

Towards Risk-Aware Resource Selection

Ilya Markov¹, Mark Carman², and Fabio Crestani¹

¹ University of Lugano (USI),
Via G. Buffi 13, Lugano 6900, Switzerland
{ilya.markov, fabio.crestani}@usi.ch

² Monash University,
Victoria 3800, Australia
mark.carman@monash.edu

Abstract. When searching multiple sources of information it is crucial to select only relevant sources for a given query, thus filtering out non-relevant content. This task is known as resource selection and is used in many areas of information retrieval such as federated and aggregated search, blog distillation, etc. Resource selection often operates with limited and incomplete data and, therefore, is associated with a certain risk of selecting non-relevant sources due to the uncertainty in the produced source ranking. Despite the large volume of research on resource selection, the problem of risk within resource selection has been rarely addressed. In this work we propose a resource selection method based on document score distribution models that supports estimation of uncertainty of produced source scores and results in a novel risk-aware resource selection technique. We analyze two distributed retrieval scenarios and show that many queries are risk-sensitive and, because of that, the proposed risk-aware approach provides a basis for significant improvements in resource selection performance.

1 Introduction

In some cases, such as federated retrieval, aggregated search or blog search, it is necessary to perform retrieval over multiple sources of information, such as federated collections, verticals or blogs. For a given user's query often only few sources contain relevant information. Therefore, it is useful to filter out all non-relevant sources and search only those that are likely to contain relevant documents. The task of selecting such sources is referred to as *resource selection* and is usually performed by a centralized broker. Resource selection was introduced and extensively studied in the area of distributed information retrieval [1,2,3,4,5].

Due to operational constraints (e.g., sources do not provide direct access to their content) and for efficiency reasons resource selection is often performed based on incomplete samples of documents obtained from each federated source [6,7,8]. This introduces uncertainty into the selection process making it risky [9]. In this paper we approach the problem of risk-aware resource selection in the following way. First, we propose to perform resource selection based on the distribution of document scores in each federated source. Second, we derive a general formula of the variance of source scores, thus capturing their uncertainty.

Third, following the risk-aware approach to IR [10], we incorporate the variance of source scores into the resource selection process and develop a risk-aware resource selection technique, where the amount of allowed risk is controlled by a risk-adjusting parameter.

We analyze two recently proposed distributed retrieval scenarios, namely federated Web search [11] and shard ranking [12], and show that many queries are risk-sensitive, i.e. require a risk-aware component when performing resource selection. We also show that the proposed risk-aware resource selection technique with appropriate values of a risk-adjusting parameter gives significant improvements of resource selection performance.

2 Related Work

Many resource selection techniques have been proposed in the literature. *Large-document* approaches (LD) represent each federated source as a concatenation of its documents and, given a user’s query, rank these large documents using standard IR ranking functions, such as INQUERY in CORI [1] and language modelling in [13]. *Small-document* techniques (SD), instead, make use of a centralized index of documents sampled from federated sources (denoted as CSI for *centralized sample index*). These documents are ranked with respect to a user’s query and sources are selected based on the number of documents they contribute to the top of such ranking [2,3,4,5,12,14]. SD techniques were shown to outperform LD methods [2,3,5] and, therefore, we will follow the small-document idea in this paper. There is also a large volume of literature on efficient, supervised and theoretical methods for resource selection [8], but we do not describe them here as they are not directly related to the current work.

Despite many studies performed on resource selection, still little is understood about its uncertainty and the risk, associated with the selection process. Markov et al. [9] addressed the uncertainty in resource selection by perturbing various components of SD techniques, such as the document retrieval function, query and ranking of sampled documents. This approach managed to improve the performance of SD methods in some cases. However, this was done at the cost of performing resource selection multiple times, thus, hurting its efficiency. In this paper we propose a resource selection technique, which allows estimation of uncertainty, associated with source scores, using a single resource selection run. It then incorporates the estimated risk directly into the selection process, resulting in risk-aware resource selection.

