Bjørn Are Odden
Ole Alexander Høyby

TDT4117 Information Retrieval - Autumn 2021

Task 1 - Relevance Feedback

Explain the main differences between:

1. Explain the difference between automatic local analysis and automatic global analysis.
   Local - The Documents retrieved are examined to automatically determine query expansion.
   The terms used in the query expansion are determined by analyzing the top ranked documents and finding correlations between terms using clusters.
   There is no need for relevance feedback in this method.

   Global – Utilizes a thesaurus to provide and select terms to use in query expansion. The statistical analysis of the thesaurus is pre-computed on the complete collection. There is no need for relevance feedback in this method.

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

   Relevance feedback: Is based on the idea that a user gives feedback on how relevant an initial query is. The algorithm then uses that feedback to perform a new query with improved results.

   Query Expansion: evaluates the query and expands it with similar terms to increase retrieval performance.

   Term Re-weighting: Assuming that each term has an initial weight, Re-weighting takes some feedback (usually from the user) and re-weights the terms, increasing weight on relevant terms and decreasing weight on non-relevant terms.

Bjørn Are Odden
Ole Alexander Høyby

Task 2 - Language Model

Explain:

1. Explain the language model, what are the weaknesses and strengths of this model?
   Language modeling (LM) is the use of various statistical and probabilistic techniques to determine the probability of a given sequence of words occurring in a sentence.
   Strengths:
   - Mathematically precise
   - Simple and intuitive

   Weaknesses:
   - Difficult to consider the users preference and needs

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:
   |d1| = 8 |d2| = 12 |d3| = 7 Total = 27
   $p(q_1|d_1)$ = [(1/8 + 2/27) / 2] = 0.100
   $p(q_1|d_2)$ = [(1/12 + 2/27) / 2] = 0.079
   $p(q_1|d_3)$ = [(0/7 + 2/27) / 2] = 0.037
   $p(q_2|d_1)$ = [(1/8 + 2/27) / 2] · [(0/8 + 0/27) / 2] = 0
   $p(q_2|d_2)$ = [(0/12 + 2/27) / 2] · [(0/12 + 0/27) / 2] = 0
   $p(q_2|d_3)$ = [(1/7 + 2/27) / 2] · [(0/7 + 0/27) / 2] = 0
   $p(q_3|d_1)$ = [(1/8 + 2/27) / 2] · [(1/8 + 2/27) / 2] = 0.01
   $p(q_3|d_2)$ = [(1/12 + 2/27) / 2] · [(0/12 + 2/27) / 2] = 0.003
   $p(q_3|d_3)$ = [(0/7 + 2/27) / 2] · [(1/7 + 2/27) / 2] = 0.004

Bjørn Are Odden
Ole Alexander Høyby

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 0 values** from queries containing terms that do not appear in the given document.
   From $p(q_3|d_2)$ we see that apple does **not** appear in the document, this would **-without** smoothing, give this a score of 0. ‘Apple’ does, however, appear twice in the collection and ‘doctor’ once in the document. The probability is therefore not zero, and with smoothing a total score of 0.003 is set, which is lower than for $p(q_3|d_1)$ (which contain both doctor and apple).

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 the fraction of retrieved documents that are relevant.
   Precision = P (relevant|retrieved)

   Can be calculated by P = TP / (TP+FP)

   Recall is the fraction of relevant documents that are retrieved.
   Recall = P (retrieved|relevant)

   Can be calculated by R = TP / (TP+FN)

   TP = True Positive (for a retrieved document)
   FP = False positive (for a retrieved document)
   FN = False negative (for a non-retrieved document)

   Using these measures in isolation can lead to some “traps” in that one can get 100 % recall simply by fetching every document for a query. Similar situation with precision, one

Bjørn Are Odden
Ole Alexander Høyby

can simply retrieve very few documents, i.e. very low recall, for a high precision result. Clearly, a balance must be struck.

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

MAP (Mean Average Precision) is obtained by taking the mean of the average precision for a set of queries.

- Pros:
  - Good measure for situations where a list of several relevant documents/items is needed
  - Rewards having relevant documents higher up on the ranking, vs. having the same amount of relevant documents further down.

- Cons
  - Not fit for fine-grained numerical ratings.
  - If for some reason you are going to be using *all* the retrieved documents, the weighting on relevant documents appearing earlier will skew the score in favor of head heavy over tail heavy rankings.

MRR (Mean Reciprocal Rank) an average of reciprocal ranks over a set of queries, where the reciprocal rank for a single query is determined by the position of the first relevant document in the retrieved list.

- Pros:
  - Scenario when there's only one "correct" document, i.e., the user is looking for a specific document, i.e. a login page.
  - Scenario where the user is looking for a fact, with one correct answer.
- Cons:
  - Only cares about the position of the highest ranked relevant item, so it doesn't matter whether the rest of the documents are relevant or not.

Bjørn Are Odden
Ole Alexander Høyby

- MRR will score a list with one relevant result the same way it scores a list with many relevant 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.

Documents retrieved	Retrieved	Relevant	Recall	Precision	Interp. precision
0	-	-	0	undefined	0.6
1	55	0	0	0	0.6
2	500	1	0.1	0.5	0.6
3	2	0	0.1	0.33	0.6
4	23	1	0.2	0.5	0.6
5	72	1	0.3	0.6	0.6
6	79	0	0.3	0.5	0.571
7	82	1	0.4	0.571	0.571
8	215	0	0.4	0.5	0.5

Task 4 - Interpolated Precision

1. What is interpolated precision?
When graphing precision vs. recall, we end up with saw tooth-shaped curves, as recall always stays at the current level if the next document is not relevant, or increases by one if it is, while precision can start at e.g. 1, drop to 0.5, then back up to 0.66. Interpolation is a way to smooth out these curves by always setting the interpolated precision at the current recall level to be the maximum precision found at any recall level greater than the current level. This also allows (interpolated) precision to be defined at recall level 0. This smoothing

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