The Interactive PRP for Diversifying Document Rankings

Guido Zuccon[*], Leif Azzopardi, and C. J. "Keith" van Rijsbergen
{guido, leif, keith}@dcs.gla.ac.uk
School of Computing Science, University of Glasgow, Glasgow, UK

ABSTRACT

The assumptions underlying the Probability Ranking Principle (PRP) have led to a number of alternative approaches that cater or compensate for the PRP's limitations. In this poster we focus on the Interactive PRP (iPRP), which rejects the assumption of independence between documents made by the PRP. Although the theoretical framework of the iPRP is appealing, no instantiation has been proposed and investigated. In this poster, we propose a possible instantiation of the principle, performing the first empirical comparison of the iPRP against the PRP. For document diversification, our results show that the iPRP is significantly better than the PRP, and comparable to or better than other methods such as Modern Portfolio Theory.

Categories and Subject Descriptors: H.3.3 [Information Storage and Retrieval] - *Information Search and Retrieval - Retrieval Models*

General Terms: Algorithms, Theory.

1. INTRODUCTION

The PRP has played a central role in the development of Information Retrieval (IR): in the context of ad-hoc retrieval the PRP has underpinned the development of most formal models. This is because, if a system upholds the PRP then its response is guaranteed theoretically to be optimal given the query. In practice, this principle has largely stood the test of time, but it relies on a number of key assumptions which have been called into question [3, 5, 6]. In particular, the PRP has been criticised because of the *independence assumption*, which assumes only document and query are sufficient to determine whether a document is relevant, ignoring the influence of other documents in the ranking [3].

In non-traditional evaluation contexts such as subtopic retrieval, where dependencies between documents are considered, a number of ranking principles and strategies alternative to the PRP have been proposed. These include Maximal Marginal Relevance (MMR) [1] and Modern Portfolio Theory (MPT) [5], along with the recently proposed Quantum PRP (qPRP) [6], as well as the untested iPRP [2]. The latter provides a theoretical framework for extending the PRP to the context of interactive IR. However, the iPRP has not been empirically tested or validated; so it is un-

clear whether it performs better than other approaches, or not. In this poster, we compare the rankings that the iPRP would initially deliver to the user in the *first pass of retrieval*, against the rankings based on other approaches (and leave the interaction to future work).

2. THE INTERACTIVE PRP

In [2], Fuhr proposes a theoretical framework for extending the PRP to the context of interactive IR. In this framework, the independence assumption is rejected since relevance is assumed to depend upon the documents the user has previously examined. Search is then modelled as situation, called a list of choices, that are presented to the user. The user moves between situations by accepting one of the choices. Once a choice has been accepted, the retrieval system is required to produce a new list of choices depending upon the previous choice. The ranking principle strives to provide the optimum ordering of the choices presented in each situation, such that for each rank position i, documents under the iPRP are ranked as follows:

$$d_i = \operatorname*{arg\,max}_{d \in \mathcal{RE} \setminus RA} \left[e + P(d) \left(b_{d,i} Q(d) + g \left(1 - Q(d)\right) \right) \right]$$

where $Q(.)$ is the probability[1] that the user does not revise their choice of selecting document d; e is the effort of examining document d; g is the additional effort required for correction if the user judges a viewed document as irrelevant; and $b_{d,i}$ is the benefit of ranking document d if relevant.

The iPRP assumes that the user would examine each document in turn. Thus at each rank, the previous documents would influence the relevance of the subsequent documents. In order to obtain the ranking for the first pass of retrieval (i.e. before any actual user interaction has transpired), we can ignore the costs associated with the user or assume them constant. So both e and g are set to zero in the first pass. The probability $Q(.)$ of a user not revising their choice can be treated as constant for all the documents, and thus it is dropped for rank equivalence reasons.

This leaves one final choice to be made about the benefit of ranking a document d at rank i. Since this is dependent upon the documents that have been previously ranked, then a reasonable approximation of the benefit would be to determine how dissimilar the considered document is to all previous documents. This can be achieved through a summation over all previously ranked documents of a measure of dissimilarity or anti-correlation, i.e. we assume the possible benefit comes from novel information. If document d is

---

[1]i.e. the probability that the user does not change their mind about the relevance of the document d after examining it.

Table 1: Performances of PRP, MMR, MPT, iPRP and qPRP for subtopic retrieval. Significant better performances (measured by t-test) over the PRP are indicated with *. No statistically significant differences are calculated between runs in the TREC 6-8 dataset due to the limited amount of topics [4].

