Chapter 10
Decision Tree Models for Ranking Data

A number of models for ranking data were introduced in Chaps. 8 and 9. However, not all of these models are designed to incorporate individual/object-specific covariates. Distance-based models discussed in Sect. 8.3 are typical examples of ranking models that are not presently designed to incorporate covariates. As these models generally assume a homogeneous population of individuals, they always give the same predicted ranking. Order statistics models discussed in Sect. 8.3 and Chap. 9 are typical examples of models that are able to incorporate covariates in a "linear model" form. However, there are only a few diagnostic procedures available to determine whether a satisfactory model is found. For instance, is it necessary to transform some of the covariates? Which variables or interaction terms should be included into the model?

For those ranking models that are able to incorporate covariates, it will be difficult to interpret the coefficients of the fitted models if nonlinearity or higher-order interactions are present. For example, Holland and Wessells (1998) applied a rank-ordered logit model with more than 20 interaction terms to predict consumer preferences for fresh salmon. They reported that the model performance has greatly improved after including the interactions but at the same time they mentioned that interpretation of the coefficients is less clear.

The use of decision trees can provide a powerful nonparametric model capable of automatically detecting nonlinear and interaction effects. This could serve as a complement to existing parametric models for ranking data. Decision trees are so called because they can be constructed by a set of rules displayed in a treelike structure. Figure 10.1 exhibits a decision tree with 3 leaf (or terminal) nodes. Since the resulting trees are easy to interpret and provide insight into the data structure, they have been popularly used for classification and regression problems by statisticians, machine learning researchers, and many other data analysts.

© Springer Science+Business Media New York 2014

M. Alvo, P.L.H. Yu, *Statistical Methods for Ranking Data*, Frontiers in Probability and the Statistical Sciences, DOI 10.1007/978-1-4939-1471-5_10

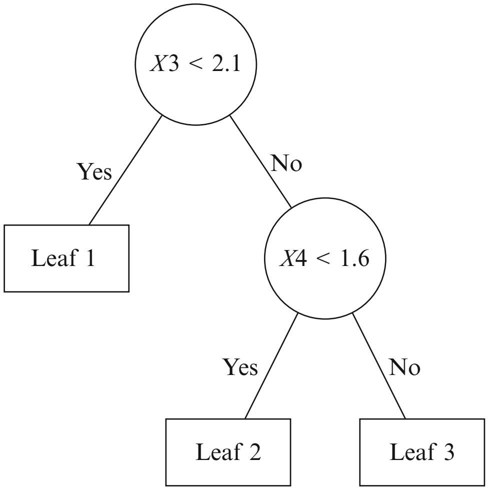

Fig. 10.1 A hypothetical decision tree

In the literature, there are many variants of tree-construction methods such as CART (Breiman et al. 1984) and C4.5 (Quinlan 1992). Many of these decision tree models are constructed in a top-down manner: starting at the root node (the entire training data set) and recursively partitioning the data into two or more child nodes in such a way that each new generation of nodes has better performance than its parent node. The most important step in tree construction is to select the best split for each internal node according to a certain splitting criterion. One approach is to search for the best split based on an "impurity" function (*impurity function approach*). An impurity function defined for each node measures the degree of impurity of the node. The most frequently used impurity functions are the entropy and the Gini index. An alternative approach to do splitting is to apply a statistical test of homogeneity to test whether the split can make the child nodes with significant different distributions of the data (*statistical test approach*). Common statistical tests are the chi-square test and likelihood ratio test for independence in a two-way contingency table. Generally speaking, construction of a decision tree comprises two stages: *tree growing* and *tree pruning*. See Appendix C for a detailed review of decision trees.

Many popular statistical and data mining software such as SAS/Enterprise Miner, Salford Predictive Modeler, and R provide modeling tools for building decision trees for discrete choice data but not for ranking data. In this chapter, we will introduce recent works by Yu et al. (2010) and Wan (2011) which developed decision tree models based on impurity function and statistical test approaches, respectively.

10.1 Impurity Function Approach

In this section, we describe a methodology for constructing a decision tree for ranking data based on the impurity function approach. First of all, the ranking data is randomly partitioned into a training set and a testing set. Following the idea of the CART method, Yu et al. (2010) developed a decision tree algorithm which consists of two stages:

- Tree growing: starting from the root node (the training set), recursively partition each node to identify the best split according to the impurity function until some split-stopping criteria are met. Once tree growing is stopped, a tree is built.
- Tree pruning: using the tree found in the tree-growing stage, find the best subtree by removing branches that does not show any significant improvement in a cost-complexity measure based on a ten-fold cross-validation.

10.1.1 Building Decision Tree for Ranking Data

10.1.1.1 Tree Growing

In growing a tree, the most important step is to search for the best splitting rule in each node based on an impurity function. Here, we will introduce four impurity functions designed for ranking data. Suppose we are given an internal node τ which contains a data set of rankings of t objects. We first define some notations.

Definition 10.1. Let $\boldsymbol{\pi}^{t,m}$ be a set of all possible rankings of m objects from the t objects. Let $\boldsymbol{\pi}^{t,m}_{\{a_1,a_2,\cdots,a_m\}}$ be a set of all possible rankings of the m objects $\{a_1, a_2, \cdots, a_m\}$ from the t objects. Let $\boldsymbol{\Omega}^t_m$ be a collection of all possible subsets of m objects from the t objects.

For example, we have $t = 4$ objects: $\{1, 2, 3, 4\}$. Then we have

$$\boldsymbol{\pi}^{4,2} = \{(1,2), (2,1), (1,3), (3,1), (1,4), (4,1), (2,3), (3,2), (2,4), (4,2), (3,4), (4,3)\}$$

$$\boldsymbol{\pi}^{4,2}_{\{1,2\}} = \{(1,2), (2,1)\}$$

$$\boldsymbol{\Omega}^4_2 = \{\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}\}.$$

Definition 10.2 (Top q notion). For $\boldsymbol{r} = (a_1, a_2, \cdots, a_q) \in \boldsymbol{\pi}^{t,q}$, let $p_T(\boldsymbol{r} \mid \tau)$ be the proportion of individuals in node τ who rank object a_1 the first, object a_2 the second, and so on until object a_q the qth. Ranks of the remaining $t - q$ objects are not considered.

Definition 10.3 (*m*-wise notion). For $\boldsymbol{r} = (a_1, a_2, \cdots, a_m) \in \boldsymbol{\pi}^{t,m}_{B_m}$, where $B_m \in \boldsymbol{\Omega}^t_m$, let $p_W(\boldsymbol{r} \mid \tau)$ be the proportion of individuals in node τ who rank object a_1 higher than object a_2, object a_2 higher than object a_3, and so on until object a_{m-1} higher than object a_m. Objects other than $a_1, a_2, \ldots, a_n$ are not taken into consideration.

In Appendix C.2.1 impurity functions for unordered categorical responses are described. We provide an extension of the Gini and entropy impurity functions to deal with ranking data.

