Top-K Ranking: An Information-theoretic Perspective

Yuxin Chen
EE, Stanford University
Email: yxchen@stanford.edu

Changho Suh
EE, KAIST
Email: chsuh@kaist.ac.kr

***Abstract*—We develop an information-theoretic framework that explores the identifiability of top-K ranked items. The goal of the problem considered herein is to recover a consistent ordering that emphasizes the top-K ranked items, based on partially revealed preferences. Under the Bradley-Terry-Luce model that postulates a set of latent preference scores underlying all items and the odds of paired comparisons depend only on the relative scores of the items involved, we characterize the fundamental limits (up to some constant gap) on the amount of information required for reliably identifying the top-K ranked items. Here we introduce an information-theoretic notion of *reliable ranking*, meaning that the probability of the estimated ranking being inconsistent with the ground truth can be made arbitrarily close to zero. We single out one significant measure that plays a crucial role in determining the limits: the separation measure that quantifies the gap of preference scores between the K^{th} and $(K+1)^{\text{th}}$ ranked items. We show that the minimum sample complexity required for reliable top-K ranking scales inversely with the separation measure.**

I. Introduction

Rank aggregation has been investigated in a wide spectrum of contexts like social choice, voting, web search, graduate admission, caching, crowdsourcing, recommendation system and sports competitions. In the big data era, however, a new challenge arises: due to the huge number of items/candidates to be ranked, observation of the complete data sets is often infeasible. This challenge brings about a series of significant questions. How can we identify a ranking of the items, given highly incomplete measurements of data sets? Is there a fundamental limit on the amount of information needed for reliably identifying the ranking? If so, is there a computationally efficient low-complexity algorithm that can achieve the limit.

In an effort to address the above questions, a natural ranking problem has been explored. Given a large collection of items, the goal of the problem is to recover a preference ordering (the ground-truth ranking if any) on the items from a subset of measurements data sets. Measurements are often of the pairwise/listwise information type: movie A is preferred over movie B; player A defeats player B. There has been a proliferation of ranking algorithms, most of which postulate the existence of ground-truth preference scores on the items that determine their ranking. Two popular paradigms are: (1) maximum likelihood estimation (MLE) [1]; (2) spectral ranking algorithms such as PageRank [2] and Rank Centrality [3].

While these algorithms are shown to often achieve a low squared loss w.r.t. the preference scores, the minimum squared loss does not necessarily imply a high ranking accuracy. Also these methods do not well respect many realistic ranking scenarios in which only a few significant items, say top-K items, are particularly emphasized. Regardless of the item importance, the algorithms simply place the same weight on a loss w.r.t. each preference score, yielding a complete ranking of the entire items. This possibly degrades the ranking accuracy w.r.t. the top ranked items of significant interest.

In this work, we seek to overcome the above challenges, developing a computationally efficient ranking algorithm that ensures the best ranking accuracy given partial measurements of data. Specifically we aim at recovery of the top-K ranked items. To this end, we consider a task-oriented performance metric that can directly measure a ranking accuracy. Unlike the minimum square error which does not necessarily implies the optimal ranking accuracy, we emphasize the probability of an error event: an estimate of the top-K ranked items is inconsistent with the ground truth. This naturally leads us to define an information-theoretic notion of *reliable ranking*, meaning that the error probability can be made arbitrarily close to zero as the number of items tends to infinity. As for observed information, we consider a popular pairwise comparison model, called the Bradley-Terry-Luce (BTL) model [4], [5], a long-standing model that has been applied in numerous applications.

Our contributions are two-fold. First, we characterize the fundamental limits (up to some constant factor gap) on the amount of pairwise information required for reliable top-K ranking. In particular, we emphasize a separation measure that quantifies the gap of preference scores between the K^{th} and $(K+1)^{\text{th}}$ ranked items. Our result demonstrates that the minimal sample complexity or quality of paired evaluation scales inversely with the separation measure.

Secondly, we propose a nearly linear-time two-stage algorithm, called *Spectral MLE*, which allows perfect top-K identification as soon as the sample complexity exceeds the limits (modulo some constant). Specifically, Spectral MLE starts by obtaining careful initial scores that are faithful in the ℓ_2 sense (e.g. via a spectral method), and then iteratively sharpens the pointwise estimates by comparing the preceding estimates with coordinate-wise MLE. This algorithm is designed primarily in an attempt to seek a score estimate with minimal pointwise loss.

978-1-4673-7852-9/15/$31.00 ©2015 IEEE

II. Model

Suppose that we observe a few pairwise evaluations over n items. We assume that the pairwise comparison outcomes are generated according to the BTL model [4], [5]. The BTL model hypothesizes on the existence of some hidden preference vector $\boldsymbol{w} = [w_i]_{1\le i\le n}$, where w_i represents the underlying preference score of item i. The outcome of each paired comparison depends only on the scores of the items involved. Without loss of generality, assume that $w_1 \ge w_2 \ge \cdots \ge w_n > 0$. It is also assumed that the dynamic range of the preference scores is fixed irrespective of n: $w_i \in [w_{\min}, w_{\max}], 1 \le i \le n$ for some positive constants $w_{\min}$ and $w_{\max}$ bounded away from 0, which amounts to the most challenging regime [3].

