davidson, 1970). We refer this general model as the BT model. Possible extensions of the BT model can be found in Marden (1995), Turner and Firth (2012), Hatzinger and Dittrich (2012), and Cattelan (2012).

3. Tree methods

In this section, we first present our proposed method. It uses a scoring system which converts paired comparison outcomes into score vectors. The decision tree is obtained after the GUIDE regression tree (Loh and Zheng, 2013) is applied to the score vectors. Later, we introduce the tree method of Strobl et al. (2011). Both methods use the covariates only for node (data) splitting. The covariates are not used for data fitting in the terminal nodes.

3.1 The proposed method

We use the GUIDE regression tree for multi-response data to build our decision tree. Other multivariate regression trees, such as CTREE, can be used as well. A brief summary of the GUIDE regression tree methodology is given as follows.

The GUIDE regression tree originally targets univariate responses (Loh, 2002). It is extended to multi-responses by Loh and Zheng (2013). Loh (2002)
considers split variable selection and split point selection separately. In the process of selecting split variables, the method analyzes the associated univariate residual pattern of each covariate or interaction of two covariates. The pattern is accessed by the Pearson’s chi-squared independence tests between the signed residuals from a model fitted to the data and the covariate in the node. The covariate associated with the smallest \( p \)-values of the corresponding tests is selected as the split variable. Loh and Zheng (2013) generalizes the method to handle multi-response data by analyzing multivariate residual pattern where the residual vector is obtained by fitting a constant (mean response vector) model. At each node, the sample mean vector is fitted to the data and the residual vectors are obtained. Each residual is assigned to the positive or non-positive group based on its sign. Thus, when the assignment is applied to the multi-response outcomes, there are \( 2^J \) possible patterns for the residual sign vector. To determine if a predictor variable \( X \) is associated with the residual pattern, a contingency table with the sign patterns as the columns and the values of \( X \) (grouped, if \( X \) is not categorical) as the rows is created and the \( p \)-value of the Pearson’s independence test is obtained. The variable associated with the smallest \( p \)-value is the split variable. Detailed description of this selection scheme is given in Algorithm 3.1 of Loh and Zheng (2013).

By default, the split set is chosen by exhaustively searching all possible values after the split variable is determined. The selected set yields maximum reduction in node impurity which is defined as the sum of squared errors (normalized, if desired). This selection procedure continues until no further split is needed. The pruning method with cross-validation is applied to choose the right size tree (Breiman et al.; 1984). It gives multivariate mean responses as the predicted vector in each terminal node. Other features includes missing values handling and variables importance ranking. The method creates binary regression trees for multi-response data and it is free of selection bias (Loh and Zheng; 2013). The GUIDE regression tree method is implemented in the GUIDE program which can be obtained from the webpage: \texttt{http://www.stat.wisc.edu/~loh/guide.html}.

For paired comparison outcomes, we first employ a scoring system to convert them into multi-response outcomes. Our scoring system allows two points for a win, one point for a tie, and zero point for a loss for each paired comparison. This
2-1-0 scoring system is used, for example, in some hockey and soccer tournaments.
For each subject, we count the accumulated points for each object and treat the score outcomes as the corresponding multi-response outcomes. Afterward, the GUIDE regression tree method is applied to these outcomes and the covariates. Hence, in each terminal node, the average scores of the objects are obtained which in terms give the preference scale of the subjects. This approach therefore results a decision tree which can be used to analyze paired comparison data. Consequently, our method can be applied directly to data where the response is the ranks of the objects. Also, other scoring system is applicable, if it is available.

The following proposition supports the uses of the 2-1-0 scoring system and the average scores to give the preference scale to the objects in each terminal node. It shows that there is a one to one mapping between a ranking sequence on the total scores vector and a ranking sequence on the worth parameter estimators for complete paired comparison data.

**Proposition 1.** Let the total scores for object \( j \) be \( s_j = \sum_{i=1}^{n} (2w_{ij} + t_{ij}) \) where \( w_{ij} \) and \( t_{ij} \) are the numbers of wins and ties for subject \( i \), respectively. Denote the total scores vector as \( S = (s_1, \ldots, s_J) \). Then, for a balanced paired comparison experiment, the ranks based on \( S \) agree with the ranks obtained from the worth parameter estimators, \( \hat{\Pi} \), under the BT model.

