Evaluating Relevance Feedback and Display Strategies for Searching on Small Displays

Vishwa Vinay[1], Ingemar J. Cox[1], Natasa Milic-Frayling[2], Ken Wood[2]

[1] Department of Computer Science, University College London, UK
v.vinay@cs.ucl.ac.uk, ingemar@ieee.org

[2] Microsoft Research Ltd, 7 J J Thomson Avenue, Cambridge, UK
{natasamf, krw}@microsoft.com

Extended Abstract

Searching information resources using mobile devices is affected by displays on which only a small fraction of the set of ranked documents can be displayed. In this study we explore the effectiveness of relevance feedback methods in assisting the user to access a predefined target document through searching on a small display device. We propose an innovative approach to study this problem. For small display size and, thus, limited decision choices for relevance feedback, we generate and study the complete space of user interactions and system responses. This is done by building a tree - the documents displayed at any level depend on the choice of relevant document made at the earlier level. Construction of the tree of all possible user interactions permits an evaluation of relevance feedback algorithms with reduced reliance on user studies. From the point of view of real applications, the first few iterations are most important – we therefore limit ourselves to a maximum depth of six in the tree.

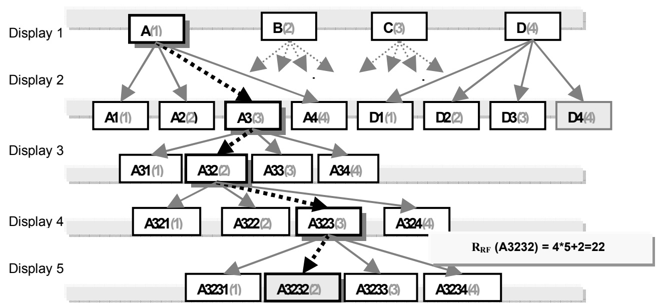

Figure 1. Decision tree for iterative relevance feedback, showing nodes in which the target document is reached, the rank of a document within each display, and the calculation of RF-rank for the target. This branch is expanded only till depth 5 because the target has been found

We use the Rocchio relevance feedback scheme in conjunction with the *tf-idf* scheme where documents and queries are represented as vectors of term weights normalized for length, and similarity is measured by the cosine distance between these vectors. We only consider relevant documents, with the Rocchio feedback weights all being 1. The search task is to *find* a randomly chosen *target* in the database using an initial *query* of four randomly chosen words from the target. The evaluation metric is the total number of documents *seen* before the target is found. The baseline is the rank of the document after the initial query (R_{Scroll}), i.e. before any relevance feedback is applied. The *minimum feedback rank* (min R_{RF}) for a given target document corresponds to the best case scenario where the user always provides the system with the optimal choice of document for relevance feedback, thus providing an upper bound on the effectiveness of relevance feedback. The *number of target document occurrences* in a tree provides a measure of the likelihood of a non-ideal user locating the target document. At each search iteration, we display K=4 documents to the user. The most obvious strategy is to display the K documents with the highest rank which is likely to result in a set of documents all very similar to one another. An alternative approach is to display a selection of documents such that a user's response maximizes the *immediate information gain* to the system and helps to minimize the number of search iterations. This is approximated by sampling K documents from the underlying distribution of similarity. In the experiments we use the Reuters-21578 collection of textual documents. Using the 19,043 documents that have non-empty "Title" and "Body" fields, we remove the stop words and create a vector representation of documents with *tf-idf* weights. Table 1 contains the statistics of *successful searches*, ie; trees which contain the target. The RF rank of an *ideal user* is the minimum path length from the root of the tree to a node with the target, whereas the mean length of all paths leading to the target represents the average performance of successful users. For the Top-K scheme, 52 of the 100 trees contained the target, whereas the corresponding number was 97 for the sampled scheme. However, 4.49% of paths in successful searches led to the target for Sampled displays as opposed to 46.67% for the Top-K.

Table 1 : Performance of Rocchio RF Algorithm based on the Initial Query

Scroll Rank Range	Number Targets	Number of Targets Found		Avg. No. of Documents seen without RF		Avg. No. of Documents seen by the 'ideal user' using RF		No. of Docs. seen with RF averaged over all successful users	
		Top-K	Sampled	Top-K	Sampled	Top-K	Sampled	Top-K	Sampled
1 – 20	45	45(100%)	45(100%)	4.37778	4.37778	4.31111	5.33333	16.5418	19.1322
21 – 40	14	6(42.8%)	14(100%)	25.5	29.7857	20.6667	13.0714	21.6236	21.919
41 – 60	5	0(0%)	5(100%)	-	54.2	-	16.6	-	21.9912
61 – 80	4	0(0%)	4(100%)	-	66.5	-	16.5	-	21.8056
81 – 100	6	0(0%)	6(100%)	-	92.8333	-	15.3333	-	21.4944
>100	26	1(3.84%)	23(89%)	367	341.304	20	18.5652	20.7828	22.1351

The results indicate that if the user's query is sufficiently accurate, then the initial rank of the target document is likely to be high and scrolling or relevance feedback with a greedy display performs almost equally well. However, if the user's initial query is poor, then scrolling is futile and relevance feedback with a display strategy that maximizes information gain is preferable. Amongst the two display strategies, the success of the greedy update relies on a good initial query, whereas the sampled update provides performance almost independent of the initial query but is very sensitive to feedback. Future work includes the examination of other display strategies, including hybrid strategies that attempt to optimally combine the exploratory properties of maximizing information gain with the exploitative properties of greedy displays, and also to verify our results with a user trial.

References

[1] Cox, I. J., Miller, M.L., Minka, T.P., Papathomas, T.V., and Yianilos, P.N. The Bayesian Image Retrieval System, PicHunter: Theory, Implementation and Psychophysical Experiments. IEEE Transactions on Image Processing, 9(1):20-37, 2000.

[2] Harman, D. Relevance feedback revisited. Proceedings of 15th annual international ACM SIGIR conference on research and development in information retrieval, Copenhagen, 1.10, 1992.

[3] Rocchio, J. Relevance feedback information retrieval. In Gerard Salton (ed.): The Smart Retrieval System — Experiments in Automatic Document Processing, pp. 313–323. Prentice-Hall, Englewood Cliffs, N.J., 1971.

[4] Robertson, S.E., Sparck Jones, K. Relevance weighting of search terms. Journal of the American Society for Information Science 27, 1976, pp. 129-146.

[5] Sparck Jones, K., Walker, S., and Robertson, S.E. A probabilistic model of information retrieval: development and comparative experiments. Information Processing and Management 36 (2000) 779-808, 809-840.