Reducing the Uncertainty in Resource Selection

Ilya Markov[1], Leif Azzopardi[2], and Fabio Crestani[1]

[1] University of Lugano, Via G. Buffi 13, 6900, Lugano, Switzerland
{ilya.markov,fabio.crestani}@usi.ch
[2] University of Glasgow, 18 Lilybank Gardens, G12 8QQ, Glasgow, UK
Leif.Azzopardi@glasgow.ac.uk

Abstract. The distributed retrieval process is plagued by uncertainty. Sampling, selection, merging and ranking are all based on very limited information compared to centralized retrieval. In this paper, we focus our attention on reducing the uncertainty within the resource selection phase by obtaining a number of estimates, rather than relying upon only one point estimate. We propose three methods for reducing uncertainty which are compared against state-of-the-art baselines across three distributed retrieval testbeds. Our results show that the proposed methods significantly improve baselines, reduce the uncertainty and improve robustness of resource selection.

1 Introduction

Distributed Information Retrieval (DIR) [5,14] involves searching over multiple sources of information such as distributed digital repositories or verticals [2,7]. Searching across multiple distributed, and usually uncooperative, sources, first requires sampling each source, then given a query, resource selection is performed, before the retrieved results are merged [5,14]. As a result, the process is inherently uncertain: starting from incomplete and possibly biased samples of federated sources, going through two levels of imprecise estimates of retrieval scores for sources and documents, ending up with a heuristic score normalization for merging. Some aspects of uncertainty within the process have been investigated. For instance, significant work has focused on improving the accuracy and representation of the samples. However, the uncertainty arising within the resource selection component has not been fully addressed. In this paper, we posit that by tackling the uncertainty of resource selection, which is a crucial DIR component, we can improve the performance and robustness of the DIR process.

Dealing with uncertainty has been a central theme within IR and as a result numerous works have tried to overcome this problem in various ways (eg. [9,10,23,25]). Recently, approaches that try to deal with the risk and uncertainty have become quite popular and have been shown to be effective for centralized IR (for example, Risk Minimization Framework [23], Portfolio Theory [21], etc). In this work, we shall adapt the method proposed by Collins-Thompson and Callan for reducing uncertainty within the query expansion process [9] to reduce uncertainty within resource selection. We develop three variants and compare

P. Serdyukov et al. (Eds.): ECIR 2013, LNCS 7814, pp. 507–519, 2013.

them against several state-of-the-art DIR testbeds. Our results show that significant improvements over baselines can be obtained. Moreover, the proposed techniques improve robustness of resource selection by increasing the agreement between different selection methods and by marginalizing out some of the factors that introduce uncertainty.

The remainder of the paper is structured as follows. In Section 2 we give an overview of DIR and uncertainty associated with each of its phases and then focus on resource selection. We discuss the ways to deal with uncertainty in resource selection in Section 3. Section 4 introduces the experimental setup that is used in Section 5 to evaluate the proposed methods. We conclude the paper by outlining our findings and directions for future work in Section 6.

2 Background

The DIR process is composed of a number of phases: (i) sampling/representation of sources, (ii) resource selection, and (iii) results merging. DIR usually operates in uncooperative environments, where no access to the content of federated sources, other than a user interface, is available. Therefore, sampling is required to obtain a representation of each source. This is accomplished by sending a number of queries to each source and storing the obtained documents in a centralized index (this is known as query-based sampling [6]). Since only a sample of the documents from each source is taken, the subsequent DIR steps, such as resource selection and results merging, have incomplete and possibly biased information to form estimates. While obtaining larger samples would significantly decrease the amount of incompleteness and uncertainty given the lack of information, cost constraints often preclude from doing so [6]. Often, a reasonable tradeoff between cost (the number of documents sampled from each source) and effectiveness (the final performance of a DIR system) needs to be found [3]. Consequently, much research within DIR has focused on obtaining the most representative samples from sources using the least number of documents, trying to increase the content similarity between a sample and a full source [6,8] and the alignment of a sample to user needs [3,16].

When a user's query is submitted to a DIR system, resource selection identifies the most relevant sources, where the query is then forwarded to, and results merging combines the obtained source-specific results. Since resource selection is the main focus of this work we will discuss it in more details in Section 2.1.

