

Clustering and Prediction of Rankings Within a Kemeny Distance Framework

Willem J. Heiser and Antonio D'Ambrosio

Abstract Rankings and partial rankings are ubiquitous in data analysis, yet there is relatively little work in the classification community that uses the typical properties of rankings. We review the broader literature that we are aware of, and identify a common building block for both prediction of rankings and clustering of rankings, which is also valid for partial rankings. This building block is the Kemeny distance, defined as the minimum number of interchanges of two adjacent elements required to transform one (partial) ranking into another. The Kemeny distance is equivalent to Kendall's τ for complete rankings, but for partial rankings it is equivalent to Emond and Mason's extension of τ . For clustering, we use the flexible class of methods proposed by Ben-Israel and Iyigun (Journal of Classification 25: 5–26, 2008), and define the disparity between a ranking and the center of cluster as the Kemeny distance. For prediction, we build a prediction tree by recursive partitioning, and define the impurity measure of the subgroups formed as the sum of all within-node Kemeny distances. The median ranking characterizes subgroups in both cases.

1 Introduction

Ranking and classification are basic cognitive skills that people use every day to create order in everything that they experience. Many data collection methods in the life and behavioral sciences often rely on ranking and classification. Grouping and ordering a set of elements is also a major communication and action device in social

W.J. Heiser (✉)

Institute of Psychology, Leiden University, 2300 RB Leiden, The Netherlands

e-mail: Heiser@Fsw.Leidenuniv.nl

A. D'Ambrosio

Department of Industrial Engineering, University of Naples Federico II, Piazzale Tecchio, 80125, Naples, Italy

e-mail: antdambr@unina.it

life, as is clear when we consider rankings of sport-teams, universities, countries, web-pages, French wines, and so on. Not surprisingly, the literature on rankings is scattered across many fields of science.

Statistical methods for the analysis of rankings can be distinguished in (1) data analysis methods based on badness-of-fit functions that try to describe the structure of rank data, (2) probabilistic methods that model the ranking process, and assume substantial agreement (or *homogeneity*) among the rankers about the underlying order of the rankings, and (3) probabilistic methods that model the population of rankers, assuming substantial disagreement (or *heterogeneity*) between them. Let us look at each of these in turn.

Two examples of data analysis methods based on badness-of-fit functions that have been applied to rankings are principal components analysis (PCA, see Cohen and Mallows 1980; Diaconis 1989; Marden 1995, Chap. 2), and multidimensional scaling (MDS) or unfolding (Heiser and de Leeuw 1981; Heiser and Busing 2004). In psychometrics, PCA on rankings was justified by what is called the *vector model for rankings*, going back to the independent contributions of Guttman (1946); Slater (1960) and Tucker (1960) and popularized by Carroll (1972, pp. 114–129) through his MDPREF method. It is also possible to perform a principal components analysis while simultaneously fitting some optimal transformation of the data that preserves the rank order (in a program called CATPCA, cf. Meulman et al. 2004). By contrast, the unfolding technique is based on the *ideal point model for rankings*, which originated with Coombs (1950, 1964, Chaps. 5–7), but his analytical procedures were only provisional and had been soon superseded by MDS methods (Roskam 1968; Kruskal and Carroll 1969). Unfortunately, however, MDS procedures for ordinal unfolding tended to suffer from several degeneracy problems for a long time (see Van Deun 2005; Busing 2009 for a history of these difficulties and state-of-the-art proposals to resolve them). One of these proposals, due to Busing et al. (2005), is available under the name PREFSCAL in the IBM-SPSS Statistics package.

Probabilistic modeling for the ranking process assuming homogeneity of rankers started with Thurstone (1927, 1931), who proposed that judgments underlying rank orders follow a multivariate normal distribution with location parameters corresponding to each ranked object. Daniels (1950) looked at cases in which the random variables associated with the ranked objects are independent. Examples of more complex *Thurstonian models* include Böckenholt (1992), Chan and Bentler (1998), Maydeu-Olivares (1999) and Yao and Böckenholt (1999). A second class of models assuming homogeneity of rankers started with Mallows (1957), and was also based upon a process in which pairs of objects are compared, but now according to the Bradley-Terry-Luce (BTL) model (Bradley and Terry 1952; Luce 1959), thus excluding intransitivities. These probability models amount to a negative exponential function of some distance between rankings, for example the distance related to Kendall's τ (see Sect. 3); hence their name *distance-based ranking models* (Fligner and Verducci 1986). A third class of models assuming homogeneity of rankers decompose the ranking process into a series of independent stages. The stages form a nested sequence, in each of which a Bradley-Terry-Luce choice process is assumed for selecting 1 out of j options, with $j = m, m - 1, \dots, 2$; hence

their name *multistage models* (Fligner and Verducci 1988). We refer to Critchlow et al. (1991) for an in-depth discussion of all of these models. Critchlow and Fligner (1991) demonstrated how both the Thurstonean models and the multistage BTL models can be seen as generalized linear models and be fitted with standard software.

