TDT4117 Information Retrieval - Autumn 2021

Assignment 2

Written by Thomas Bjerke and Trym Grande

Task 1 - Relevance Feedback

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

Automatic Global Analysis:

- Determines term similarity using pre-computed statistical analysis of the document collection.
- Computes frequency matrices showing how often similar terms occur.
- Expands query to the statistically most occuring terms.
- Requires intensive term correlation, but only once at system development time.

Automatic Local Analysis:

- Dynamically determines similar terms based on analysis of the top ranked retreived documents, at query time.
- Correlation analysis are based on only the local set of retrieved documents for a given query.
- Interprets similar (related) terms only within relevant documents.
- Requires intensive term correlation computation for **every** query at run time, but number of terms and documents is less than in global analysis.
- Gives the best results.

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

Relevance feedback is the principle of responding to a query, followed by the user giving feedback on what documents are relevant and what are non-relevant, allowing the search engine to improve relevancy of the results.

Query Expansion is the process of reformulating a given query in order to improve search results. This works by finding similar terms to the ones being queried like synonyms, antonyms etc., using stemming, spell checking, and then applying it to the query.

Term Re-weighting is re-applying the candidate terms that were achieved during query expansion to the query.

Task 2 - Language Model

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

The Statistical Language Model uses a generated statistical probability for each n number of consecutive terms given each document. This means that the probobility for a query is defined by multiplying the probability of each term sequence, where the query can be seen as a permutation of terms. A unigram language model (n=1) is often used in IR as it looks at occurrencase of individual terms.
Strength: flexible to any order n of consecutive terms.
Weakness: Needs parameter tuning and smoothing to avoid recall.

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:

$$\hat{P}(t|M_d) = (1{-}\lambda)\hat{p}_{mle}(t|M_d) + \lambda\hat{p}_{mle}(t|C), \lambda = 0.5$$

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

$$t_C = t_{d1} + t_{d2} + t_{d3} = 8 + 12 + 7 = 27$$

$$\hat{P}(t|M_d) = (1{-}\lambda)\hat{p}_{mle}(t|M_d) + \lambda\hat{p}_{mle}(t|C), \lambda = 0.5$$

$$\hat{P}(t|M_d) = \frac{1}{2}\hat{p}_{mle}(t|M_d) + \frac{1}{2}\hat{p}_{mle}(t|C)$$

$$\hat{P}(t|M_d) = \frac{\hat{p}_{mle}(t|M_d) + \hat{p}_{mle}(t|C)}{2}$$

Language model of d1:

w	**P(w|q)**
STOP	$\frac{1}{8} = 0.125$
an	$\frac{\frac{1}{8}+\frac{1}{27}}{2} = 0.0810$
apple	$\frac{\frac{1}{8}+\frac{2}{27}}{2} = 0.0995$
a	$\frac{\frac{1}{8}+\frac{1}{27}}{2} = 0.0810$
day	$\frac{\frac{1}{8}+\frac{1}{27}}{2} = 0.0810$
keeps	$\frac{\frac{1}{8}+\frac{1}{27}}{2} = 0.0810$
the	$\frac{\frac{1}{8}+\frac{4}{27}}{2} = 0.137$
doctor	$\frac{\frac{1}{8}+\frac{2}{27}}{2} = 0.0995$
away	$\frac{\frac{1}{8}+\frac{1}{27}}{2} = 0.0810$

Language model of d2:

w	**P(w|q)**
STOP	$\frac{1}{12} = 0.0833$
the	$\frac{\frac{2}{12}+\frac{4}{27}}{2} = 0.0602$
best	$\frac{\frac{1}{12}+\frac{1}{27}}{2} = 0.0602$
doctor	$\frac{\frac{1}{12}+\frac{2}{27}}{2} = 0.0787$
is	$\frac{\frac{1}{12}+\frac{1}{27}}{2} = 0.0602$
one	$\frac{\frac{1}{12}+\frac{2}{27}}{2} = 0.0787$
you	$\frac{\frac{1}{12}+\frac{1}{27}}{2} = 0.0602$
run	$\frac{\frac{1}{12}+\frac{1}{27}}{2} = 0.0602$
to	$\frac{\frac{1}{12}+\frac{1}{27}}{2} = 0.0602$

w	P(w|q)
and	$\frac{\frac{1}{12}+\frac{1}{27}}{2} = 0.0602$
can't	$\frac{\frac{1}{12}+\frac{1}{27}}{2} = 0.0602$
find	$\frac{\frac{1}{12}+\frac{1}{27}}{2} = 0.0602$

