Norwegian University of Science and Technology

Assignment Title

Assignment 2

-

Relevance Feedback, Language Model, Evaluation of IR Systems, Interpolated Precision

Course

TDT4117 Information Retrieval

Semester

FALL 2018

Date Submitted

14/10/2018

Submitted by

Dusan Jakovic

NTNU

Kunnskap for en bedre verden

Task 1: Basic Definitions

1. Explain the difference between automatic local analysis and automatic global analysis.

Local:

- In the local analysis, documents retrieved are examined to automatically determine query expansion.
- There is no need for relevance feedback.
- I will dynamically determine similar terms based on the top ranked documents that are retrieved.
- Avoiding ambiguity by determining similar terms only within relevant documents

Automatic;

- In the global, thesaurus is used to help to select terms for expansion.
- Determines term similarity through a pre-computed statistical analysis.
- Expanding queries with statistically most similar terms
- Association matrices which quantify term correlations on how frequently the co-occur

2. What is the purpose of relevance feedback? Explain the terms Query Expansion and Term Re-weighting. What separates the two?

Relevance feedback allows “searchers” to give feedback on which results that are relevant, and which aren’t. This helps the search engine to better under understand the query and improve the results.

In query expansion, users give additional input on the query words or phrases, possibly suggesting additional query terms. A search engine like e.g., Google or Yahoo!, are suggesting related queries to the one users are searching for. It is reformulating the give query to improve retrieval performance, especially recall.

Term re-weighting is the process of increasing weights on the terms in relevant documents, and decreasing the weights on terms in non relevant documents.

Task 2: Language Model

1. Explain the language model, what are the weaknesses and strengths of this model?

A separate language model is associated with each document in a collection, and the documents are ranked based on the probability of the query in the documents language model, P(d | q).

Strengths:

- Precise
- Simple
- Intuitively
- Computationally tractable

Weaknesses:

- User preferences and relevance feedback is difficult to integrate into the model
- Estimation of the document model is an issue. Such as choices of how to smooth it effectively.

2. Given the following documents and queries, build the language model according to the document collection.

d1 = An apple a day keeps the doctor away.
d2 = The best doctor is the one you run to and can't find.
d3 = One rotten apple spoils the whole barrel.

q1 = doctor
q2 = apple orange
q3 = doctor apple

Use MLE for estimating the unigram model and estimate the query generation probability using the Jelinek-Mercer smoothing:

$\widehat{P}(t|M_d) = (1 - \lambda)_{\widehat{P}mle}(t|M_d) + \lambda_{\widehat{P}mle}(t|c)\ ,\ \lambda = 0.5$

For each query, rank the documents using the generated scores.

d1 = 8, d2 = 12, d3 = 7, c = 27

P(doctor) = 2/27, P(apple) = 2/27, P(orange) = 0/27

q1:

- d1: ((0.5) * (1/8)) + (0.5 * (2/27)) = 0.0995
- d2: ((0.5) * (1/12)) + (0.5 * (2/27)) = 0.0787
- d3: ((0.5) * (0/7)) + (0.5 * (2/27)) = 0.0370

q2:

- d1: ((0.5) * (1/8)) + (0.5 * (2/27)) * ((0.5) * (0)) + (0.5 * (0)) = 0
- d2: ((0.5) * (0/12)) + (0.5 * (2/27)) * ((0.5) * (0)) + (0.5 * (0)) = 0
- d3: ((0.5) * (1/7)) + (0.5 * (2/27)) * ((0.5) * (0)) + (0.5 * (0)) = 0

q3:

- d1: ((0.5) * (1/8)) + (0.5 * (2/27)) * ((0.5) * (1/8)) + (0.5 * (2/27)) = 0.1019
- d2: ((0.5) * (1/12)) + (0.5 * (2/27)) * ((0.5) * (0/12)) + (0.5 * (2/27)) = 0.0787
- d3: ((0.5) * (0/7)) + (0.5 * (2/27)) * ((0.5) * (1/7)) + (0.5 * (2/27)) = 0.0397

3. Explain what smoothing means and how it affects retrieval scores. Describe your answer using a query from the previous subtask.

Smoothing is used to remove null-values from the search to avoid assigning zero probability to terms that are not found in the document. One technique is to move som *mass* probability from the query terms in the document to the terms not in the document.

By looking at q1 and q2, we can see that there is a term(s) that does not occur in the document but they still occur in the collection. Hence, the probability is not zero.

Task 3: Evaluation of IR Systems

1. Explain the terms Precision and Recall, including their formulas. Describe how differently these metrics can evaluate the retrieval quality of an IR system.

Precision:

The fraction of retrieved documents that are *relevant*.

Formula: P(relevant documents| retrieved documents)

Recall:

The fraction of the relevant documents that are *retrieved.*

Formula: P(retrieved relevant documents| relevant documents)

2. Given the following set of relevant documents rel = {23, 10, 33, 500, 70, 59, 82, 47, 72, 9}, and the set of retrieved documents ret = {55, 500, 2, 23, 72, 79, 82, 215}, provide a table with the calculated precision and recall at each level.

Doc. Retrieved	Doc. Relevant	Precision	Recall
55	NON REL	0/1= 0	0/10 = 0
500	REL	½ = 0.5	1/10 = 0.1
2	NON REL	1/3 = 0.3	1/10 = 0.1
23	REL	2/4= 0.5	2/10 = 0.2
72	REL	3/5 = 0.6	3/10 = 0.3
79	NON REL	3/6 = 0.5	3/10 = 0.3
82	REL	4/7 = 0.57	4/10 = 0.4
215	NON REL	4/8 = 0.5	4/10 = 0.4

Task 4: Interpolated Precision

1. What is interpolated precision?

It is a simpler version of the Precision(Recall)-function. It lets each precision be the maximum of all future points, which removes the “jiggles” in plot, and it is necessary to compute the precision at *recall-level* zero.

2. Given the example in Task 3.2, find the interpolated precision and make a graph.