Probabilistic models for the population of rankers assuming substantial heterogeneity of their rankings are of at least three types. First, there are probabilistic versions of the ideal point model involving choice data (Zinnes and Griggs 1974; Kamakura and Srivastava 1986), or rankings (Brady 1989; Van Blokland-Vogelzang 1989; Hojo 1997, 1998). Second, instead of assuming one probabilistic model for the whole population, we may move to (unknown) mixtures of subpopulations, characterized by different parameters. For example, mixtures of models of the BTL type were proposed by Croon (1989), and mixtures of distance-based models by Murphy and Martin (2003). Gormley and Murphy (2008a) provided a very thorough implementation of two multistage models with mixture components. Third, heterogeneity of rankings can also be accounted for by the introduction of *covariates*, from which we can estimate mixtures of *known* subpopulations. Examples are Chapman and Staelin (1982), Dittrich et al. (2000), Böckenholt (2001), Francis et al. (2002), Skrandal and Rabe-Hesketh (2003), and Gormley and Murphy (2008b). All of these authors use the generalized linear modeling framework.

Most methods that are mainstream in the classification community follow the first approach, that is, they use an algorithm model (e.g., hierarchical clustering, construction of phylogenetic trees), or try to optimize some badness-of-fit function (e.g., *K*-means, fuzzy clustering, PCA, MDS). Some of them analyze a rank ordering of dissimilarities, which makes the results order-invariant, meaning that order-preserving transformations of the data have no effect. However, there are very few proposals in the classification community directly addressing clustering of multiple rankings, or prediction of rankings based on explanatory variables characterizing the source of them (covariates). Our objective is to fill this gap, and to catch up with the statisticians.¹

Common to all approaches is that they have to deal with the sample space of rankings, which has a number of very specific properties. Also, most methods either implicitly or explicitly use some measure of correlation or distance among rankings. Therefore, we start our discussion with a brief introduction in the geometry of rankings in Sect. 2, and how it naturally leads to measures of correlation and distance in Sect. 3. We then move to the median ranking in Sect. 4, give a brief sketch in Sect. 5 of how we propose to formulate a clustering procedure and to build a prediction tree for rankings, and conclude in Sect. 6.

¹During the Frankfurt DAGM-GfKI-2011-conference, Eyke Hüllermeier kindly pointed out that there is related work in the computer science community under the name “preference learning” (in particular, Cheng et al. (2009), and more generally, Fürnkranz and Hüllermeier 2010).

