Dublin City University at CLEF 2006: Robust Cross Language Track

Adenike M. Lam-Adesina   Gareth J. F. Jones

School of Computing, Dublin City University, Dublin 9, Ireland

{adenike,gjones}@computing.dcu.ie

Abstract. The main focus of the DCU group's participation in the CLEF 2006 Robust Track track was to explore a new method of re-ranking a retrieved document set based on the initial query with a pseudo relevance feedback (PRF) query expansion method. The aim of re-ranking using the initial query is to force the retrieved assumed relevant set to mimic the initial query more closely while not removing the benefits of PRF. Our results show that although our PRF is consistently effective for this task, the application of the current version of our new re-ranking method has little effect on the ranked output.

1 Introduction

This paper describes the DCU experiments for the CLEF 2006 Robust Track. Our official submissions included monolingual runs for English and for Spanish, Italian and French where topics and documents were translated into English, and a bilingual run for Spanish using English topics. Our general approach was to translate non-English documents and topics into English for use as a pivot language using Systran Version: 3.0. Pseudo Relevance Feedback (PRF) using a summary-based approach, shown to be effective in our submissions to previous CLEF workshops, was applied. In addition, for this task we explored the application of a new post-retrieval re-ranking method that we are developing.

The remainder of this paper is structured as follows: Section 2 describes our system setup and the IR methods used, Section 3 presents our experimental results and Section 4 concludes the paper with a discussion of our findings.

3 System Setup

Basic Retrieval System For our experiments we used the City University research distribution version of the Okapi system retrieval system. Stopwords were removed from both the documents and search topics, and the Okapi implementation of Porter stemming algorithm [1] was applied to both the document and search terms. The Okapi system is based on the BM25 weighting scheme [2]. In our experiments values of the $k1$ and b BM25 parameters were estimated using the CLEF 2003 ad hoc retrieval task data.

Pseudo-Relevance Feedback Short and imprecise queries can affect information retrieval (IR) effectiveness. To lessen this negative impact, relevance feedback (RF) via query expansion (QE) is often employed. QE aims to improve initial query statements by addition of terms from user assessed relevant documents. Pseudo-Relevance Feedback (PRF) assuming top ranked documents are relevant, can result in a query drift if expansion terms are selected from assumed relevant document which are in fact not relevant. In our past research work [3] we discovered that although a top-ranked document might not be relevant, it often contains information that is pertinent to the query. Thus, we developed a PRF method that selects appropriate terms from document summaries. These summaries are constructed in such a way that they contain only sentences that are closely related to the initial query. Our QE method selects terms from summaries of the top 5 ranked documents. The summaries are generated using the method described in [3]. For all our experiments we used the top 6 ranked sentences as the summary of each document. From this summary we collected all non-stopwords and ranked them using a slightly modified version of the Robertson selection value (*rsv*) [2]. The top 20 terms were then selected in all our experiments. In our modified version, potential expansion terms are selected from the summaries of the top 5 ranked documents, and ranked using statistics from assuming that the top 20 ranked documents from the initial run are relevant.

3.3 Re-ranking Methodology

As part of our investigation for the CLEF 2006 robust track we explored the application of a novel re-ranking of the retrieved document list obtained from our PRF process. This reordering method attempts to ensure that retrieved documents with more matching query terms have their ranking improved, while not discarding the effect of document weighting scheme used. To this end we devised a document re-ranking formula as follows:

$$\frac{doc\_wgt}{(1-b)+(b*nmt/mmt)} \qquad (1)$$

where $doc\_wgt =$ the original document matching score
$b =$ an empirical value ranging between 0.1 and 0.5
$nmt =$ the number of original topic terms that occur in the document
$mmt =$ mean value of *nmt* for a given query over all retrieved documents

4 Experimental results

In this section we describe our parameter selection and present our experimental results for the CLEF 2006 Robust track. Results are given for baseline retrieval without feedback, after the application of our PRF method and after the further application of our re-ranking procedure. Our experiments used the Title and Description (TD) fields or Title, Description and Narrative (TDN) fields of the topics.

