Statistica Sinica **26** (2016), 429-444
doi:http://dx.doi.org/10.5705/ss.2014.164

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 Bradley-Terry model when the 2-1-0 scoring system is employed. Our tree method itself is free of selection bias. Simulation and 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, Wickelmaier, and Zeileis (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, Hornik, and Zeileis (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)), GUIDE (Loh (2002, 2009); Loh and Zheng (2013)), CTREE (Hothorn, Hornik, and Zeileis (2006)), and MOB (Zeileis, Hothorn, and Hornik (2008)). For paired comparison data, Strobl, Wickelmaier, and Zeileis (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 the 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 that gives two points for a win, one point for a tie, and no points 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.

Such a tree is illustrated with a training-delivery-mode data set in which 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 description 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. For this data set, there are three terminal nodes and they are further presented in Figure 1 (nodes 3, 4 and 5). From Figure 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

Table 1. Mean scores in the terminal nodes of the GUIDE tree for the TRDEL data (Figure 1).

	CO	TV	PA	AU	CL
node 3	5.11	2.70	4.44	2.07	5.67
node 4	5.53	4.00	3.16	3.21	4.11
node 5	5.12	3.27	4.38	1.69	5.54

5), they 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 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. In this way, 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, Wickelmaier, and Zeileis (2011) are described and contrasted in Section 3 and 4. In Section 5, simulation experiments are reported comparing these two methods.

2. The Bradley-Terry Model

We consider n subjects who compare all unordered pairs of J objects. As a result, each subject performs $k^* = J*(J-1)/2$ 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

$$p_{jk1} = \frac{\pi_j}{\pi_j + \pi_k + \nu\sqrt{\pi_j\pi_k}},$$
$$p_{jk2} = \frac{\pi_k}{\pi_j + \pi_k + \nu\sqrt{\pi_j\pi_k}},$$
$$p_{jk3} = \frac{\nu\sqrt{\pi_j\pi_k}}{\pi_j + \pi_k + \nu\sqrt{\pi_j\pi_k}},$$

