A Quantum-based Model for Interactive Information Retrieval*

Benjamin Piwowarski and Mounia Lalmas

University of Glasgow, Department of Computing Science,
Glasgow G12 8QQ, UK

Abstract Even the best information retrieval model cannot always identify the most useful answers to a user query. This is in particular the case with web search systems, where it is known that users tend to minimise their effort to access relevant information. It is, however, believed that the interaction between users and a retrieval system, such as a web search engine, can be exploited to provide better answers to users. Interactive Information Retrieval (IR) systems, in which users access information through a series of interactions with the search system, are concerned with building models for IR, where interaction plays a central role. In this paper, we propose a general framework for interactive IR that is able to capture the full interaction process in a principled way. Our approach relies upon a generalisation of the probability framework of quantum physics.

1 INTRODUCTION

In less than twenty years, search engines on the Web have revolutionised the way people search for information. The speed with which one can obtain an answer to a keyword-based query on the Web is fostering interaction between search engines and their users. Helping users to reach relevant material faster will most likely make use of such rich interaction. Another key to future search systems is the context that further defines the search, whether it be external (e.g. time of the day, location) or internal (e.g. the interests of the user).

Putting aside the problem of evaluating such contextual and interactive search, building models able to explicitly take into account both is of importance, especially since Information Retrieval (IR) models seem to have reached maturity and there is an obvious need to go beyond current state-of-the-art [2].

There are many reasons why we cannot assume that users will provide enough information to state an unambiguous Information Need (IN), such as a TREC topic description. First, users do not always know how to express their IN and

* An extended version of this paper [1] contains the discussion about related works, a tutorial section about the relationship between quantum and classical probabilities, and an example of how the quantum formalism can be used to extend the Rocchio algorithm.

they sometimes have only a vague knowledge of what they are looking for. Second, users knowledge and interests might evolve during the search, thereby modifying their IN. Therefore, it is important that implicit contextual and interaction "information" become integrated directly into IR models *and* experiments [3].

Beside standard relevance feedback models like the Rocchio algorithm [4] or the Okapi model [5], some recent works have attempted to capture context [6] or interaction [7]. However, there is not yet a principled framework that combines both, and that, equally importantly, tries to capture the different forms of possible interactions, namely, query (re)formulation, clicks, navigation. Those tasks are all performed frequently in web searches.

In this paper, we present a framework for interactive and contextual IR. We view search as a process with two different dynamics: (P1) The system tries to capture the user IN while (P2) the user cognitive state, and hence the user IN, is evolving and changing [8]. While the former could be modeled by standard probabilistic models, we claim that the latter can be better modeled by the generalisation of probability theory that has been developed in quantum physics. Moreover, the strong geometric component of the quantum probability framework is particularly important since standard IR models rely on vector spaces and on (some variants of) the cosine similarity [9]. We show how the quantum formalism generalises these latter models (Section 3.1). In particular, we believe that one strength of the geometric models in IR is that they are intuitive. Adding a probabilistic view on this geometry opens the door for new and potentially more powerful IR models.

This paper describes how the quantum probability formalism could be used to build an interactive IR framework.

2 An Information Need Space

Our working hypothesis is that a *pure*, in the sense that we know exactly what the user is looking for, user IN can be represented as a system in quantum physics, i.e. as a unit vector in a Hilbert space[1], and that this state evolves while the user is interacting with the system.

According to the quantum probability formalism, this (IN) vector generates a probability distribution over the different subspaces of the Hilbert space. We make the hypothesis that among other possible uses, such subspaces can be related to the relevance of documents, therefore enabling the computation of a relevance score for a document, and to user interactions (like typing a query or clicking on a document), making it possible to exploit them.

From a geometric perspective, using subspaces to describe "regions" of INs has been (sometimes implicitly) studied and motivated in some works relying on a vector space representation [10,11,12]. Using those IN "regions", the search process would be modelled as follows. At the very beginning of the search process, the user IN is underspecified and is a mixture of *all* possible pure INs. That is,

[1] In brief, an inner product vector space defined over the complex field, see [9] for a formal definition

without any information about the user, we can only know that the user is in one of all the possible IN states with a probability that depends e.g. on how popular this IN is.