Given a ranking data set of t objects, the extended Gini and entropy developed using the top q and m-wise notions are defined as follows:

$$\text{Top-}q\text{ Gini:} \qquad i_T^{(q)}(\tau) = 1 - \sum_{r \in \boldsymbol{\pi}^{t,q}} [p_T(r \mid \tau)]^2 \tag{10.1}$$

$$\text{Top-}q\text{ entropy:} \qquad i_T^{(q)}(\tau) = - \sum_{r \in \boldsymbol{\pi}^{t,q}} p_T(r \mid \tau) \log_2 p_T(r \mid \tau) \tag{10.2}$$

$$m\text{-wise Gini:} \qquad i_W^{(m)}(\tau) = \frac{1}{C^t_m} \sum_{\boldsymbol{B}_m \in \boldsymbol{\Omega}^t_m} \left(1 - \sum_{r \in \boldsymbol{\pi}^{t,m}_{Bm}} [p_W(r \mid \tau)]^2 \right) \tag{10.3}$$

$$m\text{-wise entropy:} \qquad i_W^{(m)}(\tau) = \frac{-1}{C^t_m} \left(\sum_{B_m \in \boldsymbol{\Omega}^t_m} \sum_{r \in \boldsymbol{\pi}^{t,m}_{Bm}} p_W(r \mid \tau) \log_2 p_W(r \mid \tau) \right) \tag{10.4}$$

The normalizing term $1/C^t_m$ is to bound $i_W^{(m)}(\tau)$ in the range of 0 and 1.

Given an impurity function $i(\tau)$ (one of the above measures), we define the goodness of (binary) split s for node τ, denoted by $\triangle i(s, \tau)$, as

$$\triangle i(s, \tau) = i(\tau) - p_L i(\tau_L) - p_R i(\tau_R).$$

$\triangle i(s, \tau)$ is the difference between the impurity measure for node τ and the weighted sum of the impurity measures for the left child and the right child nodes. The weights, p_L and p_R, are the proportions of the samples in node τ that go to the left node τ_L and the right node τ_R, respectively.

By going through all the possible splits, the best split of node τ is the one with the largest goodness of split $\triangle i(s, \tau)$. The node splitting will continue until the node size is less than a prespecified minimum node size value. In our application, the minimum node size is set to be one-tenth of the size of the training set.

10.1.1.2 Tree Pruning

After the tree is fully constructed, we proceed to the tree pruning stage where the optimal subtree is determined to improve the prediction accuracy. The basic idea is to make use of the minimal cost-complexity algorithm described in Appendix C.2.2 with ten-fold cross-validation to obtain the final tree that minimizes the misclassification cost. See Yu et al. (2010) for details of the choice of misclassification cost and the implementation procedure of tree pruning.

10.1.2 Leaf Assignment

Various approaches are proposed to make the assignment for every leaf node:

1. Calculate the mean rank of each object and deduce the predicted ranking by ordering the mean ranks.
2. Calculate the top-choice frequency of each object and decide the predicted ranking by ordering the frequency.
3. Use the most frequently observed ranking to represent the predicted ranking.
4. Look at the paired comparison probabilities of each object pair or the top-5 most frequently observed ranking responses.

The first three approaches reveal the predicted ranking of the objects. However, in some situations, the predicted rankings are not of primary concern. Instead, it is of interest to investigate the importance of the covariates to the rank order preference. For this kind of exploration, method 4 provides a more general idea of how the preference orders are distributed within a leaf node. In the following, we advocate a nonparametric procedure to examine the differences among the mean ranks of objects in each leaf node.

10.1.2.1 Nonparametric Inference of the Leaf Nodes

Given a fitted tree, the training data is partitioned into a number of relatively more homogeneous leaf nodes. In this case, we would like to understand the preference of the objects in each leaf node. It is thus of interest to examine the rank order difference among the objects so that a clearer picture can be obtained in interpreting each leaf node.

The first thing we should consider is to test for randomness on ranking of t objects in each leaf node. This is equivalent to apply the Friedman test to the ranking data in each leaf node. Under the null hypothesis H_0 of randomness in the Friedman test, the Friedman test statistic Q_χ, corrected for the observed ties, in leaf node τ with n_τ individuals is given by

$$Q_\chi = \frac{(t-1)\sum_{i=1}^{t}[r_i - n_\tau(t+1)/2]^2}{\sum_{j=1}^{n_\tau}\sum_{i=1}^{t} r_{ij}^2 - n_\tau t(t+1)^2/4},$$

where r_i is the sum of the ranks r_{ij} for object i. When the data contains no ties, Q_χ becomes the test statistic used in Sect. 2.3.

Iman and Davenport (1980) argued that the chi-square approximation of Q_χ may be undesirably conservative and suggested using the following F distribution approximation:

$$Q_F = \frac{(n_\tau - 1)Q_\chi}{n_\tau(t-1) - Q_\chi}.$$

Under H_0, Q_F follows asymptotically an F distribution with $(t-1)$ and $(n_\tau - 1)(t-1)$ degrees of freedom.

Upon rejection of H_0 in the Friedman test, it is possible to identify the difference between specific pairs of objects by the multiple comparison procedure (Conover 1999). Preferences on objects a and b are significantly different at the significance level α if the following inequality is satisfied:

$$|r_a - r_b| > t_{1-\alpha/2}\sqrt{\frac{2n_\tau\left(\sum_{i=1}^{n_\tau}\sum_{j=1}^{t} r_{ij}^2 - \sum_{j=1}^{t} r_{j/n_\tau}^2\right)}{(n_\tau - 1)(t-1)}}$$

where $t_{1-\alpha/2}$ is the value from the student-t distribution with $(n_\tau - 1)(t-1)$ degrees of freedom.

10.1.3 Performance Assessment of Decision Tree for Ranking Data

Another important issue that should be addressed is the performance assessment of the fitted decision tree. The most frequently used performance measure is misclassification rate. However, this is not a good performance measure for assessing predictive accuracy of the fitted tree because a predicted ranking can only be classified either correctly or incorrectly, overlooking the fact that the predicted ranking can be partially agreed with the observed ranking. That means some objects in the rank permutation, but not all, are in the correct ordered position.

A widely used single measure for evaluating the overall performance of a binary classifier is the area under the receiver operating characteristic (ROC) curve. It is simple and attractive because it is not susceptible to the threshold choice and it is regardless of the costs of the different kinds of misclassification and class priors

(Bradley 1997; Hand and Till 2001) . The value of AUC always falls within [0.5, 1.0] – it equals 0.5 when the instances are predicted at random and equals 1.0 for perfect accuracy.

However, standard ROC curve can be used for binary data only. Hand and Till (2001) extended it to discrete choice data and Yu et al. (2010) further generalized it to ranking data. The general idea is to first convert the observed and predicted rankings in a testing set into many binary choices with each comparing preference between a pair of objects. Using the observed and predicted binary choice outcomes for each pair of objects, an ROC curve can be drawn and the corresponding AUC can be computed using the standard procedure. Finally, the AUC of the fitted decision tree for ranking data can be obtained by taking the average of the AUCs for all pairs of objects.

10.1.4 Analysis of 1993 European Value Priority Data

The ranking data set was obtained from the International Social Service Programme (ISSP) in 1993 (Jowell et al. 1993), which is a continuing, annual program of cross-national collaboration on surveys covering a wide spectrum of topics for social science research. The survey was conducted using standardized questionnaire in 1993 at 20 countries around the world, such as Great Britain, Australia, the USA, Bulgaria, the Philippines, Israel, and Spain. It mainly focused on value orientations, attitudes, beliefs, and knowledge concerning nature and environmental issues and included the so-called Inglehart Index, a collection of four indicators of materialism/post-materialism as well. Respondents were asked to pick the most important (rank “1") and the second most important (rank “2") goals for their government from the following four alternatives:

1. Maintain order in nation (ORDER).
2. Give people more to say in government decisions (SAY).
3. Fight rising prices (PRICES).
4. Protect freedom of speech (SPEECH).

