

On Perfect Document Rankings for Expert Search

Craig Macdonald and Iadh Ounis
Department of Computing Science
University of Glasgow, Scotland, UK
{craigm,ounis}@dcs.gla.ac.uk

ABSTRACT

Expert search systems often employ a document search component to identify on-topic documents, which are then used to identify people likely to have relevant expertise. This work investigates the impact of the retrieval effectiveness of the underlying document search component. It has been previously shown that applying techniques to the underlying document search component that normally improve the effectiveness of a document search engine also have a positive impact on the retrieval effectiveness of the expert search engine. In this work, we experiment with fictitious *perfect* document rankings, to attempt to identify an upper-bound in expert search system performance. Our surprising results infer that non-relevant documents can bring useful expertise evidence, and that removing these does not lead to an upper-bound in retrieval performance.

Categories and Subject Descriptors: H.3.3 [Information Storage & Retrieval]: Information Search & Retrieval

General Terms: Performance, Experimentation

Keywords: Expert Search, Document Search

1. INTRODUCTION

An expert search system is designed to rank candidate persons in response to a query, usually using documentary evidence that represent each candidate's expertise areas. Many expert search models are based on the premise that the more a document is related to the topic of the query, the more likely that candidates associated to that document will have relevant expertise. However, the manner in which the strength of topicality - how much each document is related to the query - influences the final expert search effectiveness has seen less analytical research. Typically, experiments have shown that by applying a known technique which usually improves the retrieval performance of a document search engine, performance is also improved for the expert search engine. In particular, both Macdonald & Ounis [4] and Balog et al. [2] showed that the training of the document search component could improve the retrieval effectiveness of the expert search engine. Moreover, when field-based weighting models [4], query term proximity [3] or query expansion [3] were applied to the document ranking, the retrieval performance of the candidate ranking could again be enhanced. In contrast, in this work, we apply a different method of changing the document ranking,

namely the application of a perfect document ranking. To our best knowledge, this is the first study of the extent to which an upper-bound limit of expert search performance is achievable when in presence of perfect document rankings. The central contribution of this work is the observation that a perfect document ranking system does not necessarily lead to an upper-bound expert search performance. In doing so, we suggest that the document search systems created for use by expert search systems should have different characteristics than previously thought.

2. EXPERT SEARCH APPROACHES

Over the past few years, various approaches have been proposed for expert search, where documentary evidence of expertise for each candidate (called candidate profiles) is used to rank candidates in response to a query. In general, the most effective approaches use a ranking of documents that is mapped into a ranking of candidates. This is the approach taken by the Voting Model [4], which sees the expert search task as a voting process. In the Voting Model, the ranking of documents (denoted $R(Q)$) defines votes for candidates to be retrieved: each time a document associated with a candidate is ranked in $R(Q)$, then this is an implicit vote for that candidate to have relevant expertise to the query. In this work, we use two rank-based *voting techniques* to create rankings of candidates from documents in $R(Q)$: ApprovalVotes (which uses the presence of documents as votes) and BordaFuse (which uses the ordering of the voting documents) [4]. Other voting techniques use the scores of the voting documents (e.g. expCombMNZ and others in [4]).

The focus of this work is to investigate the role of the document ranking, and its effect on the quality of the generated ranking of candidates. At first hand, it seems intuitive that a more refined, *higher quality* document search component will allow an expert search system to attain improved retrieval performance. Indeed, various studies have examined how the application of existing or adapted information retrieval (IR) techniques - each known to enhance retrieval effectiveness on document retrieval tasks - to the document search component of the expert search system has resulted in increased retrieval performance [2, 3, 4]. In doing so, more on-topic documents are ranked higher, providing more refined expertise evidence that can be used to rank the experts.

3. PERFECT DOCUMENT RANKINGS

Given knowledge of the relevant documents for a query, a perfect ranking is easy to generate, with only and all relevant documents being retrieved. In the following, we use perfect document rankings to determine if they lead to an

Document Ranking	ApprovalVotes		BordaFuse	
	MAP	MRR	MAP	MRR
Perfect (Mean)	0.2867	0.3643	0.2858	0.3654
Perfect (StdDev)	0.0000	0.0000	0.0124	0.0114
Perfect (Max)	0.2867	0.3643	0.3028	0.3894
DLH13	0.2250	0.3178	0.2747	0.3679
DLH13 Proximity	0.2486	0.3397	0.3138	0.4198

Table 1: Maximum achievable retrieval performance by two voting techniques, when perfect document rankings are used. Comparable results using standard weighting models, using their default and trained settings are also shown.

upper-bound in expert search performance. In particular, in our experiments, we use the two evaluation tasks of the TREC 2007 Enterprise track, namely the document search task (DS), and the expert search task (ES). The aim of the DS task was to identify relevant documents for each query, particularly those which were key in understanding the topic area [1]. For the ES task, the participating systems should suggest relevant experts for each query. Importantly, the query topics used for both tasks are identical.

