On Ranking and Choice Models

Shivani Agarwal
Radcliffe Institute for Advanced Study,
Harvard University, Cambridge, MA, USA
Indian Institute of Science, Bangalore, India
shivani@csa.iisc.ernet.in

Abstract

In today's big data era, huge amounts of ranking and choice data are generated on a daily basis, and consequently, many powerful new computational tools for dealing with ranking and choice data have emerged in recent years. This paper highlights recent developments in two areas of ranking and choice modeling that cross traditional boundaries and are of multidisciplinary interest: ranking from pairwise comparisons, and automatic discovery of latent categories from choice survey data.

1 Introduction

In today's big data era, huge amounts of data are generated in the form of rankings and choices on a daily basis: restaurant ratings, product choices, employer ratings, hospital rankings, and so on. Given the increasing universality of such ranking and choice data, many powerful new computational tools for dealing with such data have emerged over the last few years in areas related to AI, including in particular machine learning, statistics, operations research, and computational social choice. Here we briefly highlight such developments in two broad areas: ranking from pairwise comparisons, and automatic discovery of latent categories from choice data.

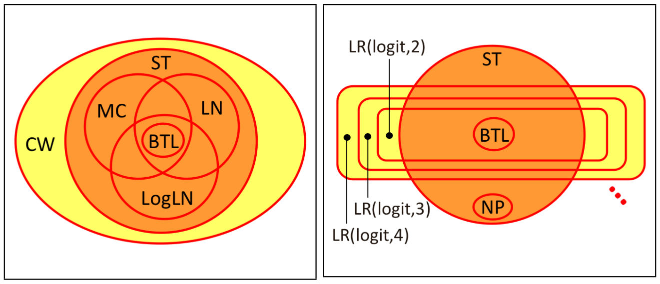
As is well known, humans generally find it easier to express preferences in the form of comparisons between two items, rather than directly ranking a large number of items. Indeed, in many domains, one is given outcomes of pairwise comparisons among some set of items (such as movies, candidates in an election, or sports teams), and must estimate a ranking of all the items from these observed pairwise comparisons. Due to the ubiquitous nature of the problem, several different algorithms have been developed for ***ranking from pairwise comparisons*** in different communities, including e.g. maximum likelihood estimation methods [Bradley and Terry, 1952; Luce, 1959], spectral ranking methods [Kendall, 1955; Keener, 1993; Negahban *et al.*, 2012], noisy sorting methods [Braverman and Mossel, 2008], and many others; however, little has been understood about when one algorithm should be preferred over another. In the first part of the paper (Section 2), we discuss recent developments in understanding these issues, including understanding the conditions under which different ranking algorithms succeed or fail, and how to design new algorithms for ranking form pairwise comparisons that achieve desirable goals under various conditions [Rajkumar and Agarwal, 2014; 2016a; Rajkumar *et al.*, 2015; Rajkumar and Agarwal, 2016b].

In marketing, when presenting products to users in a store or on a website, it is important to categorize related products together so that users can quickly find what they are looking for. These categories should ideally be based on how users themselves make choices, so that products that tend to be liked or disliked together are grouped together. In the second part of the paper (Section 3), we describe a new approach for ***automatic discovery of categories from choice data***, which brings together ideas from random utility choice models and topic models in order to automatically discover latent categories from choice survey data [Agarwal and Saha, 2016].

2 Ranking from Pairwise Comparisons

Ranking from pairwise comparisons is a ubiquitous problem that arises in a variety of applications. The basic types of questions of interest here are the following: Say there are n items, denoted $[n] = \{1, \ldots, n\}$, and we observe the outcomes of a number of pairwise comparisons among them (such as pairwise preferences among movies, pairwise judgments among job candidates, or pairwise game outcomes among sports teams). Based on these pairwise comparisons, can we find a good ranking of the n items, or identify a 'best' item or a 'good' set of items among them? How many comparisons do we need? What sorts of algorithms can we use? Under what conditions do these algorithms succeed?