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

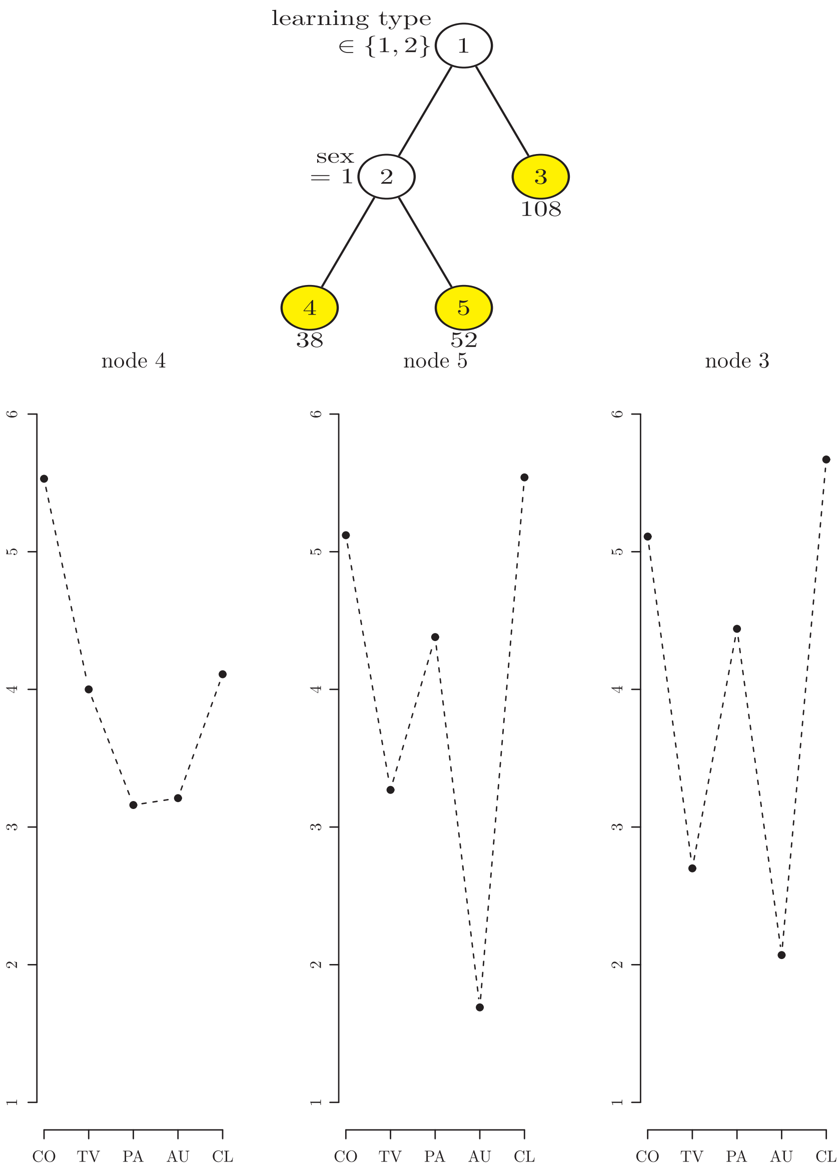

Figure 1. GUIDE decision tree for the TRDEL data. At each intermediate node, an observation goes to the left branch if and only if the condition is satisfied. Sample sizes are below nodes. The plot also shows the average scores of the objects in each terminal node (CO: computer-based, TV: TV-based, PA: paper-based, AU: audio-based, and CL: classroom-based).

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 to 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 that 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, Wickelmaier, and Zeileis (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.

The GUIDE regression tree originally targeted univariate responses (Loh (2002)). It was extended to multi-responses by Loh and Zheng (2013). 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: `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 no points 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. The GUIDE regression tree method is then applied to these outcomes and the covariates. At each terminal node, the average scores of the objects are obtained, which in turn gives the preference scale of the subjects. This approach results in a decision tree that can be used to analyze paired comparison data. This method can be applied directly to rank data, and can be used with other scoring systems.

A proposition due to Davidson (1970) ties the ranking sequence on the total scores vector to the ranking sequence on the worth parameter estimators for complete paired comparison data.

Proposition 1. *Let the total scores for object j be $s_j = \sum_{i=1}^{n}(2 * w_{ij} + t_{ij})$ 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, the ranks based on S agree with the ranks obtained from the worth parameter estimators, $\hat{\Pi}$, under the BT model.*

For the TRDEL data set, we find the worth parameter estimates (0.30, 0.12, 0.19, 0.08, 0.31) with ranks $(4, 2, 3, 1, 5)$, where our scoring system yields the average scores (5.19, 3.10, 4.18, 2.19, 5.33). The rank vectors here are the same. Some optimum properties of using the row sum ranking method are given by Huber (1963, p. 517-518).

By treating the score outcomes as multi-responses and applying the GUIDE method, a decision tree is obtained. The associated average scores can be used to give the object preference scaling of the subjects in each terminal node, and preference margins can be compared within and across terminal nodes.

3.2. The bttree method

The bttree method of Strobl, Wickelmaier, and Zeileis (2011) is based on the model-based recursive partitioning scheme (Zeileis, Hothorn, and Hornik (2008)) that fits local parametric models by partitioning the sample space. The advantages of recursively fitting local BT models as opposed to a fully parametric approach are given in Strobl, Wickelmaier, and Zeileis (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 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, Wickelmaier, and Zeileis (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. As well, the 0-SE pruning rule with cross-validation (Breiman et al. (1984)) is used. For both tree methods, the minimum node size is 5 for data analysis and is 10 in the simulation studies.

4. Data Analysis

Our proposed method and the bttree method are applied to two data sets. The methods yield the same object ranks for the same subjects, but yield different trees in these two cases, mainly because of 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 the modes (objects) computer-based (CO), TV-based (TV), paper-based (PA), audio-based (AU), and classroom-based (CL) training. Three covariates were *age* in years, *learning personality types* of accommodating, diverging, converging, and assimilating, and *sex* male or female. The accumulated frequencies

Table 2. Observed frequencies of paired comparisons for the TRDEL data.

