Expected Average Precision

Javed A. Aslam, Emine Yilmaz, Virgil Pavlu
College of Computer and Information Science
Northeastern University

October 16, 2004

For a query Q with R relevant documents, the *average precision* of a given ranked list of documents is the sum of the precisions at relevant documents within that list, divided by R.

For any document list of length k and for all i, $1 \leq i \leq k$, let $x_i \in \{0, 1\}$ denote the relevance of the document at rank i. We may then define the sum of the precisions at relevant documents as follows.

$$\begin{aligned} sp(x_1, \ldots, x_k) &= \sum_{i=1}^{k} \left(\frac{x_i}{i} \sum_{j=1}^{i} x_j \right) \\ &= \sum_{i=1}^{k} \sum_{j=1}^{i} \frac{x_i x_j}{i} \end{aligned}$$

We note that

$$\begin{aligned} sp(x_1, \ldots, x_k) &= \sum_{i=1}^{k-1} \left(\frac{x_i}{i} \sum_{j=1}^{i} x_j \right) + \frac{x_k}{k} \sum_{j=1}^{k} x_j \\ &= sp(x_1, \ldots, x_{k-1}) + \frac{x_k}{k} \sum_{j=1}^{k} x_j. \end{aligned}$$

As a consequence, we have

$$sp(x_1, \ldots, x_{k-1}, 0) = sp(x_1, \ldots, x_{k-1}) \tag{1}$$

and

$$sp(x_1, \ldots, x_{k-1}, 1) = sp(x_1, \ldots, x_{k-1}) + \frac{1}{k} \left(1 + \sum_{j=1}^{k-1} x_j \right). \tag{2}$$

Now let $p(x_1, \ldots, x_n)$ denote a joint distribution over the relevances associated with document lists of length n. The *expected* sum precision is then

$$\sum_{x_1, \ldots, x_n} sp(x_1, \ldots, x_n)\, p(x_1, \ldots, x_n).$$

Now assume that $p(x_1, \ldots, x_n)$ is a *product distribution*; i.e.,

$$p(x_1, \ldots, x_n) = p_1(x_1) \cdot p_2(x_2) \cdots p_n(x_n).$$

For notational convenience, let $p_i = p_i(1)$ for all i. In other words, p_i is the probability that the document at rank i is relevant. We now prove the following claim.

Claim 1 *Given a product distribution $p(x_1, \ldots, x_n)$ over the relevances associated with document lists of length n, the expected sum precision is*

$$\sum_{i=1}^{n} \frac{p_i}{i} \left(1 + \sum_{j=1}^{i-1} p_j \right).$$

Proof: The expected sum precision can be calculated as follows.

$$
\begin{aligned}
&\sum_{x_1,\ldots,x_n} sp(x_1,\ldots,x_n)\, p(x_1,\ldots,x_n) \\
&= \sum_{x_1,\ldots,x_n} sp(x_1,\ldots,x_n)\, p(x_1,\ldots,x_{n-1}) \cdot p_n(x_n) \\
&= \sum_{x_1,\ldots,x_{n-1}} sp(x_1,\ldots,x_{n-1},0)\, p(x_1,\ldots,x_{n-1}) \cdot p_n(0) + \\
&\quad \sum_{x_1,\ldots,x_{n-1}} sp(x_1,\ldots,x_{n-1},1)\, p(x_1,\ldots,x_{n-1}) \cdot p_n(1) \\
&= \sum_{x_1,\ldots,x_{n-1}} sp(x_1,\ldots,x_{n-1})\, p(x_1,\ldots,x_{n-1}) \cdot (1-p_n) + \\
&\quad \sum_{x_1,\ldots,x_{n-1}} \left(sp(x_1,\ldots,x_{n-1}) + \frac{1}{n}\left(1 + \sum_{j=1}^{n-1} x_j\right)\right) p(x_1,\ldots,x_{n-1}) \cdot p_n \\
&= \sum_{x_1,\ldots,x_{n-1}} sp(x_1,\ldots,x_{n-1})\, p(x_1,\ldots,x_{n-1}) \cdot (1-p_n) + \\
&\quad \sum_{x_1,\ldots,x_{n-1}} sp(x_1,\ldots,x_{n-1})\, p(x_1,\ldots,x_{n-1}) \cdot p_n + \\
&\quad \sum_{x_1,\ldots,x_{n-1}} \frac{1}{n}\left(1 + \sum_{j=1}^{n-1} x_j\right) p(x_1,\ldots,x_{n-1}) \cdot p_n \\
&= \sum_{x_1,\ldots,x_{n-1}} sp(x_1,\ldots,x_{n-1})\, p(x_1,\ldots,x_{n-1}) + \\
&\quad \frac{p_n}{n}\left(1 + \sum_{j=1}^{n-1}\left(\sum_{x_1,\ldots,x_{n-1}} x_j\, p(x_1,\ldots,x_{n-1})\right)\right) \\
&= \sum_{x_1,\ldots,x_{n-1}} sp(x_1,\ldots,x_{n-1})\, p(x_1,\ldots,x_{n-1}) + \frac{p_n}{n}\left(1 + \sum_{j=1}^{n-1} p_j\right)
\end{aligned}
$$

Thus, we have a *recurrence* for the expected sum precision. Iterating this recurrence, we obtain

$$\sum_{x_1,\ldots,x_n} sp(x_1,\ldots,x_n)\, p(x_1,\ldots,x_n) = \sum_{i=1}^{n} \frac{p_i}{i}\left(1 + \sum_{j=1}^{i-1} p_j\right).$$

□

Since the average precision is the sum precision divided by the constant R, we have that the expected average precision is

$$\frac{1}{R}\sum_{i=1}^{n} \frac{p_i}{i}\left(1 + \sum_{j=1}^{i-1} p_j\right).$$