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; IILS, Tsinghua University; Microsoft Research Asia

Background of Ranking

Model of ranking (for web search):
- $X$: Instance space.
- $I$: A set of degrees of relevancy.
- $f: X \rightarrow R$: a scoring function.
- Ranking with $f$: sorting by the scores $x_1, \ldots, x_n \rightarrow \{x'_1, \ldots, x'_{n}\}$, so that $f(x'_1) \geq \ldots \geq f(x'_{n})$.

Evaluate the performance of $f$ by NDCG:
- Given dataset $S_n = \{(x_1, y_1), \ldots, (x_n, y_n)\}$.
- Obtain ranking $x'_1, \ldots, x'_{n}$.
- Degree of relevance: $y'_1, \ldots, y'_{n}$.
- Discounted Cumulative Gain (DCG): $\text{DCG}(f, S_n) = \sum_{i=1}^{n} D_i \cdot y'_i$.
- Discount: $D_i = \frac{1}{\log(1+i)}$.
- Normalized DCG (NDCG):
  $\text{NDCG}(f, S_n) = \frac{\text{DCG}(f, S_n)}{\text{IDCG}(S_n)} \in [0, 1]$. Where $\text{IDCG}(S_n)$ is the ideal DCG value on $S_n$.
- NDCG: $\text{NDCG}(f, S_n) = \max_f \text{NDCG}(f, S_n)$.

Questions:
- Are NDCG family good ranking measures?
- How to set discount $D_i$ for NDCG?

A first observation (Motivation):

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

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

**Theorem.** Let $D_i = \frac{1}{\log(1+i)}$. Then for every ranking function $f$, $\lim_{n \rightarrow \infty} \text{NDCG}(f, S_n) \rightarrow 1$ a.s.

Consistent distinguishability of 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 $\text{neg}(N)$ and $\beta \in [0, 1]$ such that for every sufficiently large $N$, with probability $1 - \text{neg}(N)$ over the random draw of the dataset $S_n$, the inequality $M(f_0, S_n) > M(f_1, S_n)$ holds for all $n \geq N$ simultaneously.

**Main results for NDCG**

Consistent distinguishability of standard NDCG:

**Theorem.** For every pair of scoring functions $f_0, f_1$, let $\mathcal{F}(s) = \Pr[Y = 1 | f_i(X) = s], i = 0, 1$, where $f_i(X)$ is the canonical version of ranking function $f_i$ defined as $f_i(x) = \Pr[y_i|f_i(x) \leq f_i(x)]$. If $\mathcal{F}(s)$ are $\alpha$-Hölder continuous in $s$, then $f_0$ and $f_1$ are consistently distinguishable by standard NDCG unless $\mathcal{F}_0 = \mathcal{F}_1$ almost everywhere.

Consistent distinguishability of NDCG with polynomial discount:

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

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

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

**Results for NDCG@k**

Consistent distinguishability of NDCG@k

- NDCG@k: $D_i = 0$ 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 $c$.

**Summary of the results**

<table>
<thead>
<tr>
<th>Measure</th>
<th>$D(r)$</th>
<th>$\lim_{n \rightarrow \infty} \text{NDCG}(f, S_n)$</th>
<th>Distinguishability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard NDCG</td>
<td>$\frac{1}{\log(1+\beta)}$</td>
<td>$\frac{1}{(1-\beta)\int_{s(1-\beta)^{-\beta}}}$</td>
<td>Yes, if $\int_{1}^{\infty} \Delta \mathcal{F}(s) \cdot (1-s)^{-\beta} ds \neq 0$</td>
</tr>
<tr>
<td>Inverse polynomial</td>
<td>$r^{-\beta}$, $\beta \in (0, 1)$</td>
<td>$\Pr[Y = 1</td>
<td>f_i(X) = 1]$</td>
</tr>
<tr>
<td>Zipfian</td>
<td>$r^{-1}$</td>
<td>$\Pr[Y = 1</td>
<td>f_i(X) = 1]$</td>
</tr>
<tr>
<td>NERU</td>
<td>$2^{-O(r)}$</td>
<td>Does not exist</td>
<td>No</td>
</tr>
<tr>
<td>NDCG@k, $k = O(1)$</td>
<td>$\frac{1}{\log(1+\beta)}$, $r \leq k$</td>
<td>Does not exist</td>
<td>No</td>
</tr>
<tr>
<td>NDCG@k, $k = o(n)$</td>
<td>$\frac{1}{\log(1+\beta)}$, $r \leq k$</td>
<td>$\Pr[Y = 1</td>
<td>f_i(X) = 1]$</td>
</tr>
<tr>
<td>NDCG@k, $k = cn$</td>
<td>$\frac{1}{\log(1+\beta)}$, $r \leq k$</td>
<td>$\frac{c \Pr[Y = 1</td>
<td>f_i(X) \leq 1]}{\min(c, p)}$</td>
</tr>
<tr>
<td>NDCG@k, $k = cn$</td>
<td>$r^{-\beta}$, $r \leq k$</td>
<td>$\frac{1}{(1-\beta)\int_{s(1-\beta)^{-\beta}}}$</td>
<td>Yes, if $\int_{1}^{\infty} \Delta \mathcal{F}(s) \cdot (1-s)^{-\beta} ds \neq 0$</td>
</tr>
</tbody>
</table>

**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