Language model of d3:

w	P(w|q)
STOP	$\frac{1}{7} = 0.143$
one	$\frac{\frac{1}{7}+\frac{1}{27}}{2} = 0.0899$
rotten	$\frac{\frac{1}{7}+\frac{1}{27}}{2} = 0.0899$
apple	$\frac{\frac{1}{7}+\frac{2}{27}}{2} = 0.108$
spoils	$\frac{\frac{1}{7}+\frac{1}{27}}{2} = 0.0899$
the	$\frac{\frac{1}{7}+\frac{1}{27}}{2} = 0.0899$
whole	$\frac{\frac{1}{7}+\frac{1}{27}}{2} = 0.0899$
barrel	$\frac{\frac{1}{7}+\frac{4}{27}}{2} = 0.0899$

$$P(q1|M_{d1}) = \prod_{i=1}^{r} P(w_i|q)$$

q1:

$$P(q1|M_{d1}) = 0.0995 \cdot 0.125$$

$$P(q1|M_{d2}) = 0.0787 \cdot 0.0833$$

$$P(q1|M_{d3}) = (\frac{\frac{0}{7}+\frac{0}{27}}{2}) \cdot 0.143 = 0$$

$$d1 > d2 > d3$$

q2:

$$P(q2|M_{d1}) = 0.0995 \cdot (\frac{\frac{0}{8} + \frac{0}{27}}{2}) \cdot 0.125 = 0$$

$$P(q2|M_{d2}) = (\frac{\frac{0}{12} + \frac{2}{27}}{2}) \cdot (\frac{\frac{0}{12} + \frac{0}{27}}{2}) \cdot 0.0833 = 0$$

$$P(q2|M_{d3}) = 0.108 \cdot (\frac{\frac{0}{7} + \frac{0}{27}}{2}) \cdot 0.143$$

$$d3 > d1 = d2$$

q3:

$$P(q3|M_{d1}) = 0.0995 \cdot 0.0995 \cdot 0.125 = 0.00124$$

$$P(q3|M_{d2}) = 0.0787 \cdot (\frac{\frac{0}{12} + \frac{2}{27}}{2}) \cdot 0.0833 = 0.000243$$

$$P(q3|M_{d3}) = (\frac{\frac{0}{7} + \frac{2}{27}}{2}) \cdot 0.108 \cdot 0.143 = 0.000572$$

$$d1 > d3 > d2$$

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

A problem with the language model is that non-occurring terms have probability = 0, which in turn zeros out the probability for the whole query. This is solved using term smoothing, where the terms are made non-zero by a given parameter factor λ. For example, we can look at q3 for d3, where term "apple" does not occur in the document, but since it does occur in the rest of the collection, $\hat{P}(”apple”|M_{d3})$ will be smoothed out to more than 0, resulting in a non-zero result. However this does not apply for cases where a term that does not occur in the collection at all - e.g. $P(q2|M_{d2}) = 0$.

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 is a measure of how many relevant documents were retrieved divided by how many documents were retrieved. This is important for being able to quickly googling something, were you don't want to read through a lot of documents.

Recall is a measure of how many relevant documents were retrieved divided by how many relevant documents there are. This represents how many relevant documents were left behind in the result. A user would for example want low recall for finding an important document on his computer, because finding the result would matter more than looking through many results.

Something to note here, is that an IR model can achieve 100% recall by simply retrieving all documents from the collection. Also, a low recall usually gives higher precision, meaning there needs to be a balance. This balancing is called f-measure trade off.

2. Explain the terms MAP and MRR ranking methods. List two pros and cons of each of methods in information retrieval querying.

Mean Average Precition and Mean Reciprocal Rank are both measures for summarizing how the model performs in terms of results.

MAP considers whether all of the relevant items tend to get ranked highly. It does this by calculating the average precision of each document for a given query, where non-retrieved documents are counted as 0. This is counted up to a given number of documents, with the notation MAP@5 og MAP@10 for the first 5 or 10 results.
Pros:

- Worsk well for normal web searches were the user usually looks through no more than the first page of results.
- Number of documents can be varied as parameter.
  Cons:
- Returns the same as MRR when there is only one relevant document.
- Gives good single metric of precision-recall curve.

MRR is a metric often used for even more precise queries such as finding a specific website, or looking up an answer to a question.
Pros:

- Good for when there is only one relevant result e.g. voice assistants.
- Number of documents can be varied as parameter.
  Cons:
- Does not care about lower ranked documents.

- Does not work for use cases of browsing results.

3. 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.

len(rel)=10
len(ret)=8
rel ret = 500, 23, 72, 82
len(rel ret) = 4

P = rel ret / ret
R = rel ret / rel

level	precision	recall
unranked	$\frac{\text{rel ret}}{\text{ret}} = \frac{4}{8} = 0.5$	$= \frac{\text{rel ret}}{\text{rel}} = \frac{4}{10} = 0.4$
1	$\frac{0}{1} = 0$	$\frac{0}{10} = 0$
2	$\frac{1}{2} = 0.5$	$\frac{1}{10} = 0.1$
3	$\frac{1}{3} = 0.\bar{3}$	$\frac{1}{10} = 0.1$
4	$\frac{2}{4} = 0.5$	$\frac{2}{10} = 0.2$
5	$\frac{3}{5} = 0.6$	$\frac{3}{10} = 0.3$
6	$\frac{3}{6} = 0.5$	$\frac{3}{10} = 0.3$
7	$\frac{4}{7} = 0.57$	$\frac{4}{10} = 0.4$
8	$\frac{4}{8} = 0.5$	$\frac{4}{10} = 0.4$

Task 4 - Interpolated Precision

1. What is interpolated precision?

Interpolated precision is using interpolation techniques to estimate unknown values using surrounding known values. This can allow for a continuous precision graph, although it will not be as accurate.

