A Novel Re-Ranking Approach Inspired by Quantum Measurement

Xiaozhao Zhao[1], Peng Zhang[2], Dawei Song[2], and Yuexian Hou[1]

[1] School of Computer Sci & Tec, Tianjin University, China
[2] School of Computing, The Robert Gordon University, United Kingdom
{0.25eye,krete1941}@gmail.com;{p.zhang1,d.song}@rgu.ac.uk

Abstract. Quantum theory (QT) has recently been employed to advance the theory of information retrieval (IR). A typical method, namely the Quantum Probability Ranking Principle (QPRP), was proposed to re-rank top retrieved documents by considering the inter-dependencies between documents through the "quantum interference". In this paper, we attempt to explore another important QT concept, namely the "quantum measurement". Inspired by the photon polarization experiment underpinning the "quantum measurement", we propose a novel re-ranking approach. Evaluation on several TREC data sets shows that in ad-hoc retrieval, our method can significantly improve the first-round ranking from a baseline retrieval model, and also outperform the QPRP.

1 Introduction

Following van Rijsbergen's pioneering work [3], which shows the potential of quantum theory (QT) in IR, a Quantum Probability Ranking Principle (QPRP) [5] was recently proposed. QPRP captures the inter-document dependencies in the form of "quantum interference". In this paper, we aim to explore another important concept of the quantum theory, i.e., the "quantum measurement". The photon polarization [2] is one of the key experiments that support the explanation of quantum measurement. Briefly, after a couple of polarization filters are inserted between the light source (which generates the photons) and a screen, the amount of light finally on the screen can be well explained by the quantum rather than classical measurement [2].

This inspires us to make an analogy of photon polarization in IR. We view the documents as photons generated from the source, and the retrieval process as measuring all the documents by the query polarization filter. In the first-round retrieval, only the query measurement (q-measure for short) is involved to measure the initial state of each document and yield the relevance probability. In order to re-rank the retrieved documents, we insert a t polarization filter with t-measure, which measures the document relevance with respect to topmost documents. The intuition is that usually the topmost (e.g. top 5) documents are more likely to be relevant. After the t-measure, the states of documents are changed, implying their relevance probabilities with respect to the query are revised. Based on the above ideas, we propose a novel re-ranking approach, called Quantum Measurement inspired Ranking model (QMR).

2 Quantum Measurement inspired Ranking (QMR)

2.1 Introduction to Quantum Measurement

We first introduce the basic quantum measurement used in the photon polarization experiment. Please refer to [2] for the complete description for this experiment. A photon's polarization state can be modeled by a unit vector pointing to an appropriate direction. Specifically, the quantum state of any arbitrary polarization can be represented by a linear combination $a\left|\uparrow\right\rangle + b\left|\rightarrow\right\rangle$ of two orthonormal basis vectors $\left|\uparrow\right\rangle$ (vertical polarization) and $\left|\rightarrow\right\rangle$ (horizontal polarization), where the amplitudes a and b are complex numbers such that $|a|^2 + |b|^2 = 1$. The quantum measurement on a state transforms the state into one of the measuring device's associated orthonormal basis. The probability that the state is measured by a basis vector is the squared magnitude of the amplitude in the direction of the corresponding basis vector. For example, a state $\varphi = a\left|\uparrow\right\rangle + b\left|\rightarrow\right\rangle$ is measured by $\left|\uparrow\right\rangle$ with probability $|a|^2$, and by $\left|\rightarrow\right\rangle$ with probability $|b|^2$. After the measurement of $\left|\uparrow\right\rangle$, the state φ will collapse to $a\left|\uparrow\right\rangle$. Similarly, after the measurement of $\left|\rightarrow\right\rangle$, φ will collapse to $b\left|\rightarrow\right\rangle$.

