Developing the Quantum Probability Ranking Principle

Guido Zuccon
Department of Computing Science
University of Glasgow
Scotland (UK)
guido@dcs.gla.ac.uk

Leif Azzopardi
Department of Computing Science
University of Glasgow
Scotland (UK)
leif@dcs.gla.ac.uk

ABSTRACT

In this work, we summarise the development of a ranking principle based on quantum probability theory, called the Quantum Probability Ranking Principle (QPRP), and we also provide an overview of the initial experiments performed employing the QPRP. The main difference between the QPRP and the classic Probability Ranking Principle, is that the QPRP implicitly captures the dependencies between documents by means of "quantum interference". Subsequently, the optimal ranking of documents is not based solely on documents' probability of relevance but also on the interference with the previously ranked documents. Our research shows that the application of quantum theory to problems within information retrieval can lead to consistently better retrieval effectiveness, while still being simple, elegant and tractable.

1. INTRODUCTION

The idea of using quantum theory in information retrieval (IR) was formally put forward by van Rijsbergen [9] in 2004[1]. In [9], the main thesis of this seminal book is to use quantum theory as a bridge between the three mainstream IR approaches; i.e. vector space models, logic models and probability models. While this direction has been largely unexplored, recently there has been a spate of work which aims to develop quantum inspired or quantum based information retrieval models [1, 2, 3, 4, 5, 6, 8, 7, 13, 11].

In this work, we report on the the development the Quantum Probability Ranking Principle [14, 12]. The ranking principle is derived by developing an *analogy* between the famous double-slit experiment and document ranking. The double slit experiment was conducted to demonstrate that kolmogorovian probability fails to adequately describe the outcome of physical phenomena, and this motivated the development of quantum probability theory which incorporates the quantum interference between events.

In [14], it is hypothesized that this quantum interference can be used to account for the interdependence between documents and their associated judgements. In certain tasks, the relevance of a document may depend on the previous documents already assessed, for example in the novelty and diversity tracks. In sub-topic retrieval, the IR system has to provide a document ranking which covers all the possible facets (subtopics) relevant to the user's information need as soon as possible in the ranking. Consequently, following the traditional Probability Ranking Principle, where document dependence is ignored, leads to sub-optimal performance [10][2]. In [12], we perform a series of experiments that also indicate this is the case, and further show that the QPRP leads to better empirical performance. This is because within the QPRP the interdependence between documents is naturally accounted for through the *quantum interference*, and the QPRP suggests that documents ranked until position $n-1$ interfere with the degree of relevance of the document ranked at position n. Intuitively, documents expressing diverse information have higher degree of interference than documents that are similar. For the same reason, documents containing novel information might strongly interfere with documents ranked in previous positions. Even contrary information might be captured by the interference term: documents containing content contrary to the one presented at the previous rank positions might trigger a revision of user's beliefs about the topic.

The remainder of this paper is as follows: the next section briefly outlines the QPRP. Section 3 presents the main results from the study recently performed on sub-topic retrieval. Finally, we conclude in Section 4 by outlining the directions for further work using the quantum based ranking principle.

2. THE QPRP

In [14], the Quantum Probability Ranking Principle is proposed and it's derivation is based on an analogy with the famous double slit experiment. The resultant of this work was the following formulation: when ranking documents, the IR system has to maximise the total satisfaction of the user given the document ranking, achievable by maximising the total probability of the ranking. Using the quantum law of total probability, the resultant ranking strategy impose to select at each rank position a document d such that

$$d = \arg\max \big(P(d_i) + \sum_{d_x \in RA} I_{d_x,d_i} \big) \tag{1}$$

where RA is the list of documents already ranked and I_{d_x,d_i} is the interference between documents d_x and d_i. Note that

[1] Prior to this, van Rijsbergen gave talks as early as 1996 on the topic.

[2] This has led to arguments for the development of a new ranking principle.

Appears in the Proceedings of the 1st Italian Information Retrieval Workshop (IIR'10), January 27–28, 2010, Padova, Italy.
http://ims.dei.unipd.it/websites/iir10/index.html

the traditional PRP is equivalent to the QPRP when the interference is null, i.e. $I_{d_x,d_i} = 0, \forall d_x, d_i \in C$, the documents corpus. In physics, the interference indicates the amount and kind of interaction between waves. If two waves strongly interact with each other, then the absolute value of their interference is high, and vice versa low. The interaction can generate two different outcomes: either increase the effect generated by the sum of the two waves (constructive interference, $I > 0$) or decrease it (destructive interference, $I < 0$). In IR, the interference I_{d_x,d_i} could be negative or positive, and thus demote or promote a document in the ranking depending on the context. For instance in sub-topic retrieval it would be sensible if documents related to the same subtopics negatively interfere, lowering the chances to rank both of them at high positions. This scenario is discussed in the next section.