A natural statistical framework for analyzing such questions assumes that for each pair of items (i, j), there is an underlying pairwise preference probability $P_{ij} \in [0, 1]$ such that whenever items i and j are compared, item i beats item j with probability P_{ij} and j beats i with probability $P_{ji} = 1 - P_{ij}$. Collectively, these pairwise preference probabilities form an underlying *pairwise preference matrix* $\mathbf{P} \in [0, 1]^{n \times n}$ (with $P_{ii} = \frac{1}{2}\ \forall i$). Different statistical models for pairwise comparisons lead to different conditions on $\mathbf{P}$. For example, if the pairwise preference probabilities follow the well-known Bradley-Terry-Luce (BTL) statistical model for pairwise comparisons, then there is a score vector $\mathbf{w} \in \mathbb{R}^n_{++}$ such that $P_{ij} = \frac{w_i}{w_i + w_j}\ \forall i, j$ [Bradley and Terry, 1952; Luce, 1959]; if they follow the 'noisy permutation' (NP)



Condition on P	**Property satisfied by P**
Bradley-Terry-Luce (BTL)	$\exists \mathbf{w} \in \mathbb{R}^n_{++} : P_{ij} = \frac{w_i}{w_i + w_j}$
Low-noise (LN)	$P_{ij} > \frac{1}{2} \Rightarrow \sum_k P_{ik} > \sum_k P_{jk}$
Logarithmic LN (LogLN)	$P_{ij} > \frac{1}{2} \Rightarrow \sum_k \ln(\frac{P_{ik}}{P_{ki}}) > \sum_k \ln(\frac{P_{jk}}{P_{kj}})$
Markov consistency (MC)	$P_{ij} > \frac{1}{2} \Rightarrow \pi_i > \pi_j$
Stochastic transitivity (ST)	$P_{ij} > \frac{1}{2}, P_{jk} > \frac{1}{2} \Rightarrow P_{ik} > \frac{1}{2}$
Condorcet winner (CW)	$\exists i : P_{ij} > \frac{1}{2}\ \forall j \neq i$
Noisy permutation (NP)	$\exists \sigma \in \mathcal{S}_n, p < \frac{1}{2} : P_{ij} = \begin{cases} 1-p & \text{if } i \succ_\sigma j \\ p & \text{otherwise} \end{cases}$
Low rank (LR(ψ, r))	$\mathrm{rank}(\psi(\mathbf{P})) \leq r \ \ (\psi : [0,1] \to \mathbb{R}, r \in [n])$

Figure 1: Conditions on the matrix of underlying pairwise preference probabilities $\mathbf{P}$ and relationships among them. These conditions play an important role in determining the success of different algorithms for ranking from pairwise comparisons.

model, then there is a permutation $\sigma \in \mathcal{S}_n$ and noise parameter $p \in [0, \frac{1}{2})$ such that $\forall i \neq j$, $P_{ij} = 1 - p$ if $\sigma(i) < \sigma(j)$ (which we also denote as $i \succ_\sigma j$), and $P_{ij} = p$ otherwise [Braverman and Mossel, 2008; Wauthier *et al.*, 2013]. Several other conditions on $\mathbf{P}$ are also of interest: see Figure 1.