	>	<
CO : TV	157	41
CO : PA	109	89
TV : PA	82	116
CO : AU	162	36
TV : AU	129	69
PA : AU	137	61
CO : CL	86	112
TV : CL	55	143
PA : CL	72	126
AU : CL	51	147

Table 3. Worth parameter estimates in the terminal nodes of the bttree decision tree for the TRDEL data (Figure 2.)

	CO	TV	PA	AU	CL
node 2	0.38	0.13	0.14	0.10	0.25
node 3	0.25	0.10	0.22	0.06	0.37

table of 10 paired comparisons is listed in Table 2. The learning types (styles) were characterized in Kolb and Kolb (2005). The data set was analyzed by Hatzinger and Dittrich (2012) using a loglinear model and they found that learning type to be an informative factor.

We applied the bttree method to the data and the resulting tree is given in Figure 2. The worth parameter estimates in the terminal nodes are given in Table 3. From Figure 2, we observe that the method splits on gender only. For female trainees (node 3), the preference is `CL>CO>PA>TV>AU`. The preference order of the male trainees (node 2) is `CO>CL>PA>TV>AU`, differing from that of the female trainees only on the first two objects. The bttree method does not use the learning type variable to distinguish the preference patterns among the trainees, where our tree (Figure 1) shows that our method splits first on learning type and then on sex. Our method has that the learning type variable is informative.

4.2. ISSP2000 data

The dataset contains 6 items and 1,595 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 C, air pollution caused by cars, I, air pollution caused by industry, F, pesticides and chemicals used in farming, W, pollution of country's rivers, lakes and streams, T, a rise in the world's temperature, and G, modifying the genes of certain crops. The covariates are SEX : (1) male, (2) female; URB: (1) urban area, (2) suburbs of

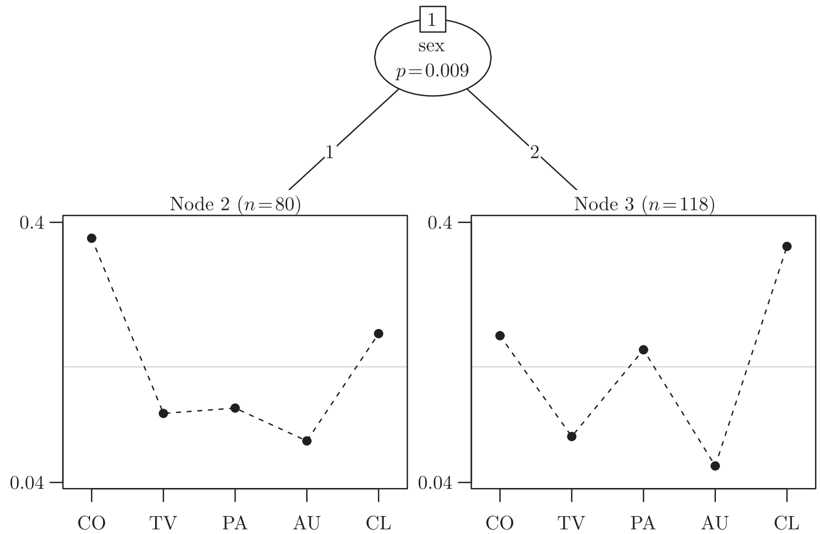

Figure 2. Bttree decision tree for the TRDEL data (CO: computer-based, TV: TV-based, PA: paper-based, AU: audio-based, and CL: classroom-based). At each intermediate node, an observation goes to the child node if and only if the stated condition is satisfied. The plot also shows the worth parameter estimates in each terminal node.

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 goals 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 3. The average scores in the terminal nodes are given in Table 4. The GUIDE decision tree indicates that the participants of Great Britain are concerned more about industrial pollution and water quality. 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, industrial pollution is also an important item. So are rising temperatures and genetic modification. 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 among all the terminal nodes.

The bttree decision tree is given in Figure 4. It splits on AGE and SEX after first splitting on CNTRY. 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 the reverse. For the Austrian respondents, I is the top concern for those people with AGE = 1. For others, women rank G and T on top and men rank I on top followed by G and T.