After removing those invalid responses, the survey gave a ranked data set of 5,737 observations with top choice and top two rankings. In addition, the data provide some judge-specific characteristics and they are applied in tree partitioning. The candidate splitting variables are summarized in Table 10.1.

Respondents can be classified into value priority groups on the basis of their top two choices among the four goals. “Materialist” corresponds to an individual who gives priority to ORDER and PRICES regardless of the ordering, whereas those who choose SAY and SPEECH will be termed “post-materialist.” The last category consists of judges giving all the other combinations of rankings, and they will be classified as holding “mixed” value orientations.

Inglehart’s thesis of generational based values has been influential in political science since the early 1970s. He has argued that value priorities were shifting

Table 10.1 Description of 1993 EVP data

Covariate	Description/Code	Type
Country	West Germany = 1, East Germany = 2, Great Britain = 3, Italy = 4, Poland = 5	Nominal
Gender	Male = 1, female = 2	Binary
Education	0–10 years = 1, 11–13 years = 2, 14 or more years = 3	Ordinal
Age	Value ranges from 15 to 91	Interval
Religion	Catholic and Greek Catholic=1, Protestant = 2, others = 3, none = 4	Nominal

Table 10.2 Summary of the best pruned subtrees by four impurity measures

Method	Avg. AUC	SE	AUC	No. of leaves	Depth
Top-2 entropy	0.61947	0.0056	0.62951	12	5
Pairwise Gini	0.61896	0.0058	0.62902	12	5
Pairwise entropy	0.61857	0.0056	0.62709	11	5
Top-2 Gini	0.61425	0.0063	0.61931	9	4

profoundly in economically developed Western countries, from concern over sustenance and safety needs toward quality of life and freedom of self-expression, thus from a materialist orientation to a post-materialist orientation. In this analysis, we study the Inglehart hypothesis in five European countries by our decision tree approach, which helps to identify the attributes that affect Europeans' value priority.

The data are divided randomly into 2 sets, 70 % to the training set for growing the initial tree and finding the best pruned subtree for each of the four splitting criteria; and 30 % to the testing set for performance assessment and selection of the splitting criterion to build the final tree.

As a decision tree is an unstable classifier, small changes in the training set can cause major changes in the fitted tree structure; we therefore repeat this procedure 50 times and compare the four splitting criteria with their averaged AUC. The final tree model is created using the entire data set for interpretation. Notice that the testing set is not involved in the tree building process and pruned subtree selection. The four splitting criteria for rankings include top-2 and pairwise measures of Gini and entropy.

The second and third columns of Table 10.2 show the averaged AUC and their standard error of the best pruned subtrees for each splitting criterion based on 50 repetitions. The tree structure and performance of the final models are also presented. Figure 10.2 displays the six ROC curves of each object pairs that arise from the top-2 entropy tree. The tree did a better job of predicting the object pair "SAY vs PRICES", but poor for "SAY vs SPEECH." The performance of the four trees is comparable and it is hard to distinguish them in the graph.

The four tree models are found to have similar node partitions. The root node is split according to whether the judges came from Poland or not. At the second

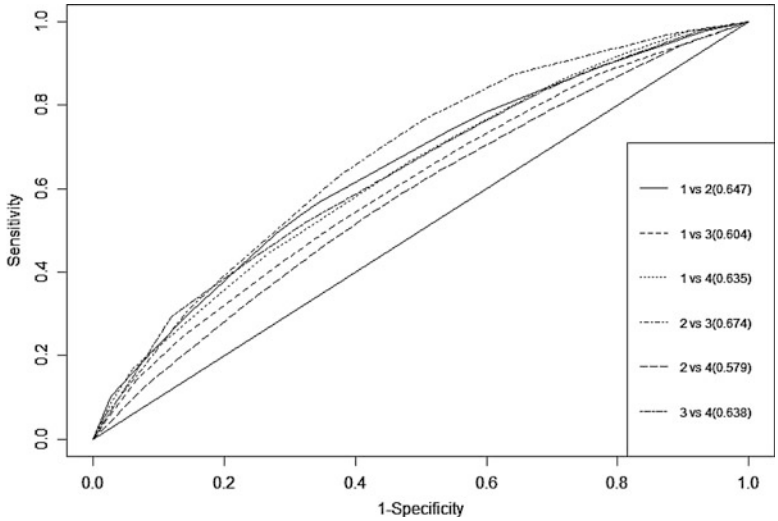

Fig. 10.2 ROC curves of top-2 entropy tree
Remark: The four value objects are coded as follows 1 = [ORDER], 2 = [SAY], 3 = [PRICES] and 4 = [SPEECH]. The 45° diagonal line connecting (0,0) and (1,1) is the ROC curve corresponding to random chance. Given next to the legends are the areas under the corresponding dashed ROC curves

level, the splits are based on age. For Polish, the respondents are divided with the rule "age<59?", while the remaining judges are split according to age < 53 or not. Further partitions involve education level, country, and age. The factors religion and gender do not seem to be influential. It is observed that in the learning phase, top-2 Gini tends to give a smaller tree, while top-2 entropy gives a more complicated tree on average. Based on the assessment criterion, the top-2 entropy tree is chosen as the best model and it is applied for further analysis.

The tree with 5 levels of depth and 12 leaves is sketched in Figure 10.3. For the sake of brevity, we do not show the other three tree structures. A summary of the leaf nodes of the final tree is reported in two tables. Table 10.3 shows the individuals' value priority, the three most frequent top two rankings, together with the proportion of six pairs of political goals in each leaf node whereas Table 10.4 lists the mean rank of the four political goals. In order to determine if the observed rankings in the leaf node imply statistically significant differences across alternatives, the Friedman test is used, and highly significant results are obtained in all leaves, indicating that the respondents had different priority to at least one political goal. The post hoc multiple comparison procedures are thus further performed and based on the results, the rankings of the four goals are deduced (see Table 10.4).

We now turn to examine the covariate and interaction effects based on the final tree model. In Poland, individuals were more likely to favor materialistic objects ORDER (in leaves 5, 8, and 9) and PRICES (in leaves 5 and 8) than the other two post-materialistic objects. In East Germany, judges appeared to support ORDER and

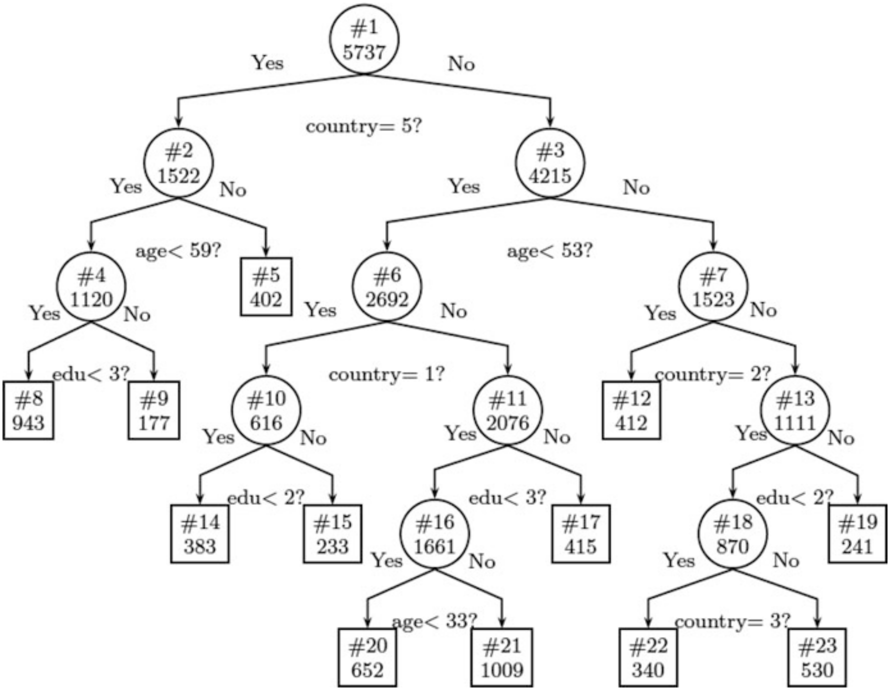

Fig. 10.3 Fitted tree based on top-2 entropy
Remark: In each node, the node ID and the number of judges are shown. The splitting rule is given under the node. The abbreviation "edu" stands for the variable education

