Linear NDCG and Pair-wise Loss

Xiao-Bo Jin and Guang-Gang Geng*[1]

xbjin9801@gmail.com; gengguanggang@cnnic.cn

Abstract

Linear NDCG is used for measuring the performance of the Web content quality assessment in ECML/PKDD Discovery Challenge 2010. In this paper, we will prove that the DCG error equals a new pair-wise loss.

Keywords: NDCG, Learning to rank, Web content quality assessment

1. Linear NDCG

In ECML Discovery Challenge 2010, the evaluation measure is a variant of the NDCG ($NDCG^\beta$). Given the sorted ranking sequence g and all ratings $\{r_i\}_{i=1}^{|S|}$, the discount function and NDCG are defined as ($r_i \in \{0, 1, \ldots, L-1\}$):

$$DCG_g^\beta = \sum_{i=1}^{|S|} r_i(|S| - i) \ , \ NDCG^\beta = \frac{1}{DCG_\pi^\beta} DCG_g^\beta, \tag{1}$$

where DCG_π^β is the normalization factor that is DCG in the ideal permutation π ($DCG_g^\beta \leq DCG_\pi^\beta$). We call $\Delta DCG^\beta = DCG_\pi^\beta - DCG_g^\beta$ as the DCG error. Specially, $DCG_\pi^\beta = mn + \frac{m(m-1)}{2}$ for the bipartite ranking. It is worth noticing that the above NDCG is different from the classical NDCG for the query-dependent ranking, where the DCG function is (for the single query):

$$DCG_g^\alpha = \sum_{i=1}^{|S|} \frac{2^{r_i} - 1}{\log_2(i+1)} \ , \ NDCG^\alpha = \frac{1}{DCG_\pi^\alpha} DCG_g^\alpha. \tag{2}$$

Consider the case of the query-dependent ranking with L ratings. For the given query, the dataset S can be divided into $\{S_i\}_{i=0}^{L-1}$ according to the

[1] Guang-Gang Geng is the corresponding author.

ratings of the instances. Generally, we can define the empirical error for the multi-partite case:

$$\hat{R}(f) = \frac{1}{Z} \sum_{0 \le a < b < L} \sum_{i=1}^{|S_a|} \sum_{j=1}^{|S_b|} (b-a) I[f(\boldsymbol{x}_i^b) < f(\boldsymbol{x}_j^a)], \tag{3}$$

where $Z = \sum_{0 \le a < b < L} |S_a||S_b|$. Specially, we also define the following unnormalized empirical error:

$$R(f) = \sum_{0 \le a < b < L} \sum_{i=1}^{|S_a|} \sum_{j=1}^{|S_b|} (b-a) I[f(\boldsymbol{x}_i^b) < f(\boldsymbol{x}_j^a)]. \tag{4}$$

2. $\boldsymbol{NDCG^{\beta}}$ and Pair-wise Loss

In this section, we will prove the following conclusion:

$$\Delta DCG^{\beta} = R(f). \tag{5}$$

Theorem 1. *For L-partite ranking problem, the unnormalized empirical error can be divided into the following form:*

$$R(f) = \sum_{0 \le a < b < L} \sum_{i=1}^{|S_a|} \sum_{j=1}^{|S_b|} (b-a) I[f(\boldsymbol{x}_i^b) < f(\boldsymbol{x}_j^a)] = \sum_{k=0}^{L-2} R_k(f), \tag{6}$$

where

$$R_k(f) = \sum_{a=0}^{k} \sum_{b=k+1}^{L-1} \sum_{i=1}^{|S_a|} \sum_{j=1}^{|S_b|} I[f(\boldsymbol{x}_i^b) < f(\boldsymbol{x}_j^a)]. \tag{7}$$

Proof. For the convenience of the description, we represent the conclusion as follows:

$$\begin{aligned} G^L(f) &= \sum_{k=0}^{L-2} R_k(f) \\ &= \sum_{k=0}^{L-2} \sum_{a=0}^{k} \sum_{b=k+1}^{L-1} \sum_{i=1}^{|S_a|} \sum_{j=1}^{|S_b|} I[f(\boldsymbol{x}_i^b) < f(\boldsymbol{x}_j^a)] \\ &= R^L(f) \end{aligned} \tag{8}$$

