Towards Predicting Relevance Using a Quantum-Like Framework*

Emanuele Di Buccio[1], Massimo Melucci[1], and Dawei Song[2]

[1] University of Padua, Italy
[2] Robert Gordon University, Aberdeen, UK

Abstract. In this paper, the user's relevance state is modeled using quantum-like probability and the interference term is proposed so as to model the evolution of the state and the user's uncertainty about the assessment. The theoretical framework has been formulated and the results of an experimental user study based on a TREC test collection have been reported.

1 Introduction

An Information Retrieval (IR) system has to decide whether a document contains information relevant to a user who interacts with the document. By means of relevance prediction, the system may propose documents, queries, advertisements or other information. In this paper, it is distinguished between *relevance assessment* (or assessment) and *relevance state* (or state) – the states are "stored" in the user's mind and cannot be observed by the system whereas the assessments can be observed only at the end of the interaction time. Thus, a state can be viewed as a superposition of assessments which "collapses" to an assessment when the user may make it explicit.

If the retrieval system monitors the interaction, it can collect interaction features (e.g., the display time) and the final assessment. The collected assessments may be used to train the system, but they are only the final outcome of a quite complex interaction in which many unserved variables have been hidden to the system. If the user is not even willing to provide an assessment, it is customary that the system implements implicit feedback based on observed interaction features [1, 2].

Differently from the assessment, the state may change while the user interacts with the document, but it is unobservable. Moreover, the state is not necessarily the final assessment because the latter is only one of potential values of the state. What the system should do is to predict the state at an arbitrary interaction instant. To this end, a probabilistic model may be used, however, either

* The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement N. 247590 and the UK's Engineering and Physical Sciences Research Council under the Renaissance project (Grant No: EP/F014708/2).

the training dataset is unavailable or only stores (final) assessments whereas a training set should store states so as to predict the state at every instant.

In this paper, an approach based on state modeling which does not rely only on interaction features is proposed. The state is modeled as a *quantum-like* probability (QLP) function which does not obey to the classical probability laws – the obedience to these laws would imply that the state *is* either relevance or non-relevance at every interaction instant. QLP instead includes an *interference term* emerging from the superposition of relevance and non-relevance. The interference term updates the prediction probability depending on the degree to which the state evolves and on the user's uncertainty about relevance.

Our hypothesis is that the superposition of relevance and non-relevance, which reflects the user's uncertainty about his assessment, affects the user's behaviour and therefore the interaction features, thus making prediction performed by the system less precise and more prone to error. If our hypothesis is true, it can have important theoretical and practical implications. For example, if interference is detected, the system may support the user to clarify his state by suggesting example documents or by presenting the results in more effective way. Moreover, if there is interference, the system may infer that the initial query is difficult and therefore invite the user to add terms.

Recently, QLP has been investigated in the context of document ranking [3], cognition [4] and dependency between topics [5] or between documents [6]. In this paper, the dependency between the time intervals of the interaction and the assessment is addressed both at the theoretical level and through an experimental user study based on a TREC test collection.

2 Probability with relevance state

Suppose the display time is divided into a given number of equally sized time intervals. The time interval is represented by an observable T such $T = t$ means that the t-th interval has been observed. Using the distributive law,

$$(T = t) = (T = t \wedge R = r) \vee (T = t \wedge R \neq r) \qquad (1)$$

where $R = r$ refers to an assessment[3]. Using classical probability,

$$\Pr(T = t) = \Pr(T = t|R = 0)\Pr(R = 0) + \Pr(T = t|R = 1)\Pr(R = 1)\ . \qquad (2)$$

Prediction requires the estimation of $\Pr(R = 1|T = t)$ by means of Bayes' theorem. However, classical probability *assumes* that the state is either relevance or non-relevance at every interaction instant, that is, that the user is concerned about relevance or non-relevance, exclusively. In fact, the exclusiveness between relevance and non-relevance is not true and a superposition state better models this uncertainty because it does not admit the distributive law [7]. Note that the same argument is valid also for non-binary relevance.

[3] For the sake of simplicity, binary assessments are supposed, $r = 0, 1$.

A superposition state ϕ can be modeled by using QLP:

$$|\phi\rangle = a_0|R=0\rangle + a_1|R=1\rangle \qquad |a_0|^2 + |a_1|^2 = 1 \qquad |a_r|^2 = \Pr(R=r)\ . \quad (3)$$

where constants and vectors are in the complex field. ϕ is neither the positive assessment, the negative assessment nor the average of the two (i.e., the expected value, the mean or the mixture) because, if it were, the existence of *any* of the predefined assessments should be admitted for *every* user. QLP is expressed as

$$q(t) = |\langle\varphi|T=t\rangle|^2 = p(t) + 2|a_0||a_1||\langle R=0|T=t\rangle||\langle T=t|R=1\rangle|\cos\theta \quad (4)$$

where $p(t) = \Pr(T=t)$, $\cos\theta$ is the real part of the complex number $\bar{a}_0 a_1 \langle R = 0|T=t\rangle\langle T=t|R=1\rangle$ and the second term of the right hand of Eq. 4 is the *interference term*.