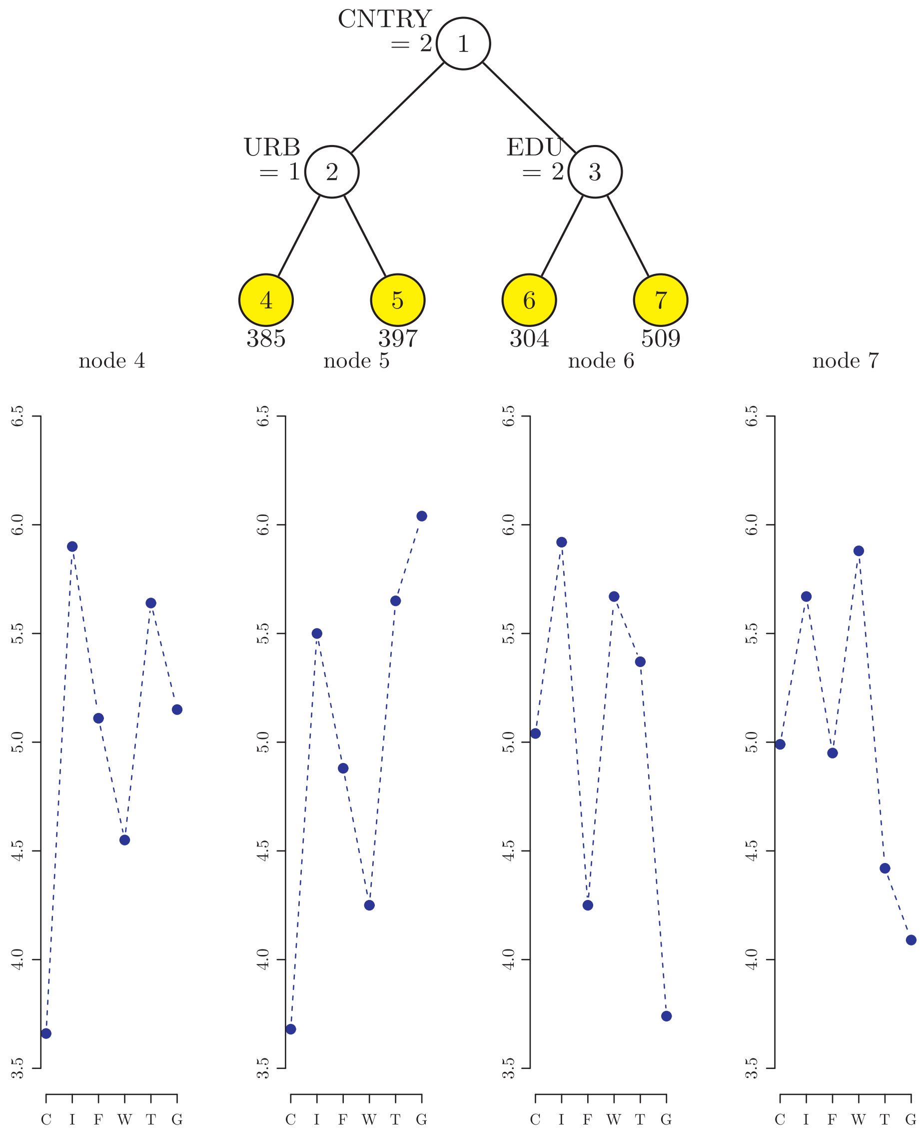

Figure 3. GUIDE decision tree for the ISSP2000 data. At each intermediate node, an observation goes to the left branch if and only if the condition is satisfied. Sample sizes are below nodes. The plot also shows the average scores of the objects in each terminal node (C: Car, I: Industry, F: Farm, W: Water, T: Temperature, and G: Gene).

The average scores vector is $(4.35, 5.73, 4.84, 5.12, 5.20, 4.76)$, by our method, and the worth parameter estimates vector is $(0.11, 0.25, 0.15, 0.17, 0.18, 0.14)$ at the root node. These vectors have the rank order: $(1, 6, 3, 4, 5, 2)$. Both trees split first on the country variable, yet the other split variables are different. The

Table 4. Mean scores in the terminal nodes of the GUIDE tree for the ISSP2000 data (Figure 3).