SAY more; particularly those older generations gave higher priority to ORDER (in leaf 12). Respondents of West Germany showed stronger emphasis on SAY. Those better educated West Germans were more post-materialist than the lower educated ones as they preferred SAY and SPEECH, rather than the other two materialist objects (in leaf 15). Mixed value orientations were anchored in British because all the related leaf nodes give us a preference prediction of ORDER $\succ$ SAY or SAY $\succ$ ORDER.

Summarizing, we note that:

(i) Despite some cross-national differences, our findings do not deviate much from Inglehart's theory, which claimed that societies embrace post-materialistic values as they move toward more economic security and affluence. The older European generations experienced economic and social insecurity in their preadult years during World War II. They thus attached more importance to materialistic values compared to younger cohorts. Younger postwar generations developed post-materialist values as they grew up during periods of relative prosperity.
(ii) There is a clear tendency in each country for the higher educated to be the more post-materialistic groups. Duch and Taylor (1993) stated that the

Table 10.3 Value priority, frequent rankings, and pairwise probabilities in leaf nodes of top-2 entropy tree

			Frequent top two ranking			Pairwise probabilities					
Node(τ)	Node Size	Value	1st	2nd	3rd	$p_w(1,2\|\tau)$ (%)	$p_w(1,3\|\tau)$ (%)	$p_w(1,4\|\tau)$ (%)	$p_w(2,3\|\tau)$ (%)	$p_w(2,4\|\tau)$ (%)	$p_w(3,4\|\tau)$ (%)
5	402	M	1,3 (40.3%)	3,1 (24.9%)	1,2 (7.5%)	83.1	60.8	87.3	21.0	53.7	83.1
8	943	M	3,1 (25.7%)	1,3 (22.5%)	3,2 (13.0%)	65.3	45.9	79.3	31.0	64.8	82.1
9	177	M	1,3 (17.3%)	3,1 (13.7%)	1,2 (12.2%)	58.8	55.9	69.2	47.7	61.9	63.8
12	412	B	1,3 (27.5%)	1,2 (22.2%)	2,1 (18.2%)	67.5	77.7	90.0	53.9	72.0	68.3
14	383	B	2,1 (17.0%)	2,3 (12.3%)	2,4 (12.0%)	27.4	32.1	41.0	37.6	40.1	36.0
15	233	P	2,4 (25.9%)	4,2 (15.7%)	2,3 (11.7%)	44.1	57.8	64.8	63.6	69.5	57.3
17	415	B	2,4 (15.7%)	2,1 (14.9%)	1,2 (14.2%)	44.5	63.1	61.0	68.3	71.1	51.0
19	241	B	1,3 (22.1%)	2,1 (14.9%)	1,2 (13.8%)	60.8	76.3	74.5	58.3	62.2	51.9
20	652	B	2,1 (20.1%)	2,3 (19.1%)	1,2 (13.4%)	41.6	59.7	70.6	68.9	77.3	61.5
21	1009	B	1,3 (18.0%)	1,2 (16.1%)	2,1 (15.8%)	52.6	64.5	74.1	60.0	72.8	62.6
22	340	B	1,3 (18.1%)	2,3 (15.9%)	1,2 (12.3%)	51.3	55.6	71.2	52.8	68.2	67.9
23	530	M	1,3 (28.2%)	1,2 (16.2%)	3,1 (9.6%)	69.0	69.1	78.7	45.7	61.8	66.8

Remark: The four political goals are labeled as 1=[ORDER], 2 = [SAY], 3 = [PRICES], and 4 = [SPEECH]. In column 3, "Value" shows the value priority group of judges in each leaf node, where B = mixed values, M = materialist, and P = post-materialist. In columns 4–6, "i, j" implies goal $i \succ$ goal j and the percentage beside indicates the proportion of instances having the corresponding top two ranking in node τ

Table 10.4 Result of Friedman test and multiple comparison procedures in leaf nodes of top-2 entropy tree

Node (τ)	Node Size	Mean rank				Friedman test statistic	Multiple comparison (mean rank difference) †						Goal priority#
		ORDER	SAY	PRICES	SPEECH		ORDER vs SAY	ORDER vs PRICES	ORDER vs SPEECH	SAY vs PRICES	SAY vs SPEECH	PRICES vs SPEECH	
5	402	1.69	3.08	1.99	3.24	285.8**	−1.40	−0.30	−1.55	1.10	−0.16	−1.25	1≻3≻2≻4
8	943	2.10	2.70	1.95	3.26	318.6**	−0.60	0.15	−1.17	0.75	−0.57	−1.31	1≻3≻2≻4
9	177	2.16	2.49	2.40	2.95	14.0**	−0.33	–	−0.79	–	−0.46	−0.55	1≻2≻4; 3≻4
12	412	1.65	2.42	2.63	3.30	188.6**	−0.77	−0.98	−1.66	−0.22	−0.89	−0.67	1≻2≻3≻4
14	383	2.33	2.11	2.64	2.92	35.0**	0.22	−0.31	−0.58	−0.53	−0.80	−0.27	2≻1≻3≻4
15	233	2.73	1.86	2.97	2.44	42.4**	0.86	−0.24	−0.29	−1.11	−0.58	0.53	2≻4≻1≻3
17	415	2.31	2.05	2.80	2.83	45.4**	0.26	−0.49	−0.52	−0.75	−0.78	–	2≻1≻3, 4
19	241	1.88	2.40	2.83	2.89	40.5**	−0.52	−0.94	−1.00	−0.43	−0.48	–	1≻2≻3, 4
20	652	2.28	1.95	2.67	3.09	127.8**	0.33	−0.39	−0.81	−0.72	−1.14	−0.42	2≻1≻3≻4
21	1009	2.09	2.20	2.62	3.10	165.9**	−0.11	−0.53	−1.01	−0.42	−0.90	−0.48	1≻2≻3≻4
22	340	2.22	2.30	2.40	3.07	39.5**	–	−0.19	−0.85	–	−0.77	−0.67	1≻3≻4; 2≻4
23	530	1.83	2.62	2.48	3.07	114.0**	−0.78	−0.65	−1.24	0.14	−0.46	−0.59	1≻3≻2≻4

Remark: ** The test is significant at 0.01 level
† Only the mean rank differences of those pairwise multiple comparison procedures at $p < 0.05$ were shown
# In the last column, the codes 1–4 represent one of the political goals: 1 = [ORDER], 2 = [SAY], 3 = [PRICES], and 4 = [SPEECH]

post-materialistic objects tap certain fundamental democratic values, such as liberty and rights consciousness. The better educated would have had more opportunity to learn to appreciate such principles, and thus they will prefer post-materialistic objects more.

