A Theoretical Analysis of NDCG Family Ranking Measures

Yining Wang, Liwei Wang,

Yuanzhi Li, Di He, Wei Chen, Tie-Yan Liu

Peking University

Tsinghua University

Microsoft Research Asia

Ranking: an example

Google | Princeton University

Princeton University - Home
www.princeton.edu/
Princeton University is a vibrant community of scholarship and learning that stands in the nation's service and in the service of all nations.
Score: 24 / 30 · 59 Google reviews · Write a review

Princeton University - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Princeton_University
Princeton University is a private Ivy League research university located in Princeton, New Jersey, United States. Princeton has a long history. Founded in 1746 ...
History - List of Princeton University ... - Eating clubs - Chapel

Princeton University evacuated after bomb threat - CNN.com
www.cnn.com/2013/06/11/us/new-jersey-princeton.../index.html
8 hours ago – Princeton University on Tuesday morning evacuated its campus after "a bomb threat to multiple unspecified campus buildings," the school said ...

Images for princeton university - Report images

Model of Ranking

- An instance space X ;
- A finite set of degree of relevancy Y ;
- A ranking (scoring) function $f: X \rightarrow \Re$;
- Ranking a set $S_n = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ with f :
  - Sorting $x_1, \ldots, x_n$ with f:

$$x_1, \ldots, x_n \xrightarrow{f} x^f_{(1)}, \ldots, x^f_{(n)}$$

$$f(x^f_{(1)}) \geq f(x^f_{(2)}) \ldots \geq f(x^f_{(n)})$$

$$y_1, \ldots, y_n \longrightarrow y^f_{(1)}, \ldots, y^f_{(n)}$$

How to evaluate the ranking function?

- Normalized Discounted Cumulative Gain (**NDCG**): widely used in web search
  - Discounted Cumulative Gain (**DCG**)

    Given a discount function $D(r),\ r = 1,2,\ldots,$ the DCG of a ranking function f on a data set $S_n = \{(x_1, y_1),\ldots,(x_n, y_n)\}$ is defined as:

$$DCG_D(f, S_n) = \sum_{r=1}^{n} y_{(r)}^{f} D(r)$$

  - **NDCG**:

$$NDCG_D(f, S_n) = \frac{DCG_D(f, S_n)}{IDCG_D(f, S_n)}, \qquad IDCG_D(f, S_n) = \max_{f'} DCG_D(f', S_n)$$

  - Commonly used discounts: logarithmic discount (**Standard NDCG**)

$$D(r) = \frac{1}{\log(1 + r)}$$

Some natural questions

- Are the NDCG family good evaluation measures?

- Why do we use the logarithmic discount? Any theoretical justifications?

- Is there any criterion to determine whether a ranking evaluation measure is good?

A first observation (motivation)

- The standard NDCG (log discount) always converges to the **same** limit.

- **Theorem.** Let $D(r) = \frac{1}{\log(1+r)}$. Assume S_n consists of i.i.d. data. Then for **every** ranking function f

$$NDCG_D(f, S_n) \xrightarrow{n \to \infty} 1, \qquad \text{a.s.}$$

Standard NDCG is difficult to differentiate ranking functions on large data sets?

Proposed criterion: Consistent Distinguishability

- Consistent Distinguishability (intuition):

For every pair of substantially different ranking functions, the ranking measure should be able to **consistently** distinguish the two functions on almost all data sets.

- Definition:

Let $(x_1, y_1), (x_2, y_2), \ldots$ be i.i.d. drawn from an underlying distribution. Let $S_n = \{(x_1, y_1), \ldots, (x_n, y_n)\}$. A pair of ranking functions f_0, f_1 is said to be **consistently distinguishable** by a ranking measure M, if there exists a negligible function $neg(N)$ and $b \in \{0,1\}$ so that for every sufficiently large N, with probability $1 - neg(N)$

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

holds for all $n > N$ simultaneously.

Main results (informal)

- **Standard NDCG** (log discount):

  Under natural conditions, every pair of substantially different ranking functions is consistently distinguishable by standard NDCG (although the measure converges to the same limits).

- **Characterization of feasible discounts of NDCG:**

  - For polynomial discount $D(r) = r^{-\beta},\ \beta \in (0,1)$, under weak conditions every pair of substantially different ranking functions is consistently distinguishable by NDCG with this discount.

  - For polynomial discount $D(r) = r^{-\beta},\ \beta > 1$, NDCG does not have consistent distinguishability. Moreover, $NDCG_D(f, S_n)$ does not converge in general.

Thanks!