	C	I	F	W	T	G
node 4	3.66	5.90	5.11	4.55	5.64	5.15
node 5	3.68	5.50	4.88	4.25	5.65	6.04
node 6	5.04	5.92	4.25	5.67	5.37	3.74
node 7	4.99	5.67	4.95	5.88	4.42	4.09

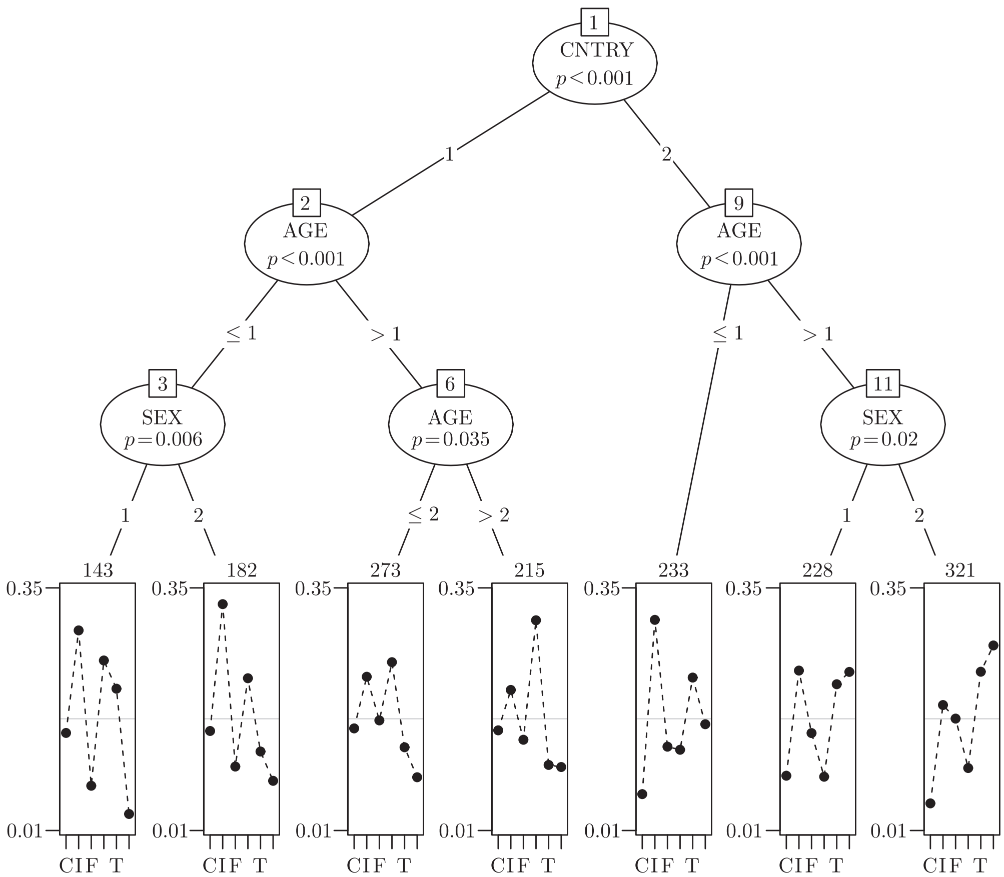

Figure 4. Bttree decision tree for the ISSP2000 data. At each intermediate node, an observation goes to the child node if and only if the stated condition is satisfied. The plot also shows the worth parameter estimates in each terminal node (C: Car, I: Industry, F: Farm, W: Water, T: Temperature, and G: Gene).

GUIDE tree gives 4 terminal nodes while the bttree method gives 7. The preference patterns found in both trees agree with Dittrich et al. (2007, p. 25), while our tree explains the difference among the subjects' preferences more simply.

Table 5. 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; U_n is uniform on the integer values $[1, n]$; C_m denotes a m-level category variable with equal probability for each category.