One of our goals in recent work has been to understand how the matrix of underlying pairwise probabilities $\mathbf{P}$ affects the ranking goals that can be achieved, the success of different algorithms in achieving those goals, and the number of pairwise comparisons that are needed. We focus here mostly on settings where item pairs to be compared are selected randomly (or fixed in advance), but similar concerns also apply when pairs to be compared are selected in an active fashion.[1]

As it turns out, the matrix $\mathbf{P}$ plays a huge role in determining the success of different algorithms for ranking from pairwise comparisons. For example, spectral ranking algorithms such as Rank Centrality perform well when $\mathbf{P}$ satisfies the BTL condition or the slightly more general Markov consistency (MC) condition, but can fail miserably under more general settings of $\mathbf{P}$; indeed, if $\mathbf{P}$ is known only to satisfy stochastic transitivity (ST), then using an SVM-based ranking algorithm or a topological sort based algorithm can be a better choice [Rajkumar and Agarwal, 2014; 2016a]. If $\mathbf{P}$ does not satisfy ST, then all these algorithms fail to even recover a good set of items at the top, but one can use other algorithms for this purpose [Rajkumar *et al.*, 2015]. When comparisons can be made among only $O(n \log n)$ non-actively sampled pairs, then under suitable conditions on $\mathbf{P}$, one can use algorithms based on low-rank matrix completion [Rajkumar and Agarwal, 2016b]. Below we summarize some of these findings and give pointers for further investigation.

2.1 Finding an Optimal Ranking

Let us start by considering the setting where all $\binom{n}{2}$ pairs are compared a fixed number of times, say K times each.[2] Thus the input pairwise comparison data here is of the form $\{y_{ij}^1, \ldots, y_{ij}^K\}_{i<j}$, with $y_{ij}^k = 1$ denoting that the k-th comparison between i and j resulted in i beating j, and $y_{ij}^k = 0$ denoting the reverse; under the statistical model discussed above, each comparison outcome y_{ij}^k is a random draw from a Bernoulli random variable with parameter P_{ij}. Given this pairwise comparison data, consider the goal of finding a good ranking or permutation of the n items, $\widehat{\sigma} \in \mathcal{S}_n$.

A natural measure of the quality of a permutation σ is its *pairwise disagreement error* w.r.t. the underlying pairwise preference probabilities $\mathbf{P}$:

$$\mathrm{dis}(\sigma, \mathbf{P}) = \frac{1}{\binom{n}{2}} \sum_{i<j} \mathbf{1}\big((P_{ij} - \tfrac{1}{2})(\sigma(j) - \sigma(i)) < 0\big)\,.$$

An *optimal* ranking or permutation σ^* is then one that minimizes this pairwise disagreement error:

$$\sigma^* \in \mathrm{argmin}_{\sigma \in \mathcal{S}_n} \mathrm{dis}(\sigma, \mathbf{P})\,.$$

Under what conditions on $\mathbf{P}$ can we find an optimal ranking from the observed pairwise comparison data? Clearly, as the number of comparisons K per pair increases, we expect to be able to construct increasingly accurate estimates of $\mathbf{P}$. Which ranking algorithms have the property that as K increases, the rankings they produce approach an optimal ranking?

In recent work [Rajkumar and Agarwal, 2014; 2016b], we show that if the underlying pairwise probabilities $\mathbf{P}$ satisfy the LN condition, then both the simple matrix Borda algorithm and the popular method of maximum likelihood estimation under a BTL model succeed in recovering (with high probability) an optimal ranking (for sufficiently large K); if $\mathbf{P}$ satisfies the LogLN condition, then the least squares ranking algorithm succeeds in recovering such an optimal ranking; and if $\mathbf{P}$ satisfies the MC condition, then the Rank Centrality algorithm recovers an optimal ranking. Outside these conditions on $\mathbf{P}$, however, all these algorithms can fail miserably. We also give alternative algorithms for ranking from pairwise comparisons, including an SVM-based ranking algorithm and a topological sort based algorithm, that provably recover an optimal ranking from pairwise comparisons under more general $\mathbf{P}$ satisfying only stochastic transitivity. See Table 1 for a summary, and [Rajkumar and Agarwal, 2014; 2016a] for further details.

[1] The literature on dueling bandits provides a nice entry point for understanding similar issues when item pairs are allowed to be actively selected; e.g. see [Busa-Fekete and Hüllermeier, 2014] for a recent survey.