3 An experimental study

Our experimental study aimed at measuring the interference term. Through a user study, the subjects were asked to interact with the WT10g test collection and to assess the relevance after browsing the documents; for example, the event that the user `user1` submitted the topic `506`, started to interact with the document `WTX074-B09-156` on the 29th of July, 2008, at 5:44:23pm and made the relevance assessment after 167 seconds, was recorded as `user1 | 506 | WTX074-B09-156 | 17:44:23 | 2008-07-29 | 0 | 167000`. The experiment protocol instructed the subjects not to provide any assessment if they believed that the document was irrelevant. The dataset and the other details were described in [2].

Each display time was divided into non-overlapping ten-second intervals, thus obtaining a number of records for each event – the majority of the records do not have any assessment because only the last interval has been associated to an assessment. An access occurs when a user is interacting with a document within a given time interval. In this way, the number of accesses can be calculated for each event; for example, the user `user1` interacted the document `WTX074-B09-156` in 17 distinct and consecutive intervals – the assessment was recorded only at the 17th interval.

Table 1 summarises our experimental results, where the interference term, i.e., $q(t) - p(t)$, is shown in the last column. It was supposed that $R = 0 \equiv r = 0$ and $R = 1 \equiv r > 0$. The table is truncated after ten intervals because the other intervals have low frequencies and most of the interactions ended before one hundred seconds.

The interference term at the early intervals when the proportion of accesses was the highest and, incidentally, when the user's interaction often ends with an assessment [8] is significant. This outcome signals that, even when it was supposed that the majority of users reach an agreement on relevance, QLP may reveal interference and then uncertainty in the user's state.

If our hypothesis is true, it can have some implications. First, the interference term can help model the evolution of the state while the user is interacting with

t	Interval		$q(t)$	$p(t)$	$p(t\|0)$	$p(t\|1)$	$p(t\|2)$	$p(t\|3)$	Interference
0	0	10	0.192	0.179	0.228	0.108	0.074	0.153	7.3%
1	10	20	0.131	0.260	0.264	0.194	0.255	0.306	-49.6%
2	20	30	0.097	0.145	0.145	0.129	0.148	0.159	-33.1%
3	30	40	0.074	0.098	0.088	0.173	0.074	0.096	-24.5%
4	40	50	0.061	0.057	0.047	0.050	0.107	0.057	7.0%
5	50	60	0.052	0.039	0.042	0.043	0.027	0.038	33.3%
6	60	70	0.044	0.033	0.032	0.043	0.040	0.025	33.3%
7	70	80	0.036	0.032	0.027	0.029	0.040	0.051	12.5%
8	80	90	0.031	0.022	0.022	0.022	0.027	0.019	40.9%
9	90	100	0.027	0.016	0.013	0.007	0.040	0.013	68.8%

Table 1. Interference across 10-second intervals.

the document. The presence of a Cosine in the interference term means that $q(t)$ may be less or greater than $p(t)$ at different t's, thus helping model the evolution of the assessment in the user's mind.

Second, the knowledge of the behaviour of the interference term is crucial to prediction. Indeed, if the system were able to measure the interference, it could predict when the state converges to the assessment.

Third, if the system could predict the interference term on the basis of some observed variables, it would be possible to predict the relevance assessment – the predicted assessment would be the closest vector $|R = r\rangle$ to the vector $|\phi\rangle$.

Finally, interference can be seen as a mass of probability that may be distributed across the $p(t|r)$'s, thus changing the $p(r|t)$'s and then the prediction outcome and the ranking of suggested items. The modeling and the prediction using the QLP will be the focus of our future work.

References

1. Liu, C., White, R.W., Dumais, S.: Understanding web browsing behaviors through weibull analysis of dwell time. Proc of SIGIR. (2010) 379–386
2. Di Buccio, E., Melucci, M., Song, D.: Exploring combinations of sources for interaction features for document re-ranking. Proc of HCIR. (2010) http://www.hcir2010.org/
3. Zuccon, G., Azzopardi, L.: Using the quantum probability ranking principle to rank interdependent documents. Proc of ECIR. (2010) 357–369
4. Bruza, P., Busemeyer, J.R., Gabora, L.: Special issue on quantum cognition. J. of Math. Psych. **53** (2009) 303–305
5. Wang, J., Song, D., Zhang, P., Yuexian, H., Bruza, P.: Explanation of relevance judgement discrepancy with quantum interference. Proc of QI. (2010) 117–124
6. Zhang, P., Song, D., Hou, Y., Wang, J., Bruza, P., Melucci, M., McCall, J.: Automata modeling for cognitive interference in users' relevance judgment. Proc of QI. (2010) 125–133
7. van Rijsbergen, C.: The geometry of information retrieval. Cambridge University Press, UK (2004)
8. White, R.W., Kelly, D.: A study on the effects of personalization and task information on implicit feedback performance. Proc of CIKM. (2006) 297–306