

Introduction to Information Retrieval (Chapter 9)

Relevance feedback and query expansion

1. Background

In most collections, the same concept may be referred to using different words. This issue, known as synonymy, has an impact on the recall of most information retrieval (IR) systems. The methods for tackling this problem split into two major classes: global methods and local methods.

- Global methods are techniques for expanding or reformulating query terms independent of the query and results returned from it, so that changes in the query wording will cause the new query to match other semantically similar terms. Global methods include:
 - Query expansion/reformulation with a thesaurus or WordNet;
 - Query expansion via automatic thesaurus generation;
 - Techniques like spelling correction.
- Local methods adjust a query relative to the documents that initially appear to match the query. The basic methods here are:
 - Relevance feedback;
 - Pseudorelevance feedback, also known as blind relevance feedback;
 - (Global) Indirect relevance feedback.

2. Relevance feedback and pseudo relevance feedback

The idea of relevance feedback (RF) is to involve the user in the IR process so as to improve the final result set.

2.1 The Rocchio algorithm for relevance feedback

We want to find a query vector, denoted as \vec{q} , that maximizes similarity with relevant documents while minimizing similarity with nonrelevant documents. The Rocchio (1971) algorithm can achieve that.

$$\vec{q}_m = \alpha \vec{q}_0 + \beta \frac{1}{|D_r|} \sum_{\vec{d}_j \in D_r} \vec{d}_j - \gamma \frac{1}{|D_{nr}|} \sum_{\vec{d}_j \in D_{nr}} \vec{d}_j \quad (1)$$

where q_0 is the original query vector; q_m is the modified query. D_r and D_{nr} are the set of known relevant and nonrelevant documents, respectively. \vec{d}_j is the similarity between query and document, which can be cosine similarity. α , β , and γ are weights attached to each term. These control the balance between trusting the judged document set versus the query: If we have a lot of judged documents, we would like a higher β and γ .

Positive feedback also turns out to be much more valuable than negative feedback, and so most IR systems set $\gamma < \beta$. Reasonable values might be $\alpha = 1$, $\beta = 0.75$, and $\gamma = 0.15$. In fact, many systems, such as the image search system, allow only positive feedback, which is equivalent to setting $\gamma = 0$. Another alternative is to use only the marked nonrelevant document that received the highest ranking from the IR system as negative feedback (here, $|D_{nr}| = 1$ in Equation (1)). This is because if several nonrelevant documents are used for feedback calculation, some of them might bear some similarity with a few relevant documents (assuming that the user marks them as nonrelevant based on a few words or sentences depicted and doesn't process sufficient knowledge about them). In such a case, some of the properties of relevant documents might not be conveyed properly to the IR system, which results in low precision output. There are low chances of such a problem if only 1 nonrelevant is used.

2.2 When does relevance feedback work?

- First, the user has to have sufficient knowledge to be able to make an initial query that is at least somewhere close to the documents they desire.
- Second, the RF approach requires relevant documents to be similar to each other. That is, they should cluster. Ideally, the term distribution in all relevant documents will be similar to that in the documents marked by the users, and the term distribution in all nonrelevant documents will be different from those in relevant documents.

2.3 Pseudo relevance feedback

Pseudo-relevance feedback, also known as blind relevance feedback, automates the manual part of RF so that the user gets improved retrieval performance without an extended interaction. The method is to do normal retrieval to find an initial set of most relevant documents, to then assume that the top k ranked documents are relevant, and finally to do RF as before under this assumption.

This automatic technique mostly works. Evidence suggests that it tends to work better than global analysis. It has been found to improve performance in the TREC ad hoc task.

3. Global methods for query reformulation

(1) **Vocabulary tools for query reformulation.** The IR system suggests search terms by means of a thesaurus or a controlled vocabulary.

(2) **Query expansion.** The most common form of query expansion is global analysis, using some form of the thesaurus. For each term t in a query, the query can be automatically expanded with synonyms and related words of t from the thesaurus.

Methods for building a thesaurus for query expansion include the following.

- Use of a controlled vocabulary that is maintained by human editors.
- A manual thesaurus.
- An automatically derived thesaurus. Word cooccurrence statistics over a collection of documents in a domain are used to automatically induce a thesaurus
- Query reformulations based on query log mining.

(3) **Automatic thesaurus generation.** There are two main approaches for automatic thesaurus generation.

- Exploit word cooccurrence.

The simplest way to compute a cooccurrence thesaurus is based on term similarities. We begin with a term-document matrix A , where each cell $A_{t,d}$ is a weighted count $w_{t,d}$ for term t and document d , with weighting so A has length-normalized rows. If we then calculate $C = AA^T$, then $C_{u,v}$ is a similarity score between terms u and v , with a larger number being better.

If A is simply a Boolean term-document matrix, then C is the cooccurrence matrix, c_{ij} represents the times of term i , and term j cooccur.

- Use a shallow grammatical analysis of the text and to exploit grammatical relations or grammatical dependencies.

4. Summary

- Relevance feedback is one of the most used and most successful approaches for synonymy problems.
- Relevance feedback can improve both recall and precision. But, in practice, it has been shown to be most useful for increasing recall in situations where recall is important.
- In Rocchio (1971) algorithm, α, β, γ controls the balance between trusting the judged document set versus the query: If we have a lot of judged documents, we would like a higher β and γ .
- Relevance feedback can improve both recall and precision. But, in practice, it has been shown to be most useful for increasing recall in situations where recall is important.

- In fact, many systems, such as the image search system, allow only positive feedback, which is equivalent to setting $\gamma = 0$. Another alternative is to use only the marked nonrelevant document that received the highest ranking from the IR system as negative feedback (here, $|D_{nr}| = 1$ in Equation (1)).
- Some experimental results have also suggested that using a limited number of terms like this may give better results (Harman 1992), although other work has suggested that using more terms is better in terms of retrieved document quality (Buckley et al. 1994). But, the long queries may result in a high computing cost for the retrieval and potentially long response times for the user.
- In general, RF has been little used in web searches.
- Pseudo-relevance feedback mostly works. Evidence suggests that it tends to work better than global analysis. It has been found to improve performance in the TREC ad hoc task.
- Implicit feedback is less reliable than explicit feedback but is more useful than pseudo RF, which contains no evidence of user judgments.
- Use of query expansion generally increases recall and is widely used in many science and engineering fields.
- Simply using word cooccurrence is more robust (it cannot be misled by parser errors) for automatic thesaurus generation, but using grammatical relations is more accurate.
- Query expansion is often effective in increasing recall. However, query expansion may also significantly decrease precision, particularly when the query contains ambiguous terms.
- Overall, query expansion is less successful than RF, although it may be as good as pseudo RF.