[2] The results we discuss in this setting also extend to settings where for each pairwise comparison to be made, a pair of items to be compared is selected randomly and independently according to some probability distribution $\boldsymbol{\mu} \in \Delta_{\binom{[n]}{2}}$ with $\mu_{ij} > 0\ \forall i < j$ (see [Rajkumar and Agarwal, 2014] for details); we consider the basic setting here, with all $\binom{n}{2}$ pairs compared K times, for simplicity.

Table 1: Properties of different ranking algorithms when all $\binom{n}{2}$ pairs are compared.

Ranking Algorithm	**Finds optimal ranking?**					**Finds Condorcet winner?**	**Finds top cycle?**	**Finds Copeland set?**
	BTL	**LN**	**LogLN**	**MC**	**ST**			
Matrix Borda	✓	✓	×	×	×	×	×	×
BTL-MLE	✓	✓	×	×	×	×	×	×
Least Squares Ranking	✓	×	✓	×	×	×	×	×
Rank Centrality	✓	×	×	✓	×	×	×	×
SVM-Based Ranking	✓	✓	✓	✓	✓	×	×	×
Topological Sort	✓	✓	✓	✓	✓	×	×	×
Matrix Copeland	✓	✓	✓	✓	✓	✓	✓	✓

In general, when $\mathbf{P}$ does not satisfy ST, finding an optimal ranking is hard even under exact knowledge of $\mathbf{P}$; indeed, this corresponds to the minimum weighted feedback arc set (MWFAS) problem, which is known to be NP-hard.[3] In such cases, one may still like to be able to find a ranking that places 'good' items at the top; we discuss this next.

2.2 Finding Good Items at the Top

As discussed above, when the underlying pairwise probabilities $\mathbf{P}$ do not satisfy stochastic transitivity, then even under exact knowledge of $\mathbf{P}$, finding an optimal ranking of all n items as above is NP-hard. In such settings, it is natural to ask that an algorithm return a ranking that places 'good' items at the top. For example, when $\mathbf{P}$ admits a Condorcet winner (an item that beats each other item with probability greater than half), it is natural to ask this item be placed at the top. More generally, when there is no Condorcet winner, one can consider various notions of *tournament solutions*, each of which gives a different way to define sets of winners in a tournament with cycles [Moulin, 1986; Brandt *et al.*, 2016], and ask that the items in the corresponding tournament solution be ranked at the top. For example, the *top cycle* is the smallest set of items $W \subseteq [n]$ for which each item in the set beats each item outside the set with probability greater than half (i.e. for which $P_{ij} > \frac{1}{2}\ \forall i \in W, j \notin W$); the *Copeland set* is the set of items $C \subseteq [n]$ that beat the largest number of items with probability greater than half (i.e. $C = \mathrm{argmax}_i \sum_j \mathbf{1}(P_{ij} > \frac{1}{2})$); and so on.

As we showed recently [Rajkumar *et al.*, 2015], when the underlying preference probabilities $\mathbf{P}$ do not satisfy ST, most commonly used algorithms for ranking from pairwise comparisons are unable to find even the Condorcet winner when it exists, and more generally, fail to rank items in various natural tournament solutions at the top (see Table 1). However, it is possible to design ranking algorithms that do find such tournament solutions at the top. For example, given enough pairwise comparisons (i.e. for large enough K), the Matrix Copeland algorithm successfully places both the top cycle and the Copeland set at the top (Table 1).[4] See [Rajkumar *et al.*, 2015] for further examples and discussion.

[3] When $\mathbf{P}$ satisfies ST, the graph induced by $\mathbf{P}$ is acyclic, and in this case the MWFAS problem under knowledge of $\mathbf{P}$ is easy.

[4] The Copeland set is always a subset of the top cycle.

Table 2: Conditions under which different ranking algorithms are known to have associated performance guarantees when only $O(n \log n)$ randomly selected pairs are compared.