The proof of Proposition 1 is given by Davidson (1970, p. 322) followed by some real examples. For the TRDEL data set, we obtain the following result.

**Application.** Suppose that the BT model is fitted to the TRDEL data. The worth parameter estimates are \((.30, .12, .19, .08, .31)\) with ranks \((4, 2, 3, 1, 5)\). Our scoring system yields the average scores \((5.19, 3.10, 4.18, 2.19, 5.33)\). We find that the rank vector of the average scores is exactly the same as that of the worth parameter estimates.

Therefore, the average scores of the objects reveal exactly the same preference ranks as if the BT model is fitted to the current data. Moreover, some optimum properties of using this total scores (row sum) ranking method are given by Huber (1963, p. 517-518).

For paired comparison data, we first convert comparison outcomes into score outcomes for each subject. By treating the score outcomes as multi-responses,
we apply the GUIDE method and a decision tree is obtained. The associated average scores are used to give the object preference scaling of the subjects in each terminal node. We can therefore identify the preference patterns of the subjects in the terminal nodes. Moreover, preference margins can be compared not only within each terminal node but also across the terminal nodes.

3.2 The bttree method

The bttree method of Strobl et al. (2011) is based on the model-based recursive partitioning scheme (Zeileis et al.; 2008) which fits local parametric models by partitioning the sample space. The advantages of recursively fitting local BT models as opposed to the fully parametric approaches are given in Strobl et al. (2011, Section 4). The method fits the BT model at each node. It then utilizes the parameter instability tests (Zeileis and Hornik; 2007) to select split variables. The split selection method is described in the following.

Denote \( \theta_j = \log \pi_j, j = 1, \ldots, J \) and \( \Theta = (\theta_1, \ldots, \theta_J) \). Given observations \( y_i \in \{1, 2, 3\}^k, i = 1, \ldots, n \), by assuming independence among the paired outcomes, the joint log-likelihood function based on the BT model is given by

\[
\log L(\Theta|y_1, \cdots, y_n) = \sum_{i=1}^{n} \sum_{j<k}^{3} \sum_{c=1}^{\ell} I(y_{i,jk} = c) \log(p_{jkc})
\]

where \( \Psi(y_i, \Theta) \) is the likelihood contribution of the \( i \)-th observation. The derivative of the likelihood contributions with respect to the parameter vector is

\[
\psi(y_i, \Theta) = \frac{\partial \Psi(y_i, \Theta)}{\partial \Theta} = \sum_{j<k}^{3} \sum_{c=1}^{\ell} I(y_{i,jk} = c) \frac{\partial \log(p_{jkc})}{\partial \Theta}.
\]

Let \( \hat{\Theta} \) be the MLE of \( \Theta \). Given \( l = 1, \ldots, m \) covariates, the deviations are cumulatively aggregated along each of the \( m \) covariates:

\[
W_l(t) = \hat{V}^{-\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{[n \cdot t]} \psi(y_{(i|l)}, \hat{\Theta}) \quad (0 \leq t \leq 1),
\]

where the index \( (i|l) \) is the \( i \)-th ordered observation with respect to the \( l \)-th
covariate, $\lfloor u \rfloor$ denotes the integer part of $u$, and $\hat{V} = \sum_{i=1}^{n} \psi(y_i, \hat{\Theta})\psi(y_i, \hat{\Theta})^T$.

Under the null hypothesis of parameter stability, the cumulative sum process, $W_l(t)$, converges to a $J$-dimensional Brownian bridge (Zeileis and Hornik; 2007). The associated $p$-value is obtained for each covariate. The covariate with the smallest $p$-value is then the candidate of the split variable. After the split variable is decided, the split set which yields two children nodes is searched exhaustively over all possible values. The one which maximizes the sum of the likelihood values associated with the two children nodes is chosen.