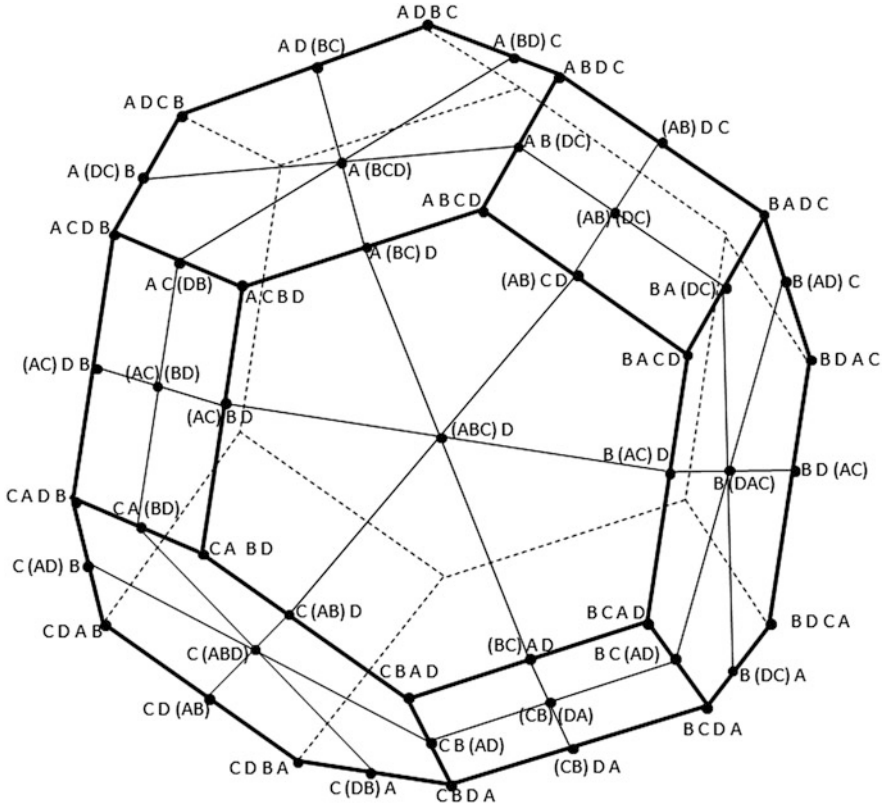


Fig. 1 Permutation polytope for all 24 full rankings of four objects, supplemented by all partial rankings with one tie-block of two or three objects, or two tie-blocks of two objects. Full rankings have equal distance towards the center; partial rankings lie strictly within this sphere. For clarity, mirror images at the back of the polytope are not labeled

2 Geometry of Rankings

The 24 full rankings that can be formed from four objects form a permutation polytope that has the shape of a truncated octahedron (cf. Thompson 1993; Heiser 2004). Thompson offered an thorough study of the permutation structure of partial rankings, showing that the 12 partial rankings with a tie in last position form a truncated tetrahedron, as do the 12 partial rankings with a tie in first position. The 12 partial rankings with a tie in middle position, however, are the intersection of a cube and an octahedron, forming a cuboctahedron. Then there are six partial rankings with two tie-blocks forming an octahedron, and finally four partial rankings with tie-blocks of three in last position or in first position, each forming a tetrahedron.

It should be noted that these generalized permutation polytopes can be connected with each other in a single graph if we introduce nodes in the original truncated octahedron that are half-way the nodes of the full rankings. This integrated graph of

all full and partial rankings is given in Fig. 1. All lines in this graph now indicate a reversal or switch from one inequality to an equality, or vice versa, except for the lines in the hexagons that connect to partial rankings with tie-blocks of three, which represent two switches. The natural graphical distance in the integrated permutation polytope is the sum of the line segments that need to be traversed along the shortest path in going from one node to another, and this distance is equivalent to the count of the minimum number of interchanges of two adjacent elements required to transform one (partial) ranking into another.

More generally, it will be clear that the sample space of rankings has the following characteristic properties: it is finite and discrete, it has many symmetries (for every ranking there is a reverse ranking), it is endowed with a graphical metric, and it intersects with a hypersphere: all full rankings are equidistant towards the zero ranking in which all objects are tied. All partial rankings lie strictly within the hypersphere. For a discussion of the consequences of this geometry for various ranking and choice models, we refer to Zhang (2004). Rankings can also arise indirectly as a consequence of doing pairwise discriminant analyses among m populations (Kamiya and Takemura 1997, 2005). Under the unfolding model, only a limited amount of rankings can occur (Coombs 1964; Kamiya et al. 2006, 2011). The probabilistic models mentioned in the Introduction describe specific distributions across the polytope.

3 Kendall's τ and the Kemeny Distance

Although there was earlier relevant work (see Kruskal 1958, Sect. 17), Kendall (1938) marks the beginnings of the first wave of contributions to the study of rankings as a separate topic in statistics. Kendall defined τ as a coefficient that “measures the closeness of correspondence between two given rankings in the sense that it measures how accurate either ranking would be if the other were objective” (Kendall 1938, p. 85). He then derived its exact sampling distribution and standard error, assuming one given order and a universe in which all the possible rankings occur an equal number of times, and he showed that this distribution is already close to normal for relatively small sample size. In Kendall (1948), he also gave a second definition of τ as a “coefficient of disarray”. Calling the minimum number of switches which transform any ranking into any other ranking of the same number of objects s , he showed that

$$\tau = 1 - \frac{2s}{\frac{1}{2}n(n-1)}.$$

This equivalence between τ and s establishes their connection with the permutation polytope, and thus their fundamental relevance for the study of rankings, because s is just the graphical distance defined in the previous section. The *minimum move metric* s is called the Kendall distance (cf. Marden 1995, p. 25).

Emond and Mason (2002) noted that there is a problem with the Kendall distance in the case of partial rankings. In that case, it is easy to show that it violates the triangle inequality (e.g., consider A(BC), ABC, and (AB)C), so it is not a proper metric. This anomaly is due to the way in which Kendall (1948, Chap. 3) defined τ when there are tied ranks.