10.2 Statistical Test Approach Based on Intergroup Concordance

In Sect. 10.1, a decision tree for ranking data was constructed in such a way that the leaf nodes are as pure as possible according to a certain impurity measure. In other words, the node splitting in the impurity function approach aims to find the best split such that the two resulting child nodes are as homogeneous as possible. However, this does not guarantee that the rankings of objects between the two child nodes are significantly different.

As mentioned earlier, there is a statistical test approach which can provide an alternative splitting measure based on a test for intergroup concordance on rankings for the selection of the best split during the tree-growing stage. In Sect. 4.2, we have seen several tests for testing for agreement or concordance between two or more groups in ranking a set of objects. Therefore, it is possible to apply those tests in constructing decision tree for ranking data.

10.2.1 Building Decision Tree Using Test for Intergroup Concordance

To construct a decision tree based on the statistical test of intergroup concordance, we follow the same methodology used in the impurity function approach in Sect. 10.1 except for the choice of splitting criterion. For the splitting criterion, the best splitting rule of a node is the one that maximizes the test statistic for testing concordance between two child nodes.

Here, we will make use of two tests of concordance based on Spearman and Kendall statistics described in Sect. 4.2. Note that the combined estimates of the covariance function is used throughout. The notation SP refers to the Spearman statistic and KW to the Kendall statistic.

The partitioning process stops when further splitting does not lead to a statistically significant result. The tree size can then be controlled by setting a threshold significance level on the test procedures. Lower threshold values tend to produce smaller trees.

After the tree is fully constructed, the cost-complexity pruning procedure is executed to avoid the problem of overfitting the ranking data. Most steps are the same as the one used in Sect. 10.1 except that an alternative cost function based on Spearman Footrule distance is considered. For the details of this pruning procedure, see Wan (2011).

10.2.2 Analysis of US General Social Survey Data on Job Value Preference

The general social survey (GSS) has been conducted in US by the National Opinion Research Center (NORC) of the University of Chicago annually since 1972 (except for the years 1979 and 1981) and biennially since 1994 (Davis and Smith 2009). Each year the GSS consisted of a 90-min in-person interview with a full-probability sample of English- or Spanish-speaking persons aged 18 years or above who lived in households. It is a multidimensional social survey that gathers sociodemographic characteristics and replicated core measurements on social and political attitudes and behaviors, plus topics of special interest. Many of the core questions have remained unchanged since 1972 to facilitate time-trend studies as well as replication of earlier findings. In relation to job value, the characteristics were measured by the GSS in an ipsative approach. The respondents were asked to rank in order of preference from (i) "most preferred," (ii) "second most important," to (v) "fifth most important" five aspects about a job:

1. High income (JINC)
2. No danger of being fired (JSEC)
3. Working hours are short, lots of free time (JHOUR)
4. Chances for advancement (JPRO)
5. Work important and gives a feeling of accomplishment (JMEAN)

Job values, as defined by Kalleberg (1977), are what individuals hold as desirable with respect to their work activity and the attitudes are central to the social psychology of work. Under many other names (including work/occupational value/attribute/characteristic), they refer to the importance people place on occupational rewards and play a key role in conditioning a range of work-related outcomes, such as job satisfaction and commitment, work centrality, and occupational choice and stability. Theoretically, the perceived job attributes have been conceptualized into two value dimensions, either *intrinsic* or *extrinsic*. Intrinsic values concern the rewards emanating directly from the work activity and experience itself (e.g., job autonomy, challenge, use of abilities, expression of interest and creativity, workplace cooperation, job useful to society). In contrast, extrinsic values involve the rewards derived from the job but external to the work itself (e.g., job security, pay, fringe benefit, prestige, promotional opportunities, pleasant working environment, good hours, no excessive amount of works). Among the five work values listed in the GSS, the first four attributes (i)–(iv) represent extrinsic factors of the job, while the last value (v) is considered intrinsic.

An ongoing interest among researchers in work value preference has been witnessed over the decades. Previous studies have shown that work values are not externally given, but rather come as a result of socialization processes throughout an individual's life (Johnson 2002; Mortimer and Lorence 1979). They are developed initially as a function of parents' social origin and socioeconomic positions, during schooling and the early years of work. Past research has consisted of comparing and explaining the differences in job value preferences between white and blue collar workers (Weaver 1975), male and female (Lacy et al. 1983), older and younger generations (Loscocco and Kalleberg 1988), as well as blacks and whites (Martin and Tuch 1993). In the following section, we exploit the data from the General Social Survey to revisit the priority of occupational values in the US using the proposed tree model.

This study utilized data from three samples ($N = 3744$) collected in 1973, 1985, and 2006, in order to examine the role of social class origins and socioeconomic characteristics in shaping one's job value orientation. Table 10.5 shows the eight individual attributes (sex, race, birth cohort, highest educational degree attained, family income, marital status, number of children that the respondent ever had, and household size) and three properties of work conditions (working status, employment status, and occupation) that are involved in the model building process. The attribute "year," referring to the year of survey, is also included to address the question of changing work values over time.

Complete rankings of the five job characteristics were obtained from the entire sample. Summarized results of their preferences by years are provided in Table 10.6. The pattern is remarkably consistent over the 30-year period. Meaningful work (JMEAN) was far more important than any other value, taking over 40 % of the top rank order in each year of survey. The next two attributes, high income (JINC) and having opportunities for advancement (JPROMO), were fairly close to each other in importance. Less than 5 % of the respondents regarded short working hours and more leisure time (JHOUR) as the most important value, while 51.3 % placed it as the least important.

Next, the ranking data are analyzed using the decision tree model described earlier. Following the methodology presented in Sect. 10.1.4, we randomly divide the data into 2 parts: (1) the learning set constituting 70 % of the data to grow the initial tree with a threshold significance level of 0.5 and search the best pruned subtree for each of the test statistic and (2) the testing set containing the remaining data to evaluate the tree performance and select the best splitting measure to build the final tree. Note that the entire sample will be included to produce the final model.

Preliminary attempts at learning from the data are not promising because the overall value placed on short working hours in all leaf nodes is found to be the least important to people than meaningful work, high income, and advancement opportunities. Given the limited significance of "JHOUR," we removed it from further analysis and reduced the number of ranked objects to four. The performance of the best pruned subtree for each statistical measure is reported in columns 2 and 3 of Table 10.7. The averaged AUC and the standard error obtained over 50 replications do not differ much among the measures. Indeed, the trees developed have similar structure; the splitting rules at the first, second, and third level are found to be the same. For this reason, we restrict attention to the SP tree model in the coming discussion.