2.2 Our Proposed Model

We define the initial quantum state of any document d as:

$$\varphi_d = \alpha_d \left|1\right\rangle + \beta_d \left|0\right\rangle \tag{1}$$

where $|\alpha_d|^2 + |\beta_d|^2 = 1$, the state $\left|1\right\rangle$ denotes the relevance basis, and the state $\left|0\right\rangle$ denotes the non-relevance basis with respect to the given query q. For the q-measure, φ_d is measured by $\left|1\right\rangle$ and then yields d's relevance probability $|\alpha_d|^2$.

In the first-round retrieval, only the q-measure is involved to compute the document relevance. Therefore, we have $|\alpha_d|^2 = p(d|q)$ and $|\beta_d|^2 = 1 - p(d|q)$, where $p(d|q)$ is the relevance probability returned by a retrieval function.

To re-rank the documents, we introduce the t-measure, which measures any document d with respect to the topmost documents. Specifically, assume a top document d_t has its quantum state $\varphi_{d_t} = \alpha_{d_t}\left|1\right\rangle + \beta_{d_t}\left|0\right\rangle$, where $|\alpha_{d_t}|^2 = p(d_t|q)$. We are interested in the state of document d after measured by φ_{d_t} (i.e. t-measure). Accordingly, the following equation needs to be solved:

$$\varphi_d = \alpha_d \left|1\right\rangle + \beta_d \left|0\right\rangle = \lambda\varphi_{d_t} + \mu\varphi_{d_t}^{-1} \tag{2}$$

where $\varphi_{d_t}^{-1}$ is orthonormal to φ_{d_t}. After formal calculations, we have

$$|\lambda| = |\alpha_d\alpha_{d_t} + \beta_d\beta_{d_t}| \tag{3}$$

After the t-measure, φ_d will collapse to the direction of φ_{d_t} and become

$$\varphi_d^t = \lambda\varphi_{d_t} = \lambda\alpha_{d_t}\left|1\right\rangle + \lambda\beta_{d_t}\left|0\right\rangle \tag{4}$$

where φ_d^t is the state vector of document d after the t-measure. Now, in order to obtain d's relevance probability with respect to the query, d's current state φ_d^t is then measured by q-measure, and the probability on the relevance basis $\left|1\right\rangle$ is

$$p(d|d_t, q) = |\lambda\alpha_{d_t}|^2 \tag{5}$$

This shows that the revised relevance probability for the document d is $p(d|d_t, q)$. If $d = d_t$, then $|\lambda| = 1$ and $p(d|d_t, q) = |\alpha_{d_t}|^2 = p(d_t|q)$, which means that d_t's relevance probability is unchanged after the t-measure.

The above t-measure only considers one topmost document. If we consider k (e.g. 5) topmost documents, denoted as a set T, the revised relevance probability of a document d can be formulated as :

$$p(d|T, q) \propto \sum_{d_t \in T} p(d|d_t, q) sim(d, d_t) \tag{6}$$

where the $sim(d, d_t)$ is the similarity between the document d and d_t, which indicates the importance of the t-measure with respect to the corresponding d_t. In our approach, the revised relevance probabilities by Eq. 6 are used to re-rank the documents. This is different from QPRP, in which the revised relevance probability is the sum of the original probability and the interference term. In addition, our QMR uses the inter-document similarity to indicate the weight of the corresponding t-measure, while QPRP integrates the inter-document similarity into the interference term.

3 Empirical Evaluation

Experiments are constructed on four TREC collections: WSJ87-92 (with topics 151-200), AP88-89 (with topics 151-200), ROBUST2004 (with topics 601-700) and WT10G (with topics 501-550). The title field of topics are used as queries. Lemur toolkit 4.7 is used for indexing and retrieval. All collections are stemmed using Porter stemmer with standard stop words removed during indexing.

