Estimating Interference in the QPRP for Subtopic Retrieval

Guido Zuccon[1], Leif Azzopardi[1], Claudia Hauff[2], and C. J. "Keith" van Rijsbergen[1]
[1]Dept. of Computing Science, University of Glasgow, Scotland (UK)
[2]Human Media Interaction, University of Twente, Netherlands
{guido, leif, keith}@dcs.gla.ac.uk, c.hauff@ewi.utwente.nl

ABSTRACT

The Quantum Probability Ranking Principle (QPRP) has been recently proposed, and accounts for interdependent document relevance when ranking. However, to be instantiated, the QPRP requires a method to approximate the "interference" between two documents. In this poster, we empirically evaluate a number of different methods of approximation on two TREC test collections for subtopic retrieval. It is shown that these approximations can lead to significantly better retrieval performance over the state of the art.

Categories and Subject Descriptors: H.3.3 Information Storage and Retrieval - *Retrieval Models*
General Terms: Theory, Experimentation
Keywords: Quantum Probability Ranking Principle, Interference estimation, Diversity

1. INTRODUCTION

The Probability Ranking Principle (PRP) provides a theoretically sound ranking strategy, that assumes the independence between document relevance [7]. To move beyond independence, a number of ranking strategies have been proposed which account for interdependent document relevance assessments [6, 8, 9]. A theoretically motivated development that also accounts for this limitation is the recently proposed Quantum Probability Ranking Principle (QPRP) [11]. However, to instantiate a retrieval model that upholds the QPRP the "interference" between documents needs to be estimated (see [10] for more details). This ranking principle prescribes that documents should be ranked according to:

$$\begin{aligned} d_i &= \arg\max \left(P(d) + \sum_{d' \in RA} I_{d,d'} \right) \qquad (1) \\ &= \arg\max \left(P(d) + \sum_{d' \in RA} \sqrt{P(d)}\sqrt{P(d')}\cos\theta_{d,d'} \right) \end{aligned}$$

where $I_{d,d'}$ is the interference between documents d and d', $P(d)$ is the estimated probability of relevance of document d, and RA is the list of documents already ranked. The angle $\theta_{d,d'}$ is the difference between the phases of the probability amplitudes associated to documents d and d'. The interference term arises because in quantum probability theory, the total probability obtained from the composition of the probabilities associated to two events is the sum of the probabilities of the events and their "interference" (i.e. $p_{AB} = p_A + p_B + I_{AB}$)[1] [4]. Under the QPRP, the documents in the ranking share relationships at relevance level, i.e. they interfere with each other, and this interference has to be taken into account when ranking documents.

An open problem in the development of the QPRP is: how to effectively estimate/approximate this interference term? In other words: how can the interference between documents be approximated, and does this translate into effective retrieval performance for tasks, such as subtopic retrieval, where it is imperative that interdependent document relevance is considered? In this poster, we empirically investigate several strategies to approximate the interference term in the QPRP on two TREC collections for subtopic retrieval.

2. APPROXIMATING INTERFERENCE

In this section we propose a number of approaches to approximate the interference term of an instantiation of the QPRP in the context of subtopic retrieval. We set $\cos\theta_{d,d'} = -f_{sim}(d, d')$, where $f_{sim}(d, d')$ is a similarity function between the (normalised) term vectors of documents d and d'.This results in demoting similar documents in the ranking while diverse and novel documents are promoted. With this substitution, the interference term becomes:

$$I_{d,d'} = -\sqrt{P(d)}\sqrt{P(d')} f_{sim}(d, d') \qquad (2)$$

In the empirical investigation proposed in our study, the components of the term vector of a document are the Okapi BM25 weights corresponding to each term occurring in the document. We test two different working hypothesis when computing f_{sim} between the documents already present in the ranking (e.g. $d_1, \ldots, d_{i-1}$) and the current candidate document d:

1. **pairwise:** the user judges the interest of the current document by comparing it to each of the previous ranked documents: in this case, the current candidate and the documents already ranked are compared using f_{sim} in a pairwise fashion;
2. **surrogates:** the user judges the interest of the current document by comparing it to the knowledge he acquired from documents $d_1, \ldots, d_{i-1}$: the current candidate is then compared against a surrogate of the documents already ranked, which is obtained interpolating[2] $d_1, \ldots, d_{i-1}$.

[1]As opposed to what happens in Kolmogorovian probability theory, i.e. $p_{AB} = p_A + p_B$, when A and B are mutually exclusive events.
[2]In this work, we linearly interpolate the documents' term vectors to

SIGIR'10, July 19–23, 2010, Geneva, Switzerland.
ACM 978-1-60558-896-4/10/07.

				Pairwise					Surrogates				
	Measure	PRP	PT	Pear.	L2	Cos.	Jac.	KL	Pear.	L2	Cos.	Jac.	KL
TREC 678	α-ndcg@10	.416	.424	.418	.431	.427	.419	.413	.415	.426	.424	.412	**.433**
	NRBP	.123	.126	.127	.128	.127	.124	.117	.127	.128	.126	.120	**.135**
	IA-P@10	.058	.062	.063	.063	.064	.061	.062	.060	.059	.060	.053	**.067**
	s-r@10	.379	.384	.387	.385	.388	.381	.389	.379	.375	.380	.360	**.402**
Clueweb	α-ndcg@10	.093	.105	.094	.099	.099	.097	**.115*†**	.106*	.100	.094	.106*	.092
	NRBP	.032	.029	.035†	.029	.034†	.035†	**.043†**	.039†	.029	.034*†	.037	.032
	IA-P@10	.033	.041*	.035	.046	.040	.038	**.047*†**	.038	.043	.037	.044	.024
	s-r@10	.151	.178*	.180*	.173*	.168*	.165*	**.190*†**	.185*	.175	.160	.184*	.113

Table 1: Overview of the results obtained over two TREC test collections. Each similarity or correlation function indicates an instantiation of the QPRP where the corresponding function is employed to estimate interference. Statistical significance over the PRP is indicated by *, while † indicates statistical significant improvements over PT.