Our method is based on the distribution of document scores, similarly to [15,16,17]. Baumgarten [16] models document scores using the Gamma distribution and estimates the parameters of the distribution assuming a particular retrieval model, namely probabilistic retrieval. The recently proposed Taily resource selection technique [15] infers score distributions from term statistics, cooperatively provided by federated sources, and also assumes a specific document retrieval function. Then both studies calculate source scores as the area under the right tail of the estimated score distribution. However, the approach

in [16] is tied to a particular retrieval model and particular score distribution, while Tally requires certain cooperation from federated sources, which is rarely available in practice. Our approach, instead, is completely uncooperative and does not depend on a specific retrieval method and/or score distribution model. But the main novelty of our technique is that it considers the risk associated with source scores and incorporates it into the resource selection process.

3 Risk-Aware Resource Selection

In this section we first describe the idea of using score distributions for resource selection (called DSDE in this paper, which stands for Document Score Distribution Estimation). Then we show how the variance of source scores can be computed and incorporated into the selection process, resulting in a risk-aware resource selection (RA-DSDE). Finally, we instantiate our general approach using a normal distribution.

3.1 Score Distributions for Resource Selection

Small-document resource selection techniques aim at selecting those sources that contain the largest number of relevant documents for a given user’s query. Since the actual relevant documents are not known, the top-ranked documents are used instead [4]. Thus, SD resource selection methods aim at selecting the sources that have the largest number of top-ranked documents for the query.

Following the SD approach, we first define a threshold τ and calculate the number of documents in a source R that have scores higher than τ (we consider these documents as high scoring):

$$\text{count}(R|q, \tau) = |d \in R : \text{score}(d|q) > \tau| \tag{1}$$

where d denotes a document and q denotes a query.

We assume that the distribution of document relevance scores in each source can be approximated by a probability distribution with a distribution function F and a density function f . The area under the density function to the right of the threshold τ is the approximate fraction of high scoring documents in R :

$$\frac{|d \in R : \text{score}(d|q) > \tau|}{|d \in R|} \approx P[\text{score}(d|q) > \tau] = \int_{\tau}^{\infty} f(t)dt = 1 - F(\tau) \tag{2}$$

By combining Equations 1 and 2 we can calculate the expected number of high scoring documents per source with a closed-form expression and use it as a score for ranking sources:

$$\text{score}(R|q, \tau) = \mathbb{E}[\text{count}(R|q, \tau)] = |R| \cdot (1 - F(\tau)) \tag{3}$$

where $|R| = |d \in R|$ is the number of documents in R and τ is the parameter of the DSDE approach, which can be used for tuning the algorithm according to user needs. Since relevance scores are comparable across different sources, the fitted models are also comparable and, therefore, we can use the same threshold τ for all sources. The value of τ used in our experiments is discussed in Section 4.

3.2 Expectation and Variance of Source Scores

Equation 3 gives an estimation of a score of a source R based on its distribution function $F(x)$. Since the exact function $F(x)$ is not known and has to be approximated for a given query, it becomes also a function of a set of parameters θ : $F = F(x; \theta)$.

The parameters θ are approximated based on the observed scores of documents sampled from R with the maximum likelihood estimation (MLE). Using the delta method, it can be shown that in this case the expectation and variance of the distribution function $F(\tau; \hat{\theta})$ are calculated as follows:

$$\mathbb{E}[F(\tau; \hat{\theta})] = F(\tau; \hat{\theta}) \quad (4)$$

$$Var[F(\tau; \hat{\theta})] = \frac{1}{n} \frac{\partial F}{\partial \theta^T}(\tau; \hat{\theta}) I(\hat{\theta})^{-1} \frac{\partial F}{\partial \theta}(\tau; \hat{\theta}) \quad (5)$$

where $I(\hat{\theta})$ is the Fisher information matrix and n is the number of documents sampled from R .

