Processing Queries in Session in a Quantum-inspired IR Framework

Ingo Frommholz[1], Benjamin Piwowarski[1], Mounia Lalmas[2], Keith van Rijsbergen[1]

[1] School of Computing Science, University of Glasgow
ingo@dcs.gla.ac.uk, benjamin@bpiwowar.net, keith@dcs.gla.ac.uk
[2] Yahoo! Research Barcelona
mounia@acm.org

Abstract. In a search session, users tend to reformulate their queries, for instance because they want to generalise or specify them, or because they are undergoing a drift in their information need. This motivates to regard queries not in isolation, but within the session they are embedded in. In this poster, we propose an approach inspired by quantum mechanics to represent queries and their reformulations as density operators. Differently constructed densities can potentially be applied for different types of query reformulation. To do so, we propose and discuss indicators that can hint us to the type of query reformulation we are dealing with.

1 Introduction

Common IR systems regard queries in isolation, albeit allowing for query expansion based on relevance feedback. However, in a search session users often reformulate their initial query, for instance because it was ill-specified or they noticed that it did not precisely represent their information need as a more specialised or generalised query would. For example, a query for "processors" may be deemed too broad when looking for processors manufactured by a specific company and thus might be reformulated into "amd processors" or "intel processors". A query reformulation might also represent a minor drift in information needs, for instance a user could seek for pubs in San Francisco after looking for hotels in San Francisco, or it may be the result of a new, independent, information need. This observation motivates that queries should be seen in a session context rather than in isolation, and motivated the new session track at TREC[3].

Using user session information has already been studied in IR. For example, Zhai [5] proposes a risk minimisation framework where each interaction between a user and the system can be captured by a profile described as a language model. However, this type of profile-based approach does not exhibit how to actually take into account query reformulation. Another type of related work is done on identifying query reformulations in query logs, where the typical usage is to propose the user potentially relevant query reformulations [1].

[3] http://ir.cis.udel.edu/sessions/index.html

Recently a quantum-inspired IR framework was introduced that assumes that there exists an "information need space", geometrically modelled as a Hilbert space (a vector space with an inner product). The system's view on the user's information need is represented as a set of *state vectors* φ_i (unit vectors in the Hilbert space) with a given probability p_i. Documents are modelled as subspaces in the information need space. A matching function is realised by taking the squared length of the projection of the state vectors and their respective probabilities [3]. Formally, the user's information need can be represented as a *density operator* $\rho = \sum_i p_i \varphi_i \varphi_i^\top$, which defines a probability distribution over the subspaces in a Hilbert space [4] in the following way. The probability $\Pr(S)$ that a document represented by a subspace S is relevant is defined to be $\mathrm{tr}(\rho \hat{S})$ where $\hat{S}$ is the projector onto the subspace S. It has been shown that by doing so the framework can compete with BM25 in an ad hoc scenario [3]. Following this line, we regard ρ as a query representation in the further considerations.

A potential but so far rather unexplored aspect of the above framework is its ability to dynamically react on different kinds of user interaction, based on the current state the system is in (i.e., the query representation ρ). Translated to queries in a session, the system is able to change its state (based on a newly arriving query or query reformulation) while considering the state it was previously in, for instance based on a previous query.

2 Processing Queries in Session

We illustrate the approach for the simple case of two queries, a query q and a consecutive query q', although it can potentially be extended to a session of more than 2 queries, or to sessions with different types of interaction (e.g., clicks, trackback).

It is desirable to detect what is the relationship between queries q and q', i.e. whether q' is related (generalisation, specialisation or information need drift) or unrelated to q. The idea is that a consecutive query q' should be processed differently, depending on its type. This requires an automatic categorisation of the query reformulation, for which we can apply the quantum formalism as well, and we present the general idea on how this can be done in Section 2.1. We then describe in Section 2.2 how query densities can be created depending on the relationship between the two queries.

The methods presented below rely, for each query q, on the computation of densities ρ_q, which can be done as described in [3]. We also rely on the definition of a subspace O_q that has the property that *any* possible information need vector of a user that has typed the query q is contained by it. This subspace is the subspace spanned by all the vectors that define the density.

