Cumulated Relative Position: A Metric for Ranking Evaluation (Extended Abstract)^{*}

Marco Angelini³, Nicola Ferro¹, Kalervo Järvelin², Heikki Keskustalo², Ari Pirkola², Giuseppe Santucci³, and Gianmaria Silvello¹

> ¹ University of Padua, Italy {ferro,silvello}@dei.unipd.it ² University of Tampere, Finland {kalervo.jarvelin,heikki.keskustalo,ari.pirkola}@uta.fi ³ "La Sapienza" University of Rome, Italy {angelini,santucci}@dis.uniroma1.it

Abstract. The development of multilingual and multimedia information access systems calls for proper evaluation methodologies to ensure that they meet the expected user requirements and provide the desired effectiveness. In this paper, we propose a new metric for ranking evaluation, the CRP.

1 Introduction and Motivations

The development of information access systems calls for proper evaluation methodologies in particular for what is concerned with the evaluation of rankings. A range of evaluation metrics, such as MAP and nDCG, are widely used and they are particularly suitable to the evaluation of *Information Retrieval (IR)* techniques in terms of the quality of the output ranked lists, and often to some degree suitable to the evaluation of user experience regarding retrieval. Unfortunately, the traditional metrics do not take deviations from optimal document ranking sufficiently into account. We think that a proper evaluation metric for ranked result lists in IR should: (a) explicitly handle graded relevance including negative gains for unhelpful documents, and (b) explicitly take into account document misplacements in ranking either too early or too late given their degree of relevance and the optimal ranking. In the present paper, we propose such a new evaluation metric, the *Cumulated Relative Position (CRP)*.

We start with the observation that a document of a given degree of relevance may be ranked too early or too late regarding the ideal ranking of documents for a query. Its relative position may be negative, indicating too early ranking, zero indicating correct ranking, or positive, indicating too late ranking. By cumulating these relative rankings we indicate, at each ranked position, the net effect of document displacements, the CRP. CRP explicitly handles: (a) graded

^{*} The extended version of this abstract has been published in [1].

relevance, and (b) document misplacements either too early or too late given their degree of relevance and the ideal ranking. Thereby, CRP offers several advantages in IR evaluation: (i) at any number of retrieved documents examined (rank) for a given query, it is obvious to interpret and it gives an estimate of ranking performance; (ii) it is not dependent on outliers since it focuses on the ranking of the result list; (iii) it is directly user-oriented in reporting the deviation from ideal ranking when examining a given number of documents; the effort wasted in examining a suboptimal ranking is made explicit.

2 Definition of Cumulated Relative Position

We define the set of *relevance degrees* as (REL, \leq) such that there is an order between the elements of *REL*. For example, for the set $REL = \{nr, pr, fr, hr\}$, *nr* stands for "non relevant", *pr* for "partially relevant", *fr* for "fairly relevant", *hr* stands for "highly relevant", and it holds $nr \leq pr \leq fr \leq hr$.

We define a function RW : $REL \to \mathbb{Z}$ as a monotonic function which maps each relevance degree ($rel \in REL$) into an *relevance weight* ($w_{rel} \in \mathbb{Z}$), e.g. RW(hr) = 3. This function allows us to associate an integer number to a relevance degree.

We define with D the set of documents we take into account, with $N \in \mathbb{N}$ a natural number, and with D^N the set of all possible vectors of length Ncontaining different orderings of the documents in D. We can also say that a vector in D^N represents a ranking list of length N of the documents D retrieved by an IR system. Let us consider a vector $\mathbf{v} \in D^N$, a natural number $j \in [1, N]$, and a relevance degree $rel \in REL$, then the ground truth function is defined as:

$$\begin{array}{l} \operatorname{GT} : D^N \times \mathbb{N} \to REL \\ \mathbf{v}[j] \mapsto rel \end{array} \tag{1}$$

Equation 1 allows us to associate a relevance degree to the document $d \in D$ retrieved at position j of the vector \mathbf{v} , i.e. it associates a relevance judgment to each retrieved document in a ranked list.

In the following, we define with $\mathbf{r} \in D^N$ the vector of documents retrieved and ranked by a run r, with $\mathbf{i} \in D^N$ the ideal vector containing the best ranking of the documents in the pool (e.g. all highly relevant documents are grouped together in the beginning of the vector followed by fairly relevant ones and so on and so forth), and with $\mathbf{w} \in D^N$ the worst-case vector containing the worst rank of the documents retrieved by the pool (e.g. all the relevant documents are put in the end of the vector in the inverse relevance order).

From function GT we can point out a set called *relevance support* defined as:

$$RS(\mathbf{v}, rel) = \{ j \in [1, N] \mid \mathrm{GT}(\mathbf{v}, j) = rel \}$$

$$\tag{2}$$

which, given a vector $\mathbf{v} \in D^N$ – it can be a run vector \mathbf{r} , the ideal vector \mathbf{i} , or the worst-case vector \mathbf{w} – and a relevance degree *rel*, contains the indexes j

of the documents of **v** with which the given relevance degree (rel) relevance is associated.

