A Topic Modeling Approach to Rank Aggregation

Weicong Ding, Prakash Ishwar, Venkatesh Saligrama
Department of Electrical and Computer Engineering, Boston University, Boston, MA
{dingwc,pi,srv}@bu.edu

Abstract

We propose a new model for rank aggregation from pairwise comparisons that captures both ranking heterogeneity across users and ranking inconsistency for each user. We establish a formal statistical equivalence between the new model and topic models. We leverage recent advances in the topic modeling literature to develop an algorithm that can learn shared latent rankings with provable statistical and computational efficiency guarantees. The method is also shown to empirically outperform competing approaches on some semi-synthetic and real world datasets.

1 Introduction

The problem of inferring prevalent total rankings from partial preferences, known as rank aggregation, has been extensively studied over the last several decades with important applications to social choice, recommendation systems, and e-commerce [1–13]. The recent explosion of web technologies has enabled us to collect partial preferences, e.g., ratings or pairwise comparisons, for large sets of items, e.g., products from Amazon, movies from Netflix, or restaurants from Yelp, from a large and diverse population of users. In this paper, we focus on pairwise comparisons which is a building block for other partial preferences.

Research to-date has largely focused on either (a) learning *one* global ranking which is "optimally" consistent with all the observations according to some metric [7–15], or (b) associating each user with only a *single* ranking scheme which is sampled from a probabilistic model with multiple constituent rankings [1–6]. In the latter setting, each user is viewed as being consistent across time in generating all pairwise comparisons. In the context of new web-scale applications, however, (i) there are typically multiple global preferences within a diverse and heterogeneous population, (ii) the comparisons generated by the same user are, in general, inconsistent across time, especially for very similar items, and (iii) the number of comparisons that we can observe from each user is typically limited. These important aspects are not fully captured by any of the existing models.

In this paper, we introduce a new generative model for pairwise comparisons that captures a large heterogeneous population of inconsistent users (Sec. 2). The essence of the model is to view *multiple comparisons* of *each* user as a *probabilistic mixture* of a few rankings that are *shared* across users. We then establish a formal statistical equivalence between this model and the probabilistic topic modeling problem where each document in a corpus is modeled as a probabilistic mixture of a few topics [16]. We then develop a computationally and statistically efficient algorithm to consistently estimate the shared rankings (Sec. 3), and demonstrate competitive performance on semi-synthetic and real-world datasets (Sec. 4).

Although originating from a very different modeling perspective, our approach is most closely related to the seminal work in [1, 2] but with major differences (see Table 1). As it turns out, our proposed generative model subsumes those proposed in [1, 2] as special cases. On the other hand, while the algorithm proposed in [1, 2] can be applied to our more general setting, our proposed algorithm has provably better computational efficiency, polynomial sample complexity, and superior empirical performance.

Table 1: Comparison to closely related work [1, 2]

Method	**Assumptions on $\boldsymbol{\sigma}$**	**Statistics used**	**Consistency proved?**	**Computational complexity**	**Sample complexity**
[1, 2]	Separability	1st order	Yes	Exponential in K	Not provided
This paper	Separability	up to 2nd order	Yes	Polynomial	Polynomial

2 A new generative model and formal connection to topic models

We now formally describe our generative model. Let $\mathcal{U} = \{1, \ldots, Q\}$ be a universe of Q items and $\mathcal{P} = \{\{i,j\} \;:\; i < j, i, j \in \mathcal{U}\}$ be the set of all the *unordered* pairs. We consider a population of M users in which each user compares $N \geq 2$ pairs of items. We assume that the unordered pairs of items $\{i,j\}$ for comparison are independently drawn from a distribution μ on $\mathcal{P}$ and $\mu_{i,j} > 0$ for all i, j pairs. We denote the n-th comparison result of user m by an *ordered* pair $w_{m,n} = (i,j)$, if user m compares item i and j and prefers i over j.

