Measure-based Metasearch
Javed A. Aslam, Virgiliu Pavlu and Emine Yilmaz

Introduction

The Metasearch Problem

Search for: chili peppers

Some Popular Metasearch Techniques

- **CombMNZ** is based on combining the normalized scores given to each document by the underlying systems.
- **Condorcet** is based on viewing the metasearch problem as a multi-candidate election where the documents are candidates and the systems are voters expressing preferential rankings among the candidates.

Evaluation Measures

- **Average precision (AP)** is the average of the precisions at relevant documents.
- **Precision at cutoff k (PC(k))** is the precision at rank k.
- **R-precision (RP)** is the precision at rank \( R \), where \( R \) is the total number of relevant documents for the query.

**FACTS:**

- Evaluation measures aim to assess how well a system retrieves relevant documents, as measured and evaluated by implicit weights.
- Retrieval systems aim to retrieve relevant documents as well as possible.

**HYPOTHESIS:**

- Evaluation measures will assign “high” weights to relevant documents when applied to the lists generated by “good” retrieval systems. Thus, the weights implicit in evaluation measures can be used to identify likely relevant documents in a list.

Methodology

Weights Implicit in Evaluation Measures

- **Precisions at standard cutoffs (PC(k))**
  
  PC(k) implicitly assigns a weight of \( 1/k \) to each of the top \( k \) documents in a list and a weight of 0 to every remaining document.

- **R-precision (RP)**
  
  R-precision is PC(\( R \)), where \( R \) is the total number of relevant documents for the query.

**Problem:** Computing the weights associated with R-precision requires a priori knowledge of \( R \)!

**Average precision (AP)**

AP is the average of the precisions at relevant documents. Hence, one can compute average precision as follows, where \( N \) is the length of the retrieved list, \( rel(i) \) is the 0-1 relevance of the document at rank \( i \), \( R \) is the number of relevant documents in the query.

\[
AP = \frac{1}{R} \cdot \sum_{i : rel(i) = 1} PC(i)
\]

\[
= \frac{1}{R} \cdot \sum_{i=1}^{N} rel(i) \cdot PC(i)
\]

\[
= \frac{1}{R} \cdot \sum_{i=1}^{N} \frac{rel(i) \cdot \sum_{j=1}^{i} rel(j)}{i}
\]

\[
= \frac{1}{R} \cdot \sum_{1 \leq i < j \leq N} \frac{1}{i} \cdot rel(i) \cdot rel(j)
\]

Thus, average precision assigns a weight of \( \frac{1}{R \cdot i} \) to each pair of ranks \((i,j)\), for all \( 1 \leq j < i \leq N \).

We can compute the weight associated with each rank \( r \) by summing the weights associated with all pairs involving \( r \):

\[
\sum_{i=1}^{r} \frac{1}{R \cdot i} + \sum_{i=r+1}^{N} \frac{1}{R \cdot i} = \frac{1}{R} \cdot \left( 1 + \frac{1}{1 + 1} + \frac{1}{1 + 2} + \cdots + \frac{1}{R} \right)
\]

\[
= \frac{1}{R} \cdot \left( 1 + H_N - H_r \right)
\]

where \( H_k \) is the \( k \)-th harmonic number.

**Note:** \( R \) simply acts as a scaling factor. We don’t need a priori knowledge of \( R \)!

Metasearch Lists Through Implicit Weights

To obtain a consensus metasearch list through the weights assigned by an evaluation measure:

- compute the average weight assigned to the document across the underlying lists.
- rank the documents according to these average scores.

Results

The table shows the mean average precisions of the metasearch algorithms associated with average precision, R-precision and precision at cutoff \( k \), where \( k = 10, 20, 100, 1000 \) and compares those with CombMNZ And Condorcet algorithms. The results show that:

- System-oriented measures, e.g. AP and RP, tend to implicitly weight the set of relevant documents higher than user-oriented measures, e.g. PC(k).
- Performance of the metasearch algorithms corresponding to AP and RP often equals or exceeds CombMNZ and Condorcet.

**Note:** The AP and RP results shown in the table constitute statistically significant improvements\(^1\) w.r.t. CombMNZ and Condorcet when labeled with a \( \bullet \) and/or \( \circ \).

\(^1\)Sign test of significance, 90% confidence level.