We believe that using an IN space can model interactive IR since users change their point of view during a search, and relevance, contrarily to topicality, is expected to evolve within a search session [8,13]. More specifically, we can identify two different types of dynamics within the search process: (P1) The IN becomes increasingly specific *from a system point of view*, e.g. when a user types some keywords or clicks on some documents, i.e. the uncertainty is reduced; and (P2) The IN changes *from a user point of view*. The IN can become more specific as the user reads some documents, or it can slightly drift as user interests do.

Whereas the first process can be easily described within a standard probabilistic framework (we restrict the IN to subspaces of the whole space), the latter would benefit from a quantum probability formalism as the INs can drift from two overlapping subspaces. We posit that the classical probabilistic framework would address the uncertainty of the system view over the retrieval process (P1) whereas the quantum probability framework addresses the changes of the user internal state (P2). As the quantum probability framework is a generalisation of the probabilistic one, we can use the same representation and evolution operators to model both processes.

3 A quantum view

Quantum probability can be thought of as an extension of classical probability theory, and relies on linear algebra in Hilbert spaces. The equivalent of a logical proposition or event A is a subspace or equivalently [9] its associated projector O_A which is called a *yes/no observable*.

All the information about the probability distribution is contained into a *density operator* ρ, and it can be shown that for *any* probability distribution over a Hilbert space there exists a corresponding density operator [9, p. 81]. A density operator ρ can be written as a mixture of projectors $\rho = \sum_O \Pr(O)\, O$ where the sum ranges over projectors O and $\Pr(O)$ sum up to 1. Note that a pure state is defined as a density which is equal to a one-dimensional projector. Denoting tr the trace operator, the probability of the event O_A for the density operator ρ is then given by

$$\Pr_\rho(O_A) \doteq \operatorname{tr}(\rho O_A) \tag{1}$$

From a practical point of view, the above description of standard probabilities with Hilbert spaces unlocks the potential of defining probabilities through geometric relationships, and permits a generalisation to a non standard probability formalism, which we describe in the next section. We posit that at this level, we are able to model the first component of the search process, which corresponds to finding the right *subspace* of the IN, i.e. in classical terms to find the subset of the IN sample space. However, it is intuitive to think that INs are not mutually exclusive. We make the hypothesis that such a non-exclusiveness is captured by the geometry of IN space, and this can be modelled within a quantum probability formalism.

3.1 Superposition, mixtures and information needs

We introduce the notion of superposition and mixture, and relate them to their use in our model of interactive IR. Said shortly, superposition relates to an ontologic uncertainty (the system state is perfectly known, but some events are true *only* with a given probability) whereas mixture relates to standard probabilistic uncertainty (the system is in one of the states with a given probability). Superposition is a salient characteristic of quantum probabilities and is important since it gives us a way to represent geometrically new INs while the quantum probability framework ensures we can still compute probabilities for the new INs. Mixture and superposition gives us more flexibility in the way we can represent our current state of knowledge of an IN.

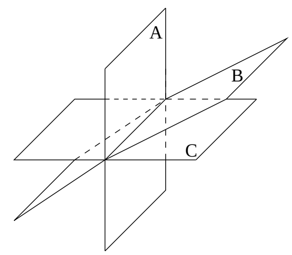

Fig. 1. Three two-dimensional subspaces (A, B, C) in a three dimensional space.

Let us illustrate this with an example. Suppose that $\omega_T = \begin{pmatrix} 1 \ 0 \end{pmatrix}^\top$ and $\omega_L = \begin{pmatrix} 0 \ 1 \end{pmatrix}^\top$ form a basis of the IN space ($\top$ denotes the transpose of a matrix). Suppose the (projector associated to the) former represents the IN of a user looking for information about tigers (T) and the latter about lions (L). In order to represent a user looking for a tigron (the offspring of a tiger and a lion), we assume that this can be represented by (the projector associated to) the vector $\omega_{TL} = \frac{1}{\sqrt{2}} (\omega_T + \omega_L)$ which is a *superposition* of two INs, where the $\frac{1}{\sqrt{2}}$ factor ensures ω_{TL} norm is one. This is a strong assumption which we will study when experimenting with the framework. Aerts and Gabora [14] worked on how to combine concepts in a (quantum) vector space, but use spaces of increasing dimensionality to do so (through the use of a tensor product). As a final remark on superposition of INs, we would like to note that complex numbers could be used to combine INs, e.g. to distinguish tigrons (the tiger is the father) from ligers (the lion is the father), and that superposition is not restricted to topicality. For instance, assuming that we know how to represent a user searching for a paragraph and a user searching for a chapter, we could imagine representing a user looking for a paragraph as a superposition of both.