The preference model for each user is modeled as arising from a *probabilistic mixture* of K total rankings over the Q items that are *shared* among the M users. If the K shared total rankings are denoted by permutations $\sigma^1, \ldots, \sigma^K$ of the Q items, and probability vector $\boldsymbol{\theta}_m$ denotes the user-specific weights over the K rankings for user m, [1] then the generative model for the pairwise comparisons from each user m, for $m = 1, \ldots, M$, can be described as follows,

1. Sample a K dimensional weight vector $\boldsymbol{\theta}_m$ from a prior distribution $\Pr(\theta)$;
2. For each comparison $n = 1, \ldots, N$:
   (a) Sample a pair of items $\{i,j\}$ from μ.
   (b) Sample $z_{m,n} \in \{1, \ldots, K\} \sim$ Multinomial($\boldsymbol{\theta}_m$).
   (c) If $\sigma^{z_{m,n}}(i) < \sigma^{z_{m,n}}(j)$, then $w_{m,n} = (i,j)$, otherwise $w_{m,n} = (j,i)$.

For convenience, we represent $\sigma^1, \ldots, \sigma^K$ by a $W \times K$ nonnegative ranking matrix $\boldsymbol{\sigma}$ whose $W = Q(Q-1)$ rows are indexed by all the ordered pairs (i,j) and $\sigma_{(i,j),k} = \mathbb{I}(\sigma^k(i) < \sigma^k(j))$. The k-th column of $\boldsymbol{\sigma}$ is an equivalent representation of the ranking σ^k. Figure 1 shows an example $\boldsymbol{\sigma}$. We denote the $K \times M$ dimensional weight matrix whose columns are the user-specific mixing weights $\boldsymbol{\theta}_m$'s by $\boldsymbol{\theta}$. Finally, we let $\mathbf{X}$ denote the $W \times M$ empirical comparisons-by-user matrix where $X_{(i,j),m}$ denotes the number of times that user m compares pair $\{i,j\}$ and prefers item i over j. The principal algorithmic problem is to estimate the ranking matrix $\boldsymbol{\sigma}$ given $\mathbf{X}$ and K.

Let $\mathbf{P}$ be a $W \times W$ diagonal matrix with the (i,j)-th diagonal component equal to $P_{(i,j),(i,j)} = \mu_{i,j}$ and define $\mathbf{B} = \mathbf{P}\boldsymbol{\sigma}$. The proposed generative model induces the following probabilities on $w_{m,n}$:

$$\Pr(w_{m,n} = (i,j)|\boldsymbol{\theta}_m) = \mu_{i,j} \sum_{k=1}^{K} \sigma_{(i,j),k}\theta_{k,m} = \sum_{k=1}^{K} B_{(i,j),k}\theta_{k,m} \tag{1}$$

Now consider a standard probabilistic topic model on a set of M documents, each composed of N words drawn from a vocabulary of size W, with a $W \times K$ *topic* matrix $\boldsymbol{\beta}$ and document-specific mixing weights $\boldsymbol{\theta}_m$ sampled from a topic prior $\Pr(\theta)$ [e.g. 17, 18]. The distribution induced on $w_{m,n}^{topic-model}$ – the n-th word in document m – has the same form as in (1):

$$\Pr(w_{m,n}^{topic-model} = l|\boldsymbol{\theta}_m) = \sum_{k=1}^{K} \beta_{l,k}\theta_{k,m} \tag{2}$$

where l is any distinct word in the vocabulary. Noting that $\mathbf{B}$ is column-stochastic, we have

Lemma 1. *The proposed generative model is statistically equivalent to a standard topic model with the topic matrix $\boldsymbol{\beta} = \mathbf{B}$ and the topic prior set to be $\Pr(\theta)$.*

Since (i) $\boldsymbol{B} = \mathbf{P}\boldsymbol{\sigma}$, (ii) $\mu_{i,j} = \mu_{j,i}$, and (iii) $\sigma_{(i,j),k} + \sigma_{(j,i),k} = 1$, therefore $\boldsymbol{\sigma}$ can be inferred directly from $\mathbf{B}$: $\sigma_{(i,j),k} = B_{(i,j),k}/(B_{(i,j),k} + B_{(j,i),k})$. Thus, the problem of estimating the ranking matrix $\boldsymbol{\sigma}$ can be solved by any approach that can learn the topic matrix $\boldsymbol{\beta}$. Our approach to is leverage a recent line of work in topic modeling [17–20] that comes with asymptotic consistency and statistical and computational efficiency guarantees and establish parallel results for ranking via