Figure 10.4 depicts the ROC curves of six object pairs that arise from the SP tree. The predictive performance of the classifier is found to be superior on the object pairs "JMEAN vs JSEC" and "JMEAN vs JINC" and inferior on the object pairs

Table 10.5 Description of US general social survey data of job value preference

Covariate	Description	Code	Type
Year	Year of survey	1973, 1985, 2006	Nominal
Sex	Sex	1 = male, 2 = female	Binary
Race	Race	1 = white, 2 = black, 3 = others	Nominal
Cohort	Birth cohort (actual year of birth)	1 = 1883–1889, 2 = 1890–1899, 3 = 1900–1909, 4 = 1910–1919, 5 = 1920–1929, 6 = 1930–1939, 7 = 1940–1949, 8 = 1950–1959, 9 = 1960–1969, 10 = 1970–1979, 11 = 1980–1989	Ordinal
Educ	Highest educational degree attained	0 = less than high school, 1 = high school, 2 = associate/junior college, 3 = bachelor's, 4 = graduate	Nominal
Finc	Family income	1 = \$1–999, 2 = \$1,000–1,999, 3 = \$2,000–2,999, 4 = \$3,000–3,999, 5 = \$4,000–4,999, 6 = \$5,000–5,999, 7 = \$6,000–6,999, 8 = \$7,000–7,999, 9 = \$8,000–8,999, 10 = \$9,000–9,999, 11 = \$10,000–10,999, 12 = \$11,000–11,999, 13 = \$12,000–14,999, 14 = \$15,000–19,999, 15 = \$20,000–24,999, 16 = \$25,000–29,999, 17 = \$30,000–39,999, 18 = \$40,000–49,999, 19 = \$50,000–74,999, 20 = \$75,000–99,999, 21 = >\$100,000	Ordinal
Marital	Marital status	1 = married, 2 = widowed, 3 = divorced, 4 = separated, 5 = never married	Nominal
Child	No. of children ever had	0, 1, 2, 3, 4, 5, 6, 7, 8, or more	Ordinal
Hsize	Household size	Value ranges from 1 to 16	Interval
Wkstat	Working status	1 = working full time, 2 = working part time, 3 = with a job, but not at work because of temporary illness, vacation, and strike, 4 = unemployed, laid off, looking for work, 5 = retired, 6 = in school, 7 = keeping house, 8 = others	Nominal
Employ	Employment status	1 = self-employed, 2 = someone else	Binary
Occ	Occupation	1 = professional, technical, and related workers, 2 = administrative and managerial workers, 3 = clerical and related workers, 4 = sales workers, 5 = service workers, 6 = agriculture, animal husbandry and forestry workers, fishermen, 7 = production and related workers, transport, equipment operators, and laborers	Nominal

Table 10.6 Importance of five job values in the US General Social Survey

Year	JMEAN	JINC	JPRO	JSEC	JHOUR	Sample size
	Top choice					
1973	620 (53.7%)	221 (18.4%)	223 (19.2%)	77 (6.4%)	59 (4.9%)	1,200
1985	643 (48.8%)	255 (19.4%)	287 (21.8%)	91 (6.9%)	41 (3.1%)	1,317
2006	506 (41.2%)	283 (23.1%)	244 (19.9%)	131 (10.7%)	63 (6.4%)	1,227
Total	1769 (47.2%)	759 (20.3%)	754 (20.1%)	299 (8.0%)	163 (4.4%)	3,744
	Last choice					
1973	90 (7.5%)	79 (6.6%)	115 (9.6%)	359 (29.9%)	557 (46.4%)	1,200
1985	67 (5.1%)	47 (3.6%)	77 (5.8%)	351 (26.7%)	775 (58.8%)	1,317
2006	85 (6.9%)	85 (6.9%)	122 (9.9%)	345 (28.1%)	590 (48.1%)	1,227
Total	242 (6.5%)	211 (5.6%)	314 (8.4%)	1055 (28.2%)	1922 (51.3%)	3,744

Table 10.7 Summary of the best pruned subtrees by two statistical significance measures

Method	Avg. AUC	SE	AUC	No. of leaves	Depth
Spearman	0.65045	0.0056	0.64689	28	9
Kendall	0.64844	0.0062	0.64229	22	12

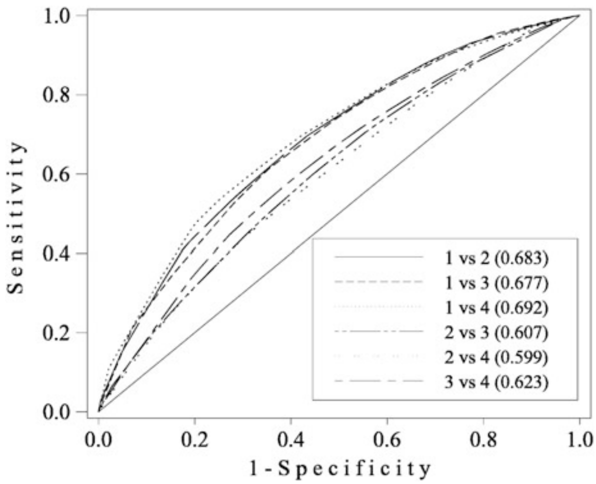

Fig. 10.4 ROC curves of SP tree
Remark: The four value objects are coded as follows: 1 = [JMEAN], 2 = [JINC], 3 = [JPRO], and 4 = [JSEC]. The 45° diagonal line connecting (0,0) and (1,1) is the ROC curve corresponding to random chance. The areas under the corresponding *dashed* ROC curves appear in brackets

"JINC vs JSEC" and "JINC vs JPRO." As displayed in Figure 10.5, the nine-level SP tree has 28 leaf nodes (in square box). Inside each node, the node ID and the number of judges are shown, whereas the splitting rule is given under the node. Race is