The superposed IN ω_{TL} is quite different to the IN of a user who is equally interested by tigers or lions. The latter would be represented as a *mixture* of the INs ω_T and ω_L. Formally, this IN would be associated with a density operator$\rho_{T\vee L} = \frac{1}{2} (\rho_T + \rho_L)$ where ρ_L and ρ_T are respectively the projectors associated with ω_T and ω_L, e.g. $\rho_T = \omega_T \omega_T^\top$. The density operator $\rho_{T\vee L}$ is to be interpreted by saying that with probability one half the IN is about tigers (or equivalently about lions).

We can see also the difference if we represent the densities by their matrices in the (ω_T, ω_L) basis. We have the mixture of IN $\rho_{T\vee L} = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$ which is different from the pure IN $\rho_{TL} = \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}$. An important observation is that these different densities imply different probabilities. Let us suppose that the relevance of a document corresponds to a yes/no observable, and that the relevance of a document about lions (respectively tigers, tigrons) are represented by the projectors (yes/no observables) O_L, O_T and O_{TL} associated with the subspaces generated by ω_T, ω_L and ω_{TL}, respectively. For example, $O_T = \omega_T\omega_T^\top$. According to Eq. (1), we can compute the probability of relevance of the different documents, which gives:

$$\Pr_{\rho_{TL}}(O_L) = \Pr_{\rho_{T\vee L}}(O_L) = \tfrac{1}{2} \text{ and } \Pr_{\rho_{TL}}(O_{TL}) = 1 \neq \Pr_{\rho_{T\vee L}}(O_{TL}) = \tfrac{1}{2}$$

Interestingly, we cannot distinguish the probability of relevance of the document about lions when the IN is about either tigers and lions or about tigrons (two first probabilities) but there are two reasons for this: In the former, the probability $\frac{1}{2}$ is caused by the discrepancy between the IN and the document, whereas in the second case the probability is due to the fact that the document only covers a part of the information need. Next, thanks to the quantum formalism the probabilities for the same INs are different when we evaluate the relevance of the document about tigrons (two last probabilities). We thus benefit from a two-dimensional space to distinguish different INs that would be expressed similarly in a standard vector space model. One consequence is that if we search for a set of documents that satisfy T or L, we would have two different types of documents (about tigers and lions, assuming each document covers one IN only) whereas one document would satisfy TL.

Mixtures are also useful to represent the IN density operator ρ_0 at the very beginning of the information retrieval process, as we do not know which state the user is in. We would define the initial IN density operator as $\rho_0 = \sum_i \Pr_i \mathbb{P}_i$ where i ranges over all the possible pure information needs $\mathbb{P}_i$ and $\Pr_i$ is the probability that a random user would have the IN i when starting a search. Using the mixture is also motivated by the fact that we deal with classical undeterminism, i.e. we know the user is in a given state but we do not know which. The mixture can also be thought as a set of vectors describing all the possible INs, each vector being associated with a probability. This representation is particularly useful in the next section where we show how this initial IN ρ_0 is transformed through interactions.

3.2 Measurement and Interaction

Beside differentiating mixture and superpositions, the quantum formalism has also consequences for computing a conditional probability. These consequences are linked to the way a measurement is performed in quantum physics. We use measurement to model interaction and describe in this section both how the measurement modifies the density operatorρ and how we link measurement to the different interactions.