Fortunately, there is a well-founded distance without these problems, called the Kemeny distance, conceived independently in the context of social choice theory (Kemeny 1959; Kemeny and Snell 1962). Kemeny had set up a set of reasonable axioms of which perhaps the most characteristic one is that the distance be *invariant under addition of equally ranked first and/or last objects*. The unique distance satisfying all axioms turns out to be:

$$d_{Kem}(R_s, R_t) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m |x_{(s)ij} - x_{(t)ij}|,$$

where R_s and R_t are any two rankings, m is the number of objects, and $x_{(s)ij}$ is defined as equal to 1 if object i is preferred to object j in ranking s , equal to -1 if the reverse is true, and equal to 0 if the two objects are tied. Clearly, the Kemeny distance is of the city-block type in the space of pair comparisons.

When there are no ties, the Kemeny distance is equal to the Kendall distance. From its definition, it is not hard to see that it counts the number of interchanges of pairs of elements required to transform one (partial) ranking into another, so it is equal to the graphical distance among any two elements in the integrated permutation polytope in Fig. 1.

4 Finding a Central Ranking: The Median Ranking

There is an extensive literature on finding a central ranking for a given set of individual rankings, also called the *social choice* problem, or the consensus problem. But when the Kemeny distance is the metric of choice, it will lead us to one specific central ranking. Consider a set of individual rankings R_s , with $s = 1, \dots, n$, and let us indicate the center to be found by \hat{S} . Then we have

$$\hat{S} = \arg \min_S \sum_{s=1}^n w_s d_{Kem}(R_s, S).$$

Here we have used a weighted version, with weights w_s for ranking R_s (one obvious choice of weights is the relative frequency with which each unique ranking occurs). Center \hat{S} so defined is usually called the *consensus ranking* in the social choice literature, as well as in discrete mathematics, and the *median ranking* in statistics. For a review of ranking models for the consensus problem, see Cook (2006).

Emond and Mason (2002) proposed a new rank correlation coefficient for the case of partial rankings, called τ_X (τ -extended), to resolve the difficulty with the Kendall distance mentioned in the previous section. It is equal to Kendall's τ for complete rankings, while for partial rankings $1 - \tau_X$ is equivalent to Kemeny distance. Maximizing the weighted sum of τ_X leads to the same median ranking. Now, it is well known that finding \hat{S} is an NP-hard problem (Barthélemy et al. 1989). Emond and Mason's reformulation has the advantage that it allows a branch-and-bound algorithm that is practical up to about 20 objects and an unlimited number of rankers, and deals correctly with partial rankings.

5 Application to Clustering and Recursive Partitioning

We will now give a brief sketch of how we are using the Kemeny distance and the median ranking for classification of multiple rankings. First, we outline a non-hierarchical clustering algorithm and next we show how to use explanatory variables (covariates) to build a prediction tree. For clustering, we follow a generalized K -means method, and for building the prediction tree, we use standard CART methodology (Breiman et al. 1984) involving a binary segmentation procedure that recursively partitions the set of rankings, with a specific impurity measure in the splitting rule. But of course, other choices are possible.

Ben-Israel and Iyigun's (2008) *probabilistic distance clustering* framework allows for probabilistic allocation of cases to classes. So it is a form of fuzzy clustering, rather than hard clustering. It is based on the principle that probability and distance are inversely related. Shepard (1987) accumulated lots of evidence for a similar principle governing contingencies of behavior. Under this principle, we define a loss function for *K-Median Cluster Component Analysis (CCA)* as follows:

$$CCA(\mathbf{P}, S_1, \dots, S_K) = \sum_{s=1}^n \sum_{k=1}^K p_k^2(R_s) d_{Kem}(R_s, S_k),$$

where $p_k(R_s)$ is the probability of allocating ranking s to cluster component k , S_k is the center of component k for $k = 1, \dots, K$, and \mathbf{P} is the $n \times K$ matrix of allocation probabilities. If we differentiate the *CCA* function with respect to $p_k(R_s)$, subject to the constraint that allocation probabilities for a given ranking sum to one, we obtain the stationary equation $p_k(R_s) d_{Kem}(R_s, S_k) = \text{constant depending on } R_s$. So the stationary equations of the *CCA* optimization problem are consistent with the principle of probability being inversely related to distance. Since the *CCA* function splits into K parts, finding S_k given some given values of the allocation probabilities \mathbf{P} reduces to finding a median ranking using the k th column of \mathbf{P} . For finding \mathbf{P} given K median rankings an explicit formula is available. A more detailed description and evaluation of *K*-median cluster component analysis is in preparation (Heiser and D'Ambrosio 2011).