Ranking Algorithm	**Condition for Performance Guarantee**
Rank Centrality	BTL
Balanced Rank Estimation	NP
Low-Rank Pairwise Ranking	LR$(\psi, r) \cap$ ST

2.3 Ranking from Comparisons of $O(n \log n)$ Pairs

The above discussion focused on the setting when all $\binom{n}{2}$ pairs can be compared. However, when n is large, comparing all pairs is impractical. A natural question that arises then is: Under what conditions on $\mathbf{P}$ can we find a good ranking from comparisons of only $O(n \log n)$ (randomly sampled) pairs? Previous work has shown this is possible under the BTL condition via the Rank Centrality algorithm [Negahban *et al.*, 2012] and under the NP condition via the Balanced Rank Estimation algorithm [Wauthier *et al.*, 2013]; in recent work, we show that this is in fact possible under a broader family of 'low-rank' conditions (which includes the BTL condition as a special case, but is considerably more general; see Figure 1), via a low-rank matrix completion based algorithm that we call Low-Rank Pairwise Ranking. See Table 2 for a summary, and [Rajkumar and Agarwal, 2016b] for further details.

3 Discovery of Categories from Choice Data

Consider now a marketing situation with n items or products, which need to be grouped into categories based on their being liked (or disliked) together. Say we have choice survey data from M users, in which each user m is shown some subset of items $S_m \subseteq [n]$ and is asked to indicate his or her preferences among these items, e.g. by indicating his or her top few choices in the set, or by assigning a rating (say 1–5) to each item in the set. Can we automatically discover meaningful categories from such choice data?

Suppose there are K unknown categories to be discovered. We model each category k via a random utility model (RUM) that associates n random variables $X_{k1}, \ldots, X_{kn}$ with the n items. In order to allow different users to have different preferences among these categories, we posit that each user m has a hidden preference vector $\boldsymbol{\theta}^m \in \Delta_K$ indicating his or her preferences among the K categories. We then posit a

Figure 2: Categories discovered from sports choice survey data in which 253 respondents each rated the following 7 sports on a scale of 1–5 based on how much they liked them: Baseball, Basketball, Cycling, Football, Jogging, Swimming, Tennis. The first category emphasizes individual sports; the second category emphasizes team sports; the third category separates out jogging. Categories can overlap: e.g. the above categories suggest cycling is likely to be co-enjoyed both with swimming and tennis and with jogging.

generative model for the observed choice data, whose parameters are the parameters of the K RUMs (i.e. parameters of the distributions of X_{ki}), as well as parameters of a shared Dirichlet prior on the preference vectors $\boldsymbol{\theta}^m$; given the observed choice data, we then fit these parameters to the data using a variational expectation-maximization procedure.

Under our generative model, given a set of items $S_m \subseteq [n]$, user m picks her first choice as follows: she draws a category k_1 according to her preference vector $\boldsymbol{\theta}^m$, then draws utility values x_i^1 for all items $i \in S_m$ according to $\{X_{k_1,i} : i \in S_m\}$ and picks an item with the highest drawn utility: $i_1 \in \arg\max_{i \in S_m} x_i^1$. Having picked her first choice i_1, in order to pick her second choice, she draws a category k_2 according to $\boldsymbol{\theta}^m$, then draws utility values x_i^2 for all items i in S_m *except the previously chosen item* i_1 according to $\{X_{k_2,i} : i \in S_m \setminus \{i_1\}\}$, and then picks an item $i_2 \in \arg\max_{i \in S_m \setminus \{i_1\}} x_i^2$. This process is repeated until the desired number of choices have been made. To model ratings data, which can be viewed as partial ranking data, we average over all choices that are consistent with the observed ratings.