For simplicity, we now use O_A to denote the related yes/no observable, subspace or projector. Since there is a one-to-one correspondence between them [9], they can be used to denote the same thing albeit in a different context. Given a system density operator ρ, if we observe O_A, the new density operator denoted $\rho \triangleright O_A$ is defined by

$$\rho \triangleright O_A = O_A \rho O_A / \mathrm{tr}\left(\rho O_A\right) \tag{2}$$

This amounts to restricting ρ to the subspace defined by O_A and ensuring that $\rho \triangleright O_A$ is still a density operator. The effect of the restriction is to project every IN of the mixture ρ onto the subspace defined by O_A (with some renormalisation to ensure the probabilities still sum up to 1). One can readily verify that the probability of O_A with respect to $\rho \triangleright O_A$ is 1. It means that when A has just be measured, we know it is true at least until further interaction (or in general, evolution) modifies the density operator. Measurement can be thought as a generalisation of conditionalisation, as we can compute the conditional probability of O_A given O_B, or more precisely of measuring O_A knowing that we have measured O_B, as $\mathrm{Pr}_\rho\left(O_B|O_A\right) = P_{\rho \triangleright O_A}(O_B)$.

In quantum theory, the order of the measurements is important, since in general the densities $\rho \triangleright O_A \triangleright O_B$ (applying two times the Eq. (2), for O_A and then for O_B) and $\rho \triangleright O_B \triangleright O_A$ are different. It is a desirable property whenever subsequent measurements of a system should yield different results, which is the case in interactive IR: The sequence of interactions represents the evolution of the user, and should be taken into account. A user drifting from an IN (e.g. hotels in Barcelona) to another (e.g. museum in Barcelona) is not the same as the reverse, which illustrates the adequacy of the quantum formalism to handle such drifts. This is illustrated by Figure 1, where visually it can be seen that measuring O_B (hotels) then O_C (museums) is different from the reverse, since in the first case the IN vectors will lie in the subspace C whereas they would lie in B in the other case.

Starting with the initial density operator ρ_0 (section 3.1), we make the assumption that each implicit or explicit interaction between the IR system and the user corresponds to a measurement, i.e. that every interaction is associated with a yes/no observable O. After the interaction, we can recompute the IN density operator using Eq. (2). For example, a user whose internal context is associated as O_{user}, who asked a query associated with O_{q_1} and deemed a document relevant (associated with O_{d_1}), would be represented by a density operator $\rho_0 \triangleright O_{\mathrm{user}} \triangleright O_{q_1} \triangleright O_{d_1}$. Among other users, this density operator can be used to predict the relevance of other documents.

3.2.1 Mapping interactions to observables In order to map interactions to observables, we restrict to the topical relevance and assume a vector space where dimensions are associated with terms. How to deal with more relevance dimensions is left for future work. We also assume we know how to compute the initial density operator ρ_0 – which could be approximated using the document representation described next.

Giving the current IN density operator ρ_t, we can compute the probability of relevance $\Pr_{\rho_t}(O_d)$ of a document d , provided O_d is the observable associated with the relevance of document d. To build such an observable, and as a first approximation, we can suppose that each paragraph p corresponds to exactly one IN ω_p, and hence that its representation is a one dimensional subspace. It is then possible to compute the subspace spanned by the vectors $\{\omega_p\}$ corresponding to the different paragraphs, and use this subspace to represent the document relevance. When a user deems a document relevant, we could use the same representation to update the current IN ρ_k. In that case, we would have the new IN density operator $\rho_{t+1} = \rho_t \triangleright O_d$.

The first possible type of interaction would be the (re)formulation of a query by a user. We would associate to a given query a subspace/observable O_q, and update the current probability density operator ρ_t to $\rho_{t+1} = \rho_t \triangleright O_q$. A representation of the query could for example be computed through pseudo-relevance feedback provided we know how to represent the documents: The subspace associated with O_q would then be the subspace spanned by the observables representing the top-ranked documents (by a standard IR algorithm). For example, in Figure 1, if A and B correspond to two different top-ranked documents for a given query, then O_q would correspond to the whole three dimensional space (i.e. the join of subspaces A and B). Another way to compute the query observable O_q, without relying on an external model, would be the union of the subspaces representing the paragraphs where each term of the query appears.