For all runs we present the precision at both 10 and 30 documents cutoff (P10 and P30), standard TREC average precision results (AvP), the number of relevant documents retrieved out of the total number of relevant in the collection (RelRet), and the change in number of RelRet compared to Baseline runs.

4.1 Selection of System Parameters

To set appropriate system parameters development runs were carried out using the training topics provided. The topics provided were taken from the CLEF 2003 The Okapi parameters were set as follows k1=1.2 b=0.75. For all our PRF runs, 5 documents were assumed relevant for term selection and document summaries comprised the best scoring 6 sentences in each case. Where the number of sentences was less than 6, half of the total number of sentences were chosen. The *rsv* values to rank the potential expansion terms were estimated based on the top 20 ranked assumed relevant documents. The top 20 ranked expansion terms were added to the original query in each case. Based on results from our previous experiments, the original topic terms are upweighted by a factor of 3.5.

4.2 Experimental Results

Table 1 summarises the results of our experiments. Results are shown for the following runs:

Baseline – baseline results without PRF using TDN topics fields
f20narr – feedback results using the TDN topic fields with 20 terms added
f20re-ranked - same as f20narr, but documents are re-ranked using equation (1)
f20desc – feedback results using the TD topics fields with.20 terms added

Comparing the Baseline and f20narr runs it can be seen that application of PRF improves all the performance measures for all runs with the exception of the RelRet for Spanish monolingual where there is a small reduction. By contrast for the Spanish bilingual run there is a much larger improvement in RelRet than is observed for any of the other runs. Application of the re-ranking method to the f20narr list produces little change in the ranked output. The only notable change is a further improvement in the RelRet for the Spanish bilingual task. Varying the value of the b factor in equation 1 made only a small difference to the results. We are currently investigating the reasons for these results, and exploring approaches to the re-ranking method which will have a greater impact on the output ranked lists.

The PRF using only the TD topic fields show that the addition of the N field produces an improvement in retrieval effectiveness in call cases.

Table 1. Results for Baseline, PRF and re-ranked runs results for the CLEF 2006 Robust track.

	Run-ID	English	French	Spanish	Italian	Spanish bi
Baseline (TDN)	P10	0.422	0.395	0.485	0.382	0.357
	P30	0.265	0.269	0.351	0.262	0.266
	Av.P	0.544	0.470	0.445	0.388	0.314
	RelRet	1496	2065	4468	1736	3702
F20narr (TDN)	P10	0.436	0.425	0.507	0.434	0.413
	P30	0.276	0.294	0.375	0.296	0.300
	Av.P	0.558	0.504	0.478	0.459	0.357
	RelRet	1508	2091	4413	1779	3856
	Chg RelRet	+12	+26	-55	+43	+154
F20re-ranked (TDN)	P10	0.433	0.424	0.509	0.434	0.407
	P30	0.276	0.295	0.377	0.296	0.298
	Av.P	0.558	0.508	0.480	0.459	0.358
	RelRet	1507	2092	4426	1783	3900
	Chg RelRet	+11	+27	-42	+47	+198
F20desc (TD)	P10	0.396	0.370	0.450	0.398	0.386
	P30	0.261	0.272	0.358	0.279	0.288
	Avep	0.494	0.452	0.435	0.419	0.343
	RelRet	1493	2074	4474	1778	3759
	Chg RelRet	+3	+9	+6	+42	+57

5 Conclusions

This paper has presented a summary of our results for the CLEF 2006 Robust Track. The results show that our summary-based PRF method is consistently effective across this topic set. We also explored the use of a novel post-retrieval re-ranking method. Application of this procedure led to very little modification in the ranked lists, and we are currently exploring alternative variations on this method.

References

1. M.F. Porter. An algorithm for suffix stripping. Program, 14:10-137, 1980.
2. S.E Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu and M. Gatford, Okapi at TREC-3. In D.K. Harman, editor, Proceedings of the Third Text REtrieval Conference (TREC-3), pages 109-126. NIST, 1995.
3. A.M. Lam-Adesina and G.J.F. Jones. Applying Summarization Techniques for Term Selection in Relevance Feedback. In Proceedings of the ACM SIGIR 2001, pages 1-9, New Orleans, 2001. ACM.