Based on the above discussion, Equation 3 can be rewritten as follows:

$$\widehat{score}(R|q) = |R| \cdot (1 - F(\tau; \hat{\theta})) \quad (6)$$

This means that $\widehat{score}(R|q)$ is just an approximation of the true $score(R|q)$. Combining this equation with Equations 4 and 5 we can calculate the expectation and variance of source score estimates:

$$\begin{aligned} \mathbb{E}[\widehat{score}(R|q)] &= \mathbb{E}[|R| \cdot (1 - F(\tau; \hat{\theta}))] = |R| \cdot (1 - \mathbb{E}[F(\tau; \hat{\theta})]) \\ &= |R| \cdot (1 - F(\tau; \theta)) = score(R|q) \end{aligned} \quad (7)$$

$$\begin{aligned} Var[\widehat{score}(R|q)] &= Var[|R| \cdot (1 - F(\tau; \hat{\theta}))] = |R|^2 \cdot Var[F(\tau; \hat{\theta})] \\ &= \frac{|R|^2}{n} \frac{\partial F}{\partial \theta^T}(\tau; \hat{\theta}) I(\hat{\theta})^{-1} \frac{\partial F}{\partial \theta}(\tau; \hat{\theta}) \end{aligned} \quad (8)$$

In particular, Equation 7 means that the proposed source score estimate is unbiased and asymptotically converges to the true score.

3.3 Risk-Aware Resource Selection

Following the risk-aware approach to IR [10], we combine expectation and variance of source scores in the following way:

$$\widetilde{score}(R|q) = \mathbb{E}[\widehat{score}(R|q)] - b \cdot Var[\widehat{score}(R|q)] \quad (9)$$

Here b is a risk-adjusting parameter, where $b > 0$ produces the risk-averse ranking and $b < 0$ gives the risk-inclined ranking of sources.

The final formula for calculating risk-aware source scores based on score distribution models is the following:

$$\widetilde{score}(R|q) = |R| \cdot \left(1 - F - b \cdot \frac{|R|}{n} \cdot \frac{\partial F}{\partial \theta^T} I^{-1} \frac{\partial F}{\partial \theta} \right) \quad (10)$$

where the values of F , $\frac{\partial F}{\partial \theta}$ and I^{-1} are calculated at the point $(\tau; \hat{\theta})$. Note that this formula is generic with regards to a score distribution model F and any appropriate distribution can be used in practice.

3.4 Choosing a Score Distribution Model

Various combinations of probability distributions were proposed to model scores of relevant and non-relevant documents: two normal, two exponential, two Poisson, two Gamma and two log-normal distributions [18]. Also a number of mixtures were proposed, where the scores of non-relevant documents were modeled by an exponential or a Gamma distribution, while the scores of relevant documents were modeled with a normal, a Gamma or a mixture of normals [18]. However, the most widely used score distribution model is a mixture of an exponential and normal distributions [18,19]. Therefore, we use this model in our work.

Since our resource selection method is model-agnostic and does not necessarily require a mixture of distributions, we also use a number of single distributions to model document scores, because they have less parameters, are easier to estimate and simpler to work with. In particular, we use a Gamma, a normal and an exponential distributions. Preliminary experiments showed that the mixture, Gamma and normal models performed similarly to each other, while the exponential distribution performed significantly worse than others. This is because we model the scores of top-ranked documents and, therefore, the Gamma and normal provided better fit than the exponential model, while the mixture reduced to a single normal in most cases. Since most of the chosen models provide similar performance, we will use a normal distribution in this work, as it is the simplest to operate with. Note, however, that the following discussion is fully applicable to other models that provide a good fit of document scores.