Here, we give one illustration of the usefulness of the quantum formalism for an interactive IR framework. The query observable O_q (or the document observable O_d) can be used to detect if a user's change of mind is too important to be a simple drift, an important feature an interactive IR system should have [13]. Within the quantum framework, we use the same geometric representation to both update the density operator knowing an event and to compute the probability of this event. Indeed, when at time t the user types a new query q', we can compute the probability of the query according to the current IN density operator ρ_t, i.e. compute $\Pr_{\rho_k}(O_{q'})$. Based on this value, our IR system would decide that the user switched to a new IN, and react accordingly.

4 Conclusion

We proposed a new interactive IR framework, which exploits the strong connection between geometry and probabilities present in the quantum probability formalism. Our framework allows for a principled and geometric mapping of user interactions into an IR model. In particular, we show how to handle click/relevance feedback and query reformulation. How to use the latter information has not been explored in IR so far, beside providing query recommendation. Other forms of interaction (e.g. navigation) would fit our framework, through the definition of associated subspaces. Beside measurement, the quantum framework is powerful enough to provide other types of evolution of the IN density operator. This would provide a way to predict how a user might evolve, e.g. in order to predict that users looking for hotels might look for museums in a town.

Acknowledgments This research was supported by an Engineering and Physical Sciences Research Council grant (Grant Number EP/F015984/2). Mounia Lalmas is currently funded by Microsoft Research/Royal Academy of Engineering.

References

1. Piwowarski, B., Lalmas, M.: A Quantum-based Model for Interactive Information Retrieval (extended version). ArXiv e-prints (0906.4026) (2009)
2. Sparck Jones, K.: What's the value of TREC: is there a gap to jump or a chasm to bridge? SIGIR Forum **40**(1) (2006) 10–20
3. Ingwersen, P., Järvelin, K.: The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series). Springer-Verlag, Secaucus, NJ, USA (2005)
4. Rocchio, J.J.: Relevance feedback in information retrieval. In Salton, G., ed.: The SMART retrieval system: experiments in automatic document processing. Prentice-Hall, Englewood Cliffs, USA (1971) 313–323
5. Walker, S., Robertson, S.E.: Okapi/keenbow at TREC-8. In Voorhees, E.M., Harman, D.K., eds.: NIST Special Publication 500-246: The Eighth Text REtrieval Conference (TREC-8), Gaithersburg, Maryland, USA (November 1999)
6. Melucci, M.: A basis for information retrieval in context. ACM Transactions On Information Systems **26**(3) (June 2008) 1–41
7. Shen, X., Tan, B., Zhai, C.: Implicit user modeling for personalized search. In: Proceedings of the 14th ACM international conference on Information and knowledge management, Bremen, Germany, ACM (2005) 824–831
8. Xu, Y.: The dynamics of interactive information retrieval behavior, part i: An activity theory perspective. Journal of the American Society for Information Science and Technology **58**(7) (2007) 958–970
9. van Rijsbergen, C.J.: The Geometry of Information Retrieval. Cambridge University Press, New York, NY, USA (2004)
10. Wang, X., Fang, H., Zhai, C.: A study of methods for negative relevance feedback. In Myaeng, S.H., Oard, D.W., Sebastiani, F., Chua, T.S., Leong, M.K., eds.: Proceedings of the 31st Annual International ACM SIGIR, New York, NY, USA, ACM (July 2008) 219–226
11. Widdows, D.: Orthogonal negation in vector spaces for modelling word-meanings and document retrieval. In: Proceedings of the 41st ACL conference, Morristown, NJ, USA, Association for Computational Linguistics (2003) 136–143
12. Zuccon, G., Azzopardi, L., van Rijsbergen, C.J.: Semantic spaces: Measuring the distance between different subspaces. In Bruza, P., Sofge, D., Lawless, W., van Rijsbergen, C.J., Klusch, M., eds.: Proceedings of the Third Quantum Interaction Symposium. Volume 5494 of Lecture Notes in Artificial Intelligence., Springer (March 2009)
13. Xie, H.I.: Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. Journal of the American Society for Information Science **51**(9) (2000) 841–857
14. Aerts, D., Gabora, L.: A theory of concepts and their combinations II: A Hilbert space representation. Kybernetes **34** (2005)