

Updates and Corrections to

Incorporating User Expectations and Behavior into the Measurement of Search Effectiveness

ACM Transactions on Information Systems, 35(3):24:1–24:38, June 2017

Page 14: Definition of ERR (Expected Reciprocal Rank)

We wrote:

Different metrics can then be defined via the use of one (and hence all) of these three equivalent forms. For example,

$$C_{ERR}(i) = (1 - r_i)$$

is a definition for Expected Reciprocal Rank (ERR) (Chapelle et al. 2009), taking the gain per document (mapped to the range zero to one) as a probability of halting at this point in the ranking.

In fact, it isn't quite that easy. The problem is that in the original model for ERR the r_i value is taken as the probability of deriving *full* gain at depth i , that is, the probability that the user stops at that rank and says, “aha, found something suitable”. And while it might *look* like this must be the same as saying $C_{ERR}(i) = (1 - r_i)$, it doesn't quite work out that way, because in the C/W/L metrics the $W(i)$ values get multiplied by the r_i values (second time r_i comes into it) and then summed, whereas in ERR there is a multiplier of $1/i$ to be found from somewhere.

Nor does there seem to be a solution that does describe ERR in the C/W/L framework. For example, consider a gain vector in which every value $r_i = \alpha$, for some $0 < \alpha < 1$. No matter what $C(i)$ function gets chosen, every C/W/L metric score on that gain vector has to be α , because the $W(i)$ values – whatever they are – by definition must sum to one. But for these gain vectors

$$ERR = \sum_{i=1}^{\infty} \left[\left(\prod_{j=1}^{i-1} (1 - r_j) \right) \frac{r_i}{i} \right] = \sum_{i=1}^{\infty} \left[(1 - \alpha)^{i-1} \frac{\alpha}{i} \right] = \alpha + \sum_{i=2}^{\infty} \left[(1 - \alpha)^{i-1} \frac{\alpha}{i} \right] > \alpha.$$

There might be a C/W/L metric that has very similar properties to ERR and hence can be used as a surrogate for it, but it does look like ERR is simply not describable in the C/W/L framework. This discovery was made in the course of conversations with Alfian Wicaksono.

This original note was written in September 2019; we then explored this area more carefully in a subsequent paper, see Azzopardi et al. (2021).

Page 17: Definition of INST

The definition of $Rel(i)$ immediately after Equation 3 has a typo, the subscript should be a j , as in: $Rel(i) = \sum_{j=1}^i r_j$. This was noted by the author in December 2021.

References

- L. Azzopardi, J. Mackenzie, and A. Moffat. ERR is not C/W/L: Exploring the relationship between expected reciprocal rank and other metrics. In *Proc. International Conf. on Theory of Information Retrieval (ICTIR)*, pages 231–237, 2021.
- O. Chapelle, D. Metzler, Y. Zhang, and P. Grinspan. Expected reciprocal rank for graded relevance. In *Proc. ACM International Conf. on Information and Knowledge Management (CIKM)*, pages 621–630, 2009.

Alistair Moffat,
ammoffat@unimelb.edu.au,
December 9, 2021