3.5 Normal-Based Instantiation

The normal distribution is defined with a probability density function f and a distribution function F as follows:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad F(x; \mu, \sigma^2) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x-\mu}{\sqrt{2\sigma^2}} \right) \right] \quad (11)$$

where $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is known as the *error function*. The MLE estimators of the mean μ and variance σ^2 for each source are the following:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \operatorname{score}(d_i|q), \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\operatorname{score}(d_i|q) - \hat{\mu})^2 \quad (12)$$

The Fisher information of a normal distribution and its inverse are given by the following matrices:

$$I = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}, \quad I^{-1} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \quad (13)$$

Finally, the partial derivatives of the distribution function F with respect to the parameters μ and σ are the following:

$$\frac{\partial F}{\partial \mu}(x) = -f(x), \quad \frac{\partial F}{\partial \sigma}(x) = -\frac{x - \mu}{\sigma} f(x) \quad (14)$$

Substituting the generic expressions in Equation 10 with the above formulas for the normal distribution, we obtain the following equation for source scores:

$$\widetilde{score}(R|q) = |R| \left(1 - F(\tau) - b \frac{|R|}{n} f^2(\tau) \hat{\sigma}^2 (1 + 2(\tau - \hat{\mu})^2) \right) \quad (15)$$

Below we evaluate the proposed risk-aware formula against its basic version ($b = 0$) and other state-of-the-art resource selection techniques.

4 Experimental Methodology

In this section we describe evaluation scenarios, testbeds, parameter values and the overall DIR setup used in our experiments.

Scenarios and Testbeds. We consider the following state-of-the-art cases, where resource selection has a direct application: federated Web search [11] and shard ranking [12].

In order to simulate these scenarios we adopt three testbeds, as in [20], which are the different splits of the TREC GOV2 dataset and contain about 22 million documents. The largest 1000 domains of GOV2 are used as 1000 Web search engines, constituting the gov2.1000 testbed. Since the GOV2 dataset is the crawl of the real Web, gov2.1000 can be used to simulate the federated Web search scenario. The domains of gov2.1000 are clustered into 250 partitions using the average-link agglomerative clustering (the gov2.250 testbed), making them more topically homogeneous. Therefore, gov2.250 can be seen as an intermediate scenario between federated Web search and shard ranking. Then gov2.1000 is also clustered into 30 partitions, creating the gov2.30 testbed. This testbed has topically homogeneous sources and, therefore, represents the shard ranking scenario. We use the titles of the TREC topics 701-850 from the Terabyte Track 2004-2006 as queries.

Parameter Settings. We compare the DSDE and RA-DSDE techniques to ReDDE [4] and CRCS [3], which were shown to be effective unsupervised SD resource selection methods [3,5]. We consider the exponential and linear versions of CRCS denoted as CRCS(e) and CRCS(l) respectively. Based on our preliminary experiments, the original papers and recent resource selection studies, the parameters of the methods are set as follows: ReDDE considers the top 20 documents to be relevant, CRCS(e) uses $\beta = 0.28$, CRCS(l) considers $\gamma = 20$ documents to be relevant.

DSDE needs to estimate the distribution of document scores for each federated source. For this purpose we use the top 100 documents retrieved from a

centralized sample index (CSI) for a given query for each source. The threshold τ is set to the score of the 10th document in a centralized ranking produced by CSI. The risk-adjusting parameter b of the RA-DSDE approach will be discussed in detail in Section 5.

DIR Setup. Both the federated Web search and shard ranking scenarios assume a low-cost (and sometimes uncooperative) environment, where the full centralized index of all documents is not available. In this case a number of documents are sampled from each federated source and stored in CSI. In cooperative scenarios, such as shard ranking for topically partitioned collections, a uniform sample of documents can be obtained [12]. In uncooperative environments the query-based sampling technique [21] can be used instead and we follow this approach here. Since the size of the three considered testbeds is the same, i.e. 22 million documents, we make their CSIs to be also of the same size, i.e. 300K documents for each testbed. This results in 300 sampled documents from each source of the gov2.1000 testbed, 1200 documents from each source of gov2.250 and 10000 documents from each source of gov2.30.