The procedure is applied recursively until the subsample size is too small (default value is 10) or the instability tests in a node are not significant (default value is .05). The tree method is implemented in the R package psychotree function bttree (Strobl et al.; 2011).

The default options of the bttree function are used in our study. The 2-1-0 scoring system is applied to obtain the scores. The default options of the GUIDE program are used. Besides, the 0-SE pruning rule with cross-validation (Breiman et al.; 1984) is used. For both tree methods, the minimum node size is 5 in real data analysis and it is 10 in the simulation studies.

4. Real data analysis

Our proposed method and the bttree method are applied to two real data sets in this section. From Proposition 1, we know that if the 2-1-0 scoring system is employed, our preference ranks are exactly the same as those given by the BT model. Thus, as far as the preference ranks are concerned, both methods yield the same object ranks for the same subjects. However, both methods yield different trees in these two cases. The differences are mainly because that the GUIDE and the bttree methods use different split selection methods.

4.1 TRDEL data

This data set was the result of a paired comparison study on training delivery modes and is included in the R package prefmod (Hatzinger and Dittrich; 2012). Recall that the five modes (objects) containing computer-based (CO), TV-based (TV), paper-based (PA), audio-based (AU) and classroom-based (CL) training were compared in couples by 198 trainees (subjects). Each participant was unemployed in the labor market training of the Austrian labor market service. Along with the paired comparison outcomes, three subject-specific covariates are
Table 4.1: Observed frequencies of paired comparisons for the TRDEL data.

<table>
<thead>
<tr>
<th></th>
<th>&gt;</th>
<th>&lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO : TV</td>
<td>157</td>
<td>41</td>
</tr>
<tr>
<td>CO : PA</td>
<td>109</td>
<td>89</td>
</tr>
<tr>
<td>TV : PA</td>
<td>82</td>
<td>116</td>
</tr>
<tr>
<td>CO : AU</td>
<td>162</td>
<td>36</td>
</tr>
<tr>
<td>TV : AU</td>
<td>129</td>
<td>69</td>
</tr>
<tr>
<td>PA : AU</td>
<td>137</td>
<td>61</td>
</tr>
<tr>
<td>CO : CL</td>
<td>86</td>
<td>112</td>
</tr>
<tr>
<td>TV : CL</td>
<td>55</td>
<td>143</td>
</tr>
<tr>
<td>PA : CL</td>
<td>72</td>
<td>126</td>
</tr>
<tr>
<td>AU : CL</td>
<td>51</td>
<td>147</td>
</tr>
</tbody>
</table>

collected to identify reasons for the trainees’ preference. They are age whose values are numeric in years, learning personality type: (1) accommodating, (2) diverging, (3) converging, (4) assimilating and sex: (1) male, (2) female. Its accumulated frequencies table of 10 paired comparisons is listed in Table 4.1. The learning types (styles) were characterized in Kolb and Kolb (2005). The data set is analyzed by Hatzinger and Dittrich (2012) using the loglinear model and they find that learning type is an informative factor.

We apply the bttree method to the data and the resulting tree is given in Figure 4.1. The worth parameter estimates in the terminal nodes are given in Table 4.2. From Figure 4.1, we observe that the method splits on gender only. For female trainees (node 3), their preference is in the order of CL>CO>PA>TV>AU. The preference order of the male trainees (node 2) is CO>CL>PA>TV>AU. Their preference order differs from that of the female trainees only on the first two objects. Nevertheless, the bttree method does not use the learning type variable to distinguish the preference patterns among the trainees. By contrast, our resulting tree (Figure 1.1) shows that our method splits on learning type first and on sex afterward. Our method discovers that the learning type variable is informative.

4.2 ISSP2000 data

The Dataset contains 6 items and 1595 respondents from Austria and Great Britain who were asked about their perception of environmental dangers. The data are included in the R package prefmod. The objects are CAR, air pollu-
Figure 4.1: Bttree decision tree for the TRDEL data (CO: computer-based, TV: TV-based, PA: paper-based, AU: audio-based and CL: classroom-based). At each intermediate node, an observation goes to the child node if and only if the stated condition is satisfied. The plot also shows the worth parameter estimates in each terminal node.