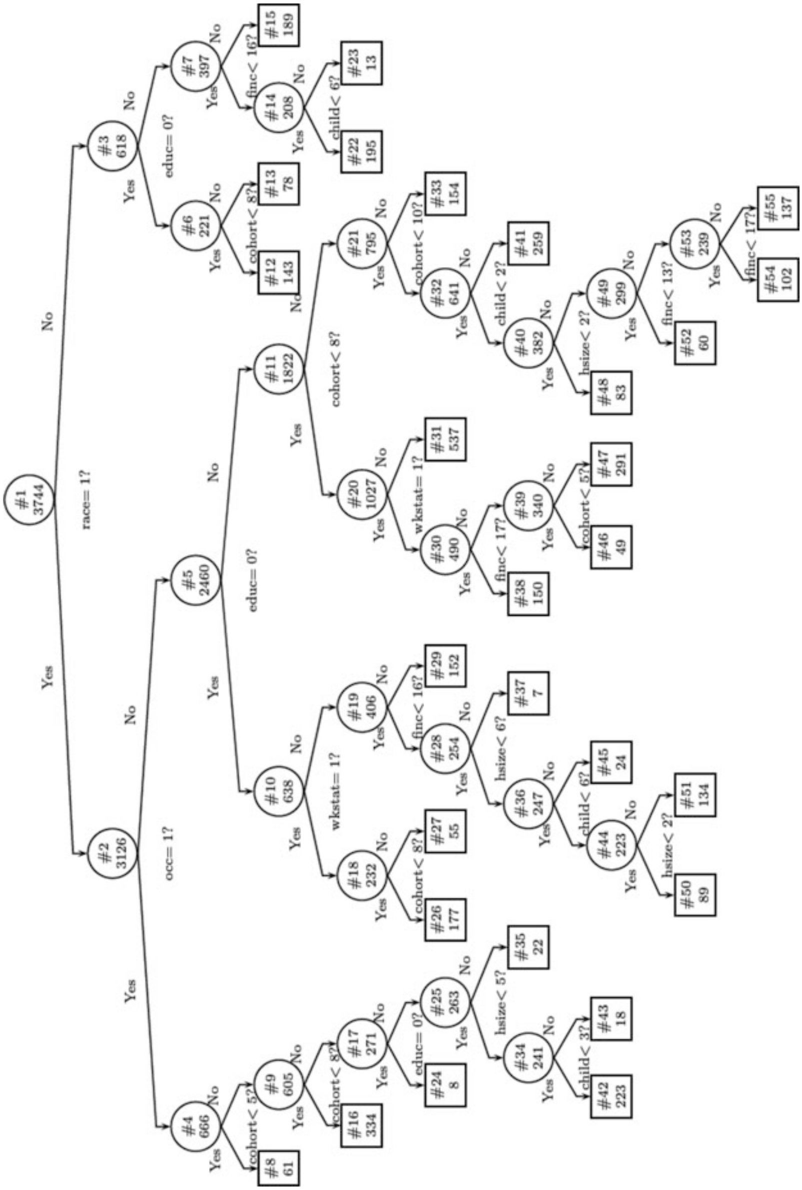

Fig. 10.5 Fitted SP tree

found to be the most important factor, splitting the entire sample into white (race=1) and black (race=2) Americans. At the second level, white Americans are separated based on their occupation (occ=1 vs occ≠1), while for black Americans, they are divided according to their highest education level attained (educ=0 or not). Other predictor variables appearing in the partitions in lower levels include birth cohort, family income, number of children ever had, working status, and household size. Surprisingly, the demographic variable sex has no contribution in the classification problem, while marital status and employment status are found not to be influential as well. In addition, the factor "year" does not appear in any splitting rule. This suggests that after controlling for the personal and work-related characteristics, the work value orientations have not varied over the past few decades.

Leaf statistics of the SP tree are given in Tables 10.8 and 10.9, which present the mean rank, the three most frequent rankings and the pairwise probability of the six object pairs. For every leaf node, the Friedman test is applied to examine if the rank order differences are significant across four job values and the test results are summarized in column 7 of Table 10.9. It is found that all tests performed are statistically significant at 0.1 level, indicating that at least one value tends to be ranked higher. We conducted multiple comparisons between all pairs of values to determine differences in terms of preferences. The induced ranking in the 28 leaf nodes is provided in the last column of Table 10.9.

It appears that black Americans placed more emphasis on the extrinsic rewards than whites, especially among those less educated (in leaf 12) and with lower family income (in leaf 22). These racial differences may be attributed to their persistent disadvantaged economic status and wage gap in the labor market (Martin and Tuch 1993). On the other hand, resulting from anti-discrimination legislation and improvements in the economic and social conditions in the US since the 1980s, a greater emphasis on intrinsic job value was found among the younger black cohort (in leaf 13).

For white Americans, the occupation effect is important. The result indicates that professional, technical, and related workers attach relatively higher importance to the intrinsic value "JMEAN" than workers in other occupational groups (in leaves 8, 16, 35, and 42). Given their greater skills and credentials, they felt more secure in their ability to seek alternative employment and, thus, showed less favor to the value "JSEC." Nevertheless, professionals with more children have less propensity to take risk in the labor market (in leaf 43).

Consistent with a hierarchy of needs perspective, nonprofessional part-time working white Americans who are less educated, earn less, and have a large household size have higher valuation on extrinsic value of job security and high income when compared to their counterparts (in leaf 7). Meanwhile, the younger cohort in their early stages of working careers tended to desire a job with good prospects for advancement (in leaves 13, 24, and 27).

In summary, the study reveals that significant racial disparities in extrinsic and intrinsic job value preferences existed in US throughout the last three decades. The race effect in conjunction with other variables including education level, family income, age, household size, number of children, and working status explains

Table 10.8 Frequent rankings and pairwise probabilities in leaf nodes of SP tree

Node (τ)	Node size	Frequent ranking			Pairwise probabilities					
		1st	2nd	3rd	$p_W(1,2\mid\tau)$ (%)	$p_W(1,3\mid\tau)$ (%)	$p_W(1,4\mid\tau)$ (%)	$p_W(2,3\mid\xi)$ (%)	$p_W(2,4\mid\tau)$ (%)	$p_W(3,4\mid\tau)$ (%)
8	61	1,3,4,2 (19.7%)	1,3,2,4 (16.4%)	1,2,3,4 (14.8%)	91.8	63.9	86.9	23.0	54.1	82.0
12	143	2,4,3,1 (12.6%)	2,4,1,3 (8.4%)	2,3,4,1 (7.7%)	33.6	50.3	52.4	65.7	69.9	54.5
13	78	3,1,2,4 (9.0%)	4,3,1,2 (9.0%)	2,3,1,4 (9.0%)	37.2	23.1	50.0	43.6	64.1	67.9
15	189	1,2,3,4 (14.8%)	1,3,2,4 (14.3%)	2,3,1,4 (9.5%)	57.7	55.6	76.7	48.7	71.4	74.1
16	334	1,3,2,4 (29.0%)	1,2,3,4 (24.3%)	1,3,4,2 (8.4%)	83.2	82.6	94.6	46.1	80.5	83.2
22	195	2,3,1,4 (9.7%)	2,3,4,1 (9.7%)	3,2,1,4 (9.7%)	37.9	31.8	60.0	52.3	79.0	74.9
23	13	1,3,4,2 (23.1%)	3,1,4,2 (15.4%)	3,1,2,4 (15.4%)	76.9	38.5	53.8	15.4	30.8	69.2
24	8	3,1,2,4 (25.0%)	3,2,1,4 (12.5%)	1,3,2,4 (12.5%)	62.5	37.5	75.0	37.5	87.5	87.5
26	177	1,3,2,4 (10.7%)	1,2,3,4 (10.2%)	2,1,3,4 (8.5%)	54.2	59.3	70.6	58.8	69.5	63.8
27	55	3,2,1,4 (16.4%)	2,3,1,4 (10.9%)	2,4,3,1 (7.3%)	29.1	30.9	58.2	49.1	70.9	67.3
29	152	1,3,2,4 (16.4%)	3,1,2,4 (9.9%)	1,2,3,4 (7.9%)	63.2	54.6	75.0	40.8	67.8	72.4
31	537	1,3,2,4 (26.3%)	1,2,3,4 (14.2%)	1,3,4,2 (11.5%)	78.6	69.6	87.5	34.8	70.0	81.8
33	154	2,1,3,4 (9.7%)	1,3,4,2 (9.7%)	3,2,1,4 (9.1%)	48.1	51.3	68.2	50.6	65.6	69.5
35	22	1,3,4,2 (27.3%)	1,2,3,4 (22.7%)	1,3,2,4 (13.6%)	77.3	77.3	86.4	40.9	68.2	66.8
37	7	4,2,1,3 (28.6%)	2,4,3,1 (28.6%)	2,4,1,3 (14.3%)	14.3	57.1	14.3	85.7	42.9	14.3
38	150	1,3,2,4 (14.7%)	1,2,3,4 (10.7%)	1,2,4,3 (8.7%)	65.3	59.3	71.3	48.7	65.3	60.7