Meas.	TREC 6-8					TREC 18 (Clueweb09)				
	PRP	MMR	MPT	qPRP	iPRP	PRP	MMR	MPT	qPRP	iPRP
NRBP	0.127	**0.132** (+3.94%)	0.127 (0.00%)	0.134 (+5.51%)	**0.136** (+7.09%)	0.044	**0.074*** (+68.18%)	**0.074*** (+68.18%)	0.060* (+36.36%)	0.048 (+9.09%)
α-NDCG@10	0.426	0.457 (+7.28%)	0.426 (0.00%)	0.433 (+1.64%)	**0.461** (+8.22%)	0.097	0.137* (+41.24%)	**0.151*** (+55.67%)	0.144* (+48.45%)	0.132* (+36.08%)
IA-P@10	0.063	0.065 (+3.17%)	0.063 (0.00%)	0.065 (+3.17%)	**0.068** (+7.94%)	0.064	0.064 (0.00%)	**0.076*** (+18.75%)	0.074* (+15.63%)	0.070 (+9.38%)

similar to the previous documents, then the correlation will be low, or negative, leading to a low total benefit. Similar documents are then demoted in the ranking, while documents that are more diverse are promoted, giving rise to the following objective function:

$$d_i = \arg\max_{d \in \mathcal{RE} \setminus RA} (P(d)b_{d,i}) =$$
$$\arg\max_{d \in \mathcal{RE} \setminus RA} \left(P(d) \frac{\sum_{d' \in RA} \rho_{d,d'}}{|RA|} \right)$$

Under the iPRP dependencies between documents are incorporated through multiplication, providing a completely different approach to alternative strategies. In contrast, MPT, MMR and qPRP combine relevance with diversity in an additive fashion.

3. EXPERIMENTS

In this paper we conducted an empirical study in the context of subtopic retrieval, with the aim of comparing the iPRP with previously proposed principles and strategies. In particular, we compare the empirical effectiveness of iPRP against those of PRP, MMR, MPT, and qPRP. We refer to [1, 5, 6] for the formulation of these approaches.

For this study, we employed the TREC 6-8 Interactive subtopic collection and the TREC 18 on the Clueweb09 (part B) dataset. Rankings are evaluated using IA-P[2], NRBP, and α-NDCG (with $\alpha = 0.5$). All approaches were implemented using Lemur[3]. Each collection was indexed where common stop-words were removed and Porter stemmer was applied. For each alternative ranking approach, re-ranking of the top[4] 100, 200, 500 and 1000 documents was performed. The Pearson's correlation between pairs of documents' term vectors was used to estimate dependencies between documents. Approaches were instantiated as follows.
PRP. The PRP is implemented employing Okapi BM25 scoring schema, where the model's parameters are set to standard values. Scores obtained by this method are used to generate the relevance estimates required by the other approaches, thus serving as baseline.
MMR. We investigated the effect on retrieval performances of MMR's parameter λ, varying λ in $[0, 1]$ with steps of 0.1.
MPT. No variance is associated with the relevance scores of Okapi BM25, and thus we resort to treat variance as adjunctive parameter, setting a constant variance value amongst documents. We investigated the optimal value of the variance σ_d^2 in the range $[10^{-10}, 10^{-1}]$ and combine it with the value of the parameter b, ranging in $[-10, 10]$.

qPRP. The implementation of the qPRP does not require extensive parameter tuning procedures. Without a method to estimate complex probability amplitudes, we resort to an approximation of phases by using Pearson's correlation between the two documents' term vectors, as suggested in [6].

Regarding parameters settings, here we present the performances of those runs that delivered the highest value of α-NDCG@10 over the whole topic set.
Results. The results of our empirical investigation are reported in Table 1. The best value of α-NDCG@10 on the TREC 6-8 for MPT is obtained when $\sigma_d^2 \leq 10^{-7}$, regardless of the value of b, and the obtained ranking is equivalent to that of PRP. This suggests that MPT's ranking formula reduces to the PRP one, when parameters are tuned so as to optimise α-NDCG@10 on the whole set of query topics for this collection. This is not the case however for TREC 18. The best performing model depends on the collection employed. For example, iPRP delivers the most effective ranking, with respect to α-NDCG@10, on the TREC 6-8 subtopics collection, while MPT delivers the best on TREC 18. Further analysis and testing are required to determine whether the differences depend upon the types of documents, needs, or number of subtopics.

4. CONCLUSIONS

In this poster, we have provided an initial attempt at instantiating the iPRP to provide the initial ranking of documents. This was evaluated in the context of subtopic retrieval. These findings suggest that the iPRP can be effective; significantly outperforming the PRP and providing comparable or better performance over other diversity approaches. While we have only considered one possible interpretation of the iPRP, further work will explore other possible instantiations, and in particular what happens when there is interaction.

5. REFERENCES

[1] J. Carbonell and J. Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *SIGIR '98*, pages 335–336, 1998.
[2] N. Fuhr. A probability ranking principle for interactive information retrieval. *JIR*, 12(3):251–265, June 2008.
[3] M. D. Gordon and P. Lenk. When is the probability ranking principle suboptimal. *JASIS*, 43(1):1–14, 1999.
[4] C. J. van Rijsbergen. *Information Retrieval, 2nd Ed.* Butterworth, 1979.
[5] J. Wang and J. Zhu. Portfolio theory of information retrieval. In *SIGIR '09*, pages 115–122, 2009.
[6] G. Zuccon and L. Azzopardi. Using the quantum probability ranking principle to rank interdependent documents. In *ECIR '10*, pages 357–369, 2010.

---

[2] Where intents are considered to be equally important or probable.

[3] The Lemur Toolkit, http://www.lemurproject.org/

[4] We only report performances for re-rankings on the top 100 due to space limits. Performances were similar for alternative configurations.