The last phase of the DIR process, results merging, aims at combining results coming from selected sources by normalizing document scores or ranks across sources. CORI results merging uses source scores produced by resource selection to weight results [4]. SSL fits a linear function between source-specific and centralized document scores, but requires an overlap between sample and source-specific results [17]. SAFE fits a function between source-specific document scores and centralized document ranks [15]. Note that these methods depend on the representativeness of a sample, the quality of source-specific and centralized retrieval systems, the parameters of merging techniques and other factors that all introduce uncertainty into the merging process and the final results. However, we leave the study of uncertainty in results merging for future work and instead concentrate on resource selection that is discussed in more details below.

2.1 Resource Selection

Early resource selection techniques combine documents sampled from a particular source into one large document. These large documents, each representing one source, are ranked for a query using a document ranking function. The best known techniques of this category are CORI and cluster-based language modelling approach [4,22].

Better performing second generation resource selection methods, such as ReDDE [18] and CRCS [13], do not combine sampled documents, but store them in a centralized index. The sampled documents are then ranked for a user's query and a score for a source R is calculated based on its documents that appear in the top n of a sample ranking:

$$s(R|q) = \frac{|R|}{|S|} \sum_{d \in R} s(d|q), \tag{1}$$

where $|R|$ is a source size, $|S|$ is a sample size and $s(d|q)$ is a score assigned to a document d for a query q by a centralized system [18]. This formula is used under the assumption that each document in a sample corresponds to $|R|/|S|$ documents in a full source.

ReDDE assigns constant scores to all documents in the top n, i.e. $s(d|q) = c$ [18]. ReDDE.top uses document relevance scores as they are calculated in a centralized sample index [1]. CRCS(e) and CRSC(l) use exponential and linear functions of a sample document rank [13] such that:

$$s_{exp}(d|q) = \alpha e^{-\beta r(d|q)}, \tag{2}$$

$$s_{lin}(d|q) = n - r(d|q). \tag{3}$$

There are many factors affecting the quality of resource selection and which increase the uncertainty. These include:

- the quality of source samples,
- the centralized ranking of sampled documents (which depends upon the retrieval system used),
- the top n documents considered from this centralized ranking, and,
- the parameters of a particular resource selection technique.

Since the first factor is due to the source sampling phase, we focus on trying to reduce the uncertainty arising from the later three. In particular, we will marginalize out the factors that contribute to the uncertainty, i.e. the effect of a retrieval system and resource selection parameters. We will show that, apart from improving performance, this will also increase the robustness of resource selection.

3 Dealing with the Uncertainty of Resource Selection

In [9], Collins-Thompson and Callan examined how uncertainty could be reduced in the context of query expansion (QE). Considering that information retrieval in general and QE in particular are uncertain, the authors argued that the output of a QE technique should be treated as a random variable whose posterior distribution needs to be estimated. To this end a number of candidate feedback models were sampled from a QE method by varying its input and the posterior mean was considered as an enhanced feedback model. By doing so, possible variations in a baseline QE method were smoothed out, thus reducing the uncertainty of QE. In their experiments, Collins-Thompson and Callan showed improvements not only in retrieval performance, but also in the robustness of the query expansion method, where robustness was defined as the number of queries helped by expansion. They observed that less queries suffered a drop in performance, while the performance of more queries was improved when compared to baseline QE methods.

Since the above approach does not assume anything particular about QE techniques involved, the same idea can be applied to deal with uncertainty in other IR contexts. We posit that we can apply Collins-Thompson and Callan's method to reduce the uncertainty in the resource selection problem.

In particular, the input to a second generation resource selection method is a ranked list of sampled documents. In order to vary this input, we create k ranked lists of documents by applying three perturbation techniques (discussed in details below). The output of resource selection is a set of source scores $s(R|q)$. We treat each $s(R|q)$ as a random variable and aim at estimating its posterior distribution by passing k inputs to a resource selection technique and obtaining a number of point estimates:

$$s_1(R|q), s_2(R|q), ..., s_k(R|q). \tag{4}$$

Then we approximate $s(R|q)$ by its posterior mean as follows:

$$s(R|q) = \frac{1}{k} \sum_{i=1}^{k} s_i(R|q). \tag{5}$$

This way resource selection aggregates more information to make a less risky decision about what sources should be selected, in contrast to state-of-the-art second generation techniques that use a single ranking of sampled documents ($k = 1$).