[1] $\sigma^k(i)$ is the position of item i in the ranking σ^k and item i is preferred over item j if $\sigma^k(i) < \sigma^k(j)$.

the equivalency result of Lemma 1. These topic modeling algorithms exploit the second-order moments of the columns of $\mathbf{X}$, i.e., a co-occurrence matrix of pairwise comparisons [17, 18, 20]. For instance, by combining Lemma 1 with results in [18] for topic modeling, the following result can be immediately established:

Lemma 2. *If $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{X}}'$ are obtained from $\mathbf{X}$ by first splitting each user's comparisons into two independent copies and then re-scaling the rows to make them row-stochastic, then*

$$M\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}^\top \xrightarrow[\text{almost surely}]{M\to\infty} \bar{\boldsymbol{B}}\bar{\mathbf{R}}\bar{\boldsymbol{B}}^\top =: \mathbf{E}, \tag{3}$$

where $\bar{\boldsymbol{B}} = \operatorname{diag}^{-1}(\boldsymbol{B}\mathbf{a})\boldsymbol{B}\operatorname{diag}(\mathbf{a})$, $\boldsymbol{B} = \mathbf{P}\boldsymbol{\sigma}$, $\bar{\mathbf{R}} = \operatorname{diag}^{-1}(\mathbf{a})\mathbf{R}\operatorname{diag}^{-1}(\mathbf{a})$, and $\mathbf{a}$ and $\mathbf{R}$ are, respectively, the $K \times 1$ expectation and $K \times K$ correlation matrix of the weight vector $\boldsymbol{\theta}_m$.

3 Estimating separable ranking matrices

A key ingredient of the recent topic modeling approaches [17–20] is the so-called *separability condition* on the topic matrix. When translated to the ranking matrix $\boldsymbol{\sigma}$ this corresponds to:

Definition 1. *The ranking matrix $\boldsymbol{\sigma}$ is separable if for each $k = 1, \dots, K$, there exists at lease one row, i.e., ordered pair (i, j), such that $\sigma_{(i,j),k} > 0$ and $\sigma_{(i,j),l} = 0$, $\forall\, l \neq k$.*

In plain words, for each ranking, there exists at least one "novel" pair of items $\{i, j\}$ such that i is uniquely preferred over j in that ranking while j is ranked higher than i in all the other rankings. For example, the ranking matrix $\boldsymbol{\sigma}$ in Fig. 1 is separable. In this example, the ordered pair $(1, 3)$ is novel to ranking $\boldsymbol{\sigma}^1$, the pair $(2, 1)$ to $\boldsymbol{\sigma}^2$, and the pair $(3, 2)$ to $\boldsymbol{\sigma}^3$. The separability assumption has also appeared, albeit implicitly in a different form, in [1, 2] in the context of rank aggregation. It is satisfied with high probability when $K \ll Q$ underlying rankings are sampled uniformly from the set of all $Q!$ permutations [2]. In our experiments, we have observed that the separability property holds true, at least approximately.

$\sigma^1: 1 > 2 > 3$
$\sigma^2: 2 > 3 > 1$
$\sigma^3: 3 > 1 > 2$

$$\begin{array}{c} \\ (1,2)\\ (1,3)\\ (2,1)\\ (2,3)\\ (3,1)\\ (3,2) \end{array}
\begin{array}{c}
\begin{array}{ccc} \text{Rank 1} & \text{Rank 2} & \text{Rank 3} \end{array}\\
\begin{bmatrix} 1 & 0 & 1\\ 1 & 0 & 0\\ 0 & 1 & 0\\ 1 & 1 & 0\\ 0 & 1 & 1\\ 0 & 0 & 1 \end{bmatrix}\\
\boldsymbol{\sigma}
\end{array}$$

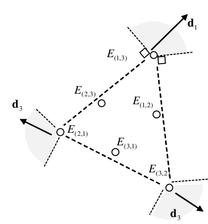

Figure 1: A separable ranking matrix $\boldsymbol{\sigma}$ with $K = 3$ rankings over $Q = 3$ items, and the underlying geometry for the row vectors of $\mathbf{E}$. Here, $(1, 3)$, $(2, 1)$, and $(3, 2)$ are novel pairs.