3. THE QPRP IN SUBTOPIC RETRIEVAL

In [12], the QPRP is empirically tested and validated on the subtopic retrieval task. The ranking under the QPRP was compared with the rankings of models which upholds the PRP, and also against state-of-the-art strategies for subtopic retrieval, i.e. MMR and Portfolio Theory (PT) [10].

The main point of this experimentation was to determine whether the inherent document interdependence could be accounted for by the interference component within the QPRP. Intuitively, the interference component depends upon both the inter-document dependencies and the document's relevance probabilities. Since it is not possible to estimate the interference component directly from the text statistics, for the experiments reported in [12], we have used the Pearson's correlation between interfering documents to compute the interference. We performed the empirical investigation over the TREC subtopic retrieval track, which includes documents from the Financial Times of London contained in TREC 6,7 and 8 collections and 20 ad-hoc retrieval topics, composed of subtopics, from the TREC interactive tracks. We retrieved documents and generated the initial probability distribution using Okapi BM25: this represented the PRP ranking. Afterwards we re-ranked the documents according to three different strategies: our QPRP method and two state-of-the-art techniques for subtopic retrieval, i.e. MMR and PT, which required parameters tuning. The experiments were repeated varying the level of retrieval cutoff and the length of the queries.

From the experimental results[3], we found that (1) the QPRP improves upon PRP baselines for all levels of S-precision and S-recall, (2) the QPRP outperforms MMR and PT across most levels, (3) the QPRP consistently outperforms other strategies across all topics when considering S-MRR@100%, meaning that on each topic the QPRP returns complete coverage of all subtopics at a rank lower than all the other strategies. And, unlike MMR and PT, no tuning or training is required!

4. CONCLUSIONS

In this paper we have reported about the recent introduction of a novel ranking strategy, the QPRP, based on quantum probability and inspired by an analogy with the double slits experiment in physics. The QPRP naturally encodes the interdependence between documents through quantum interference. The new ranking strategy has been empirically investigated, showing that the QPRP consistently outperforms both the PRP and state-of-the-art approaches, i.e. MMR and PT, without requiring parameter tuning. This suggests that the use of Quantum Theory to model processes within information retrieval can lead to substantial improvements in retrieval effectiveness.

Future work examining the utility and applications of the Quantum Probability Ranking Principle will be directed towards:

- impact of the Pearson's correlation coefficient as a mean to approximate interference;
- alternative estimations of the interference;
- how to derive a complex amplitude distribution from the document corpus;
- the relationships between interference in the quantum probability framework and conditional probabilities in Kolmogorovian probability theory; and,
- how to apply the QPRP paradigm to other retrieval tasks, e.g. ad-hoc retrieval.

Acknowledgements: We would like to thank Keith van Rijsbergen for his collaboration, support and mentoring, and Massimo Melucci for his comments and suggestions. This work has been funded by the EPSRC Renaissance project (EP/F014384/1) and the Royal Society International Joint Project JP080734.

[3]Experimental results are available online at `http://www.dcs.gla.ac.uk/~guido/qprpresults.html`

5. REFERENCES

[1] S. Arafat. *Foundations research in information retrieval inspired by quantum theory.* PhD thesis, University of Glasgow, December 2007.

[2] S. Arafat and C. J. van Rijsbergen. Quantum theory and the nature of search. In *QI '07*, pages 114–121, 2007.

[3] C. Flender, K. Kitto, and P. Bruza. Beyond ontology in information systems. In *QI 2009*, pages 276–288, 2009.

[4] Y. Hou and D. Song. Characterizing pure high-order entanglements in lexical semantic spaces via information geometry. In *QI 2009*, pages 237–250, 2009.

[5] A. F. Huertas-Rosero, L. A. Azzopardi, and C. J. van Rijsbergen. Eraser lattices and semantic contents. In *QI 2009*, pages 266–275, 2009.

[6] M. Melucci. A basis for information retrieval in context. *ACM TOIS*, 26(3):1–41, June 2008.

[7] B. Piwowarski and M. Lalmas. A quantum-based model for interactive information retrieval. In *ICTIR '09*, pages 224–231, 2009.

[8] B. Piwowarski and M. Lalmas. Structured information retrieval and quantum theory. In *QI '09*, pages 289–298, March 2009.

[9] C. J. van Rijsbergen. *The Geometry of Information Retrieval.* Cambridge University Press, 2004.

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

[11] G. Zuccon, L. Azzopardi, and C. J. van Rijsbergen. Revisiting logical imaging for information retrieval. In *SIGIR '09*, pages 766–767, 2009.

[12] G. Zuccon and L. A. Azzopardi. Using the Quantum Probability Ranking Principle to Rank Interdependent Documents. In *ECIR 2010*, 2010. to appear.

[13] G. Zuccon, L. A. Azzopardi, and C. J. van Rijsbergen. A formalization of logical imaging for information retrieval using quantum theory. In *DEXA '08*, pages 3–8, 2008.

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