Different ranked lists of documents can be obtained by perturbing queries, documents (or directly their ranking) and a retrieval system [11]. Here we adapt the query and the document ranking perturbation methods from [9] and, in addition, propose to use a retrieval system perturbation technique. Below we discuss each perturbation method in details.

Query Perturbation. Query can be modified in various ways including query expansion, query generation based on retrieved documents [24] and exploring subqueries [9]. The first two approaches move the original query closer to already retrieved documents, while here we would like to explore the space of possible rankings that is not necessarily centred around the original one. The latter approach explores query perturbation techniques in order to capture the uncertainty of query expansion and suggests two strategies for generating subqueries. The *leave-one-out* strategy removes one query term at a time, assuming that is a noise term. The *single-term* strategy uses each query term as a subquery, assuming it represents the main query concept. Each subquery, produced with the above strategies, is combined with the original query with equal weights. The leave-one-out strategy was reported to have consistently high performance [9] and, therefore, we use it here as a query perturbation technique.

Retrieval System Perturbation. Retrieval system produces different rankings of sampled documents depending on a particular retrieval approach, its parameters, preprocessing steps and other factors. Each sample ranking, in turn, gives different source scores and results in a different set of selected sources. In order to remove this effect, we use a number of different retrieval systems and aggregate source scores produced by each one of them.

Ranking Perturbation. Parameters of resource selection methods play an important role in selection performance and have to be optimized for each dataset or even for each query. Following the above line of thinking one would perform resource selection varying its parameters and then aggregate the obtained source scores in order to reduce the effect of variations within a particular selection technique. However, we treat each resource selection method as a black box and, therefore, cannot modify a method itself. Instead we address the variation within resource selection by varying its input, i.e. a ranking of sampled documents.

In particular, we sample with replacement n documents from the ranking, where documents are weighted by their relevance scores, i.e. high scoring documents are sampled more often. More formally, the probability of a document d with a relevance score $s(d|q)$ being sampled is $\frac{s(d|q)}{\sum_i s(d_i|q)}$. Although sampling is itself uncertain, this process, repeated many times, does not introduce and even reduces uncertainty by exploring the space of possible (similar) rankings of documents.

Below we evaluate resource selection based on the above perturbation techniques against state-of-the-art baselines. We will show that the proposed methods improve performance, reduce uncertainty and increase robustness of resource selection.

4 Experimental Setup

The following setup was used to evaluate the proposed perturbation-based resource selection methods against baseline techniques.

Testbeds. We used three testbeds, created based on the TREC GOV2 dataset: gov2.1000, gov2.250 and gov2.30 [1]. The gov2.1000 testbed includes 1000 largest domains of GOV2 (each as a separate source) and contains 22M documents.

In the gov2.250 and the gov2.30 testbeds the 1000 sources of gov2.1000 are clustered into 250 and 30 larger and more homogeneous sources. We used the titles of TREC topics 701-850 as queries.

DIR Setup. Query-based sampling [6] was used to sample 300 documents from each of 1000 sources of the gov2.1000 testbed, 1200 documents from gov2.250 and 10000 documents from gov2.30. We selected these sizes so that the sample index is about the same for all testbeds, i.e. 300K, because the total size of the testbeds is also the same, i.e. 22M. This means that for gov2.250 and gov2.30 there is less and less uncertainty coming from the sample. For combining source-specific results on the results merging phase we used CORI [4] as described in [12], because it was shown to outperform SSL [17] and SAFE [15] in preliminary experiments.

Resource Selection Settings. As resource selection baselines we used ReDDE [18] and CRCS [13] (both linear and exponential versions, denoted as CRCS(l) and CRCS(e) respectively). These have been shown to be the most effective resource selection techniques [13,19]. BM25 was used to rank documents in the centralized sampled index, where we used the top 50 documents for resource selection as suggested in [19].