Since the actual size of sources may be unknown in uncooperative environments, we estimate it using the multiple capture-recapture algorithm with 80 queries [22]. The Terrier³ implementation of BM25 is used as a document scoring function for CSI. After resource selection is performed, the query is sent to the selected sources and 100 documents are retrieved from each of them. The obtained results are merged using CORI [1] as presented in [23], because it showed the best performance in preliminary experiments. P@10 and MAP are used to evaluate the quality of merged results.

5 Results and Discussion

In this section we perform an empirical evaluation of the proposed risk-aware resource selection technique. We start by analyzing the optimal values of the risk-adjusting parameter b and studying the sensitivity of queries to risk. We then compare the performance of the RA-DSDE approach to basic DSDE and other state-of-the-art resource selection methods, namely ReDDE and CRCS.

5.1 Risk-Adjusting Parameter

Intuitively, the sensitivity to risk varies across queries: some queries tend to be risk-averse, i.e. require reducing the risk ($b > 0$), while others are risk-inclined, i.e. take risks and gain performance improvement ($b < 0$) [24]. Examples of different behaviors are given in Figure 1. Here the improvement of p@10 is plotted against the values of b on a log-scale for the shard ranking scenario (the gov2.30 testbed).

Figure 1 shows that the topic 819, “1890 Census”, exhibits a risk-averse behavior and has higher p@10 when the uncertainty is reduced. On the other

³ <http://terrier.org>

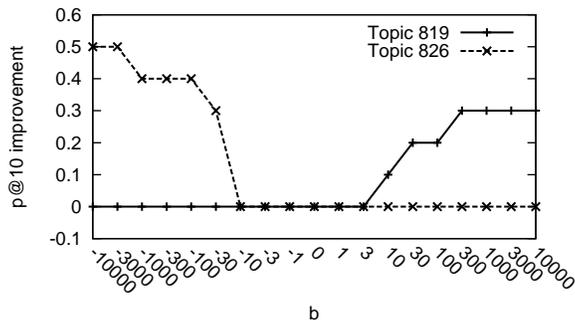


Fig. 1. Examples of risk-averse and risk-inclined topics for the shard ranking scenario (gov2.30).

Table 1. Query statistics for optimal values of b .

Measure	Testbed	Number of queries		
		$b \neq 0$	$b > 0$ (%)	$b < 0$ (%)
p@10	gov2.1000	63	41 (65%)	22 (35%)
	gov2.250	75	44 (59%)	31 (41%)
	gov2.30	87	54 (62%)	33 (38%)
MAP	gov2.1000	97	52 (54%)	45 (46%)
	gov2.250	125	65 (52%)	60 (48%)
	gov2.30	133	71 (53%)	62 (47%)

hand, the topic 826, “Florida Seminole Indians”, shows better performance for negative values of b . The differences in risk preferences can be attributed to the specificity of query terms and the overall difficulty of a query.

Table 1 presents the number of risk-sensitive queries, i.e. queries that benefit from a risk-aware component in terms of p@10 and MAP⁴. When considering p@10 as a performance measure, the federated Web search scenario (the gov2.1000 testbed) benefits from accounting for the risk in 63 cases out of 150 (42%). The intermediate scenario (the gov2.250 testbed) has half of queries being risk-sensitive, while in the shard ranking scenario (the gov2.30 testbed) 87 out of 150 queries (58%) benefit from a risk-aware component. Thus, when optimizing for p@10, on average half of the queries are risk-sensitive across various scenarios. The shard ranking scenario, having larger sources, tend to have more risk-sensitive queries than the federated Web search scenario. This is because larger sources contain more relevant documents and, therefore, are more sensitive to reranking.

Table 1 shows that more queries are risk-sensitive with respect to the MAP performance measure compared to p@10. 97 queries out of 150 (65%) benefit from a risk-aware component in terms of MAP for the federated Web search

⁴ The total number of analyzed queries is 150.