Now consider the case in which we have a set of explanatory variables (or covariates) giving one point z_s in predictor space for each ranking R_s . The aim is to predict the differences between the rankings. Tree-based methods partition the predictor space into a set of rectangular regions parallel to the coordinate axes (i.e., the explanatory variables), and fit a simple model in each of them (Hastie et al. 2001). During the recursive partitioning process in which we form a nested sequence of subsamples, we have to determine, for each possible split along the coordinate axis of any variable, the impurity of the subsamples formed. The impurity measure $Q_l(T)$ that we choose for a subsample in subtree T at node l representing a region G_l containing the profiles of n_l rankings is

$$Q_l(T) = \frac{1}{\frac{1}{2}n_l(n_l-1)} \sum_{z_s \in G_l} \sum_{z_t \in G_l}^{n_l} d_{Kem}(R_s, R_t), \text{ with } s > t.$$

Alternatively, we could have chosen the weighted sum of Kemeny distances towards the median ranking, but that would force us to solve a hard combinatorial problem many times when growing the tree. Our pruning strategy is cost-complexity pruning (Hastie et al. 2001, p. 270; also see: Mingers (1989); Cappelli et al. 2002). For the pruned tree, we calculate in each terminal node the consensus ranking as described in Sect. 4 and its corresponding τ_X , and determine for the internal nodes of the tree the weighted average τ_X . For a more detailed description and evaluation of our distance-based prediction tree, we refer to D'Ambrosio and Heiser (2011), which is based on earlier work of D'Ambrosio (2007).

In one of our test applications, on a real dataset with 500 rankings of 15 objects and 128 explanatory variables, we first obtained a maximum tree with 24 terminal nodes. In Fig. 2, the top panel shows how the impurity in the training sample (bottom line) goes down monotonically, while in the test sample (upper line) the impurity goes up when tree size passes 11, which is the size of the pruned tree. The bottom panel of Fig. 2 shows the average τ_X weighted by node size, which gives a better interpretable scale. At the root node, overall $\tau_X = 0.387$, a moderate correlation, which reaches $\tau_X = 0.489$ on average for the maximum tree. Some of the terminal nodes in the pruned tree even reach $\tau_X = 0.510$, but others are lower.

6 Concluding Remarks

Kemeny distance is the natural graphical distance on the permutation polytope, which is the sample space of rankings. The polytope can be extended to accommodate partial rankings. It provides a standard for other approaches that use more assumptions or proceed by first embedding the polytope in Euclidean space. Minimizing the sum of Kemeny distances leads to the median ranking as a center. For full rankings, one minus Kendall's τ is equivalent to the Kemeny distance. Often the median ranking has ties, or the data are partial rankings to start with. In that

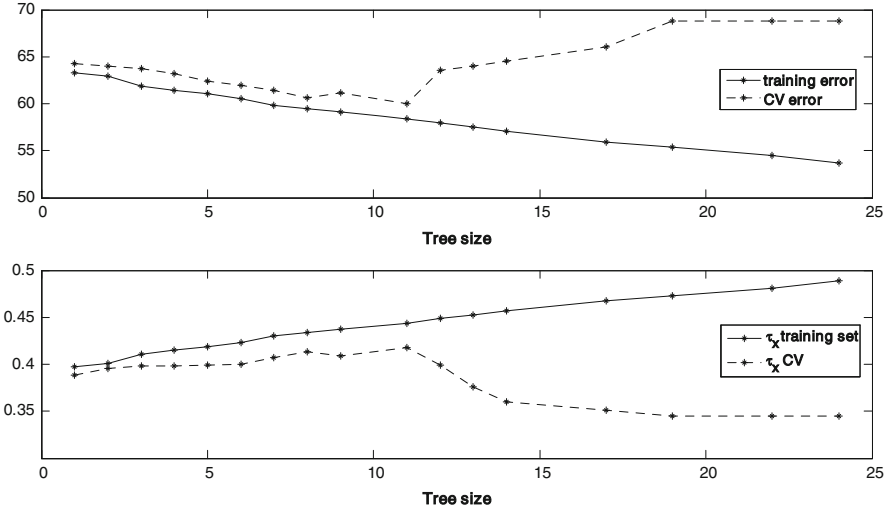


Fig. 2 Pruning sequence to decide on the depth of the tree. Training error rate is based on 350 rankings, cross-validated error rate is based on 150 rankings (using tenfold cross-validation). *Top panel* gives error rates (impurity), *bottom panel* gives the average τ_X

case, one minus Kendall's τ is faulted as a distance, because it no longer satisfies the metric axioms. Emond and Mason (2002) provided a different definition of τ for partial rankings, called τ_X , for which $1 - \tau_X$ is equal to the Kemeny distance. The new definition is welcome, because the scale of τ_X is easier to interpret than a distance scale: it is comparable across different numbers of objects.

