

Appendix: Experiments with the 2004 TREC Novelty Track collection

David Parapar and Álvaro Barreiro

IR Lab, Department of Computer Science,
University of A Coruña, A Coruña, Spain
dparapar@udc.es, barreiro@udc.es

The 2004 collection of the TREC Novelty Track consists of 50 topics specifically created for the 2004 track [1]. Two kinds of topics were created: 25 five topics about events, for news articles, and and 25 topics about opinions for articles about controversial subjects. For each topic, the documents were retrieved from the AQUAINT collection using the WebPRISE, NIST's IR system. Some irrelevant documents retrieved by the system were included in the document sets for each topic. This means that irrelevant documents that are close matches are included in the collection. The authors of the topics, primary assessors, made the relevance judgments of their own topics and a secondary group of assessors provided a second set of judgments. The percentage of relevant sentences from the primary assessors, that was taken as the official for this task is 19.2%.

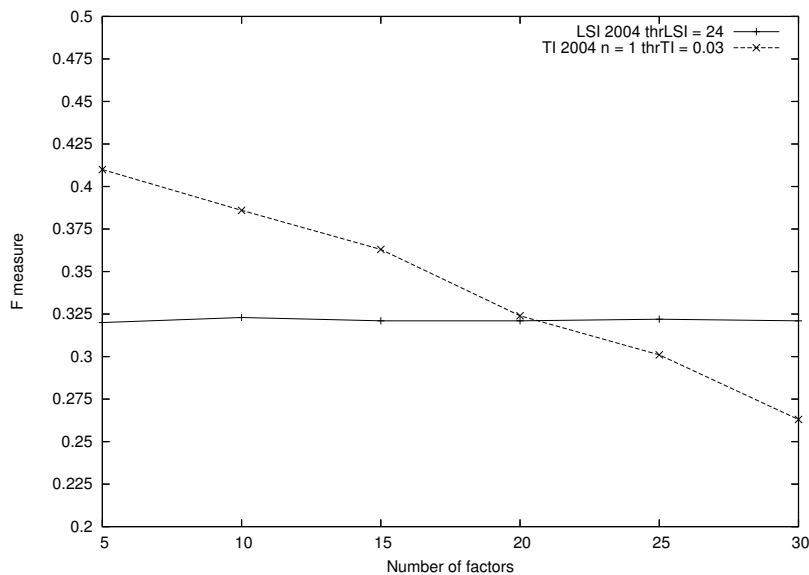


Fig. 1. F measure for the best runs in function of the number of factors (k), 2004 collection

In figure 1 the best experiments for the 2004 collection with both methods are shown. Again, for this collection, the results of the LSI method are worse than the results of the topic identification method. The best results were obtained using 10 factors ($k = 10$) in the truncation of the SVD, although there are similar results with k varying in the range from 10 to 25. The higher F value was 0.32 with THR_{LSI} in the range between 18 and 24.

In the same collection the topic identification method improves substantially the performance. The best results were obtained using 5 factors ($k = 5$), the similarity threshold THR_{TI} has a value of 0.03 and the parameter n for the number of query components has a value of 1. The best F obtained was 0.41. These results are similar to the ones obtained by the best runs submitted to the track with an F value around 0.42. The F obtained with the second human judgments is 0.6.

References

1. Soboroff, I.: Overview of the TREC 2004 Novelty Track. NIST Special Publication 500-261: The Thirteenth Text REtrieval Conference. (2004)