The first-round retrieval is carried out by the query-likelihood (QL) model [4]. The smoothing method for document language model is the Dirichlet prior with the fixed $\mu = 700$. QL is set as the baseline method. The top n retrieved documents by the QL are involved in the re-ranking process. The normalized QL scores of these retrieved documents are used to indicate the relevance probabilities, i.e., $p(d|q)$. We report the rank performance of top $n = 50$ and $n = 70$ documents, while we have similar observations when $n = 30$ and $n = 90$.

The aim of this evaluation is to test the performance of two quantum-inspired re-ranking methods, i.e., QPRP and QMR. In QPRP, for the estimation of the interference term, we adopt $\sqrt{p(d|q)p(d'|q)}\rho(d, d')$, rather than $-\sqrt{p(d|q)p(d'|q)}\rho(d, d')$ as used in [5], since the positive interference performs better in our experiments. In QMR, we adopt the Cosine function to measure the similarity between the $tf \times idf$ vectors of two documents with the query words removed, the parameter k (the number of topmost documents used) is selected from $\{5, 10\}$, and the best performance is reported. We adopt Mean Average Precision (MAP) as the primary evaluation metric and the Wilcoxon signed-rank test as the statistical significance test method.

The Evaluation results are summarized in Table 1. We can observe that both QMR and QPRP achieve significant improvement over the QL in most cases. In addition, the proposed QMR outperforms the QPRP. This is possibly because

Table 1. Evaluation Results on Top n Documents

MAP% (+chg%)	#Doc $n = 50$			#Doc $n = 70$		
	QL	QPRP	QMR	QL	QPRP	QMR
WSJ9872	21.57	22.57(+4.64)	23.77(+10.2*)	23.71	25.00(+5.44)	26.07(+9.95*)
AP8889	18.49	19.65(+6.27*)	20.74(+12.2*)	20.61	21.95(+6.50*)	23.20(+12.6*)
ROBUST04	22.78	24.26(+6.50*)	24.68(+8.34*)	24.28	26.13(+7.62*)	26.70(+9.97*)
WT10G	12.63	14.03(+8.83*)	14.32(+12.3*)	13.71	15.35(+11.9*)	15.56(+13.5*)

Significant improvements (at level 0.05) over QL are marked with *.

that in QPRP, all the previously ranked documents are used to interfere with the current document. On the other hand, in QMR, only the topmost (e.g. top 5) documents, which are more likely to be relevant, are involved to measure the current document.

4 Conclusion and Future Work

In this paper, we explore the application of the quantum measurement in the document re-ranking process and propose a novel re-ranking approach, called Quantum Measurement inspired Ranking (QMR). Evaluation results show that QMR can significantly improve the rank performance of top n documents retrieved by a typical language modeling approach. We also compared QMR with the recently proposed quantum interference based model QPRP, and results show that QMR outperforms the QPRP. In the future, we will compare QMR with other quantum inspired models, e.g. the Hilbert subspace based model in [1].

Acknowledgments

This research is funded in part by the UK's EPSRC (Grant No: EP/F014708/2), the China's NSFC (Grant No: 61070044), the EU's Marie Curie Actions-IRSES (Grant No: 247590), and the NSFC of Tianjin, China (Grant No: 09JCYBJC00200).

References

1. Piwowarski B., Frommholz I., Lalmas M., and van Rijsbergen C. J. . What can quantum theory bring to information retrieval. In *CIKM*, pages 59–68, 2010.
2. Rieffel E. G. and Polak W. An introduction to quantum computing for non-physicists. *ACM Comput.Surveys.*, 32:300–335, 2000.
3. van Rijsbergen C. J. *The Geometry of Information Retrieval.* Cambridge University Press, New York, NY, USA, 2004.
4. Zhai C. and Lafferty J. D. A study of smoothing methods for language models applied to ad hoc information retrieval. In *SIGIR*, pages 334–342, 2001.
5. Zuccon G. and Azzopardi L. Using the quantum probability ranking principle to rank interdependent documents. In *ECIR*, pages 357–369, 2010.