Authoritative Re-Ranking of Search Results

Toine Bogers and Antal van den Bosch

ILK / Language and Information Sciences, Tilburg University,
P.O. Box 90153, NL-5000 LE Tilburg, The Netherlands
{A.M.Bogers,Antal.vdnBosch}@uvt.nl

Abstract. We examine the use of authorship information in information retrieval for closed communities by extracting expert rankings for queries. We demonstrate that these rankings can be used to re-rank baseline search results and improve performance significantly. We also perform experiments in which we base expertise ratings only on first authors or on all except the final authors, and find that these limitations do not further improve our re-ranking method.

1 Introduction

Professionals have several options available to fulfill their information needs. One particularly rich source of useful information is the combined body of publications of the workgroup that the professional is a part of. Colleagues share both interests and vocabulary, and publications of colleagues are also often considered to be more trustworthy compared to random documents found in libraries and on the WWW. Workgroup members are bound by the common research focus of the workgroup, but each member also has separate interests and may be the group's expert on specific topics. By adopting a wider perspective and by disregarding institutional proximity, scientific communities or collectives of people who publish articles in a specific journal can also be considered a workgroup; in the remainder of this paper we use "workgroup" to refer to both meanings.

In this paper we present *authoritative re-ranking*, a novel method of re-ranking search results which utilizes information on topical expertise of workgroup members to improve the information retrieval process within these workgroups. We assume that we can estimate the expertise of each member of the workgroup from the aggregated content of his or her publications. Based on this, we estimate how well a term or phrase points to a certain expert, by calculating the author-term co-occurrence weights. We describe a method to create expertise rankings of the workgroup members for a query, and use these rankings to re-rank the search results produced by a baseline system. We performed experiments to determine which authors contribute the most to this re-ranking.

Constructing rankings of author expertise is a relatively new subfield of information retrieval research. TREC 2005 marked the introduction of the 'Expert Search Task', aimed at solving the problem of identifying employees who are the experts on a certain topic or in a certain situation [6]. Campbell et al. [2] performed similar experiments on a corpus of e-mail messages sent between people

in the same company. Neither approach uses these expertise rankings to improve search results. A considerable amount of research has been devoted to improving the search results of information retrieval systems. Among the more successful approaches are query expansion [7] and using cluster analysis [5] or citation analysis for re-ranking purposes [4].

2 Authoritative Re-Ranking

Our re-ranking approach was designed to be used on top of a basic TF·IDF vector space model of information retrieval. In our experiments, we used the formulas for document weights and query weights as determined by [3]. We incorporated some tried and tested low-level NLP-techniques into our baseline system, such as stop word filtering and stemming. We also experimented with statistical and syntactic phrases and optimized the use of these techniques for every test collection, as recommended by [1].

We partitioned the documents into one-vs-all data sets for each author and then calculated the co-occurrence weights of each author-term pair for each term (words and phrases) that occurred in the collection. The weights were determined using the following feature selection metrics from text categorization: Information Gain, Chi-Square, and Mutual Information [8]. We also tested using the average TF·IDF value as a measure of term informativeness; collection terms that did not occur in the author's documents were assigned a score of zero.

Combining these term weights for each author yielded a matrix of term-author weights for each of these metrics. For each query-author combination an expert score was calculated that signified the expertise of the author on the query topic. Calculating the expert scores is based on the straightforward assumption that if terms characteristic for author A occur in query Q, A is likely to be more of an expert on Q. For each author separately, the informativeness weights were collected for each of the query terms and combined into an expert score. We experimented with taking an unweighted average of the weights and an average weighted by the TF·IDF values of the query terms.

Re-ranking the baseline results using these expert rankings was the final step in authoritative re-ranking. It is based on the premise that the documents authored by the experts on the current query topic are more likely to be relevant to the query, i.e. more *suitable* to fulfill the query. Since many documents have multiple authors, the expert scores associated with each document had to be combined. Early experimentation showed that weighting the expert scores with the total number of publications of each author gave the best performance. We also investigated abating the influence of high numbers of publications with the square root and the natural logarithm of these counts as weighting factors. After computing this 'suitability' score, which is computed for each query–document combination, it is combined with the original baseline similarity score to form a new score on the basis of which the authoritative re-ranking is performed.