Given the ideal vector \mathbf{i} and a relevance degree rel, we can define the *minimum rank* in \mathbf{i} as the first position in which we find a document with relevance degree equal to rel. In the same way, we can define the *maximum rank* in \mathbf{i} as the last position in which we find a document with relevance degree equal to rel. In formulas, they become:

$$\min_{\mathbf{i}}(rel) = \min\left(RS(\mathbf{i}, rel)\right)$$

$$\max_{\mathbf{i}}(rel) = \max\left(RS(\mathbf{i}, rel)\right)$$
(3)

Given a vector **v** and a document at position $j \in [1, N]$, we can define the Relative Position (RP) as:

$$\operatorname{RP}(\mathbf{v},j) = \begin{cases} 0 & \operatorname{if} \min_{\mathbf{i}} \left(\operatorname{GT}(\mathbf{v},j) \right) \leq j \leq \max_{\mathbf{i}} \left(\operatorname{GT}(\mathbf{v},j) \right) \\ j - \min_{\mathbf{i}} \left(\operatorname{GT}(\mathbf{v},j) \right) & \operatorname{if} j < \min_{\mathbf{i}} \left(\operatorname{GT}(\mathbf{v},j) \right) \\ j - \max_{\mathbf{i}} \left(\operatorname{GT}(\mathbf{v},j) \right) & \operatorname{if} j > \max_{\mathbf{i}} \left(\operatorname{GT}(\mathbf{v},j) \right) \end{cases}$$
(4)

RP allows for pointing out misplaced documents and understanding how much they are misplaced with respect to the ideal case **i**. Zero values denote documents which are within the ideal interval, positive values denote documents which are ranked below their ideal interval, and negative values denote documents which are above their ideal interval. Note that the greater the absolute value of $\operatorname{RP}(\mathbf{v}, j)$ is, the bigger is the distance of the document at position jfrom its ideal interval. From equation 4, it follows that $\operatorname{RP}(\mathbf{i}, j) = 0, \forall j \in [1, N]$.

Given a vector \mathbf{v} and a document at position $j \in [1, N]$, we can define the Cumulated Relative Position (CRP) as:

$$CRP(\mathbf{v}, j) = \sum_{k=1}^{j} RP(\mathbf{v}, k)$$
(5)

For each position j, CRP sums the values of RP up to position j included. From equation 5, it follows that $CRP(\mathbf{i}, j) = 0, \forall j \in [1, N].$

We can point out the following properties for CRP:

- CRP can only be zero or negative before reaching the rank of the recall base (R);
- the faster the CRP curve goes down before R, the worse the run is;
- after R the CRP curve is non-decreasing;
- after that the last relevant document has been encountered, CRP remains constant;
- the sooner we reach the x-axis (balance point: $b_{\mathbf{r}}$), the better the run is.

In Figure 1 we can see a sketch of the CRP for a topic of a run. For a given topic there are two fixed values which are the rank of recall base (R) and the



Fig. 1. Cumulative Relative Position sketch for a topic of a given run (on the left) and the CRP curve of a real run taken from TREC7.

number of retrieved documents (N); this allows us to compare systems on the R basis.

The principal indicator describing the CRP curve of a topic for a given run which is the *recovery value* (ρ) defined as the ratio between R and $b_{\mathbf{r}}$: $\rho = \frac{R}{b_{\mathbf{r}}}$.

The recovery-value is always between 0 and 1 ($0 < \rho \leq 1$) where $\rho = 1$ indicates a perfect ranking and $\rho \to 0$ a progressively worse ranking. Please note that $\rho \to 0$ when $b_{\mathbf{r}} \to \infty$.

3 Final Remarks

We think that the CRP offers several advantages in IR evaluation because (a) it is obvious to interpret and it gives an estimate of ranking performance as a single measure; (b) it is independent on outliers since it focuses on the ranking of the result list; (c) it directly reports the effort wasted in examining suboptimal rankings; (d) it is based on graded relevance.

Acknowledgements The work reported in this paper has been supported by the PROMISE network of excellence (contract n. 258191) project as a part of the 7th Framework Program of the European commission (FP7/2007-2013).

References

 M. Angelini, N. Ferro, K. Järvelin, H. Keskustalo, A. Pirkola, G. Santucci, and G. Silvello. Cumulated Relative Position: A Metric for Ranking Evaluation. In Information Access Evaluation meets Multilinguality, Multimodality, and Visual Analytics. Proc. of the 3rd Int. Conf. of the CLEF Initiative (CLEF 2012). Lecture Notes in Computer Science (LNCS) 7488, Springer, Heidelberg, Germany, 2012.