RankDCG: Rank-Ordering Evaluation Measure

Denys Katerenchuk, Andrew Rosenberg
dkaterenchuk@gradcenter.cuny.edu
The Graduate Center, The City University of New York

Abstract

Ranking is used for a wide array of problems, most notably information retrieval (search). Kendall's τ, Average Precision, and nDCG are a few popular approaches to the evaluation of ranking. When dealing with problems such as user ranking or recommendation systems, all these measures suffer from various problems, including the inability to deal with elements of the same rank, inconsistent and ambiguous lower bound scores, and an inappropriate cost function. We propose a new measure, a modification of the popular nDCG algorithm, named rankDCG, that addresses these problems. We release a publicly available rankDCG evaluation package (http://www.dk-lab.com/tools/).

Problem and Constraints

Problem:

Given a list of elements A = [x1, x2, x3, ..., xn], the objective is to find list B = [rel(x1) ≥ rel(x2) ≥ rel(x3) ≥ ... ≥ rel(xn)] where rel(x) is a rank function that returns rank r with r ∈ N and n is a number of elements.

Constraints:

1. correctly work with multiple ties
2. address non-normal rank value distribution
3. emphasize correct ordering of high rank elements
4. produce consistent and meaningful scoring range

Common Metrics

1. F-measure

$$F = 2 * \frac{p * r}{p + r}$$

2. Average Precision and Mean Average Precision

$$AP = \sum_{k=1}^{n} P(k) * \Delta R(k)$$

$$MAP = \frac{\sum_{q \in Q} AP(q)}{|Q|}$$

3. Kendall's τ

$$\tau = \frac{c - d}{\frac{1}{2}n(n-1)}$$

4. Discounted Cumulative Gain

$$DCG = \sum_{i=1}^{n} \frac{rel(x_i)}{log_2(i+1)}$$

$$nDCG = \frac{DCG}{IDCG}$$

RankDCG

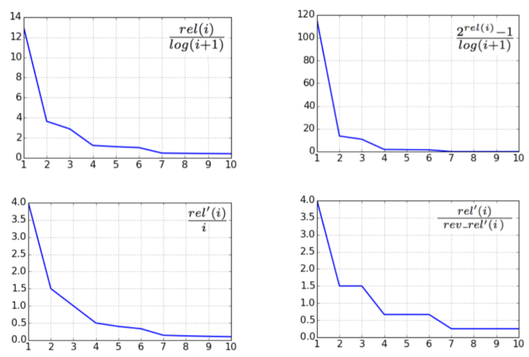
$$L = [9_1, 4_2, 4_3, 2_4, 2_5, 2_6, 1_7, 1_8, 1_9, 1_{10}]$$

$$L' = [4_1, 3_2, 3_3, 2_4, 2_5, 2_6, 1_7, 1_8, 1_9, 1_{10}]$$

$$D = [1_1, 2_2, 2_3, 3_4, 3_5, 3_6, 4_7, 4_8, 4_9, 4_{10}]$$

$$DCG' = \sum_{i=1}^{n} \frac{rel'(i)}{rev\_rel'(i)}$$

$$rankDCG = \frac{max(DCG') - DCG'}{max(DCG') - min(DCG')}$$



Test Results

#	Hypothesis List	Kendall's τ	AveP	nDCG	rankDCG
1	[9, 4, 4, 2, 2, 2, 1, 1, 1, 1]	1.0	1.0	1.0	1.0
2	[9, 4, 4, 2, 2, 1, 2, 1, 1, 1]	0.8	0.887	0.998	0.975
3	[4, 4, 2, 9, 2, 2, 1, 1, 1, 1]	0.742	0.454	0.825	0.65
4	[1, 4, 4, 2, 2, 2, 9, 1, 1, 1]	0.285	0.659	0.688	0.325
5	[1, 4, 4, 2, 2, 2, 1, 1, 1, 9]	0.285	0.697	0.667	0.325
6	[1, 1, 1, 1, 2, 2, 2, 4, 4, 9]	-0.8	0.149	0.571	0.0

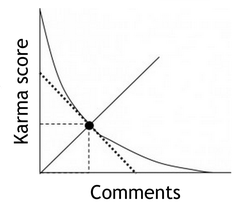
Data

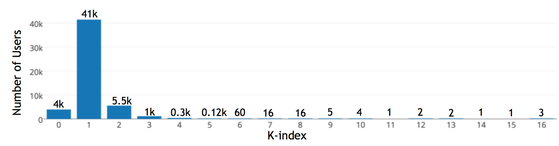
Reddit website (www.reddit.com):

- Users post comments
- Comments form threads
- ~120 participants per thread
- Comments receive karma
- Politics subreddit

Hierarchy:

- K-index - similar to h-index





Real Data Results

#	Kendall's τ	AveP	nDCG	rankDCG
1	nan	0.79	0.883	0.0
2	0.197	0.668	1.188	0.32
3	0.136	0.585	1.318	0.347
4	0.5	1	1	1

Acknowledgements

This material is based on research sponsored by DARPA under agreement number FA8750-13- 2-0041. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government.