We believe that loss-function based methods enjoy general advantages compared to methods based on probability models. They do not depend on assumptions that may be unrealistic for certain data. For rankings, in particular, the probability rationale often refers to replicated judgment processes, which is not so relevant for ranking the States of the United States (O'Leary Morgan and Morgan 2010), where the raw data are rates or percentages in the population. Note that in our use of probabilistic distance clustering, the term "probabilistic" merely expresses the uncertainty in the allocation of rankings to clusters, and does not imply an assumption about the data generating process, as in probability models.

Loss-function based methods generally tend to lead to better understood computational processes. Inclusion of weights in loss functions allows greater flexibility and generality, and in our case we profit from it in the median ranking and in the clustering algorithm. But weights can also be useful to emulate maximum likelihood estimation or to down-weight unreliable parts of the data. Some people hold, for example, that the beginning and the end of a ranking is more reliable than the middle.

Our clustering method could be compared with probabilistic models like Croon (1989), Murphy and Martin (2003), and Gormley and Murphy (2008a). Note that

when we cluster rankings, we are clustering variables, not objects. For applications where objects are to be clustered on the basis of ordinal variables, a method like GROUPALS (Van Buuren and Heiser 1989) would be a good possibility. The here adopted framework also gives us a way to adjust for cluster size (Iyigun and Ben-Israel 2008), or to develop semi-supervised learning techniques (Iyigun and Ben-Israel 2010). Our distance-based prediction tree method enjoys the general advantages of CART-like methods, such as easy interpretability and well-understood computational processes. It could be compared to methodology known under the name *hierarchical mixtures of experts*, based on probability models. An example of the mixture of experts approach is Gormley and Murphy (2008b). Another competitor for our method would be the ordinal unfolding approach with restrictions on the ideal points (Busing et al. 2010).

References

- Barthelémy, J. P., Guénoche, A., & Hudry, O. (1989). Median linear orders: Heuristics and a branch and bound algorithm. *European Journal of Operational Research*, 42, 313–325.
- Ben-Israel, A., & Iyigun, C. (2008). Probabilistic distance clustering. *Journal of Classification*, 25, 5–26.
- Böckenholt, U. (1992). Thurstonian representation for partial ranking data. *British Journal of Mathematical and Statistical Psychology*, 45, 31–49.
- Böckenholt, U. (2001). Mixed-effects analysis of rank-ordered data. *Psychometrika*, 77, 45–62.
- Bradley, R. A., & Terry, M. A. (1952). Rank analysis of incomplete block designs, I. *Biometrika*, 39, 324–345.
- Brady, H. E. (1989). Factor and ideal point analysis for interpersonally incomparable data. *Psychometrika*, 54, 181–202.
- Breiman, L., Froedman, J.H., Olshen, R.A., & Stone, C.J. (1984). *Classification and regression trees*. Wadsworth Publishing Co., Inc, Belmont, CA.
- Busing, F. M. T. A. (2009). *Some advances in multidimensional unfolding*. Doctoral Dissertation, Leiden, The Netherlands: Leiden University.
- Busing, F. M. T. A., Groenen, P., & Heiser, W. J. (2005). Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika*, 70, 71–98.
- Busing, F. M. T. A., Heiser, W. J., & Cleaver, G. (2010). Restricted unfolding: Preference analysis with optimal transformations of preferences and attributes. *Food Quality and Preference*, 21, 82–92.
- Cappelli, C., Mola, F., & Siciliano, R. (2002). A statistical approach to growing a reliable honest tree. *Computational Statistics and Data Analysis*, 38, 285–299.
- Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard et al. (Eds.), *Multidimensional scaling, Vol. I theory* (pp. 105–155). New York: Seminar Press.
- Chan, W., & Bentler, P. M. (1998). Covariance structure analysis of ordinal ipsative data. *Psychometrika*, 63, 369–399.
- Chapman, R. G., & Staelin, R. (1982). Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research*, 19, 288–301.
- Cheng, W., Hühn, J., & Hüllermeier, E. (2009). *Decision tree and instance-based learning for label ranking*. In: Proceedings of the 26th international conference on machine learning (pp. 161–168). Montreal, Canada.
- Cohen, A., & Mellows, C. L. (1980). *Analysis of ranking data (Tech. Rep.)*. Murray Hill: Bell Telephone Laboratories.