If $\boldsymbol{\sigma}$ (hence $\mathbf{B}$) is separable, then as illustrated in Fig. 1, the rows of $\mathbf{E}$ that correspond to novel pairs are the extreme points of the convex hull formed by all the row vectors of $\mathbf{E}$. Therefore, the novel pairs can be identified through an extreme point finding algorithm. Once the novel pairs of K distinct rankings are detected, the ranking matrix $\boldsymbol{\sigma}$ can be estimated in a straightforward manner by expressing the non-novel rows of $\mathbf{E}$ as convex combinations of the novel rows via least squares [17, 20].

Algorithm 1 Ranking Recovery (Main Steps) (see[18, 20] for more details)

Input: Pairwise comparisons $\widetilde{\mathbf{X}}, \widetilde{\mathbf{X}}'(W \times M)$; Number of rankings K; Number of projections P; Tolerance parameters $\zeta, \epsilon > 0$.

Output: Ranking matrix estimate $\widehat{\boldsymbol{\sigma}}$.

1: Set of Novel Pairs $\mathcal{I} \leftarrow$NovelPairDetect($\widetilde{\mathbf{X}}, \widetilde{\mathbf{X}}', K, P, \zeta$) (see Alg. 5 and 6 in [20])
2: $\widehat{\boldsymbol{B}} \leftarrow$EstimateRankings($\mathcal{I}, \mathbf{X}, \epsilon$) (see Alg. 4 in [18])
3: $\widehat{\boldsymbol{\sigma}}_{(i,j),k} \leftarrow \frac{\widehat{\boldsymbol{B}}_{(i,j),k}}{\widehat{\boldsymbol{B}}_{(i,j),k} + \widehat{\boldsymbol{B}}_{(j,i),k}}, \forall i, j \in \mathcal{U}, \forall k$
4: $\widehat{\boldsymbol{\sigma}}_{(i,j),k} \leftarrow \text{Round}[\widehat{\boldsymbol{\sigma}}_{(i,j),k}], \forall i, j \in \mathcal{U}, \forall k$

We adopt the approach proposed in [18, 20] to efficiently detect all the novel pairs using random projections. To exclude redundant rankings and ensure unique identifiability, we assume $\mathbf{R}$ to be full rank. The main steps are outlined in Alg. 1 and details can be found in [20, 21]. At a high level, steps 1 and 2 produce an estimate of $\mathbf{B}$ while steps 3 and 4 further process it to obtain an estimate of $\boldsymbol{\sigma}$. The proposed approach inherits the consistency and efficiency properties as in [20]. Formally,

Theorem 1. *Let $\boldsymbol{\sigma}$ be separable and $\mathbf{R}$ be full rank. Then Alg. 1 runs in $\mathcal{O}(MNK + Q^2K^3)$ time and consistently recovers $\boldsymbol{\sigma}$ up to a column permutation as the number of users $M \to \infty$ and the number of projections $P \to \infty$. Furthermore, $\forall \delta > 0$, if*

$$M \geq \max\left\{ 40\frac{\log(3W/\delta)}{N\rho^2\eta^4},\ 320\frac{W^{0.5}\log(3W/\delta)}{N\eta^6\lambda_{\min}^2} \right\} \text{ and } P \geq 16\frac{\log(3W/\delta)}{q_\wedge^2}$$

then, Alg. 1 fails with probability at most δ. The model parameters are defined as $\eta = \min_w [\mathbf{Ba}]_w$, $\rho = \min\{\frac{d}{8}, \frac{\pi d_2 q_\wedge}{4W^{1.5}}\}$, $d_2 = (1-b)\lambda_{\min}$, $d = (1-b)^2\lambda_{\min}^2/\lambda_{\max}$, $b = \max_{j,k:\bar{B}_{j,k}\neq 1} \bar{B}_{j,k}$, $\lambda_{\max}$ and $\lambda_{\min}$ are, respectively, the minimum and maximum eigenvalues of $\bar{\mathbf{R}}$, and $q_\wedge$ is the minimum normalized solid angle of the extreme points of the convex hull of the rows of $\mathbf{E}$ (see Eq. (1) in [20]).

4 Experimental validation

We conducted experiments on both semi-synthetic and real-world datasets to evaluate the performance of the proposed algorithm. We compared our algorithm (denoted by RP) against the algorithm proposed in [1, 2] (denoted by FJS). Due to space limitations we only summarize the salient points below and defer detailed explanations to a longer version of this paper [21].