The three perturbation approaches: (i) query, (ii) retrieval system and (iii) ranking were examined. Each approach produces a different number of document rankings. For (i), given a query of length l, query perturbation gives $l+1$ rankings, because each term is removed from the query once and the original query is also considered. For (ii), we used 10 different document scoring functions: BM25, tf-idf (Terrier and Lemur versions), language modeling (original and with Dirichlet smoothing), and a number of DFR-based functions (BM25, BB2, IFB2, InL2 and PL2)[1], thus producing 10 rankings of sampled documents. Finally, for (iii), k rankings were produced by permutating the original document ranking. In preliminary experiments, where we sampled $k \in [5..500]$ rankings, we observed that the performance converged around $k = 50$. We used $k = 100$ to ensure we were well past the point of convergence.

Performance Measures. To measure the performance of the DIR system, we used p@10 and bpref. We chose to report bpref values because in DIR there are typically a lot of un-assessed documents in the rankings that are promoted due to resource selection. However, it should be noted that the findings based on map were similar and led to the same findings. Statistical significance was measured by the paired t-test at 0.05 level. Similarly to previous studies, we selected 1 and 5 sources for each query and report the results for these settings [13].

5 Results and Discussion

In this section we first compare the retrieval performance of the proposed methods against the baseline resource selection techniques. Then we experiment with

[1] We use Terrier toolkit for indexing and retrieval: `http://terrier.org`

Table 1. Perturbation-based resource selection, gov2.30

Method	Modification	1 selected		5 selected	
		bpref	p@10	bpref	p@10
ReDDE	baseline	0.082	0.309	0.145	0.343
	ranking perturb.	0.089	0.339†	0.171†	0.372†
	query perturb.	0.085	0.306	0.159†	0.338
	system perturb.	0.086	0.311	0.164†	0.356
CRCS(l)	baseline	0.080	0.311	0.175	0.347
	ranking perturb.	0.084	0.336†	0.180†	0.361
	query perturb.	0.074	0.280	0.175	0.328∘
	system perturb.	0.077	0.312	0.178†	0.358
CRCS(e)	baseline	0.086	0.319	0.147	0.343
	ranking perturb.	0.089	0.341†	0.170†	0.364†
	query perturb.	0.082	0.301	0.161†	0.352
	system perturb.	0.087	0.319	0.166†	0.357

Table 2. Perturbation-based resource selection, gov2.250

Method	Modification	1 selected		5 selected	
		bpref	p@10	bpref	p@10
ReDDE	baseline	0.051	0.244	0.089	0.256
	ranking perturb.	0.052	0.247	0.106†	0.275†
	query perturb.	0.056	0.248	0.104†	0.257
	system perturb.	0.053	0.247	0.100†	0.269
CRCS(l)	baseline	0.054	0.210	0.144	0.278
	ranking perturb.	0.058	0.239†	0.140	0.278
	query perturb.	0.056	0.224	0.141	0.257
	system perturb.	0.056	0.218	0.152†	0.270
CRCS(e)	baseline	0.051	0.237	0.088	0.266
	ranking perturb.	0.053	0.244	0.102†	0.282†
	query perturb.	0.053	0.223	0.104†	0.251
	system perturb.	0.052	0.243	0.097†	0.266

combinations of the perturbation-based approaches and, finally, evaluate uncertainty and robustness of the proposed techniques.

Retrieval Performance. Tables 1 – 3 show the retrieval performance (bpref and p@10) for each of resource selection methods (when the number of selected sources is 1 and 5) for each testbed, respectively. In these tables, "†" denotes a statistically significant improvement over a baseline while "∘" shows a significant decrease in performance.

