A Theoretical Analysis of Normalized Discounted Cumulative Gain (NDCG) Ranking Measures

Yining Wang, Liwei Wang, Yuanzhi Li, Di He, Wei Chen and Tie-Yan Liu

School of EECS, Peking University; IIIS, Tsinghua University; Microsoft Research Asia

Background of Ranking

Model of ranking (for web search):

- $\mathcal{X}$: Instance space.
- $\mathcal{Y}$: A set of degrees of relevancy.
- $f: \mathcal{X} \to \mathbb{R}$; a scoring function.
- Ranking with f: sorting by the scores. $(x_1, \ldots, x_n) \longrightarrow (x^f_{(1)}, \ldots, x^f_{(n)})$, so that $f(x^f_{(1)}) \geqslant \ldots \geqslant f(x^f_{(n)})$.

Evaluate the performance of f by NDCG:

- Given dataset $S_n = \{(x_1, y_1), \ldots, (x_n, y_n)\}$.
- Obtain ranking $x^f_{(1)}, \ldots, x^f_{(n)}$
- Degree of relevancy: $y^f_{(1)}, \ldots, y^f_{(n)}$.
- Discounted Cumulative Gain (DCG): $\mathrm{DCG}_D(f, S_n) = \sum_{r=1}^{n} D_r \cdot y^f_{(r)}$. D_r: discount function (e.g. $D_r = \frac{1}{\log(1+r)}$).

- **Normalized DCG (NDCG)**: $\mathbf{NDCG_D(f, S_n)} = \frac{\mathrm{DCG}_D(f,S_n)}{\mathrm{IDCG}_D(S_n)} \in [0, 1]$. IDCG: the ideal DCG value on S_n. $\mathrm{IDCG}_D(S_n) := \max_f \mathrm{DCG}_D(f, S_n)$.

Questions:

- Are NDCG family good ranking measures?
- How to set discount D_r for NDCG?

A First Observation (Motivation)

Standard NDCG: logarithmic discount $D_r = \frac{1}{\log(1+r)}$. (Widely used in applications.)

Suppose S_n consists of i.i.d. data, we have:

Theorem. Let $D_r = \frac{1}{\log(1+r)}$. Then for every ranking function f,

$$\mathbf{NDCG_D(f, S_n) \to 1} \text{ a.s.}$$

Standard NDCG cannot distinguish good and bad ranking functions on large dataset?

Consistent Distinguishability

We propose a criterion to determine whether a given ranking measure is good.

Definition (Consistent Distinguishability) A pair of scoring functions f_0 and f_1 is said to be consistently distinguishable by a ranking measure M, if there exists a negligible function $\mathrm{neg}(N)$ and $b \in \{0, 1\}$ such that for every sufficiently large N, with probability $1 - \mathrm{neg}(N)$ over the random draw of the dataset S_n, the inequality

$$M(f_b, S_n) > M(f_{1-b}, S_n)$$

holds for all $n \geqslant N$ simultaneously.

Main results for NDCG

Consistent distinguishability of standard NDCG:

Theorem. For every pair of scoring functions f_0, f_1, let $\bar{y}^{f_i}(s) = \Pr[Y = 1 \mid \tilde{f}_i(X) = s]$, $i = 0, 1$, where $\tilde{f}_i(X)$ is the canonical version of ranking function f_i defined as $\tilde{f}_i(x) := \Pr[f_i(X) \leqslant f_i(x)]$. If $\bar{y}^{f_i}(s)$ are Hölder continuous in s, then f_0 and f_1 are consistently distinguishable by standard NDCG unless $\bar{y}^{f_0} = \bar{y}^{f_1}$ almost everywhere.

Consistent distinguishability of NDCG with polynomial discount:

- For NDCG with discount $D_r = r^{-\beta}$, $0 < \beta < 1$, every pair of substantially different scoring functions are consistently distinguishable.

- For NDCG with discount $D_r = r^{-\beta}$, $\beta > 1$, there is no consistent distinguishability. Moreover, in general $\mathrm{NDCG}_D(f, S_n)$ does not converge as $n \to \infty$!

- For NDCG with discount $D_r = r^{-1}$ (so-called Zipfian), it is not clear whether consistent distinguishability holds. $\mathrm{NDCG}_D(f, S_n)$ converges as $n \to \infty$.