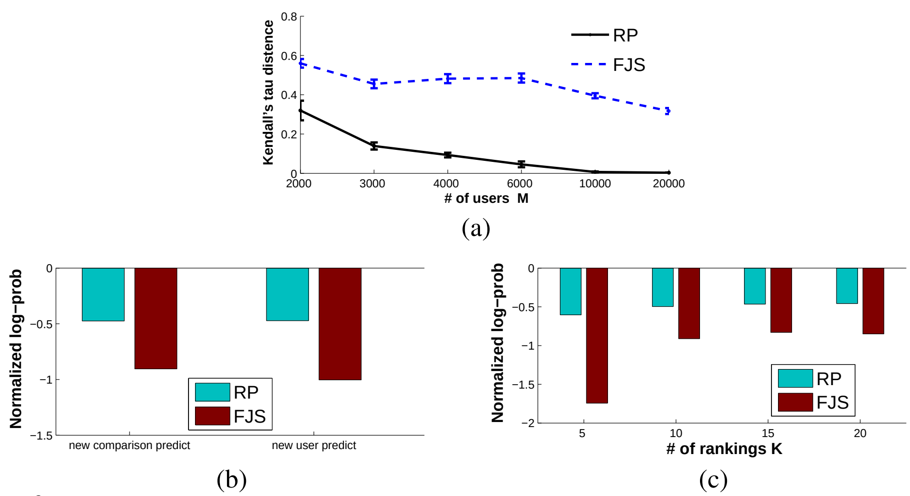

Figure 2: (a) Normalized *Kendall's tau* distance between the estimated and the ground truth rankings as a function of M on the semi-synthetic dataset ($K = 10$). (b) The normalized log-probability ($(1/\text{Total test pairs})\sum_{i,m} p(w_{\text{test},i}|\widehat{\boldsymbol{\sigma}}, \text{Trg. pairs of user } m)$) for new comparison prediction and new user prediction on the real-world Movielens dataset for $K = 10$. (c) The normalized log-likelihood for new comparison prediction on the real-world Movielens dataset for various K.

Semi-synthetic datasets. We first conducted experiments on a semi-synthetic dataset to validate the performance of our proposed algorithm when the model assumptions are satisfied. We generated semi-synthetic pairwise comparisons using the Movielens [22] benchmark movie *rating* dataset in order to match the dimensionality and other characteristics of real-world examples. We evaluated the reconstruction error between the learned rankings $\widehat{\boldsymbol{\sigma}}$ and the ground truth $\boldsymbol{\sigma}$ using the standard *Kendall's tau distance* between rankings [23]. The results in Fig. 2(a) show the superior performance of RP over FJS.

Movielens. We next conducted experiments on the real-world Movielens dataset in order to demonstrate that the proposed model can indeed effectively capture real-world variability. We considered two tasks on the real-world Movielens dataset: (1) new comparison prediction and (2) new user prediction for the $Q = 100$ most frequently rated movies. We use the standard *held-out log-likelihood* to measure the performance [4, 24]. The results are summarized in Fig 2(b) and (c). RP is clearly seen to have better performance under various parameter settings.

References

[1] S. Jagabathula and D. Shah. Inferring rankings under constrained sensing. In *Advances in Neural Information Processing Systems*, pages 753–760. Vancouver, Canada, Dec. 2008.

[2] V. Farias, S. Jagabathula, and D. Shah. A data-driven approach to modeling choice. In *Advances in Neural Information Processing Systems*, pages 504–512. Vancouver, Canada, Dec. 2009.

[3] P. Awasthi, A. Blum, O. Sheffet, and A .Vijayaraghavan. Learning mixtures of ranking models. In *Advances in Neural Information Processing Systems*. Montreal, Canada, Dec. 2014.

[4] T. Lu and C. Boutilier. Learning Mallows models with pairwise preferences. In *Proceedings of the 28th International Conference on Machine Learning*, pages 145–152, Bellevue, WA, USA, Jun. 2011.

[5] H. Azari Soufiani, H. Diao, Z. Lai, and D. C Parkes. Generalized Random Utility Models with multiple types. In *Advances in Neural Information Processing Systems*, pages 73–81. Lake Tahoe, NV, USA, Dec. 2013.