Tables 1 – 3 show that in most cases the ranking perturbation method significantly improves the performance of the baselines when 5 sources are selected: it achieves up to 18% improvement of bpref and 7-11% improvement of p@10 across all testbeds. The improvements for 1 selected source were lower for bpref (5-7%) but somewhat higher for p@10 (7-14%, statistically significant for the gov2.30 and the gov2.250 testbeds). In general, we observed that perturbation

Table 3. Perturbation-based resource selection, gov2.1000

Method	Modification	1 selected		5 selected	
		bpref	p@10	bpref	p@10
ReDDE	baseline	0.048	0.273	0.091	0.295
	ranking perturb.	0.048	0.259	0.099†	0.296
	query perturb.	0.047	0.219∘	0.093	0.254∘
	system perturb.	0.047	0.264	0.102†	0.295
CRCS(l)	baseline	0.040	0.188	0.103	0.210
	ranking perturb.	0.042	0.201	0.111†	0.233†
	query perturb.	0.039	0.159	0.102	0.179∘
	system perturb.	0.043	0.189	0.106	0.230†
CRCS(e)	baseline	0.045	0.243	0.090	0.292
	ranking perturb.	0.044	0.222	0.106†	0.292
	query perturb.	0.042	0.199∘	0.089	0.244∘
	system perturb.	0.042	0.231	0.098	0.288

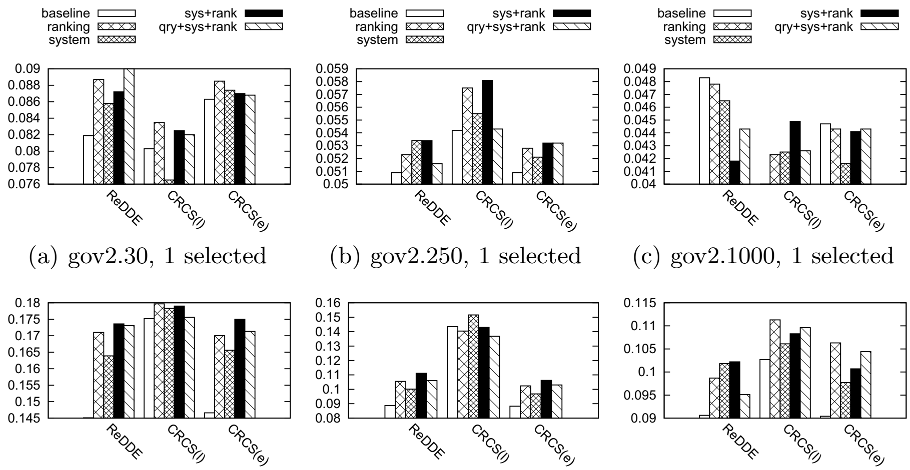
approaches affected the selection of the best source less frequently than the selection of 5 best sources.

An interesting observation is that the ranking perturbation method particularly helped poor selection baselines. See for example, ReDDE and CRCS(e) in Tables 1 and 2 when 5 sources are selected: bpref of ReDDE and CRCS(e) was increased for about 16% while bpref of better performing CRCS(l) stayed almost the same.

Query perturbation, however, provided mixed results: bpref was improved for 10-18% on the gov2.30 and the gov2.250 testbeds, while p@10 was decreased for 8-10%. Query perturbation also decreased p@10 on the gov2.1000 testbed (for up to 20%), where the drop in performance was statistically significant. In a per-query analysis, we observed that this perturbation approach tended to improve the performance of queries which had at least 3 terms, but would typically decrease the performance of queries with less than 3 terms. In the latter case, only 2 subqueries of one term in length are issued – so we only have a couple of perturbations for these queries, and since they are single term queries, they are much broader in scope and deviate from the meaning of the original query. In [9], Collins-Thompson and Callan also observed the same problem (which is thus a limitation of this approach).

Finally, the performance of the retrieval system perturbation method was similar to ranking perturbation and also significantly outperformed baselines on most occasions when 5 sources were selected (up to 12% improvement of bpref and 5% improvement of p@10 across all testbeds). In contrast to the ranking perturbation approach, this method only used $k = 10$ samples given the retrieval systems used, which may explain why the performances was not as good. Perhaps, if more (and possibly more diverse) retrieval systems were used the method could be improved. However, we leave this for future work.

In sum, the ranking and the retrieval system perturbation approaches deliver improvements over the baselines, while query perturbation shows varying performance.