As one example, Figure 2 shows the results obtained by applying our method to sports choice survey data in which 253 respondents each rated 7 sports on a scale of 1–5 based on how much they enjoyed each sport.[5] As can be seen, the method accurately discovers two categories corresponding to individual and team sports, and identifies a third category for jogging; this is consistent with previous studies that have found that people tend to have strong likes or dislikes for jogging independent of their liking for other sports [Marden, 1995]. Further examples and details can be found in [Agarwal and Saha, 2016].

4 Conclusion

Given the increasing availability of ranking and choice data, there is considerable need for new and innovative computational approaches to ranking and choice modeling. In this brief paper, we have highlighted some recent developments in ranking from pairwise comparisons and in automatic discovery of categories from choice data. There are many fascinating questions that remain open, and ranking and choice models will likely continue to be active areas of research in AI and related fields for many years to come.

[5] Here we modeled each category via a multinomial logit choice model, which corresponds to taking X_{ki} to be independent Gumbel random variables [McFadden, 1974].

Acknowledgments

Parts of the work described here are joint with Arun Rajkumar, Suprovat Ghoshal, Lek-Heng Lim, and Aadirupa Saha. Parts of this work have been supported by a Ramanujan Fellowship from DST; an Indo-US Joint Center Award from the Indo-US Science & Technology Forum; and the William & Flora Hewlett Foundation Fellowship at the Radcliffe Institute for Advanced Study, Harvard University.

References

[Agarwal and Saha, 2016] S. Agarwal and A. Saha. Automatic discovery of categories from choice data. *Forthcoming*, 2016.

[Bradley and Terry, 1952] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, pages 324–345, 1952.

[Brandt *et al.*, 2016] F. Brandt, M. Brill, and P. Harrenstein. Tournament solutions. In *Handbook of Computational Social Choice*. Cambridge University Press. *Forthcoming*, 2016.

[Braverman and Mossel, 2008] M. Braverman and E. Mossel. Noisy sorting without resampling. In *SODA*, 2008.

[Busa-Fekete and Hüllermeier, 2014] R. Busa-Fekete and E. Hüllermeier. A survey of preference-based online learning with bandit algorithms. In *ALT*, 2014.

[Keener, 1993] J. P. Keener. The Perron-Frobenius theorem and the ranking of football teams. *SIAM Review*, 35(1):80–93, 1993.

[Kendall, 1955] Maurice G. Kendall. Further contributions to the theory of paired comparisons. *Biometrics*, 11(1):43–62, 1955.

[Luce, 1959] R. D. Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.

[Marden, 1995] J. I. Marden. *Analyzing and Modeling Rank Data*. Chapman & Hall, 1995.

[McFadden, 1974] D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, 1974.

[Moulin, 1986] H. Moulin. Choosing from a tournament. *Social Choice and Welfare*, 3(4):271–291, 1986.

[Negahban *et al.*, 2012] S. Negahban, S. Oh, and D. Shah. Iterative ranking from pair-wise comparisons. In *NIPS*, 2012.

[Rajkumar and Agarwal, 2014] A. Rajkumar and S. Agarwal. A statistical convergence perspective of algorithms for rank aggregation from pairwise data. In *ICML*, 2014.

[Rajkumar and Agarwal, 2016a] A. Rajkumar and S. Agarwal. Ranking from pairwise comparisons: The role of the pairwise preference matrix. *Forthcoming*, 2016.

[Rajkumar and Agarwal, 2016b] A. Rajkumar and S. Agarwal. When can we rank well from comparisons of $O(n \log n)$ non-actively chosen pairs? *Forthcoming*, 2016.

[Rajkumar *et al.*, 2015] A. Rajkumar, S. Ghoshal, L.-H. Lim, and S. Agarwal. Ranking from stochastic pairwise preferences: Recovering Condorcet winners and tournament solution sets at the top. In *ICML*, 2015.

[Wauthier *et al.*, 2013] F. L. Wauthier, M. I. Jordan, and N. Jojic. Efficient ranking from pairwise comparisons. In *ICML*, 2013.