- Cook, W. D. (2006). Distance-based and ad hoc consensus models in ordinal preference ranking. *European Journal of Operational Research*, 172, 369–385.
- Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, 57, 145–158.
- Coombs, C. H. (1964). *A theory of data*. New York: Wiley.
- Critchlow, D. E., & Fligner, M. A. (1991). Paired comparison, triple comparison, and ranking experiments as generalized linear models, and their implementation on GLIM. *Psychometrika*, 56, 517–533.
- Critchlow, D. E., Fligner, M. A., & Verducci, J. S. (1991). Probability models on rankings. *Journal of Mathematical Psychology*, 35, 294–318.
- Croon, M. A. (1989). Latent class models for the analysis of rankings. In G. De Soete et al. (Eds.) *New developments in psychological choice modeling* (pp. 99–121). North-Holland, Elsevier.
- D'ambrosio, A. (2007). *Tree-based methods for data editing and preference rankings*. Doctoral dissertation. Naples, Italy: Department of Mathematics and Statistics.
- D'ambrosio, A., & Heiser, W. J. (2011). Distance-based multivariate trees for rankings. Technical report.
- Daniels, H. E. (1950). Rank correlation and population models. *Journal of the Royal Statistical Society, Series B*, 12, 171–191.
- Diaconis, P. (1989). A generalization of spectral analysis with application to ranked data. *The Annals of Statistics*, 17, 949–979.
- Dittrich, R., Katzenbeisser, W., & Reisinger, H. (2000). The analysis of rank ordered preference data based on Bradley-Terry type models. *OR-Spektrum*, 22, 117–134.
- Emond, E. J., & Mason, D. W. (2002). A new rank correlation coefficient with application to the consensus ranking problem. *Journal of Multi-Criteria Decision Analysis*, 11, 17–28.
- Fligner, M. A., & Verducci, J. S. (1986). Distance based ranking models. *Journal of the Royal Statistical Society, Series B*, 48, 359–369.
- Fligner, M. A., & Verducci, J. S. (1988). Multistage ranking models. *Journal of the American Statistical Association*, 83, 892–901.
- Francis, B., Dittrich, R., Hatzinger, R., & Penn, R. (2002). Analysing partial ranks by using smoothed paired comparison methods: An investigation of value orientation in Europe. *Applied Statistics*, 51, 319–336.
- Fürnkranz, J., & Hüllermeier, E. (Eds.). (2010). *Preference learning*. Heidelberg: Springer.
- Gormley, I. C., & Murphy, T. B. (2008a). Exploring voting blocs within the Irish electorate: A mixture modeling approach. *Journal of the American Statistical Association*, 103, 1014–1027.
- Gormley, I. C., & Murphy, T. B. (2008b). A mixture of experts model for rank data with applications in election studies. *The Annals of Applied Statistics*, 2, 1452–1477.
- Guttman, L. (1946). An approach for quantifying paired comparisons and rank order. *Annals of Mathematical Statistics*, 17, 144–163.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.
- Heiser, W. J. (2004). Geometric representation of association between categories. *Psychometrika*, 69, 513–546.
- Heiser, W. J., & Busing, F. M. T. A. (2004). Multidimensional scaling and unfolding of symmetric and asymmetric proximity relations. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 25–48). Thousand Oaks: Sage.
- Heiser, W. J., & D'ambrosio, A. (2011). *K-Median cluster component analysis*. Technical report.
- Heiser, W. J., & De Leeuw, J. (1981). Multidimensional mapping of preference data. *Mathématiques et Sciences Humaines*, 19, 39–96.
- Hojo, H. (1997). A marginalization model for the multidimensional unfolding analysis of ranking data. *Japanese Psychological Research*, 39, 33–42.
- Hojo, H. (1998). Multidimensional unfolding analysis of ranking data for groups. *Japanese Psychological Research*, 40, 166–171.
- Iyigun, C., & Ben-Israel, A. (2008). Probabilistic distance clustering adjusted for cluster size. *Probability in the Engineering and Informational Sciences*, 22, 603–621.

