

Negations as a high precision mechanism

David E. Losada

University of Santiago de Compostela
Intelligent Systems Group
Dept. of Electronics & Computer Science
Campus Sur s/n
15782 Santiago de Compostela
Spain
dlosada@dec.usc.es

Alvaro Barreiro

University of A Coruña
AILab
Dept. of Computer Science
Campus Elviña s/n
15071 A Coruña
Spain
barreiro@udc.es

ABSTRACT

In this paper we present preliminar large-scale experimental results on the effect of logical negations in the context of a feedback cycle. Terms are selected from judged non-relevant documents and included in the new query as negated literals. This inclusion yields similar average precision ratios but improves precision at low recall levels. For negative term selection purposes, we also ran some experiments in which the set of judged non-relevant document is not considered in its entirety. This evaluation scenario revealed that negations perform better if selected from a few top ranked non-relevant documents.

KEYWORDS: Logic-based Information Retrieval, Relevance feedback

INTRODUCTION

Relevance feedback methods have been mainly focused on moving the query towards the set of judged relevant documents whereas the impact of negative feedback is underexploited. Due to expressiveness limitations, the role of negative feedback is often reduced to decrease the weights of (positive) terms candidate to expand the query. This is an important limitation because non-relevant documents are a valuable source of information which should be exhaustively used. Moreover, there are situations in which relevant documents are not available (e.g. because the original query did not retrieve any relevant item) and, hence, the input to the feedback mechanism is only composed of negative examples.

The importance of negative feedback was already pointed out by Belkin et.al. (1997) in the context of interactive retrieval, Hoashi et.al. (2000) for filtering and Losada and Barreiro (2001a) in the framework of a logical model of information retrieval. Nevertheless, few attempts have been done to conduct *negative-intensive* experiments against large collections of documents, which is an objective here. Losada & Barreiro (2001a) reported significant improvements in average precision for four small collections and, hence, it is interesting to confront these findings with more realistic document collections.

Furthermore, in this work, we show that using top ranked non-relevant documents (*near positives*) yields better results than using all the judged non-relevant documents. This was already advanced by Singhal et.al. (1997) for a routing task but they applied the Rocchio's algorithm on the vector-space model and, thus, negative feedback effect is limited.

MODEL AND EVALUATION

Documents and queries are represented as propositional logic formulas and similarity is measured through belief revision techniques (PLBR model, Losada and Barreiro (2001b)). The PLBR similarity includes idf for matching terms but only a binary notion of term frequency can be handled. Wall Street Journal documents in TREC vols 1&2

(173k docs) and TREC-3 #151-#200 topics were used for evaluation purposes. Original logical queries are simply a conjunction of positive literals coming from stopped and stemmed terms of the TREC topic. Top 20 documents are used for feedback. Queries are expanded with positive (from judged relevant docs) and negative terms (from judged non-relevant docs) which are selected using postings. A residual evaluation methodology was applied. First, observe table 1, cols 1-4, where we contrast the baseline residual (first retrieval with top 20 docs removed) with the best results expanding with only positive terms (we tried out expansions with 10, 20, 30, 40 & 50 positive terms being 40 the optimum value). Next, we fixed the number of positive expanded terms and we tried out expansions with 2, 3, 5 and 10 negated terms. Negative terms were selected either from all the judged non-relevant documents or from the x non-relevant docs in a higher position in the rank (near positives), with $x=10, 5, 3$ and 2 . Our first finding was that the best selection of negative terms is obtained when only 3 or 5 near positives are considered. This circumstance was recurrent across all sizes of negative expansions. Indeed, to choose negative terms from all non-relevant documents was always the worst approach. In terms of overall performance, to expand with 40 positives and 10 negatives (selected from the 5 top ranked non-relevant documents) yielded the best results (table 1, cols 5&6).

Baseline residual		Positive feedback (40+)		Negative feedback (40+, 10-)	
Avg.prec.	Prec.at rec 0	Avg.prec.	Prec.at rec 0	Avg.prec.	Prec.at rec 0
18.57%	56.1%	23.64%	65.5%	23.39%	67.94%

Table 1. Baseline residual vs Positive feedback

For all the experiments involving negative terms average precision was roughly the same than the average precision obtained with positive terms. On the contrary, regarding precision at recall level 0, experiments with negative terms always improved this ratio. Although the overall improvement is modest (3.7% for the best case) it was consistent across all experiments. This suggests that negations can become a valuable precision-oriented mechanism. We believe that negative feedback performance is very sensible to the quality and discriminative power of negated terms and, hence, future efforts will be directed to revisit the negative term selection method (perhaps postings is not the most adequate way for negative feedback). Furthermore, the combined use of negated terms and logical disjunctions (e.g. to represent several *views* of a query) will also receive our attention in the near future.

ACKNOWLEDGEMENTS

This work was supported by projects TIC2002-00947 (from “Ministerio de Ciencia y Tecnología”) and PGIDT03PXIC10501PN (from “Xunta de Galicia”). The first author is supported in part by “Ministerio de Ciencia y Tecnología” and in part by FEDER funds through the “Ramón y Cajal” program.

REFERENCES

- Belkin, N., Perez Carballo, J., Cool, C., Lin, S., Park, S., Rieh, S., Savage, P., Sikora, C., Xie, H. and Allan, J. (1997) Rutgers’ TREC-6 interactive track experience. *Proc. of TREC’6*, Gaithersburg, USA, 597-610.
- Hoashi, K., Matsumoto, K., Inoue, N. and Hashimoto, K. (2000) Document filtering method using non-relevant information profile. *Proc. of ACM SIGIR ’00*, Athens, Greece, 176-183.
- Losada, D.E. and Barreiro, A. (2001a) An homogeneous framework to model relevance feedback. *Proc. of ACM SIGIR ’01*, New Orleans, USA, 422-423.
- Losada, D.E. and Barreiro, A. (2001b) A logical model for information retrieval based on propositional logic and belief revision. *The Computer Journal* **44**(5), 410-424.
- Singhal, A., Mitra, M. and Buckley, C. (1997) Learning routing queries in a query zone. *Proc. of ACM SIGIR ’97*, Philadelphia, USA, 25-32.