(a) gov2.30, 1 selected (b) gov2.250, 1 selected (c) gov2.1000, 1 selected

(d) gov2.30, 5 selected (e) gov2.250, 5 selected (f) gov2.1000, 5 selected

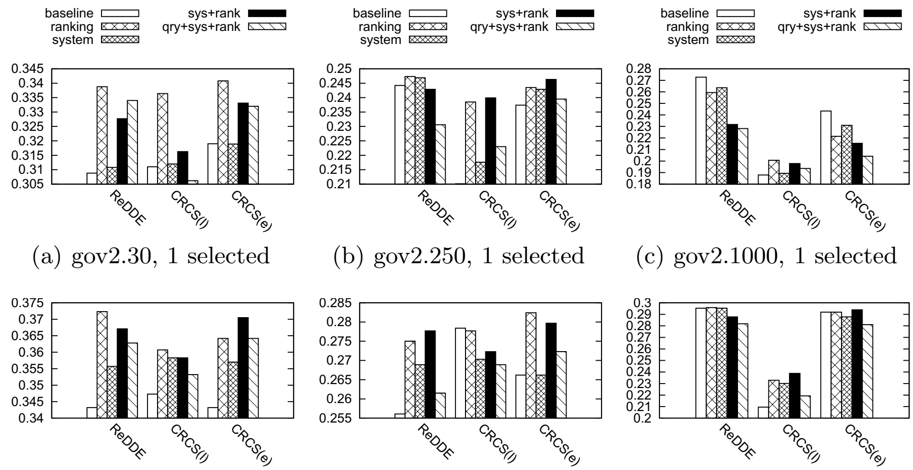
Fig. 1. Combinations of perturbation approaches, bpref

Combining Perturbation Approaches. Given the experimental results above, it is natural to wonder if applying all the perturbations together (in a cross-product manner) would lead to further improvements. For example, one can obtain a number of document rankings by perturbing a retrieval system and then perturb each ranking further by sampling documents from it. Figures 1 and 2 show the results of combining retrieval system and ranking perturbations (having $10 \cdot k$ perturbed rankings) and combing all three techniques together (having $10 \cdot k \cdot (l+1)$ rankings, where l is a query length). For clarity, we do not present results for query perturbation and its combinations, because they do not lead to significant improvements in performance.

The results show that the performance of different perturbation approaches usually did not sum up and sometimes even degraded compared to individual perturbation techniques. This can be explained by the fact that once uncertainty is captured and reduced by one method, it cannot be reduced any further by applying another one.

Uncertainty of Resource Selection In order to capture the uncertainty of resource selection we measured the variance of retrieval accuracy between different methods under the assumption that similar accuracy is achieved when most methods agree on selected sources and, thus, the uncertainty of resource selection is low. In particular, we measured the variance of bpref and p@10 across 3 selection techniques, i.e. ReDDE, CRCS(l) and CRCS(e).

In most cases the retrieval system and the ranking perturbation methods reduced the variance across selection techniques by an order of magnitude from $10^{-3} - 10^{-5}$ to $10^{-4} - 10^{-6}$, while improving performance. This means that the perturbation-based approaches agreed on selecting same high quality sources



(a) gov2.30, 1 selected (b) gov2.250, 1 selected (c) gov2.1000, 1 selected

(d) gov2.30, 5 selected (e) gov2.250, 5 selected (f) gov2.1000, 5 selected

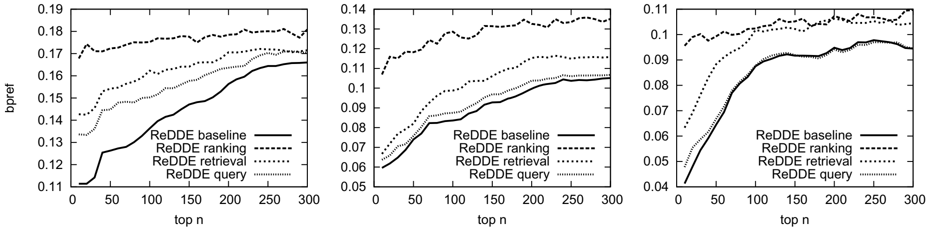
Fig. 2. Combinations of perturbation approaches, p@10

more often than the baseline methods, and thus were less uncertain, and due to that – more effective. Query perturbation again showed mixed results, but usually reduced the variance while keeping the same performance.