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

5. Simulation Experiments

While the GUIDE tree method with the 2-1-0 scoring system and the bt-tree method give the same preference ranks at each node, they yield different preference values of the subjects at the terminal nodes. Hence, we study their performance directly on rank data. We focus on the selection power of the two methods when the response depends on some covariates, then we compare their predictive power. Rank samples are simulated directly, then converted into paired comparison outcomes so that we can investigate selection and predictive powers. 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 a Kendall's tau distance-based model. For J objects with label $1, \ldots, J$, let π 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

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

where $\lambda \geq 0$ is the dispersion parameter, $d(\pi, \pi_0)$ is the Kendall tau distance function between rank function π and π_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 is π_0, the higher the probability of occurrence ranking. The distribution of ranks is more concentrated around π_0 for smaller λ. Four mutually independent covariates were generated and their distributions are given in Table 5.

5.1. Selection power

In these experiments, the ranking outcomes depend on some of the covariates. The distribution of the ranks follows the KDM model and its relationship with some covariates is given in Table 6. For example, under Model A_1, if $X_2 \leq 8$,

Table 6. 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.

Model	Child node	π_0
A_1	$X_2 \leq 8$	$1,2,3,4$
	$X_2 > 8$	$4,1,3,2$
A_2	$X_2 \leq 8$	$1,2,3,4$
	$X_2 > 8$	$4,3,2,1$
B_1	$\|X_2 - 8.5\| \leq 4$	$1,2,3,4$
	$\|X_2 - 8.5\| > 4$	$4,1,3,2$
B_2	$\|X_2 - 8.5\| \leq 4$	$1,2,3,4$
	$\|X_2 - 8.5\| > 4$	$4,3,2,1$

a distance-based model with the modal ranking $\pi_0 = (1,2,3,4)$ is generated. Otherwise, the modal ranking is $\pi_0 = (4,1,3,2)$. Two scenarios were considered and two different modal rankings ($\pi_0 = (4,1,3,2)$ and $\pi_0 = (4,3,2,1)$) for the right child node were used. In Scenario A, models with a single change point on X_2 were considered; while in Scenario B, models with a pair of change points on X_2 were considered. For each model, we generated 200 random samples. The number of times in 500 repetitions each covariate was selected by the two methods was recorded. The selected probabilities of X_2 for various λ values are given in Figures 5 and 6.

For Scenario A, the mean shift on X_2 affects the out-coming KDM models. In Figure 5, the bttree 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. In Figure 6, the proposed method outperforms the bttree method under Model B_1 and B_2. Each tree method has its own strength in selecting informative covariates.

5.2. Prediction

Two simulated tree models were used to compare the predictive power of the proposed tree method and the bttree method.

The tree model in Figure 7 was simulated. In each terminal node, a KDM model with specified π_0 and $\lambda = 0.51$ was used to generate the data. The π_0 parameter vectors $(1,2,3,4), (4,1,3,2)$ and $(4,3,2,1)$ were taken to generate data. A learning sample of size 400 was generated first. The bttree and our method were applied to the learning sample and the corresponding trees were obtained. A test sample of size 1,000 was then generated using the same model (Figure 7), and was used to measure the prediction errors of the resulting trees. This procedure was repeated 100 times and the results are summarized in Table 7. Similarly, the tree model in Figure 8 was simulated. It has four terminal nodes

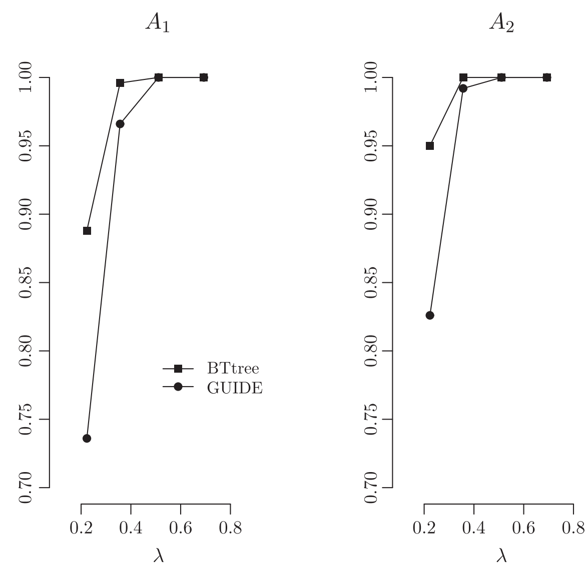

Figure 5. The estimated probability of X_2 selected by the bttree method (BTtree) and the proposed method (GUIDE) under Model A_1 and A_2.

