Information Retrieval

Lecture 8 - Relevance feedback and query expansion

Seminar für Sprachwissenschaft
International Studies in Computational Linguistics
Wintersemester 2007

Introduction

▷ An information need may be expressed using different keywords (synonymy)
   — impact on recall
   — examples: ship vs boat, aircraft vs airplane
▷ Solutions: refining queries manually or expanding queries (semi) automatically
▷ Semi-automatic query expansion:
  • local methods based on the retrieved documents and the query (ex: Relevance Feedback)
  • global methods independent of the query and results (ex: thesaurus, spelling corrections)

Overview

About Relevance Feedback
The Rocchio algorithm
About probabilistic relevance feedback
When to use Relevance Feedback
Relevance Feedback and the web
Evaluation of Relevance Feedback strategies
Other local methods for query expansion
Global methods for query expansion

About Relevance Feedback

Feedback given by the user about the relevance of the documents in the initial set of results
About Relevance Feedback (continued)

- Based on the idea that:
  1. defining good queries is difficult when the collection is (partly) unknown
  2. judging particular documents is easy
- Allows to deal with situations where the user’s information needs evolve with the checking of the retrieved documents
- Example: image search engine
  http://nayana.ece.ucsb.edu/imsearch/imsearch.html
Relevance feedback example 2
Query: New space satellite applications
+ 1. 0.539, 08/13/91, NASA Hasn’t Scrapped Imaging Spectrometer
+ 2. 0.533, 07/09/91, NASA Scratches Environment Gear From Satellite Plan
3. 0.526, 04/04/90, Science Panel Backs NASA Satellite Plan, But Urges Launches of Smaller Probes
4. 0.526, 09/09/91, A NASA Satellite Project Accomplishes Incredible Feat: Staying Within Budget
5. 0.525, 07/24/90, Scientist Who Exposed Global Warming Proposes Satellites for Climate Research
6. 0.524, 08/22/90, Report Provides Support for the Critics Of Using Big Satellites to Study Climate
7. 0.516, 04/13/87, Arianespace Receives Satellite Launch Pact From Telesat Canada
+ 8. 0.509, 12/02/87, Telecommunications Tale of Two Companies

The Rocchio algorithm
- Standard algorithm for relevance feedback (SMART, 70s)
- Integrates a measure of relevance feedback into the Vector Space Model
- Idea: we want to find a query vector $\mathbf{q}_{opt}$
  - maximizing the similarity with relevant documents while
  - minimizing the similarity with non-relevant documents
  $$\mathbf{q}_{opt} = \arg\max_{\mathbf{q}} \left[ \frac{\text{sim}(\mathbf{q}, C_r) - \text{sim}(\mathbf{q}, C_{nr})}{|C_r|} \right]$$
  With the cosine similarity, this gives:
  $$\mathbf{q}_{opt} = \frac{1}{|C_r|} \sum_{d_i \in C_r} \mathbf{d}_i - \frac{1}{|C_{nr}|} \sum_{d_i \in C_{nr}} \mathbf{d}_i$$
The Rocchio algorithm (continued)

- Problem with the above metrics: the set of relevant documents is unknown
- Instead, we produce the modified query \( m \):

\[
q_m = \alpha q_0 + \beta \frac{1}{|D_r|} \sum_{d \in D_r} d_j - \gamma \frac{1}{|D_{nr}|} \sum_{d \in D_{nr}} d_j
\]

where:
- \( q_0 \) is the original query vector
- \( D_r \) is the set of known relevant documents
- \( D_{nr} \) is the set of known non-relevant documents
- \( \alpha, \beta, \gamma \) are balancing weights (judge vs system)

Remarks:
- Negative weights are usually ignored
- Rocchio-based relevance feedback improves both recall and precision
- For reaching high recall, many iterations are needed
- Empirically determined values for the balancing weights:
  \( \alpha = 1, \beta = 0.75, \gamma = 0.15 \)
- Positive feedback is usually more valuable than negative feedback: \( \beta > \gamma \)

Rocchio algorithm: exercise

Consider the following collection (one doc per line):
good movie trailer shown
trailer with good actor
unseen movie
a dictionary made of the words movie, trailer and good, and an IR system using the standard \( \text{tf} - \text{idf} \) weighting (without normalisation).

Assuming a user judges the first 2 documents relevant for the query \textit{movie trailer}. What would be the Rocchio-revised query?

About probabilistic relevance feedback

- Alternative to the Rocchio algorithm, use a document classification instead of a Vector Space Model-based retrieval

\[
P(x_t = 1|R = 1) = \frac{|VR_t|}{|VR|}
\]
\[
P(x_t = 0|R = 0) = \frac{n_t - |VR_t|}{N - |VR|}
\]

where:
- \( N \) is the total number of documents
- \( n_t \) is the number of documents containing \( t \)
- \( VR \) is the set of known relevant documents
- \( VR_t \) is the set of known relevant documents containing \( t \)
- Problem: no memory of the original query
When to use Relevance Feedback

- Relevance Feedback does not work when:
  - the query is misspelled
  - we want cross-language retrieval
  - the vocabulary is ambiguous
  - the users do not have sufficient initial knowledge
  - the query concerns an instance of a general concept (e.g. felines)
  - the documents are gathered into subsets each using a different vocabulary
  - the query has disjunctive answer sets (e.g. "the pop star that worked at KFC")
  - there exist several prototypes of relevant documents

- Practical problem: refining leads to longer queries that need more time to process