Table 4.2: Worth parameter estimates in the terminal nodes of the bttree decision tree for the TRDEL data.

<table>
<thead>
<tr>
<th></th>
<th>CO</th>
<th>TV</th>
<th>PA</th>
<th>AU</th>
<th>CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>node 2</td>
<td>.38</td>
<td>.13</td>
<td>.14</td>
<td>.10</td>
<td>.25</td>
</tr>
<tr>
<td>node 3</td>
<td>.25</td>
<td>.10</td>
<td>.22</td>
<td>.06</td>
<td>.37</td>
</tr>
</tbody>
</table>
Table 4.3: Mean scores in the terminal nodes of the GUIDE tree for the ISSP2000 data (Figure 4.2).

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>I</th>
<th>F</th>
<th>W</th>
<th>T</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>node 4</td>
<td>3.66</td>
<td>5.90</td>
<td>5.11</td>
<td>4.55</td>
<td>5.64</td>
<td>5.15</td>
</tr>
<tr>
<td>node 5</td>
<td>3.68</td>
<td>5.50</td>
<td>4.88</td>
<td>4.25</td>
<td>5.65</td>
<td>6.04</td>
</tr>
<tr>
<td>node 6</td>
<td>5.04</td>
<td>5.92</td>
<td>4.25</td>
<td>5.67</td>
<td>5.37</td>
<td>3.74</td>
</tr>
<tr>
<td>node 7</td>
<td>4.99</td>
<td>5.67</td>
<td>4.95</td>
<td>5.88</td>
<td>4.42</td>
<td>4.09</td>
</tr>
</tbody>
</table>

...tion caused by cars, IND, air pollution caused by industry, FARM, pesticides and chemicals used in farming, WATER, pollution of country's rivers, lakes and streams, TEMP, a rise in the world's temperature and GENE, modifying the genes of certain crops. The covariates are SEX: (1) male, (2) female, URB: (1) urban area, (2) suburbs of large cities, small town, county seat (3) rural area, AGE: (1) < 40 years, (2) 41-59 years, (3) 60+ years, CNTRY: (1) Great Britain, (2) Austria, and EDU: (1) below A-level/matrice, (2) A-level/matrice or higher.

One of the goal in the study is to find the relative importance of the objects among the respondents (Dittrich et al.; 2007).

Our resulting tree is displayed in Figure 4.2. The average scores in the terminal nodes are given in Table 4.3. The GUIDE decision tree indicates that the participants of Great Britain (CNTRY=1) concern more about industrial pollution (I) and water quality (W). People in this group with EDU=1 rate water quality on top (node 7). Otherwise, they are more serious about industrial pollution (node 6). For the respondents from Austria (CNTRY=2), industrial pollution is also an important item. So are rising temperatures (T) and genetic modification (G). In particular, people in this group with URB=2 or 3 (node 5), rank G and T at the forefront. They have the largest average score for G and T respectively among all the terminal nodes.

The bttree decision tree is given in Figure 4.3. The bttree method splits on AGE and SEX variables after splitting on CNTRY variable first. It shows that the respondents from Great Britain put I and W at the forefront. Among them, younger people (AGE ≤ 1) rank I over W. Otherwise, it is W in front of I. For the Austria respondents, I is the top concern for those people with age less than 40 (AGE ≤ 1). Within the group of people with age more than 40, women rank G and T on top and men rank I on top followed by G and T.
Figure 4.2: GUIDE decision tree for the ISSP2000 data. At each intermediate node, an observation goes to the left branch if and only if the condition is satisfied. Sample sizes are below nodes. The plot also shows the average scores of the objects in each terminal node (C: Car, I: Industry, F: Farm, W: Water, T: Temperature, and G: Gene).
The average scores vector is \((4.35, 5.73, 4.84, 5.12, 5.20, 4.76)\) given by our method and the worth parameter estimates vector is \((.11, .25, .15, .17, .18, .14)\) at the root node. These two vectors have the exact the same ranking order: \((1, 6, 3, 4, 5, 2)\) and it is guaranteed by Proposition 1. Both trees split on the country variable first. The other split variables are different. The GUIDE tree gives 4 terminal nodes while the bttree method gives 7 terminal nodes. The preference patterns found in both trees agree with what Dittrich et al. (2007, p. 25) had concluded:

*It is interesting to observe in our analysis that in Great Britain it is concerns about industrial pollution and water quality—natural resource issues—which are mostly at the forefront, whereas in Austria, concern about industrial pollution is as important as concerns about the future (rising temperatures and genetic modification).*

Consequently, our tree is easier to explain the difference among the subjects’ preferences.
Figure 4.3: Btree decision tree for the ISSP2000 data. At each intermediate node, an observation goes to the child node if and only if the stated condition is satisfied. The plot also shows the worth parameter estimates in each terminal node (C: Car, I: Industry, F: Farm, W: Water, T: Temperature, and G: Gene).
5. Simulation experiments

Both our GUIDE tree method with the 2-1-0 scoring system and the bttree method give the same preference ranks in each node, even though two tree methods yield different preference values of the subjects in each terminal node. Hence, we study their performance directly on rank data. First, we focus on the selection power of the two methods when the response depends on some covariates. Later, we compare the predictive power of the two methods. In order to study their performance, rank samples are simulated directly. The rank samples are then converted into paired comparison outcomes so that we can investigate the selection power as well as the predictive power of both methods. The prediction error is defined as the average Kendall’s distance between the true ranks and the predicted ranks given by the tree.

Rank samples are generated by the following Kendall’s tau distance-based model. For $J$ objects with label $1, \ldots, J$, let $\pi$ be a rank function from $\{1, \ldots, J\}$ onto $\{1, \ldots, J\}$ where $\pi(j)$ is the rank of object $j$. The Kendall’s tau distance-based model for rank data proposed by Diaconis (1988) is

$$\Pr(\pi|\lambda, \pi_0) = e^{-\lambda d(\pi, \pi_0)} \times C(\lambda)^{-1},$$  

(KDM)

where $\lambda \geq 0$ is the dispersion parameter, $d(\pi, \pi_0)$ is the Kendall tau distance function between rank function $\pi$ and $\pi_0$,

$$d(\pi, \pi_0) = \sum_{i<j} I\{[\pi(i) - \pi(j)][\pi_0(i) - \pi_0(j)]\},$$

and $C(\lambda)$ is a proportionality constant. The closer to the modal ranking $\pi_0$ is, the higher probability of occurrence ranking has. The distribution of ranks will be more concentrated around $\pi_0$ for smaller $\lambda$. Four mutually independent covariates are generated and their distributions are given in Table 5.1.

5.1 Selection power

In the following experiments, the ranking outcomes depend on some of the covariates are simulated. The distribution of the ranks follows the KDM model and its relationship with some covariates is given in Table 5.2. For example, under Model $A_1$, if $X_2 \leq 8$, a distance-based model with the modal ranking $\pi_0 = (1, 2, 3, 4)$ is generated. Otherwise, the modal ranking becomes $\pi_0 = (4, 1, 3, 2)$. 
Table 5.1: Distributions of $X$ variables used in the simulation studies. $Z$, $U_{16}$, $C_2$, and $C_{10}$ are mutually independent; $Z$ is a standard normal variable; $U_n$ is a uniformly distributed variable on the integer values of $[1, n]$; $C_m$ denotes a $m$-level category variable with equal probability for each category.

$X_1 \sim Z$
$X_2 \sim U_{16}$
$X_3 \sim C_2$
$X_4 \sim C_{10}$

Table 5.2: Models for selection power studies of the two tree methods. The ranking outcomes based on the KDM model are generated. The distributions of $X$’s are given in Table 5.1.