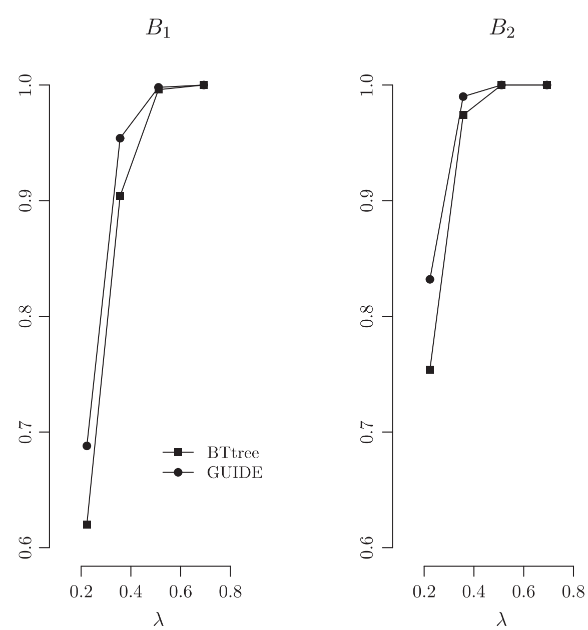

Figure 6. The estimated probability of X_2 selected by the bttree method (BTtree) and the proposed method (GUIDE) under Model B_1 and B_2.

and the first two splits depend on the X_2 covariate. The same computational procedure was conducted and the results are given in Table 8. For the first

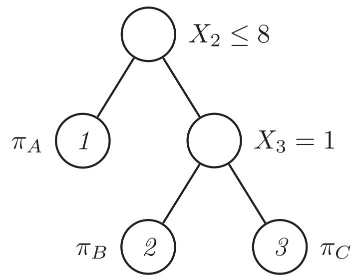

Figure 7. Simulated tree model I. The split beside each intermediate node channels each case to the left node, if it is satisfied; otherwise to the right. Beside each terminal node is the associated π_0 parameter in the KDM model which was used to generate the data in the node.

Table 7. The prediction results for the bttree and the GUIDE methods. A learning sample of size 400 and a test sample of size 1,000 were generated by the tree model in Figure 7. The paired t statistic was computed for the prediction difference between the bttree and the GUIDE methods over 100 repetitions.

π_A	π_B	π_C	t statistic	p-value
(1,2,3,4)	(4,1,3,2)	(4,3,2,1)	-3.21	0.002
(4,1,3,2)	(1,2,3,4)	(4,3,2,1)	-0.29	0.772
(4,3,2,1)	(1,2,3,4)	(4,1,3,2)	-1.38	0.170

tree model, the first split is on X_2 and 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 nodes 2 and 3. For the second tree model, the first two splits channel the data into three parts where nodes 1 and 2 contain data following the same KDM model; the variance change on X_2 ($|X_2 - 8.5| > 4$ vs. $|X_2 - 8.5| \leq 4$) separates the data into nodes 1, 2 and the other nodes.

In Table 7, the bttree method performs better than the GUIDE method in one scenario when the p-value is less than .05. In the other two scenarios, the differences are insignificant. On the other hand, In Table 8, the GUIDE method is better than the bttree method in the last case, but worse in the first case. Overall, the two methods are competitive in terms of prediction power.

Acknowledgements

The authors are very grateful to the Co-Editor, the Associate Editor and the two referees for the helpful comments and suggestions. The authors thank Prof. Wei-Yin Loh for the comments on the draft which initiate the study. This research is partly supported by Taiwan MOST grant 102-2118-M-194-002. The work was done partially while YS was visiting the Institute for Mathematical

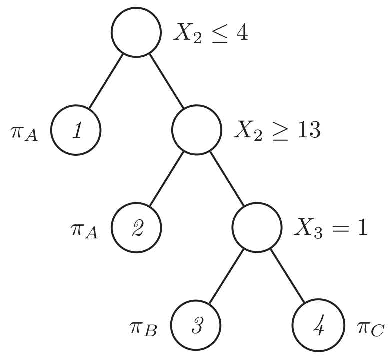

Figure 8. Simulated tree model II. The split beside each intermediate node channels each case to the left node, if it is satisfied; otherwise to the right. Beside each terminal node is the associated π_0 parameter in the KDM model which was used to generate the data in the node.