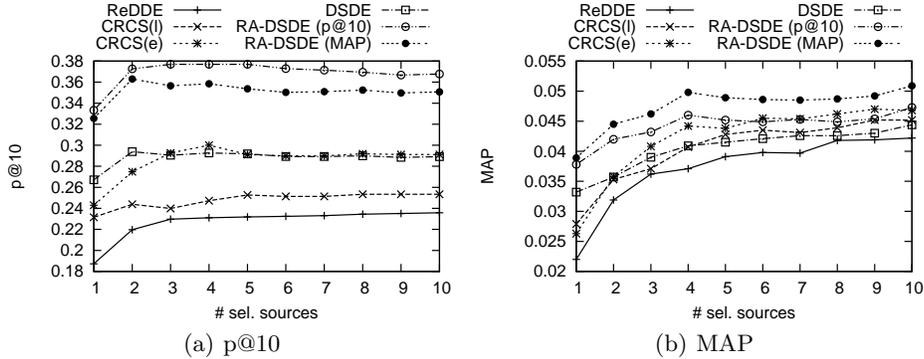


Fig. 2. P@10 and MAP for the federated Web search scenario (the gov2.1000 testbed).

scenario, 125 queries (83%) – for the intermediate case and 133 queries (89%) – for the shard ranking scenario. This behavior is intuitive, because even if accounting for risk does not affect precision at rank 10, it often improves precision at other ranks, thus improving MAP. This suggests that aggregate performance measures are more sensitive to risk and may benefit a lot from risk-aware resource selection. Also note that here, similar to p@10, the number of risk-sensitive queries increases for testbeds with larger sources.

The number of risk-averse ($b > 0$) and risk-inclined ($b < 0$) queries are also shown in Table 1. For p@10, on average, around 60% of risk-sensitive queries are risk-averse and 40% are risk-inclined. This means that more queries benefit from reducing the uncertainty in source ranking. Still there is a large portion of queries that improve performance by taking risks. For MAP these numbers are similar having slightly more risk-inclined queries. This can be explained by the aggregate nature of MAP: the risk-inclined behavior, even if hurting p@10, may improve precision at other ranks, thus improving MAP.

5.2 Retrieval Results

Resource selection performance is shown in Figures 2 – 4. Here we plot p@10 and MAP against the number of selected sources. The performance of the RA-DSDE approach is given when optimized for p@10 and for MAP.

First, note that the basic DSDE approach ($b = 0$) outperforms the baseline methods for the shard ranking scenario in terms of p@10 in almost all cases (see Figure 4). It also shows better p@10 when 1 and 2 sources are selected for the intermediate scenario (Figure 3) and the federated Web search scenario (Figure 2). However, the improvements are minor and not statistically significant.

On the other hand, as expected the optimized RA-DSDE technique significantly improves p@10 and MAP in all cases. This shows the potential of the risk-aware approach and the amount of improvement that can be achieved by

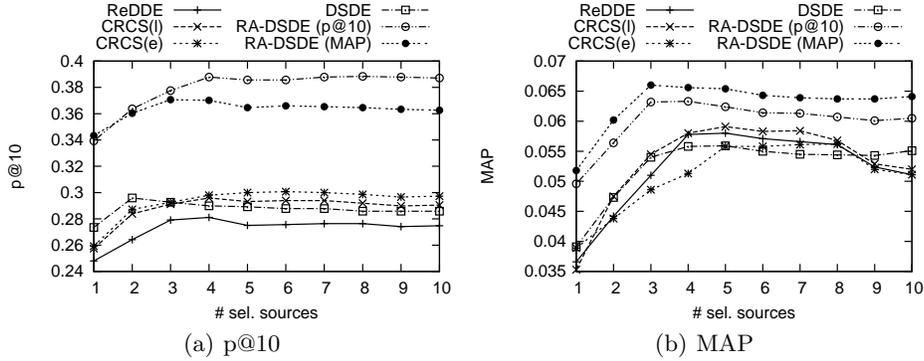


Fig. 3. P@10 and MAP for the intermediate scenario (the gov2.250 testbed).