<table>
<thead>
<tr>
<th>Model</th>
<th>Child node</th>
<th>$\pi_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>$X_2 \leq 8$</td>
<td>1, 2, 3, 4</td>
</tr>
<tr>
<td></td>
<td>$X_2 &gt; 8$</td>
<td>4, 1, 3, 2</td>
</tr>
<tr>
<td>$A_2$</td>
<td>$X_2 \leq 8$</td>
<td>1, 2, 3, 4</td>
</tr>
<tr>
<td></td>
<td>$X_2 &gt; 8$</td>
<td>4, 3, 2, 1</td>
</tr>
<tr>
<td>$B_1$</td>
<td>$</td>
<td>X_2 - 8.5</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>X_2 - 8.5</td>
</tr>
<tr>
<td>$B_2$</td>
<td>$</td>
<td>X_2 - 8.5</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>X_2 - 8.5</td>
</tr>
</tbody>
</table>

Two scenarios are considered and two different modal rankings ($\pi_0 = (4, 1, 3, 2)$ and $\pi_0 = (4, 3, 2, 1)$) for the right child node are used. In Scenario A, models with a single change point on $X_2$ are considered; while in Scenario B, models with a pair of change points on $X_2$ are considered. For each model, we generated 200 random samples. The number of times of each covariate selected by the two methods are recorded in 500 repetitions. The selected probabilities of $X_2$ for various $\lambda$ values are given in Figure 5.1 and 5.2 respectively.

For Scenario A, the mean shift on $X_2$ affects the out-coming KDM models. Figure 5.1 reveals that the btree method performs better than the proposed method under Model $A_1$ and $A_2$. For Scenario B, the variance change on $X_2$ has an effect on the out-coming KDM models. From Figure 5.2, we find that the proposed method outperforms the btree method under Model $B_1$ and $B_2$. In summary, no one tree method can outperform the other. Each tree method has its own strength in selecting the informative covariate.

5.2 Prediction
Figure 5.1: The estimated probability of $X_2$ selected by the btTree method (BTtree) and the proposed method (GUIDE) under Model $A_1$ and $A_2$.

Figure 5.2: The estimated probability of $X_2$ selected by the btTree method (BTtree) and the proposed method (GUIDE) under Model $B_1$ and $B_2$. 
Two simulated tree models are used to compare the predictive power of the proposed tree method and the bttree method in the following.

The tree model shown in Figure 5.3 is simulated. In each terminal node, a KDM model with specified $\pi_0$ and $\lambda = 0.51$ is used to generate the data. Three $\pi_0$ parameter vectors $(1, 2, 3, 4), (4, 1, 3, 2)$ and $(4, 3, 2, 1)$ used in Section 5.1 are taken to generate data. A learning sample of size 400 is generated first. Both the bttree and our proposed methods are applied to the learning sample and the corresponding trees are obtained. A test sample of size 1000 is then generated using the same model (Figure 5.3) and it is used to measure the prediction errors of the two resulting trees. This procedure is repeated 100 times and the results are summarized in Table 5.3. Similarly, a tree model shown in Figure 5.4 is simulated. It differs from Figure 5.3 in that it has four terminal nodes and the first two splits depend on the $X_2$ covariate. The same computational procedure is conducted and the results are given in Table 5.4. For the first tree model, the first split is on $X_2$ and it is the mean difference on $X_2$ ($X_2 \leq 8$ vs. $X_2 > 8$) separates the data into node 1 and the other nodes. The $X_3$ values further divide the data into node 2 and 3. For the second tree model, the first two splits channel the data into three parts where both node 1 and 2 contains data following the same KDM model. It is the variance change on $X_2$ ($|X_2 - 8.5| > 4$ vs. $|X_2 - 8.5| \leq 4$) separates the data into node 1, 2 and the other nodes.

Table 5.3 shows the bttree method performs better than the GUIDE method in one scenario where the $p$-value is less than .05. In the other two scenarios, the differences are insignificant. On the other hand, Table 5.4 shows that the GUIDE method is better than the bttree method in the last case, but worse in the first case. Overall, we find that these two methods are competitive in terms of prediction power.

6. Conclusion

Decision trees are excellent data exploratory tools. They provide intuitive insights into the data pattern at hand. In this article, we propose a new method of constructing decision trees on finding preference patterns of paired comparison data. It implements a scoring system on the paired comparisons and object scores are obtained for each subject. It then uses the scores as the responses. The GUIDE regression tree is applied to these multi-response outcomes to obtain
Figure 5.3: Simulated tree model I. The split beside each intermediate node channels each case to the left node, if it is satisfied; otherwise to the right. Beside each terminal node is the associated $\pi_0$ parameter in the KDM model which is used to generate the data in the node.