Denote by $\mathcal{G} = ([n], \mathcal{E})$ the comparison graph such that items i and j are compared if and only if (i, j) belongs to the edge set $\mathcal{E}$. We assume that $\mathcal{G}$ is drawn from the Erdős–Rényi model $\mathcal{G} \sim \mathcal{G}_{n,p_{\rm obs}}$ for some observation factor $p_{\rm obs}$. For each $(i, j) \in \mathcal{E}$, we observe L paired comparisons between items i and j. The outcome of the $l^{\rm th}$ comparison between them, denoted by $y_{i,j}^{(l)}$, is generated as per the BTL model:

$$y_{i,j}^{(l)} = \begin{cases} 1, & \text{w.p. } \frac{w_i}{w_i+w_j}, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where $y_{i,j}^{(l)} = 1$ indicates a win by i over j. We adopt the convention that $y_{j,i}^{(l)} = 1 - y_{i,j}^{(l)}$. It is assumed that conditional on $\mathcal{G}$, $y_{i,j}^{(l)}$'s are jointly independent across all l and $i > j$.

Given these pairwise observations, one wishes to see whether or not the top-K ranked items are identifiable. To this end, we consider the probability of error $P_{\rm e}$ in isolating the *set* of top-K ranked items, i.e. $P_{\rm e}(\psi) := \mathbb{P}\{\psi \ne [K]\}$ where ψ is any ranking scheme that returns a set of K indices. We aim to characterize the fundamental limits on $(L, p_{\rm obs})$ where reliable top-K ranking is feasible, i.e. $P_{\rm e}$ can be vanishingly small as n grows.

III. Main Result

We explore the fundamental limits from a minimax perspective, which centers on the design of *robust* ranking schemes that guard against the worst case in probability of error. The most challenging component of top-K rank aggregation hinges upon distinguishing the two items near the decision boundary, i.e. the $K^{\rm th}$ and $(K+1)^{\rm th}$ ranked items. Due to the random nature of the acquired finite-bit comparisons, the information concerning their relative preference could be obliterated by noise, unless their latent preference scores are sufficiently separated. In light of this, we single out a preference separation measure as follows

$$\Delta_K := \frac{w_K - w_{K+1}}{w_{\max}}. \tag{2}$$

As will be seen, this measure plays a crucial role in determining information integrity for top-K identification.

To model partial observation, we employ the Erdős–Rényi model $\mathcal{G} \sim \mathcal{G}_{n,p_{\rm obs}}$. As already noted by [1], if the comparison graph $\mathcal{G}$ is not connected, then there is absolutely no basis to determine relative preferences between two disconnected components. Therefore, a reasonable necessary condition that one would expect is the connectivity of $\mathcal{G}$, which requires $p_{\rm obs} > \log n \,/\, n$. The result in this paper will operate under this assumption.

The main result of this paper is to characterize an order-wise tight sufficient and necessary condition for top-K identifiability, as stated below.

Theorem 1: Suppose that $\mathcal{G} \sim \mathcal{G}_{n,p_{\rm obs}}$ with $p_{\rm obs} > \log n/n$. The minimal sample complexity for reliable top-K ranking is:

$$\binom{n}{2} p_{\rm obs} L = \Theta\left(\frac{n \log n}{\Delta_K^2}\right). \tag{3}$$

Proof: See the full version of this paper [6]. ■

Remark 1: Theorem 1 also determines the scaling of the fundamental ranking boundary on L: $L = \Theta(\frac{\log n}{np_{\rm obs}\Delta_K^2})$. It also characterizes the fundamental condition on the underlying score vector through the single measure Δ_K:

$$\Delta_K = \Theta\left(\sqrt{\frac{\log n}{np_{\rm obs}L}}\right). \tag{4}$$

This result is intuitive, as it says that larger Δ_K, less number of measurements is required, meaning that top-K ranking recovery is easier. □

Our ranking algorithm is to take two stages: (1) obtaining an estimate that concentrates around the ground truth in the squared loss sense (e.g., via spectral methods such as RankCentrality [3]); (2) iteratively refining the estimate in a pointwise manner via coordinate MLE. The purpose of the second stage is to detect outliers whose estimates are far from the ground-truth scores, and then control corresponding point-wise errors, thus yielding a higher ranking accuracy. Encouragingly, both RankCentrality and MLE are linear-time algorithms; hence, our proposed algorithm that synergetically exploits advantageous features of the two algorithms having the low-complexity nature in common is also computationally efficient. So there is no computational barrier away from the statistical limit characterized above.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIP) (No. Grant Number: 2015R1C1A1A02036561)

References

[1] L. R. Ford, "Solution of a ranking problem from binary comparisons," *American Mathematical Monthly*, 1957.

[2] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer networks and ISDN systems*, vol. 30, no. 1, pp. 107–117, 1998.

[3] S. Negahban, S. Oh, and D. Shah, "Rank centrality: Ranking from pairwise comparisons," 2012.

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

[5] R. D. Luce, *Individual choice behavior: A theoretical analysis*. Wiley, 1959.

[6] Y. Chen and C. Suh, "Spectral MLE: Top-K rank aggregation from pairwise comparisons," in *ICML*, 2015.