Has Portfolio Theory Got Any Principles?

Guido Zuccon, Leif Azzopardi, and C. J. "Keith" van Rijsbergen
Dept. of Computing Science, University of Glasgow
Glasgow, Scotland (UK)
{guido, leif, keith}@dcs.gla.ac.uk

ABSTRACT

Recently, Portfolio Theory (PT) has been proposed for Information Retrieval. However, under non-trivial conditions PT violates the original Probability Ranking Principle (PRP). In this poster, we shall explore whether PT upholds a different ranking principle based on Quantum Theory, i.e. the Quantum Probability Ranking Principle (QPRP), and examine the relationship between this new model and the new ranking principle. We make a significant contribution to the theoretical development of PT and show that under certain circumstances PT upholds the QPRP, and thus guarantees an optimal ranking according to the QPRP. A practical implication of this finding is that the parameters of PT can be automatically estimated via the QPRP, instead of resorting to extensive parameter tuning.

Categories and Subject Descriptors: H.3.3 Information Storage and Retrieval - *Retrieval Models*
General Terms: Theory, Experimentation
Keywords: Portfolio Theory for IR, Quantum Probability Ranking Principle, interdependent document relevance

1. INTRODUCTION

Inspired by financial models used in economics, a new model for retrieval has been developed: Portfolio Theory [2, 3]. The intuition behind the model is as follows: estimates of the document's relevance can be improved by accounting for the variance and risk of the estimate in relation to the other documents. Key to the approach is the assumption that document relevance is dependent upon other documents and this must be taken into consideration; violating the PRP. Thus, PT deviates from traditional retrieval models which assume independence between documents. In this poster, we analytically examine PT in the context of the PRP and its Quantum counterpart, explaining the relationships that exist between the different principles and PT.

2. RANKING WITH PORTFOLIO THEORY

Portfolio Theory ranks documents by balancing the probability estimates returned by a retrieval model with the variance of their estimates. This accounts for the risk associated with ranking documents under uncertainty. Specifically, the resulting ranking criteria combines the estimated document relevance with: (i) an additive term which synthesises the risk inclination of the user, (ii) the uncertainty (variance) associated with the probability estimation, and (iii) the sum of correlations between the candidate document and documents ranked in previous positions. In PT, for each rank position i, documents are selected according to:

$$d_i = \arg\max \left(P(d) - \overbrace{b(w_d \sigma_d^2 - 2 \sum_{d' \in RA} w_{d'} \sigma_d \sigma_{d'} \rho_{d,d'})}^{\text{additive term}} \right) \quad (1)$$

where $P(d)$ is the estimated probability of relevance of document d, parameter b encodes the risk propensity of the user, σ_d^2 is the variance associated to the probability estimation of document d, w_d is a weight inversely proportional to the rank position, which expresses the importance of the rank position itself, $\rho_{d,d'}$ is the correlation between document d and document d', and RA is the list of documents already ranked. Given this model we now compare it analytically to the two ranking principles.

Probability Ranking Principle: Under the PRP, the optimal ranking would be obtained by taking, at each rank position i, the document d that maximises $P(d)$. In relation to PT then, when the user parameter b is zero, or documents' variance is null, the additive component of Eq. 1 is zero. In this case, PT upholds the PRP. This guarantees the optimality of the ranking in tasks such as ad-hoc retrieval. But this is a trivial case. As soon as $|b|$ increases, the influence of the additive term will perturb the ranking, and PT will begin to violate the PRP (the greater the $|b|$ the further PT departs from the PRP)[1]. This is because documents will not be strictly ordered according to their decreasing probability of (independent) relevance as prescribed by the PRP.

Quantum Probability Ranking Principle: While, we have shown that PT does not uphold the PRP in non-trivial cases, here we explore whether PT upholds a different ranking principle: the QPRP [6]. This ranking principle stems from the use of quantum probability theory [1] within IR through an analogy between the ranking scenario and the double slit experiment. In [6], the idea is that the interference between particles is analogous to the interference between document relevance. Essentially, the interference can be thought to represent interdependent document relevance

[1] Assuming that the other PT parameters are non zero.

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

and it is estimated from documents features and relationships. The resultant ranking principle, the QPRP, retrieves at rank i a document such that:

$$\begin{aligned} d_i &= \arg\max \left(P(d) + \sum_{d' \in RA} I_{d,d'} \right) \\ &= \arg\max \left(P(d) + \sum_{d' \in RA} \sqrt{P(d)}\sqrt{P(d')}\cos\theta_{d,d'} \right) \end{aligned} \tag{2}$$

where $I_{d,d'}$ is the interference between documents d and d' and is equivalent to $\sqrt{P(d)}\sqrt{P(d')}\cos\theta_{d,d'}$. The interference 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. $p_{AB} = p_A + p_B + I_{AB}$)[2] [1]. The angle $\theta_{d,d'}$ is the phase difference between the probability amplitudes associated to documents d and d' (see [6] for further details and [5] for way to estimate this component).