Table 5.3: The prediction results for the bttree and the GUIDE methods. A learning sample of size 400 and a test sample of size 1000 are generated by the tree model in Figure 5.3. The paired $t$ statistic is computed for the prediction difference between the bttree and the GUIDE methods over 100 repetitions.

<table>
<thead>
<tr>
<th>$\pi_A$</th>
<th>$\pi_B$</th>
<th>$\pi_C$</th>
<th>$t$ statistic</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,2,3,4)</td>
<td>(4,1,3,2)</td>
<td>(4,3,2,1)</td>
<td>-3.21</td>
<td>0.002</td>
</tr>
<tr>
<td>(4,1,3,2)</td>
<td>(1,2,3,4)</td>
<td>(4,3,2,1)</td>
<td>-0.29</td>
<td>0.772</td>
</tr>
<tr>
<td>(4,3,2,1)</td>
<td>(1,2,3,4)</td>
<td>(4,1,3,2)</td>
<td>-1.38</td>
<td>0.170</td>
</tr>
</tbody>
</table>

Figure 5.4: Simulated tree model II. The split beside each intermediate node channels each case to the left node, if it is satisfied; otherwise to the right. Beside each terminal node is the associated $\pi_0$ parameter in the KDM model which is used to generate the data in the node.
Table 5.4: The prediction results for the bttree and the GUIDE methods. A learning sample of size 400 and a test sample of size 1000 are generated by the tree model in Figure 5.4. The paired $t$ statistic is computed for the prediction difference between the bttree and the GUIDE methods over 100 repetitions.

<table>
<thead>
<tr>
<th>$\pi_A$</th>
<th>$\pi_B$</th>
<th>$\pi_C$</th>
<th>$t$ statistic</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,2,3,4)</td>
<td>(4,1,3,2)</td>
<td>(4,3,2,1)</td>
<td>-7.29</td>
<td>$&lt; 10^{-10}$</td>
</tr>
<tr>
<td>(4,1,3,2)</td>
<td>(1,2,3,4)</td>
<td>(4,3,2,1)</td>
<td>1.65</td>
<td>0.102</td>
</tr>
<tr>
<td>(4,3,2,1)</td>
<td>(1,2,3,4)</td>
<td>(4,1,3,2)</td>
<td>2.84</td>
<td>0.005</td>
</tr>
</tbody>
</table>

the decision tree. Within each terminal node, the average scores of the objects are used to give the preference scale of the subjects. It retains the nature of easy interpretation without worrying about the split selection bias.

Our tree method uses average scores to give the preference scale to the objects in each terminal node. By contrast, the bttree decision tree yields the preference scale by using the worth parameter values of the fitted BT model. Proposition 1 shows that if the 2-1-0 scoring system is implemented, our preference ranks are exactly the same as those given by the bttree method. Even though the two methods give the same preference ranks to the objects for the same subjects, they differ in their split selection methods. We find that, for the TRDEL data, our method is able to select the informative covariate which is not in the bttree model. For the ISSP2000 data, the two methods give different trees and the GUIDE trees are more easier for us to explain the preference patterns. In terms of selection power, our method is competitive with the bttree method in our simulation studies. Furthermore, through simulation experiments, we find that our tree method performs as well as the bttree method in terms of predictive power measured by the Kendall’s distance.

A benefit of using scores is that the GUIDE method is applicable to missing values in paired comparison outcomes. The 2-1-0 scoring system can be changed to the other scoring system as long as it is available. As a result, our approach may provide a paradigm to construct decision trees for paired comparison data.
Bibliography


Department of Mathematics, National Chung Cheng University, Taiwan
E-mail: (lksxdd@gmail.com)

Department of Mathematics, National Chung Cheng University, Taiwan
E-mail: (yshih@math.ccu.edu.tw)