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

New interpolated table using an optimistic approach since precision is generally decreasing with recall:

level	precision	recall
0	0.6	0
0.5	0.6	0
1	0.6	0
1.5	0.6	0
2	0.6	0.1
2.5	0.6	0.1
3	0.6	0.1
3.5	0.6	0.1
4	0.6	0.2
4.5	0.6	0.2
5	0.57	0.3
5.5	0.57	0.3
6	0.57	0.3
6.5	0.57	0.3
7	0.5	0.4
7.5	0.5	0.4
8	0.5	0.4
8.5	0.5	0.4

Plotted graph using matplotlib:

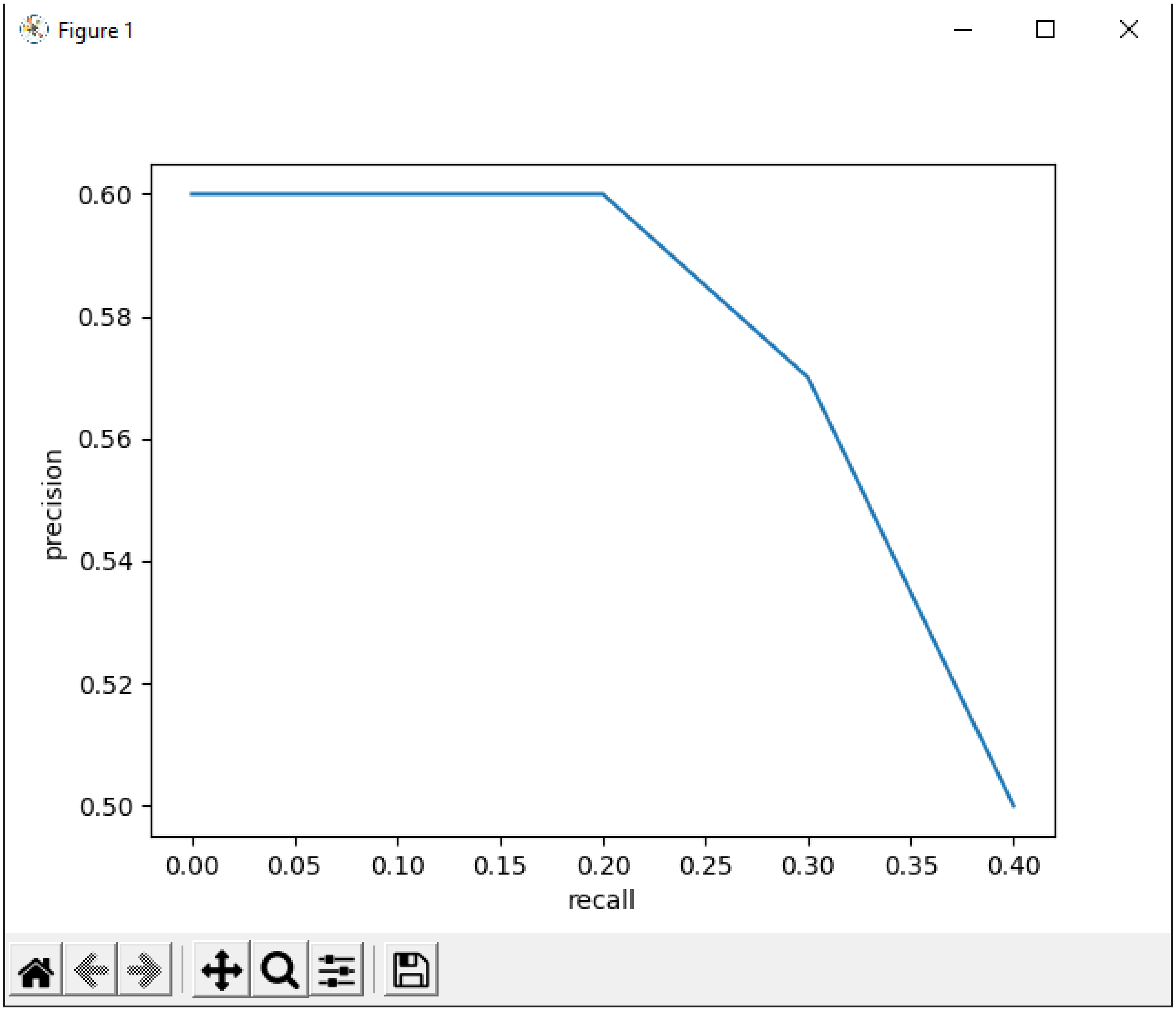
Figure 1
precision
0.60
0.58
0.56
0.54
0.52
0.50
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
recall