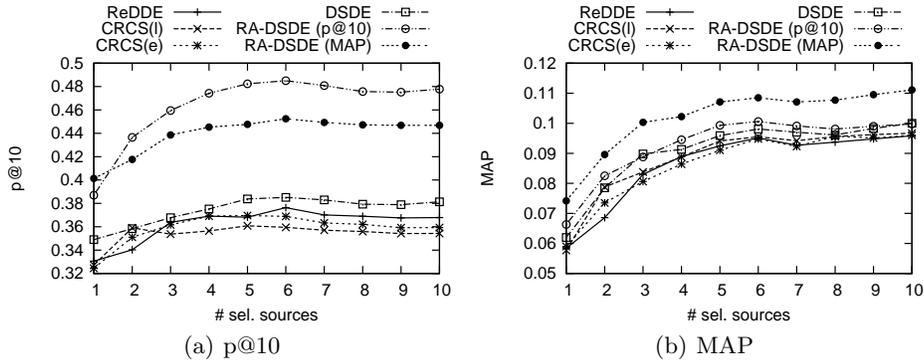


Fig. 4. P@10 and MAP for the shard ranking scenario (the gov2.30 testbed).

considering the risk in resource selection. In particular, the average improvement of RA-DSDE(p@10) over the basic DSDE method in terms of p@10 is 28% for the gov2.1000 testbed, 32% for gov2.250 and 24% for gov2.30. In terms of MAP, the RA-DSDE(MAP) method gives 17% improvement for the gov2.1000 testbed, 20% improvement for gov2.250 and 13% for gov2.30 over the DSDE baseline. These results show that both in the federated Web search and shard ranking scenarios significant improvements are possible by considering the risk-aware component during resource selection and optimizing the risk-adjusting parameter b .

Finally, note that RA-DSDE optimized for MAP also performs well in terms of p@10, although not as good as RA-DSDE(p@10). However, the opposite is not true, i.e. RA-DSDE(p@10) does not improve the performance of DSDE in terms of MAP in about half of the cases. These results are intuitive, as MAP

optimizes the average performance of resource selection, including precision at rank 10, while $p@10$ optimizes only early precision. This suggests that aggregate performance measures, such as MAP, must be used for tuning the risk-adjusting parameter and optimizing the RA-DSDE approach.

6 Conclusions and Future Work

In this paper we approached the problem of risk in resource selection. To this end we performed resource selection based on the distribution of document scores in each federated source. This method, being as effective as state-of-the-art resource selection techniques, provides a closed-form solution for the variance of source scores. Following the risk-aware approach to IR, we incorporated the variance into the source score estimates, thus, developing a risk-aware resource selection method. The amount of risk in this method is controlled by a risk-adjusting parameter that has to be set on a per query basis. Finally, we showed how to instantiate the proposed approach based on a normal distribution. Note, however, that since our method is model-agnostic, any other suitable distribution model can be used in a similar way.

The empirical study of the federated Web search and shard ranking scenarios revealed that many queries are, in fact, risk-sensitive. The experimental results suggest that aggregate measures, such as MAP, are more sensitive to risk, than measures like precision at a certain rank. Also the number of risk-sensitive queries varies slightly across testbeds and is higher for those with larger sources. The experimental evaluation also showed that by considering risk in resource selection its performance could be improved up to 32% in terms of $p@10$ and up to 20% in terms of MAP.

As for future work, we plan to study what features affect the risk preference of a particular query. Then those features can be used to predict an optimal value of the risk-adjusting parameter for a given user's query. The second direction is to apply portfolio theory to resource selection [25]. To this end, we plan to estimate the correlation between different sources and incorporate this correlation into the resource selection process.