[6] S. Oh and D. Shah. Learning mixed Multinomial Logit Model from ordinal data. In *Advances in Neural Information Processing Systems*, Montreal, Canada, Dec. 2014.

[7] M. N. Volkovs and R. S. Zemel. New learning methods for supervised and unsupervised preference aggregation. *Journal of Machine Learning Research*, 15:1135–1176, 2014.

[8] A. Rajkumar and S. Agarwal. A statistical convergence perspective of algorithms for Rank Aggregation from pairwise data. In *Proceedings of the 31st International Conference on Machine Learning*, pages 118–126, Beijing, China, Jun. 2014.

[9] F. L. Wauthier, M. I. Jordan, and N. Jojic. Efficient ranking from pairwise comparisons. In *Proceedings of the 30th International Conference on Machine Learning*, pages 109–117, Atlanta, GA, USA, Jun. 2013.

[10] S. Negahban, S. Oh, and D. Shah. Iterative ranking from pair-wise comparisons. In *Advances in Neural Information Processing Systems*, pages 2474–2482. Lake Tahoe, NV, USA, Dec. 2012.

[11] B. Osting, C. Brune, and S. Osher. Enhanced statistical rankings via targested data collection. In *Proceedings of the 30th International Conference on Machine Learning*, pages 489–497, Atlanta, GA, USA, Jun. 2013.

[12] H. Azari Soufiani., W. Chen, D. Parkes, and L. Xia. Generalized Method-of-Moments for Rank Aggregation. In *Advances in Neural Information Processing Systems*, pages 2706–2714. Lake Tahoe, NV, USA, Dec. 2013.

[13] D. F. Gleich and L.-H. Lim. Rank aggregation via nuclear norm minimization. In *Proceedings of the 17th ACM International Conference on Knowledge Discovery and Data Mining*, pages 60–68, San Diego, CA, USA, 2011.

[14] J. Huang, A. Kapoor, and C. Guestrin. Efficient probabilistic inference with partial ranking queries. In *Proceedings of the 27th conference on Unvertainty in Artificial Intelligence*, Barcelona, Spain, Jul. 2011.

[15] T. Qin, X. Geng, and T.-Y. Liu. A new probabilistic model for rank aggregation. In *Advances in Neural Information Processing Systems*, pages 1948–1956. Vancouver, Canada, Dec. 2010.

[16] D. Blei. Probabilistic topic models. *Commun. of the ACM*, 55(4):77–84, 2012.

[17] S. Arora, R. Ge, Y. Halpern, D. Mimno, A. Moitra, D. Sontag, Y. Wu, and Michael Zhu. A practical algorithm for topic modeling with provable guarantees. In *Proceedings of the 30th International Conference on Machine Learning*, pages 280–288, Atlanta, GA, USA, Jun. 2013.

[18] W. Ding, M. H. Rohban, P. Ishwar, and V. Saligrama. Topic discovery through data dependent and random projections. In *Proceedings of the 30th International Conference on Machine Learning*, pages 1202–1210, Atlanta, GA, USA, Jun. 2013.

[19] S. Arora, R. Ge, and A. Moitra. Learning topic models – going beyond SVD. In *Proceedings of the IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 1–10, New Brunswick, NJ, USA, Oct. 2012.

[20] W. Ding, M. H. Rohban, P. Ishwar, and V. Saligrama. Efficient distributed topic modeling with provable guarantees. In *Proceedings ot the 17th International Conference on Artificial Intelligence and Statistics*, pages 167–175, Reykjavik, Iceland, Apr. 2014.

[21] W. Ding, P. Ishwar, and V. Saligrama. A topic modeling approach to rank aggregation. Boston University (BU) Center for Information and Systems Engineering (CISE) Technical Report `http://www.bu.edu/systems/publications/`, Oct. 2014.

[22] `http://grouplens.org/datasets/movielens/`.

[23] M. Kendall. A new measure of rank correlation. *Biometrika*, 30:81–89, 1938.

[24] H. M. Wallach, I. Murray, R. Salakhutdinov, and D. Mimno. Evaluation methods for topic models. In *Proceedings of the 26th International Conference on Machine Learning*, pages 1105–1112, Montreal, Canada, Jun. 2009.