Like PT, the QPRP reduces to the PRP when the interference between documents is null, i.e. documents are not interdependently related. And also like PT, the QPRP is characterised by an additive ranking formula, which interpolates relevancy and document dependencies. Previously, we have shown that PT violates the PRP in non trivial circumstances. This is actually desirable, since PT aims to overcome PRP's assumption of independent document relevance.

But, does PT uphold the QPRP? To answer this, we consider a particular situation. We instantiate QPRP approximating the interference term with a function of the Pearson's correlation ρ between documents term vectors, i.e. $\cos\theta_{d,d'} = -\rho_{d,d'}$, as suggested in [4]. The QPRP's ranking formula can be written as:

$$d_i = \arg\max \left(P(d) - \sum_{d' \in RA} \sqrt{P(d)}\sqrt{P(d')}\rho_{d,d'} \right) \tag{3}$$

Similarly, the Pearson's correlation ρ can be employed to measure the correlation in Eq. 1, as proposed in [3]. Moreover, since σ_d is assumed to be a constant for each document in the collection[3], Eq. 1 can be re-stated as

$$\begin{aligned} d_i &= \arg\max \left(P(d) - b\sigma_d^2 (w_d + 2 \sum_{d' \in RA} w_{d'}\rho_{d,d'}) \right) \\ &= \arg\max \left(P(d) - \sum_{d' \in RA} 2b\sigma_d^2 w_{d'}\rho_{d,d'} \right) \end{aligned} \tag{4}$$

where w_d is dropped for rank equivalence reasons, i.e. whatever the d under consideration, w_d is constant and so is $b\sigma_d^2 w_d$. When instantiating PT in these particular circumstances, b and σ_d^2 can be treated as parameters to be tuned. In particular, PT delivers the same ranking of QPRP, i.e. theoretical optimal performances under QPRP's assumptions, when:

$$\sum_{d' \in RA} \sqrt{P(d)}\sqrt{P(d')}\rho_{d,d'} = \sum_{d' \in RA} 2b\sigma_d^2 w_{d'}\rho_{d,d'} \tag{5}$$

[2] 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.

[3] This assumption is realistic in the case relevance probabilities are estimated using the Okapi BM25 scoring schema, and has been already introduced in [3].

or when the two quantities are proportional (this is justified by rank equivalences). This relation can be exploited to estimate PT's parameters and thus guaranteeing optimality under the QPRP. In fact, from Eq. 5 and focusing on a particular d', these will be characterised by the pairs (b, σ_d^2) and the function $w_{d'}$ such that:

$$b\sigma_d^2 = \frac{\sqrt{P(d)}\sqrt{P(d')}}{2w_{d'}} \tag{6}$$

While the parameterization of Portfolio Theory means that the ranking strategy is more general and configurable than the QPRP, this introduces the complexity and burden of having to estimate these parameters. By using this relationship with the QPRP it is possible to directly estimate the parameters of PT without requiring training data and parameter estimation problems.

3. DISCUSSION AND CONCLUSION

In this poster, we have shown that Portfolio Theory for IR upholds the different ranking principles under particular conditions. While the fact that PT upholds the PRP in only a trivial case is not very useful, the fact that PT can uphold the QPRP in certain non-trivial settings is potentially very useful. This is an important contribution because it shows that PT, under particular circumstances, upholds the Quantum Probability Ranking Principle. This implies that the parameters of PT can be automatically estimated via the relationship with the QPRP and, by doing so, guarantees theoretically optimal performance under the QPRP. This has the added benefit that no expensive parameter tuning is required. It may also be the case that developments within the QPRP, specifically how the angle between documents is approximated, could also be transferred to PT, further improving the method.

Future work will be directed towards empirically investigating whether estimating the parameters of PT given this relationship with the QPRP lead to effective approximations that validate these findings.

Acknowledgments. The authors would like to thank Alvaro Huertas-Rosero. This work is partially funded by EPSRC EP/F014384/.

4. REFERENCES

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

[2] J. Wang. Mean-variance analysis: A new document ranking theory in information retrieval. In *ECIR '09*, pages 4–16, 2009.

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

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

[5] G. Zuccon, L. Azzopardi, C. Hauff, and C. J. van Rijsbergen. Estimating interference in the QPRP for subtopic retrieval. In *SIGIR '10*, 2010. to appear.

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