We performed additional experiments to test and fine-tune the ways in which similarity scores and suitability scores can be combined. The most successful

collection–author selection	re-ranked	baseline	
CACM	0.313	0.233	(+34.3%)
CACM–first	0.302		(+20.2%)
CACM–m1	0.304		(+30.5%)
CISI	0.206	0.203	(+1.5%)
CISI–first	0.206		(+1.5%)
CISI–m1	0.206		(+1.5%)
ILK	0.649	0.647	(+0.3%)
ILK–first	0.650		(+0.5%)
ILK–m1	0.656		(+1.4%)

Table 1. Comparison of the re-ranking approaches in terms of R-precision scores. The underlined scores are statistically significant improvements.

combinations involved multiplying the original similarity score with the (normalized) suitability score S and transforming the original similarity score by multiplying it with $1 + S$. Experiments showed that the optimal re-ranking settings were collection-dependent, so the settings were optimized for each collection, similar to the NLP techniques used in the baseline [1].

3 Evaluation

Investigating the merits of authoritative re-ranking in workgroups required testing our approach on test collections that (1) contain information about the authors of each document, and (2) are a realistic representation of a workgroup. We used two well-known test collections, **CACM** (3204 documents, 52 queries, and 2963 unique authors) and **CISI** (1460-76-1486), and we created a third collection called **ILK** (147-80-89), because to our knowledge no real workgroup test collections exist. **ILK** contains 147 document titles and abstracts of publications of current and ex-members of the ILK workgroup. The paper topics focus mainly on machine learning for language engineering and linguistics[1]. We also performed some experiments to determine which author rank contributes most to expertise re-ranking. We created special versions of each corpus where only the primary authors were included (**CACM–first**, **CISI–first**, and **ILK–first**), and versions where the last author was removed from the author listings (**CACM–m1**, **CISI–m1** and **ILK–m1**). Our hypothesis was that, on average, first authors contribute the most to a paper and final authors the least.

We evaluated the performance of our approach using R-precision, the precision at the cut-off rank of the number of relevant documents for a query. It emphasizes the importance of returning more relevant documents earlier. The reliability of the comparisons between our baseline system and the re-ranking approach was determined by performing paired t-tests.

[1] Publicly available at http://ilk.uvt.nl/~tbogers/ilk-collection/.

Table 1 shows the results of our experiments. Authoritative re-ranking produced statistically significant performance improvements on the **CACM** and **CISI** collections, ranging from +1.5% to +34.3%. The improvements seem to be dependent on the corpus, but even the optimal performance on the **ILK** collection yielded very small improvements. A possible reason for the differences in performance might be the topical diversity of the test collections: **CACM** seems to have a more diverse range of topics than **CISI** and **ILK** which might make it easier for different fields of expertise to be recognized.

The experiments with different author selections did not confirm our initial hypothesis. Using the expertise of all authors associated with a document yields the best results and using less authors did not increase performance significantly.

4 Conclusions

Under optimized settings, authoritative re-ranking is able to significantly boost R-precision, with the exact performance increase dependent on the document collection. The technique appears to be suited for collections with a fair topical heterogeneity, such as publications in a journal, and perhaps less so for collections of workgroups with more topical coherence among publications. Furthermore, optimal re-ranking performance requires using the expertise of all the authors associated with a document.

Acknowledgements. This research was funded by the IOP-MMI-program of SenterNovem / The Dutch Ministry of Economic Affairs, as part of the À Propos project. The authors would like to thank Frank Hofstede (Intelli-Gent B.V.) for the fruitful discussions and useful comments.

References

1. Brants, T. Natural Language Processing in Information Retrieval. In *Proc. of CLIN 2004*, Antwerp, Belgium, pp. 1–13, 2004
2. Campbell, C.S., Maglio, P.P., Cozzi, A., and Dom B. Expertise Identification using Email Communications. In *Proc. of CIKM2003*, New Orleans, LA, pp. 528–531, 2003
3. Chisholm, E. and Kolga, T.G. *New Term Weighting Formulas for the Vector Space Method in Information Retrieval*. Technical report ORNL/TM-13756, Computer Science and Mathematics Division, Oak Ridge National Laboratory, 1999
4. Giles, C.L., Bollacker, K., and Lawrence, S. CiteSeer: An Automatic Citation Indexing System. In *Proc. of Digital Libraries 98*, Pittsburgh, PA, pp. 89–98, 1998
5. Lee, K.-S., Park, Y.-C., and Choi, K.-S. Re-ranking model based on document clusters. In *Information Processing & Management*, vol. 37, no. 1, pp. 1–14, 2001
6. TREC. *TREC Enterprise Track*. http://www.ins.cwi.nl/projects/trec-ent/, 2005
7. Xu, J. and Croft, W.B. Query Expansion Using Local and Global Document Analysis. In *Proc. of SIGIR'96*, Zurich, Switzerland, pp. 4–11, 1996
8. Zheng, Z. and Srihari, R. Optimally Combining Positive and Negative Features for Text Categorization. In *Workshop for Learning from Imbalanced Datasets II, Proc. of the ICML*, Washington, DC, 2003