References

1. Callan, J.P., Lu, Z., Croft, W.B.: Searching distributed collections with inference networks. In: Proceedings of SIGIR. pp. 21–28 (1995)
2. Paltoglou, G., Salampasis, M., Satratzemi, M.: Integral based source selection for uncooperative distributed information retrieval environments. In: Proceeding of workshop on LSDS for IR. pp. 67–74 (2008)
3. Shokouhi, M.: Central-rank-based collection selection in uncooperative distributed information retrieval. In: Proceedings of ECIR. pp. 160–172 (2007)
4. Si, L., Callan, J.: Relevant document distribution estimation method for resource selection. In: Proceedings of SIGIR. pp. 298–305 (2003)
5. Thomas, P., Shokouhi, M.: Sushi: scoring scaled samples for server selection. In: Proceedings of SIGIR. pp. 419–426 (2009)

6. Callan, J.: *Advances in Information Retrieval*, chap. Distributed Information Retrieval, pp. 127–150. Kluwer Academic Publishers (2000)
7. Crestani, F., Markov, I.: Distributed information retrieval and applications. In: *Proceedings of ECIR*. pp. 865–868 (2013)
8. Shokouhi, M., Si, L.: Federated search. *Foundations and Trends in Information Retrieval* 5, 1–102 (2011)
9. Markov, I., Azzopardi, L., Crestani, F.: Reducing the uncertainty in resource selection. In: *Proceedings of ECIR*. pp. 507–519 (2013)
10. Zhu, J., Wang, J., Cox, I.J., Taylor, M.J.: Risky business: modeling and exploiting uncertainty in information retrieval. In: *Proceedings of SIGIR*. pp. 99–106 (2009)
11. Nguyen, D., Demeester, T., Trieschnigg, D., Hiemstra, D.: Federated search in the wild: the combined power of over a hundred search engines. In: *Proceedings of CIKM*. pp. 1874–1878 (2012)
12. Kulkarni, A., Tigelaar, A.S., Hiemstra, D., Callan, J.: Shard ranking and cutoff estimation for topically partitioned collections. In: *Proceedings of CIKM*. pp. 555–564 (2012)
13. Xu, J., Croft, W.B.: Cluster-based language models for distributed retrieval. In: *Proceedings of SIGIR*. pp. 254–261 (1999)
14. Markov, I., Crestani, F.: Theoretical, qualitative, and quantitative analyses of small-document approaches to resource selection. *ACM Transactions on Information Systems* 32(2), 9:1–9:37 (2014)
15. Aly, R., Hiemstra, D., Demeester, T.: Taily: shard selection using the tail of score distributions. In: *Proceedings of SIGIR*. pp. 673–682 (2013)
16. Baumgarten, C.: A probabilistic solution to the selection and fusion problem in distributed information retrieval. In: *Proceedings of SIGIR*. pp. 246–253 (1999)
17. Markov, I.: Modeling document scores for distributed information retrieval. In: *Proceedings of SIGIR*. pp. 1321–1322 (2011)
18. Arampatzis, A., Robertson, S.: Modeling score distributions in information retrieval. *Information Retrieval* 14(1), 26–46 (2011)
19. Manmatha, R., Rath, T., Feng, F.: Modeling score distributions for combining the outputs of search engines. In: *Proceedings of SIGIR*. pp. 267–275 (2001)
20. Arguello, J., Callan, J., Diaz, F.: Classification-based resource selection. In: *Proceedings of CIKM*. pp. 1277–1286 (2009)
21. Callan, J., Connell, M.: Query-based sampling of text databases. *ACM Transactions on Information Systems* 19(2), 97–130 (2001)
22. Shokouhi, M., Zobel, J., Scholer, F., Tahaghoghi, S.M.M.: Capturing collection size for distributed non-cooperative retrieval. In: *Proceedings of SIGIR*. pp. 316–323 (2006)
23. Markov, I., Arampatzis, A., Crestani, F.: On cori results merging. In: *Proceedings of ECIR*. pp. 752–755 (2013)
24. Zuccon, G., Azzopardi, L., van Rijsbergen, K.: Back to the roots: Mean-variance analysis of relevance estimations. In: *Proceedings of ECIR*. pp. 716–720 (2011)
25. Wang, J., Zhu, J.: Portfolio theory of information retrieval. In: *Proceeding of SIGIR*. pp. 115–122 (2009)