2.1 Query Categorisation

To categorise a query reformulation, the idea is to use the densities and their associated subspaces. We consider the probabilities of observing an information need corresponding to the query q (resp. q') given that the query was q' (resp. q):

$$\Pr_{q'}(O_q) = \mathrm{tr}(\rho_{q'} O_q) \text{ and } \Pr_{q}(O_{q'}) = \mathrm{tr}(\rho_q O_{q'})$$

The relationship between the two probabilities indicates what is the relationship between the two queries. Let us illustrate this idea by assuming we have documents which are about "intel processors" and documents about "amd processors". Given the way document subspaces and query densities are constructed in the quantum-based framework [3], this would mean our information need space is contained in a two-dimensional Hilbert space with the two dimensions representing "intel processors" and "amd processors", respectively. Let us assume that $q =$ "intel processors" and $q' =$ "processors", so q' is a generalisation of q. O_q would then be a 1-dimensional subspace made of the "intel processors" dimension, whereas $\rho_{q'}$ would span over both dimensions and would not be contained in O_q, leading to a lower $\Pr_{q'}(O_q)$. $O_{q'}$, on the other hand, would likely comprise both dimensions, whereas ρ_q would only contain the "intel processors" dimension. In this case ρ_q would be fully contained in $O_{q'}$, and hence $\Pr_q(O_{q'})$ would be 1.

Generalising, our hypothesis is that firstly, if $\Pr_{q'}(O_q)$ is lower than $\Pr_q(O_{q'})$, this is an indicator for generalisation (and vice versa for specialisation). Secondly, low values for both $\Pr_{q'}(O_q)$ and $\Pr_q(O_{q'})$ may mean that q and q' are independent, as the respective densities are now at a higher distance to each other. Finally, high but comparable values for $\Pr_{q'}(O_q)$ and $\Pr_q(O_{q'})$ may indicate an information need drift since it means that both densities are closely together, but do not have a specialisation/generalisation relationship.

2.2 Query Processing

Having categorised the query reformulation q' as proposed above, we can now process it according to its type.

For independence, we propose to simply use $\rho_{q'}$ as the representation the query, since there is no obvious way to use the information contained in the previous query q.

In the case of a specialisation or a topic drift, we would choose a strategy and compute a density $\rho_{q'|q}$ that represents a density ρ_q restricted to the region of the information need space defined by $O_{q'}$. The idea is that if the user has submitted the query q', it is because he or she believes that the information need has to be restricted or changed, but supposes that the previous interactions, in our case the submission query q, defines the set of initial potential information needs. In the "San Francisco" example, the interest would be that "San Francisco" has already been disambiguated, and we just need to drift from "hotels" to "pubs". A similar process would occur in the case of "intel processors", for example if the interest of the user about processors (e.g., power consumption) had already been specified.

In order to do so, we use the projection operator as defined in [2, p. 100]. This is analogous to a measurement in quantum mechanics, where the subspace

$O_{q'}$ represents an observed event and its measurement changes the state of the system. Interpreted in an IR scenario, the current state of the system, represented by the density ρ_q, is changed as we observed that the user submitted the query q'.

Finally, when q' generalises q, we might use $\rho_{q'}$ as in the case of independence. It might however also be interesting to extract some of information about the previous query, but this is more prospective. The idea would be, using ontologies, lexical resources or user logs, to generalise the query q *until* the transformed query q_g generalises q'. Then, we would apply the previous method to compute $\rho_{q'|q_g}$. For example, going from "intel processor" to "processor" should still retain the fact that the user is interested by micro-processors (and not food processors). In order to do so, "intel processor" would have to be generalised into "micro-processors".

3 Conclusion

We have discussed how a query can be represented in a quantum-inspired framework, taking into account previous queries that appeared in the same session. We proposed how we can detect within this framework the type of query reformulation we are dealing with, and, based on this type, how to transform the query representation. The various hypotheses formulated in this poster need to be verified. For our evaluation, we plan to use the test collection from TREC 2010 session track, where approaches using a query and consecutive query reformulation are evaluated.

Acknowledgement

This research was supported by an Engineering and Physical Sciences Research Council grant (Grant Number EP/F015984/2). This paper was written when Mounia Lalmas was a Microsoft Research/RAEng Research Professor.

References

1. Boldi, P., Bonchi, F., Castillo, C., Donato, D., Gionis, A., Vigna, S.: The query-flow graph: model and applications. In: CIKM (2008)
2. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Interaction. Cambridge University Press, Cambridge, UK (2000)
3. Piwowarski, B., Frommholz, I., Lalmas, M., Van Rijsbergen, K.: What can Quantum Theory Bring to Information Retrieval? In: CIKM (Oct 2010)
4. van Rijsbergen, C.J.: The Geometry of Information Retrieval. Cambridge University Press, New York, NY, USA (2004)
5. Zhai, C., Lafferty, J.: A risk minimization framework for information retrieval. Information Processing & Management 42(1) (2006)