Now we prove the conclusion $G^n(f) = R^n(f)$ with the mathematical induction on the variable n. If $n = 2$, the conclusion trivially holds. Assume that the equation is true for n, then we will prove the conclusion for $n+1$. We have

$$\begin{aligned} G^{n+1}(f) &= G^n(f) + \sum_{k=0}^{n-2}\sum_{a=0}^{k}\sum_{i=1}^{|S_a|}\sum_{j=1}^{|S_b|} I[f(\boldsymbol{x}_j^n) < f(\boldsymbol{x}_i^a)] \\ &\quad + \sum_{a=0}^{n-1}\sum_{i=1}^{|S_a|}\sum_{j=1}^{|S_b|} I[f(\boldsymbol{x}_j^n) < f(\boldsymbol{x}_i^a)] \\ &= G^n(f) + \sum_{k=0}^{n-1}\sum_{a=0}^{k}\sum_{i=1}^{|S_a|}\sum_{j=1}^{|S_b|} I[f(\boldsymbol{x}_j^n) < f(\boldsymbol{x}_i^a)] \end{aligned} \tag{9}$$

and

$$R^{n+1}(f) = R^n(f) + \sum_{a=0}^{n-1}(n-a)\sum_{i=1}^{|S_a|}\sum_{j=1}^{|S_b|} I[f(\boldsymbol{x}_j^n) < f(\boldsymbol{x}_i^a)]. \tag{10}$$

Finally, we can prove by the mathematical induction that the second item of the right side in (9) equals to the corresponding item in (10). We can see that for $n = 1$ it is trivially hold.

It follows that $G^L(f) = R^L(f)$ for all natural number with $L > 1$.

Lemma 1. *For the bipartite ranking problems, any sorted ranking sequence from $S = \{S_+, S_-\}$ can be obtained by exchanging at most $k = \min\{|S_+|, |S_-|\}$ times from the ideal ranking sequence.*

Proof. Given that there are $r(r \le m)$ negative instances in the first m positions and $s(s \le n)$ positive instances in the remain n positions.

Now we prove $s = r$ indirectly through the apagoge. If $s \ne r$, without loss of generality, we assume $r > s$. It is known that there are $r - s$ negative instances in the first m positions after s exchanges. The exchanges occur among s negative instances in the first m positions and s positive instances in the remain n positions. Then the fact that we will get $r - s + n$ negative instances is in contradiction to n negative instances. Finally, we can conclude that $r = s \le \min\{|S_+|, |S_-|\}$.

Next, we will prove

Theorem 2. *For the bipartite ranking problem, DCG errors with 1 equals the unnormalized expected losss $R(f)$:*

$$\Delta DCG^{\beta} = R(f) = \sum_{i=1}^{m}\sum_{j=1}^{n} I[f(\boldsymbol{x}_i^+) < f(\boldsymbol{x}_j^-)]. \tag{11}$$

Proof. We know that any ranking sequence can be obtained by the exchange operations from the ideal ranking sequence according to Prop. 1. Let $\{i_1, i_2, \cdots, i_k\}(1 \le i_1 < i_2 < \cdots < i_k \le m)$ and $\{j_1, j_2, \cdots, j_k\}(1 \le j_1 < j_2 < \cdots < j_k \le n)$ be the exchanged positions in the first m positions and the remain n positions, respectively. As depicted in Fig. 1, without loss of generality, we exchange i_r and j_r for the r-th time. First, we will compute the decrement relative to the ideal ranking sequence for the r-th time

$$\begin{aligned} \Delta_r DCG &= (m+n-i_r) - (m+n-(m+j_r)) \\ &= m + j_r - i_r >= 1. \end{aligned} \tag{12}$$