Overall, the results showed that perturbation-based resource selection methods increased the agreement between selection techniques compared to baseline methods, while being more effective than the latter.

Robustness of Resource Selection. To measure the robustness of resource selection techniques we studied how stable they were to changes in parameters. The parameter that all selection methods, like ReDDE and CRCS, share, is n – the number of documents considered from a centralized sample ranking. In Figure 3 we show how the baseline and the perturbation-based resource selection methods perform when n varies from 10 to 300. Since perturbation techniques affect the selection of 1 best source less often, we show the results when 5 sources are selected. Also for clarity reasons we present the plots only for ReDDE. However, the results for CRCS were similar.

Figure 3 shows that ReDDE performance increased with n. The query and the retrieval system perturbation methods exhibited similar trend but usually outperformed the baseline where system perturbation showed greater stability and better performance. Ranking perturbation, on the other hand, stayed almost flat being considerably higher than other techniques. This means that the ranking perturbation approach marginalized out the effect of resource selection parameters, thus eliminating one of the causes of uncertainty.



(a) gov2.30 (b) gov2.250 (c) gov2.1000

Fig. 3. ReDDE accuracy (bpref) for varying top n when 5 sources are selected

6 Conclusions and Future Work

In this work we attempted to address the uncertainty within resource selection, which is a critical step in the DIR process. To reduce the uncertainty, a number of rankings from the sampled documents were obtained using three different methods: query perturbation, retrieval system perturbation and ranking perturbation. By aggregating over the different rankings, we hoped to obtain a better estimates of source scores by reducing the uncertainty of these estimates. We found that the ranking perturbation approach showed the best performance and resulted in a significant improvement over baselines. Furthermore, it also reduced the variance of retrieval accuracy across different selection methods and improved the robustness of resource selection in terms of stability to changes in parameters.

With the retrieval system perturbation approach we witnessed a similar improvement in performance and reduction in uncertainty, but not to the same extent as ranking perturbation. We suspect that this is because we did not use enough retrieval systems, or retrieval systems with particularly different scoring functions, to produce better estimates. In future work, we plan to investigate using more systems to see whether further improvements can be achieved.

Finally, query perturbation approach was the least successful in reducing the uncertainty of resource selection and in increasing performance (often degrading performance). We suspect that this is because the set of possible subqueries introduced too much noise by creating topic drift, introducing different documents, and by only having a very limited number of re-rankings (proportional to the length of the query). Our findings suggest that this approach is unsuitable for reducing the uncertainty in resource selection.

In future work we will continue exploring how uncertainty in resource selection and DIR in general can be reduced. In particular, we will examine how much influence sampling, retrieval methods and selection parameters have on the uncertainty of resource selection and determine how much improvement can be made if this uncertainty is reduced. Furthermore, we will explore different ways to measure the uncertainty within the process. Lastly, we plan to address the uncertainty of results merging phase, that was shown to be the weakest DIR component [20].

Acknowledgments. This work was partially supported by the COST Action IC1002. The authors would like to thank Ian Ruthven, Djoerd Hiemstra and Mostafa Keikha for their comments on various parts of this work.