- Iyigun, C., & Ben-Israel, A. (2010). Semi-supervised probabilistic distance clustering and the uncertainty of classification. In A. Fink et al. (Eds.), *Advances in data analysis, data handling and business intelligence* (pp. 3–20). Heidelberg: Springer.
- Kamakura, W. A., & Srivastava, R. K. (1986). An ideal-point probabilistic choice model for heterogeneous preferences. *Marketing Science*, 5, 199–218.
- Kamiya, H., & Takemura, A. (1997). On rankings generated by pairwise linear discriminant analysis of m populations. *Journal of Multivariate Analysis*, 61, 1–28.
- Kamiya, H., & Takemura, A. (2005). Characterization of rankings generated by linear discriminant analysis. *Journal of Multivariate Analysis*, 92, 343–358.
- Kamiya, H., Orlik, P., Takemura, A., & Terao, H. (2006). Arrangements and ranking patterns. *Annals of Combinatorics*, 10, 219–235.
- Kamiya, H., Takemura, A., & Terao, H. (2011). Ranking patterns of unfolding models of codimension one. *Advances in Applied Mathematics*, 47, 379–400.
- Kemeny, J. G. (1959). Mathematics without numbers. *Daedalus*, 88, 577–591.
- Kemeny, J. G., & Snell, J. L. (1962). Preference rankings: An axiomatic approach. In J. G. Kemeny & J. L. Snell (Eds.), *Mathematical models in the social sciences* (pp. 9–23). New York: Blaisdell.
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30, 81–93.
- Kendall, M. G. (1948). *Rank correlation methods*. London: Charles Griffin.
- Kruskal, W. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53, 814–861.
- Kruskal, J. B., & Carroll, J. D. (1969). Geometrical models and badness-of-fit functions. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (Vol. 2, pp. 639–671). New York: Academic.
- Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- Mallows, C. L. (1957). Non-null ranking models, I. *Biometrika*, 44, 114–130.
- Marden, J. I. (1995). *Analyzing and modeling rank data*. New York: Chapman & Hall.
- Maydeu-Olivares, A. (1999). Thurstonian modeling of ranking data via mean and covariance structure analysis. *Psychometrika*, 64, 325–340.
- Meulman, J. J., Van Der Kooij, A. J., & Heiser, W. J. (2004). Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 49–70). Thousand Oaks: Sage.
- Mingers, J. (1989). An empirical comparison of pruning methods for decision tree induction. *Machine Learning*, 4, 227–243.
- Morgan, K. O., & Morgan, S. (2010). *State rankings 2010: A statistical view of America*. Washington, DC: CQ Press.
- Murphy, T. B., & Martin, D. (2003). Mixtures of distance-based models for ranking data. *Computational Statistics and Data Analysis*, 41, 645–655.
- Roskam, Ed. E. C. I. (1968). *Metric analysis of ordinal data in psychology: Models and numerical methods for metric analysis of conjoint ordinal data in psychology*. Doctoral dissertation, Voorschoten, The Netherlands: VAM.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237, 1317–1323.
- Skrondal, A., & Rabe-Hesketh, S. (2003). Multilevel logistic regression for polytomous data and rankings. *Psychometrika*, 68, 267–287.
- Slater, P. (1960). The analysis of personal preferences. *British Journal of Statistical Psychology*, 13, 119–135.
- Thompson, G. L. (1993). Generalized permutation polytopes and exploratory graphical methods for ranked data. *The Annals of Statistics*, 21, 1401–1430.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34, 273–286.
- Thurstone, L. L. (1931). Rank order as a psychophysical method. *Journal of Experimental Psychology*, 14, 187–201.

- Tucker, L. R. (1960). Intra-individual and inter-individual multidimensionality. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications* (pp. 155–167). New York: Wiley.
- Van Blokland-Vogeleesang, A. W. (1989). Unfolding and consensus ranking: A prestige ladder for technical occupations. In G. De Soete et al. (Eds.), *New developments in psychological choice modeling* (pp. 237–258). The Netherlands\North-Holland: Amsterdam.
- van Buuren, S., & Heiser, W. J. (1989). Clustering n objects into k groups under optimal scaling of variables. *Psychometrika*, 54, 699–706.
- Van Deun, K. (2005). *Degeneracies in multidimensional unfolding*. Doctoral dissertation, Leuven, Belgium: Catholic University of Leuven.
- Yao, G., & Böckenholt, U. (1999). Bayesian estimation of Thurstonian ranking models based on the Gibbs sampler. *British Journal of Mathematical and Statistical Psychology*, 52, 79–92.
- Zhang, J. (2004). Binary choice, subset choice, random utility, and ranking: A unified perspective using the permutahedron. *Journal of Mathematical Psychology*, 48, 107–134.
- Zinnes, J. L., & Griggs, R. A. (1974). Probabilistic, multidimensional unfolding analysis. *Psychometrika*, 39, 327–350.

Algorithms from and for Nature and Life

Classification and Data Analysis

Lausen, B.; Van den Poel, D.; Ultsch, A. (Eds.)

2013, XX, 547 p. 111 illus., Softcover

ISBN: 978-3-319-00034-3