41	259	1,3,2,4 (17.4%)	1,2,3,4 (14.3%)	2,1,3,4 (8.1%)	57.1	66.4	76.1	54.8	76.8	71.4
42	223	1,3,2,4 (22.0%)	1,2,3,4 (20.2%)	1,2,4,3 (12.6%)	77.6	83.0	87.4	58.3	75.8	66.8
43	18	1,2,4,3 (22.2%)	1,4,2,3 (22.2%)	1,2,3,4 (16.7%)	88.9	94.4	88.9	72.2	50.0	33.3
45	24	1,3,4,2 (20.8%)	1,3,2,4 (12.5%)	1,4,3,2 (8.3%)	70.8	54.2	70.8	16.7	37.5	66.7
46	49	1,3,2,4 (24.5%)	1,3,4,2 (12.2%)	1,2,4,3 (12.2%)	71.4	73.5	71.4	42.9	65.3	67.3
47	291	1,3,2,4 (23.4%)	1,2,3,4 (18.9%)	2,1,3,4 (8.3%)	67.7	72.2	86.6	51.9	81.4	78.4
48	83	1,3,2,4 (18.1%)	1,2,3,4 (16.9%)	1,2,4,3 (7.2%)	66.3	72.3	79.5	53.0	72.3	65.1
50	89	3,4,1,2 (11.2%)	1,3,4,2 (10.1%)	1,3,2,4 (10.1%)	64.0	47.2	56.2	36.0	48.3	62.9
51	134	1,3,2,4 (15.7%)	3,1,2,4 (9.0%)	1,4,3,2 (7.5%)	58.2	49.3	67.2	38.8	67.2	61.2
52	60	1,3,2,4 (18.3%)	1,2,4,3 (8.3%)	1,3,4,2 (6.7%)	68.3	68.3	85.0	46.7	75.0	73.3
54	102	1,3,4,2 (10.8%)	3,1,2,4 (9.8%)	3,2,1,4 (9.8%)	51.0	43.1	71.6	42.2	69.6	79.4
55	137	1,3,2,4 (20.4%)	1,2,3,4 (13.9%)	1,3,4,2 (9.5%)	65.0	59.9	80.3	41.6	73.7	81.8

Remark: The four job values are labeled as 1 = [JMEAN], 2 = [JINC], 3 = [JPRO], and 4 = [JSEC]. In columns 3–5, "i, j" implies value $i \succ$ value j and the percentage beside indicates the proportion of instances having the corresponding ranking in node τ

Table 10.9 Result of Friedman test and multiple comparison procedures in leaf nodes of SP tree

Node (τ)	Node size	Mean rank				Friedman test statistic T_F	Multiple comparison (mean rank difference) †						Job value preference#
		JMEAN	JINC	JPRO	JSEC		JMEAN vs JINC	JMEAN vs JPRO	JMEAN vs JSEC	JINC vs JPRO	JINC vs JSEC	JPRO vs JSEC	
8	61	1.57	3.15	2.05	3.23	73.66**	−1.57	−0.48	−1.66	1.10	–	−1.18	1≻3≻2, 4
12	143	2.64	1.98	2.62	2.77	32.24**	0.66	–	–	−0.64	−0.79	–	2≻1, 3, 4
13	78	2.90	2.29	1.99	2.82	26.48**	0.60	0.91	–	–	−0.53	−0.83	3, 2≻1, 4
15	189	2.10	2.38	2.30	3.22	83.46**	−0.28	–	−1.12	–	−0.85	−0.92	1≻2; 1, 3, 2≻4
16	334	1.40	2.57	2.46	3.58	481.28**	−1.17	−1.06	−2.19	–	−1.02	−1.13	1≻3, 2≻4
22	195	2.70	2.07	2.09	3.14	93.91**	0.64	0.61	−0.44	–	−1.07	−1.05	2, 3≻1≻4
23	13	2.31	3.31	1.85	2.54	8.72*	−1.01	–	–	1.46	–	–	3, 1≻2
24	8	2.25	2.38	1.88	3.50	7.05*	–	–	−1.25	–	–	−1.63	3, 1≻4
26	177	2.16	2.26	2.54	3.04	49.64**	–	−0.38	−0.88	−0.28	−0.78	−0.50	1, 2≻3≻4
27	55	2.82	2.09	2.13	2.96	20.54**	0.73	0.69	–	–	−0.87	−0.84	2, 3≻1, 4
29	152	2.07	2.55	2.23	3.15	62.19**	−0.47	–	−1.08	0.32	−0.61	−0.92	1, 3≻2≻4
31	537	1.64	2.74	2.23	3.39	535.98**	−1.09	−0.58	−1.75	0.51	−0.66	−1.17	1≻3≻2≻4
33	154	2.32	2.32	2.32	3.03	114.0**	–	–	−0.71	–	−0.71	−0.71	1, 2, 3≻4
35	22	1.59	2.68	2.27	3.45	24.05**	−1.09	−0.68	−1.86	–	−0.77	−1.18	1≻3, 2≻4
37	7	3.14	1.86	3.29	1.71	8.66*	1.29	–	1.43	−1.43	–	1.57	4, 2≻1, 3
38	150	2.04	2.51	2.47	2.97	39.29**	−0.47	−0.43	−0.93	–	−0.46	−0.50	1≻3, 2≻4

41	259	2.00	2.25	2.50	3.24	133.44**	−0.25	−0.49	−1.24	−0.24	−0.99	−0.75	1≻2≻3≻4
42	223	1.52	2.43	2.74	3.30	222.74**	−0.91	−1.22	−1.78	−0.31	−0.87	−0.56	1≻2≻3≻4
43	18	1.28	2.67	3.33	2.72	24.47**	−1.39	−2.06	−1.44	−0.67	–	–	1≻2≻3; 1≻4
45	24	2.04	3.17	2.04	2.75	13.35**	−1.13	–	−0.71	1.13	–	−0.71	1, 3≻4, 2
46	49	1.84	2.63	2.49	3.04	22.05**	−0.80	−0.65	−1.20	–	–	−0.55	1≻3≻4; 1≻2
47	291	1.74	2.34	2.46	3.46	268.89**	−0.61	−0.72	−1.73	–	−1.12	−1.01	1≻2, 3≻4
48	83	1.82	2.41	2.60	3.17	46.27**	−0.59	−0.78	−1.35	–	−0.76	−0.57	1≻2, 3≻4
50	89	2.33	2.80	2.20	2.67	12.71**	−0.47	–	–	0.60	–	−0.47	3, 1≻2; 3≻4
51	134	2.25	2.52	2.27	2.96	25.88**	–	–	−0.70	–	−0.43	−0.69	1, 3, 2≻4
52	60	1.78	2.47	2.42	3.33	43.78**	−0.68	−0.63	−1.55	–	−0.87	−0.92	1≻3, 2≻4
54	102	2.34	2.39	2.06	3.21	44.62**	–	0.28	−0.86	0.33	−0.81	−1.15	3≻1, 2≻4
55	137	1.95	2.50	2.20	3.36	92.97**	−0.55	–	−1.41	0.30	−0.86	−1.16	1, 3≻2≻4

Remark: ** The test is significant at 0.01 level; * the test is significant at 0.1 level

† Only the mean rank differences of those pairwise multiple comparison procedures at $p < 0.05$ were shown

In the last column, the code 1–4 represent each of the political goals: 1 = [JMEAN], 2 = [JINC], 3 = [JPRO], and 4 = [JSEC]

preferences on work values among Americans. Individuals with advantaged academic experiences, more income, larger household size, less children, and full-time work were more concerned with the intrinsic value of jobs. All these interaction effects have a stronger relationship to job attribute priority than gender.

Although the data utilized in this study covers a 30-year span, the cross-sectional nature of the data limits our work to an incomplete picture from individuals' transition in occupational values over time. Of course, the use of panel data will help to address the changing trajectories in a life course perspective. Another limitation of this study stems from the unavailability of time-series data. With only three time points, it is difficult to confirm whether our conclusion of social and work influences on job value preferences is short term or evidence of longer trends in society. Collecting such data will help to place the findings in a larger perspective and hence should be a goal of subsequent research.