3. EMPIRICAL STUDY AND RESULTS

To study the effectiveness of the strategies proposed in the previous section, we evaluate the QPRP in the context of subtopic retrieval, employing the TREC 678 interactive collection with the subtopics judgements described in [9], and the recent ClueWeb collection (part B only), along with the topics defined for the Web diversity track. Using the Lemur Toolkit 4.10 we indexed these collections, where stemming was applied and stop words removed. For the QPRP, we test both pairwise and surrogate comparisons, and examine a number of similarity functions to act as f_{sim}: L1 and L2 norms, cosine similarity, Jaccard coefficients, Pearson's correlation, Bayesian correlation score, KL, JSD, skew divergences [5]. The best performing approximations are reported here. Each of the different QPRP approximations were compared against the default PRP baseline, i.e. Okapi BM25 and against the state of the art Portfolio Theory (PT) [8][3]. For PT, we tuned the parameters in each collection to maximise α-ndcg@10. All the models have been implemented in C++/Lemur and code is available on request. We evaluated each model employing the diversity measures suggested in [1, 2, 3, 9].

In Table 1 we report the results for the PRP (i.e. BM25), PT and a subset of the approximations we tried for the QPRP models. First of all we note that the KL based QPRP model performs the best on each collection when surrogates are used in the TREC 678 and pairwise comparisons are used on ClueWeb. These differences might be due to the limited number of topics available for the TREC 678 collection (20 in total), but also because of the different kind of documents in these collections (newswire articles vs. Web pages). Furthermore, while no significance can be calculated for the improvements on TREC 678 due to the number of topics, improvements over PRP and PT obtained on ClueWeb are statistically significant. The Pearson based QPRP consistently provides excellent retrieval performance regardless of the comparison method - and while this is not always better than the optimised PT, it is not significantly worse, and in fact is significantly better on several measures. It should also be noted that since the QPRP based methods do not require extensive parameter tuning like PT, the KL and Pearson instantiations of the QPRP are highly competitive, simple and attractive alternatives.

form the surrogate. Alternative approaches might perform weighted interpolations of these vectors, in order to simulate user's memory effects (e.g. documents retrieved at early ranks are weighted less than documents at ranks close to i) or estimated importance of documents (e.g. documents ranked at early positions contribute more in generating the surrogate than lower ranked ones).

[3] Note, we have treated the variance of a document, σ^2, as a parameter of the PT model and conducted a grid search of the parameter space b (the user propensity to risk) by σ^2 to select the optimal run of PT on each employed collection.

4. CONCLUSIONS AND FUTURE WORKS

In this poster we have investigated a number of strategies to approximate the interference term in the QPRP. Our results show that excellent retrieval performances can be consistently obtained when employing the Pearson based QPRP, while, the KL based QPRP provides the best subtopic retrieval performances overall. Future work will examine what type of comparison (i.e. surrogate or pairwise) should be employed given the data collection, along with incorporating other types of approximations of interference within the QPRP.

Acknowledgments. This work is partially funded by EPSRC EP/F014384/.

5. REFERENCES

[1] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. In *WSDM '09*, pages 5–14, 2009.

[2] C. L. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon. Novelty and diversity in information retrieval evaluation. In *SIGIR '08*, pages 659–666, 2008.

[3] C. L. Clarke, M. Kolla, and O. Vechtomova. An effectiveness measure for ambiguous and underspecified queries. In *ICTIR '09*, pages 188–199, 2009.

[4] R. P. Feynman. The Concept of Probability in Quantum Mechanics. In *Proc. 2nd Berk. Symp. on Math. Stat. and Prob.*, pages 533–541, 1951.

[5] L. Lee. Measures of distributional similarity. In *Proc. ACL'99*, pages 25–32, 1999.

[6] F. Radlinski, R. Kleinberg, and T. Joachims. Learning diverse rankings with multi-armed bandits. In *ICML '08*, pages 784–791, 2008.

[7] S. E. Robertson. The probability ranking principle in IR. *Journal of Documentation*, 33:294–304, 1977.

[8] J. Wang and J. Zhu. Portfolio theory of information retrieval. In *SIGIR '09*, pages 115–122, 2009.

[9] C. X. Zhai, W. W. Cohen, and J. Lafferty. Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In *SIGIR '03*, pages 10–17, 2003.

[10] G. Zuccon and L. Azzopardi. Using the Quantum Probability Ranking Principle to Rank Interdependent Documents. In *ECIR '10*, pages 357–369, 2010.

[11] G. Zuccon, L. Azzopardi, and C. J. van Rijsbergen. The quantum probability ranking principle for information retrieval. In *ICTIR '09*, pages 232–240, 2009.