Relevance Feedback and the web

- Few web IR systems use relevance feedback
  - hard to explain to users
  - users are mainly interested in fast retrieval (i.e. no iterations)
  - users usually are not interested in high recall

- Nowadays: clickstream-based feedback (which links are clicked on by users)
  - implicit feedback from the writer rather than feedback from the reader

Evaluation of Relevance Feedback strategies

- Note that improvements brought by the relevance feedback decrease with the number of iterations, usually one round gives good results

- Several evaluation strategies:
  (a) comparative evaluation
     - query $q_0$ → prec/recall graph
     - query $q_m$ → prec/recall graph
     - usually +50% of mean average precision
     (partly comes from the fact that known relevant documents are higher ranked)

- Evaluation strategies (continued)
  (b) residual collection
     - same technique as above but by looking at the set of retrieved documents - the set of assessed relevant documents
     - the performance measure drops
  (c) using two similar collections
     - collection #1 is used for querying and giving relevance feedback
     - collection #2 is used for comparative evaluation
     - $q_0$ and $q_m$ are compared on collection #2
Evaluation of Relevance Feedback strategies (continued)

- Evaluation strategies:
  
  (d) user studies
  
  - e.g. time-based comparison of retrieval, user satisfaction, etc.
  - user utility is a fair evaluation as it corresponds to real system usage

Overview

About Relevance Feedback
The Rocchio algorithm
About probabilistic relevance feedback
When to use Relevance Feedback
Relevance Feedback and the web
Evaluation of Relevance Feedback strategies
Other local methods for query expansion
Global methods for query expansion

Pseudo Relevance Feedback

- Aka blind relevance feedback
- No need of an extended interaction between the user and the system
- Method:
  - normal retrieval to find an initial set of most relevant documents
  - assumption that the top $k$ documents are relevant
  - relevance feedback defined accordingly
- Works with the TREC Ad Hoc task
  - Inc.ltc (precision at $k = 50$): no-RF 62.5 %, RF 72.7 %
- Problem: distribution of the documents may influence the results

Indirect Relevance Feedback

- Uses evidences rather than explicit feedback
- Example: number of clicks on a given retrieved document
- Not user-specific
- More suitable for web IR, since it does not need an extra action from the user
Overview
About Relevance Feedback
The Rocchio algorithm
About probabilistic relevance feedback
When to use Relevance Feedback
Relevance Feedback and the web
Evaluation of Relevance Feedback strategies
Other local methods for query expansion

Vocabulary tools for query reformulation
Tools displaying:
- a list of close terms belonging to the dictionary
- information about the query words that were omitted (cf stop-list)
- the results of stemming
→ ≈ debugging environnement

Query logs and thesaurus
Users select among query suggestions that are built either from query logs or thesaurus
Replacement words are extracted from thesaurus according to their proximity to the initial query word
Thesaurus can be developed:
- manually (e.g. biomedicine)
- automatically (cf below)
NB: query expansion
(i) increases recall
(ii) may need users’ relevance on query terms (̸= documents)

Automatic thesaurus generation
Analyze of the collection for building the thesaurus automatically:
1. Using word co-occurrences (co-occurring words are more likely to belong to the same query field)
   → may contain false positives (example: apple)
2. Using a shallow grammatical analyzes to find out relations between words
   example: cooked, eaten, digested → food
Note that co-occurrence-based thesaurus are more robust, but grammatical-analyzes thesaurus are more accurate
Building a co-occurrence-based thesaurus

- We build a term-document matrix $A$ where $A_{t,d} = w_t,d$ (e.g. normalized $tf - idf$)
- We then calculate $C = A A^T$

$$C = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{pmatrix}$$

$c_{ij}$ is the similarity score between terms $i$ and $j$

Automatically built thesaurus

<table>
<thead>
<tr>
<th>Word</th>
<th>Nearest neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>absolutely</td>
<td>absolutely, wholly, exactly, nothing</td>
</tr>
<tr>
<td>bottomed</td>
<td>bottomed, lowest, lowest, lowest</td>
</tr>
<tr>
<td>captivating</td>
<td>captivating, attractive, charm</td>
</tr>
<tr>
<td>clandestine</td>
<td>clandestine, secret, secret, secret</td>
</tr>
<tr>
<td>makeup</td>
<td>makeup, make up, make up, make up</td>
</tr>
<tr>
<td>nourishing</td>
<td>nourishing, nourishing, nourishing, nourishing</td>
</tr>
<tr>
<td>paraphrases</td>
<td>paraphrases, sentences, sentences, sentences</td>
</tr>
</tbody>
</table>

Conclusion

- Query expansion using either local methods:
  - Rocchio algorithm for Relevance Feedback
  - Pseudo Relevance Feedback
  - Indirect Relevance Feedback
- or global ones:
  - Query logs
  - Thesaurus
- Thesaurus-based query expansion increases recall but may decrease precision (cf ambiguous terms)
- High cost of thesaurus development and maintenance
- Thesaurus-based query expansion is less efficient than Rocchio Relevance Feedback but may be as good as Pseudo Relevance Feedback

References

- C. Manning, P. Raghavan and H. Schütze, Introduction to Information Retrieval
- Chris Buckley and Gerard Salton and James Allan
  The effect of adding relevance information in a relevance feedback environment (1994)
  http://citeseer.ist.psu.edu/buckley94effect.html
- Ian Ruthven and Mounia Lalmas
  A survey on the use of relevance feedback for information access systems (2003)