Table 8. The prediction results for the bttree and the GUIDE methods. A learning sample of size 400 and a test sample of size 1,000 were generated by the tree model in Figure 8. The paired t statistic was computed for the prediction difference between the bttree and the GUIDE methods over 100 repetitions.

π_A	π_B	π_C	t statistic	p-value
(1,2,3,4)	(4,1,3,2)	(4,3,2,1)	-7.29	$< 10^{-10}$
(4,1,3,2)	(1,2,3,4)	(4,3,2,1)	1.65	0.102
(4,3,2,1)	(1,2,3,4)	(4,1,3,2)	2.84	0.005

Sciences, National University of Singapore in 2014. The visit was supported by the Institute.

References

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* **39**, 324-345.

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth, Belmont.

Cattelan, M. (2012). Models for paired comparison data: A review with emphasis on dependent data. *Statist. Sci.* **27**, 412-433.

Davidson, R. R. (1970). On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. *J. Amer. Statist. Assoc.* **65**, 317-328.

Diaconis, P. (1988). *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics.

Dittrich, R., Francis, B., Hatzinger, R. and Katzenbeisser, W. (2007). A paired comparison approach for the analysis of sets of Likert-scale responses. *Statist. Modelling* **7**, 3-28.

Hatzinger, R. and Dittrich, R. (2012). prefmod: An R package for modeling preferences based on paired comparisons, rankings, or ratings. *J. Statist. Software* **48**, 1-31.

Hothorn, T., Hornik, K. and Zeileis, A. (2006). Unbiased recursive partitioning: a conditional inference framework. *J. Comput. Graph. Statist.* **15**, 651-674.

Huber, P. (1963). Pairwise comparison and ranking: optimum properties of the two sum procedure. *Ann. Math. Statist.* **34**, 511-520.

Kolb, A. Y. and Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Acad. Manag. Learn. Education* **4**, 193-212.

Loh, W.-Y. (2002). Regression trees with unbiased variable selection and interaction detection. *Statist. Sinica* **12**, 361-386.

Loh, W.-Y. (2009). Improving the precision of classification trees. *Ann. Appl. Statist.* **3**, 1710-1737.

Loh, W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Rev.: Data Mining and Knowledge Discovery* **1**, 14-23.

Loh, W.-Y. and Shih, Y.-S. (1997). Split selection methods for classification trees. *Statist. Sinica* **7**, 815-840.

Loh, W.-Y. and Zheng, W. (2013). Regression trees for longitudinal and multiresponse data. *Ann. Appl. Statist.* **7**, 495-522.

Marden, J. I. (1995). *Analyzing and Modeling Rank Data.* Chapman & Hall.

Strobl, C., Wickelmaier, F. and Zeileis, A. (2011). Accounting for individual differences in Bradley-Terry models by means of recursive partitioning. *J. Educational and Behavioral Statist.* **36**, 135-153.

Turner, H. and Firth, D. (2012). Bradley-Terry models in R: The BradleyTerry2 package. *J. Statist. Software* **48**, 1-21.

Zeileis, A. and Hornik, K. (2007). Generalized M-fluctuation tests for parameter instability. *Statist. Neerlandica* **61**, 488-508.

Zeileis, A., Hothorn, T. and Hornik, K. (2008). Model-based recursive partitioning. *J. Comput. Graph. Statist.* **17**, 492-514.

Department of Mathematics, National Chung Cheng University, Chiayi 621, Taiwan.

E-mail: lksxdd@gmail.com

Department of Mathematics, National Chung Cheng University, Chiayi 621, Taiwan.

E-mail: yshih@math.ccu.edu.tw

(Received May 2014; accepted December 2014)