Results for NDCG@k

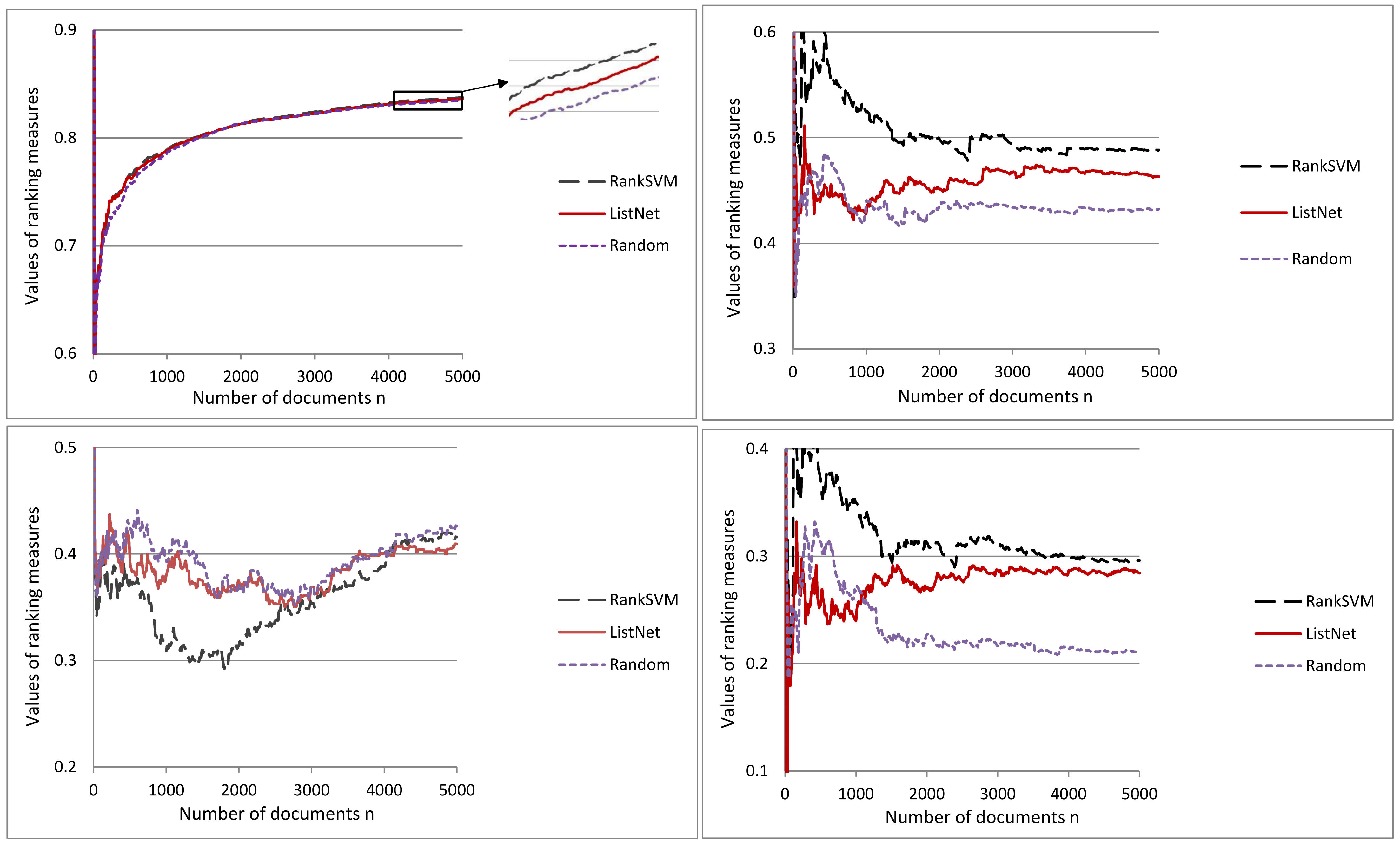
Consistent distinguishability of NDCG@k

- NDCG@k: $D_r = 0$ for $r > k$. Only consider top-k ranking results.
- Consistent distinguishability depends not only on the discount, but on the choice of k.
- Three choices of k: k is a constant; or $k = o(n)$; or $k = cn$ for some constant c.

Summarization of the results

Measure	$D(r)$	$\lim_{n\to\infty} \mathrm{NDCG}_D(f, S_n)$	Distinguishability
Standard NDCG	$\frac{1}{\log(1+r)}$	1	Yes
Inverse polynomial	$r^{-\beta}$, $\beta \in (0,1)$	$\frac{(1-\beta)\int_0^1 \bar{y}^f(s)\cdot(1-s)^{-\beta}ds}{p^{1-\beta}}$	Yes, if $\int_0^1 \Delta\bar{y}(s) \cdot (1-s)^{-\beta} ds \neq 0$.
Zipfian	r^{-1}	$\Pr[Y = 1 \mid \tilde{f}(X) = 1]$	Unknown
NERU	$2^{-O(r)}$	Does not exist	No
NDCG@k, $k = O(1)$	$\frac{1}{\log(1+r)}$, $r \leqslant k$	Does not exist	No
NDCG@k, $k = o(n)$	$\frac{1}{\log(1+r)}$, $r \leqslant k$	$\Pr[Y = 1 \mid \tilde{f}(X) = 1]$	Unknown
NDCG@k, $k = cn$	$\frac{1}{\log(1+r)}$, $r \leqslant k$	$\frac{c\cdot\Pr[Y=1 \mid \tilde{f}(X)\geqslant 1-c]}{\min\{c,p\}}$	Yes
NDCG@k, $k = cn$	$r^{-\beta}$, $r \leqslant k$	$\frac{(1-\beta)\int_{1-c}^1 = \bar{y}^f(s)\cdot(1-s)^{-\beta}ds}{(\min\{c,p\})^{1-\beta}}$	Yes, if $\int_{1-c}^1 \Delta\bar{y}(s) \cdot (1-s)^{-\beta} ds \neq 0$.

Experiments



Left to right, top to bottom: standard, polynomial with $\beta = 0.5$, NERU and standard NDCG@k with $k = 0.2n$.

antoniowyn@gmail.com wanglw@cis.pku.edu.cn