We generate 10 perfect document rankings for each query, using the DS task relevance assessments. Each document ranking is different, as a different ranking of the relevant documents may have an impact on the effectiveness of the voting techniques that consider the ordering of documents. However, the MAP, MRR, P@10, Recall, etc. of each document ranking is 1.0, as all relevant documents are retrieved, and no irrelevant ones are retrieved. The size of the document ranking is not limited, i.e. all and only relevant documents are retrieved, giving an average of 147.2 (relevant) documents retrieved per query.

We then use candidate-document associations to map document votes from the perfect rankings into candidate votes, using ApprovalVotes and BordaFuse voting techniques (score-based voting techniques require documents to have scores). For comparison, we also experiment with sub-perfect rankings, by using the top 1,000 documents retrieved by a standard document weighting model, namely DLH13 [4], as well as the integration of a query-term proximity model [3]. For these experiments, the CERC collection, including anchor text, is indexed using Terrier¹, removing standard stopwords and applying the first two steps of Porter’s English stemmer.

The results are presented in Table 1. In particular, for each (candidate ranking) evaluation measure and voting technique, we report the mean, standard deviation and maximum of the evaluation measure over the candidate rankings generated by the 10 perfect document rankings.

From the results in Table 1, we note that the two voting techniques perform very similarly over the 10 perfect document rankings applied. It is of note that ApprovalVotes is not dependant on the order of documents in the document ranking, and hence there is no variation across the 10 permutations of perfect rankings. The highest MAP achieved by the BordaFuse voting technique on a perfect document ranking is 0.3028. Table 1 also contains the results using rankings created by the DLH13 weighting model. Comparing across the results, we note that, for the ApprovalVotes technique, the perfect recall of the perfect document rankings ensures that the MAP & MRR achieved using the perfect document ranking are higher than those achieved using various standard document weighting models and techniques.

¹<http://terrier.org/>

However, for BordaFuse, the mean expertise retrieval performance achieved using the perfect document rankings is actually lower than some of the results of the sub-perfect document rankings (e.g. when using proximity). On checking using the Wilcoxon signed-rank test, we find that the perfect-based expert search rankings are not significantly better or worse than those created by the DLH13 weighting model.

These rather surprising results allow us to postulate that not all relevant on-topic documents may be good indicators of expertise evidence, and their exact ordering has an impact on the retrieval performance achievable by the BordaFuse voting technique. In this case, the ordering of documents which would give the most effective candidate ranking would have the strongest evidence for the relevant candidates first, followed by the less strong evidence for the relevant candidates, followed by tangential evidence for the relevant candidates. Documents also associated to irrelevant candidates should be pushed down in the document ranking.

Overall, the results here allow us to assert that it is not just the presence or ordering of relevant documents which have an impact on the accuracy of a ranking of candidates. Instead, documents which are retrieved but not relevant to the topic could have a positive bearing on the accuracy of the ranking of results. For instance, these documents may not exactly be on-topic (so would have been judged irrelevant during document judging), but they are about the same general topic area, and are associated to relevant candidate(s). In retrieving these documents, a document search engine may bring more evidence of expertise than the perfect IR systems simulated here. Simply irrelevant documents which happen to be associated to relevant candidates could also have a positive impact on expert search performance. However, developing an IR system to find such documents is likely to be difficult.

4. CONCLUSIONS

In this work, we used perfect document rankings as input to expert search systems. However, the use of these perfect document rankings did not produce a significant increase in the retrieval accuracy of two rank-based voting techniques over document rankings produced by a standard retrieval system. This interesting observation shows that the ordering of relevant documents may be important, and also suggests that documents that are irrelevant, but somehow related to the topic area could also have a positive bearing on the retrieval performance of the voting techniques, if associated to relevant candidates.

We propose as future work to empirically determine the benefit for expert search of each document in a document ranking, and then comparing this to the known relevance of each document. Such an analysis will provide additional insights into the usefulness of relevant and irrelevant documents for expert search.

5. REFERENCES

- [1] P. Bailey, N. Craswell, A. P. de Vries, and I. Soboroff. Overview of the TREC-2007 Enterprise Track. In *Proceedings of TREC-2007*.
- [2] K. Balog, L. Azzopardi, and M. de Rijke. Formal models for expert finding in enterprise corpora. In *Proceedings of SIGIR 2006*, pages 43–50.
- [3] C. Macdonald. *The voting model for people search*. PhD thesis, Univ. of Glasgow, 2009.
- [4] C. Macdonald and I. Ounis. Voting for candidates: adapting data fusion techniques for an expert search task. In *Proceedings of CIKM 2006*, pages 387–396.