Now, we give a detailed explanation about the increment of the unnormalize expected loss which is related to the position i_r and j_r. The increment due to the variation in the position i_r will be $m - i_r + r$ because there are $m - i_r$ positive instances in the first m positions and r positive instances in the remain n instances. As for the position j_r, the increment should be $j_r - r$ since there are $j_r - 1 - (r-1)$ negative instances in the remain n instances before j_r. In summary, we obtain the increment $\Delta_r R(f) = m + j_r - i_r$. As a result, we conclude that

$$\Delta DCG^{\beta} = \sum_{r=1}^{k} \Delta_r DCG = \sum_{r=1}^{k} \Delta_r R(f) = \Delta R(f). \tag{13}$$

Notice that the initial value of $R(f)$ (the ideal ranking sequence) is zero, this proves the theorem.

Fig. 2 gives an example to verify the conclusion $\Delta DCG^{\beta} = R(f) = 4$. The following theorem shows that the conclusion $\Delta DCG^{\beta} = R(f)$ still holds when extending to the multi-partite ranking problem.

Theorem 3. *For L-partite ranking problem, the DCG errors with Eqn. (1) equals $R(f)$:*

$$\Delta DCG^{\beta} = \sum_{0 \le a < b < L} \sum_{i=1}^{|S_a|}\sum_{j=1}^{|S_b|} (b-a) I[f(\boldsymbol{x}_i^b) < f(\boldsymbol{x}_j^a)). \tag{14}$$

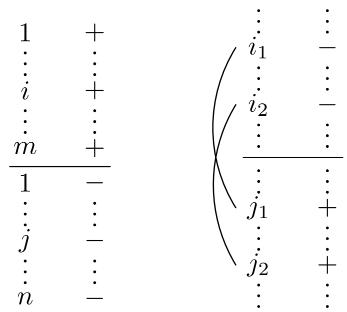

Figure 1: The ideal ranking sequence with its transformation. Left: the ideal ranking sequence, right: the ranking sequence with multiple exchanges

π	+	+	+	−	−	−
i	1	2	3	4	5	6
g	+	−	−	+	+	−

Figure 2: The example on the bipartite ranking shows $\Delta DCG^{\beta} = R(f) = 4$, where $DCG_{\pi} = 12$ and $DCG_{g} = 8$.

PROOF. From 1, we know that

$$R(f) = G(f) = \sum_{k=0}^{L-2} R_k(f). \tag{15}$$

Then we will show that DCG in L-partite problem can be written as the sum of the DCG measures of $L-1$ bipartite problems. We divide DCG_β into

$$\begin{aligned} DCG_\beta &= \sum_{i=1}^{|S|} r_i(|S|-i) \\ &= \sum_{i=1}^{|S|}\sum_{k=0}^{L-2} I[k<r_i](|S|-i) \\ &= \sum_{k=0}^{L-2} DCG_k, \end{aligned} \tag{16}$$

where $DCG_k = \sum_{i=1}^{|S|} I[k < r_i](|S|-i)$. For given k, we can assign the instances with r_i $(k < r_i)$ to the ranking 1 and the others to the ranking 0 to obtain a bipartite ranking problem with the unnormalized empirical error

$$R_k(f) = \sum_{a=0}^{k} \sum_{b=k+1}^{L-1} \sum_{i=1}^{|S_a|} \sum_{j=1}^{|S_b|} I[f(\mathbf{x}_i^b) < f(\mathbf{x}_j^a)]. \tag{17}$$

From 2, $\Delta DCG_k = R_k(f)$ holds. We have $\Delta DCG = \sum_{k=0}^{L-2} \Delta DCG_k = \sum_{k=0}^{L-2} R_k(f) = R(f)$.

π	2	2	1	0	0	0
i	1	2	3	4	5	6
g	2	0	2	1	0	0

Figure 3: The example on the multipartite ranking shows $\Delta DCG_\beta = R(f) = 3$, where $DCG_\pi^\beta = 21$ and $DCG_g^\beta = 18$.

The example in 3 supports our conclusion about the DCG error and the unnormalized expected loss in the multipartite ranking problem.