References

1. Arguello, J., Callan, J., Diaz, F.: Classification-based resource selection. In: Proceedings of the ACM CIKM, pp. 1277–1286 (2009)
2. Arguello, J., Diaz, F., Callan, J., Crespo, J.F.: Sources of evidence for vertical selection. In: Proceedings of the ACM SIGIR, pp. 315–322 (2009)
3. Azzopardi, L., Baillie, M., Crestani, F.: Adaptive query-based sampling for distributed ir. In: Proceedings of the ACM SIGIR, pp. 605–606 (2006)
4. Callan, J.P., Lu, Z., Croft, W.B.: Searching distributed collections with inference networks. In: Proceedings of the ACM SIGIR, pp. 21–28 (1995)
5. Callan, J.: Distributed Information Retrieval. In: Advances in Information Retrieval, ch. 5, pp. 127–150. Kluwer Academic Publishers (2000)
6. Callan, J., Connell, M.: Query-based sampling of text databases. ACM Transactions of Information Systems 19(2), 97–130 (2001)
7. Callan, J., Crestani, F., Nottelmann, H., Pala, P., Shou, X.M.: Resource selection and data fusion in multimedia distributed digital libraries. In: Proceedings of the ACM SIGIR, pp. 363–364 (2003)
8. Caverlee, J., Liu, L., Bae, J.: Distributed query sampling: a quality-conscious approach. In: Proceedings of the ACM SIGIR, pp. 340–347 (2006)
9. Collins-Thompson, K., Callan, J.: Estimation and use of uncertainty in pseudo-relevance feedback. In: Proceedings of the ACM SIGIR, pp. 303–310 (2007)
10. Crestani, F., Lalmas, M.: Logic and Uncertainty in Information Retrieval. In: Agosti, M., Crestani, F., Pasi, G. (eds.) ESSIR 2000. LNCS, vol. 1980, pp. 179–206. Springer, Heidelberg (2001)
11. Hauff, C.: Predicting the effectiveness of queries and retrieval systems. SIGIR Forum 44(1), 88–88 (2010)
12. Markov, I., Arampatzis, A., Crestani, F.: Improving Cori for Results Merging and Score Normalization. In: Serdyukov, P., Braslavski, P., Kuznetsov, S.O., Kamps, J., Rüger, S., Agichtein, E., Segalovich, I., Yilmaz, E. (eds.) ECIR 2013. LNCS, vol. 7814, Springer, Heidelberg (2013)
13. Shokouhi, M.: Central-Rank-Based Collection Selection in Uncooperative Distributed Information Retrieval. In: Amati, G., Carpineto, C., Romano, G. (eds.) ECIR 2007. LNCS, vol. 4425, pp. 160–172. Springer, Heidelberg (2007)
14. Shokouhi, M., Si, L.: Federated search. Foundations and Trends in Information Retrieval 5, 1–102 (2011)
15. Shokouhi, M., Zobel, J.: Robust result merging using sample-based score estimates. ACM Trans. Inf. Syst. 27(3), 1–29 (2009)
16. Shokouhi, M., Zobel, J., Tahaghoghi, S.M.M., Scholer, F.: Using query logs to establish vocabularies in distributed information retrieval. Information Processing & Management 43(1), 169–180 (2007)
17. Si, L., Callan, J.: Using sampled data and regression to merge search engine results. In: Proceedings of the ACM SIGIR, pp. 19–26 (2002)
18. Si, L., Callan, J.: Relevant document distribution estimation method for resource selection. In: Proceedings of the ACM SIGIR, pp. 298–305 (2003)

19. Thomas, P., Shokouhi, M.: Sushi: scoring scaled samples for server selection. In: Proceedings of the ACM SIGIR, pp. 419–426 (2009)
20. Thomas, P., Shokouhi, M.: Evaluating Server Selection for Federated Search. In: Gurrin, C., He, Y., Kazai, G., Kruschwitz, U., Little, S., Roelleke, T., Rüger, S., van Rijsbergen, K. (eds.) ECIR 2010. LNCS, vol. 5993, pp. 607–610. Springer, Heidelberg (2010)
21. Wang, J., Zhu, J.: Portfolio theory of information retrieval. In: Proceeding of the ACM SIGIR, pp. 115–122 (2009)
22. Xu, J., Croft, W.B.: Cluster-based language models for distributed retrieval. In: Proceedings of the ACM SIGIR, pp. 254–261. ACM (1999)
23. Zhai, C., Lafferty, J.D.: A risk minimization framework for information retrieval. Information Processing & Management 42(1), 31–55 (2006)
24. Zhou, Y., Croft, W.B.: Query performance prediction in web search environments. In: Proceedings of the ACM SIGIR, pp. 543–550 (2007)
25. Zhu, J., Wang, J., Taylor, M., Cox, I.J.: Risk-Aware Information Retrieval. In: Boughanem, M., Berrut, C., Mothe, J., Soule-Dupuy, C. (eds.) ECIR 2009. LNCS, vol. 5478, pp. 17–28. Springer, Heidelberg (2009)